Spartan'26

for Windows, Macintosh and Linux

Tutorial and User's Guide

May 11, 2026

Wavefunction, Inc.
18401 Von Karman Avenue, Suite 435 Irvine, CA 92612 U.S.A.
support@wavefun.com • https://www.wavefun.com/

Wavefunction, Inc., Japan Branch Office
3-5-2, Kouji-machi, Suite 608
Chiyoda-ku, Tokyo, Japan 102-0083
japan@wavefun.com • https://www.wavefun.jp/

Copyright © 2026 by Wavefunction, Inc.

All rights reserved in all countries. No part of this book may be reproduced in any form or by any electronic or mechanical means including information storage and retrieval systems without permission in writing from the publisher, except by a reviewer who may quote brief passages in a review.

Prepared in the United States of America

Dedication

The latest ***Spartan*** release (***Spartan'26***), is the first in more than 37 years since its initial conception in 1989 (2 years before Wavefunction Inc. was founded) without ***Spartan's*** founding architect, Warren J. Hehre.

Our founder had a knack for the transformative, he was instinctively drawn towards it. From the underpinnings of molecular orbital theory, learned as Roald Hoffman's first research assistant (as an undergraduate at Cornell University) to supporting John Pople at the Mellon institute in building the ("Pople-Hehre") basis sets that went on to form backbone of modern quantum chemical calculations, Warren Hehre was there. When SGI first introduced graphics capable of a facsimile of his vision, the application ***Spartan*** was born.

Early utility as a graphical interface to the *Gaussian* computational code (which, along with John Pople, Warren also helped establish) led to a large initial adoption by computational chemists. Warren's view, however, was somewhat *more* inclusive and encompassing. He envisioned a tool (or set of tools) that the mainstream chemistry community could depend upon, combining the power of molecular modeling with the rapidly accelerating technology of the personal computer, and that these tools would be accessed via graphical interface designed to democratize computational chemistry.

Molecular visualization exploded in the early 1990's making accurate computational approaches like density functional theory *more* practical. In the early 2000's Wavefunction partnered with Q-Chem establishing an ongoing collaboration bringing the computational pipeline (including significant contributions from Martin Head-Gordon, Anna Krylov, Peter Gill, John Herbret, and others) from numerous computational groups to chemists across numerous disciplines: Analytical & Instrumentation chemists, Synthetic & Process chemists, Natural Products chemists, Organometallic chemists, Chemistry educators, and those involved in Cheminformatics.

Warren's attention to the practical is seen throughout the ***Spartan*** user interface but also in the computational developments he pursued while leading Wavefunction Inc. These include a host of molecular visualization options (notably the electrostatic potential map and the polar and accessible area properties based upon ESP), the T1 thermochemical recipe providing rapid accurate heats of formation, and the host of NMR workflows for flexible organic molecules including the 2019 QM-NMR protocol and the more recently the MLXD and MLHF-NMR protocols. Recent efforts are representative of that *knack* for the transformative and leverage machine learning approaches in pursuit of structure elucidation.

Warren J. Hehre submitted his final calculations in January of 2026. He was in the middle of several projects–some of these available for the first time in ***Spartan'26***. Eternally missed, perpetually channeled, and never forgotten; the world is a less clever place without him.

Acknowledgements

The ***Spartan'26*** Tutorial and User's Guide was prepared by Sean Ohlinger with help from Jurgen Schnitker and Phil Klunzinger and from a foundation laid by ***Spartan's*** original architect and father, Warren J. Hehre (1945–2026). Warren deserves credit for this document's utility, while others are primarily to blame for its limitations and inaccuracies. Several additional members of Wavefunction contributed, including Bernard Deppmeier, Andy Driessen, Jeffrey Johnson, Connor Ohlinger, Noritaka Uchida, Yuusuke Watanabe and Tom Hehre. Special thanks to Pamela Ohsan for initial formatting and for years of turning sloppy manuscripts into (almost) finished works. Updates and additional revisions to follow.

Notes for v.1 release: In support of the initial ***Spartan'26*** release, this manual update contains hundreds of changes from the previous Spartan'24 version documentation. Like the initial Spartan'26 release itself, this reference builds upon the foundation laid in previous documentation. We attempt to highlight a number of new machine learning (neural network) options, representing several years of development effort from Wavefunction's current team. A set of tutorials focused *only* on new features is included upfront. Subsequent molecular mechanics and quantum chemical-based tutorials remain largely unchanged but have been expanded where leveraging machine learning options either reduces computational time or adds flexibility (or both).

Beyond the new feature tutorials, the following sections have received updates from the previous manual: Table of Contents, Section II (Operations and Overview), Sections III and IV (Basic and Advanced Tutorials) Chapter 16–20 and Glossary. The remaining chapters (21–27) and appendices will be revised in the next maintenance update (planned within 90 days of initial release).

Additional Spartan Manual updates will be made available via our online repository: https://www.wavefun.com/spartan-documentation - along with future version updates.

Wavefunction, Inc.
Irvine, CA
May 2026

Scope of this Guide

This guide provides a general working reference for ***Spartan'26*** for Windows, Macintosh and Linux. Following an introduction which presents ***Spartan*** as a tool for exploring organic, bioorganic and medicinal, inorganic and organometallic chemistry by way of molecular mechanics, quantum chemical calculations, and machine learning approaches, together with an array of graphical models for conveying the results of these calculations, it comprises **27** chapters grouped into five sections (along with several appendices).

Section I (New Feature Tutorials) provides a set of tutorials featuring use of the new machine learning (neural network) routines and their application to energy, structure, and multi-step conformational searching and NMR protocols.

Section II (**Operations and Overview**; **Chapters 1-2**) describes the operation of ***Spartan***'s graphical user interface and presents a "walk through" illustrating basic interface components.

Section III (**Organic Molecules**; **Chapters 3-7**) comprises a set of tutorials that illustrate ***Spartan***'s building capabilities (both 2D and 3D), show how information can be retrieved from the Spartan Spectra and Properties Database (SSPD) and how quantum chemical calculations can be set up, submitted, and the results accessed/interpreted.

Section IV (**Advanced Tutorials**; **Chapters 8-15**) provides additional tutorials that extend coverage to inorganic and organometallic molecules as well as molecules of interest to medicinal chemists. These tutorials also address a number of capabilities not covered in the previous section, including multi-molecule documents and associated spreadsheet and plotting functions and procedures for identifying low-energy conformers and establishing conformer distributions, for obtaining NMR spectra of organic molecules with several degrees of conformational freedom, for finding reaction transition states and assigning reaction regio and stereochemistry based on matching calculated and experimental NMR spectra and for assessing the "similarity" of molecules, either in terms of structure or of chemical function descriptors (CFDs).

Finally, a set of tutorials is provided to illustrate the access/use of the Cambridge Structural Database from within ***Spartan***. The tutorials in this section are much closer to "research grade" than the earlier tutorials especially with regard to the use of NMR spectroscopy.

Section V (**Features and Functions**; **Chapters 16-27**) describes in detail the functions available from the menus incorporated into the graphical user interface for ***Spartan***. The focus is on graphical input and manipulation of structure, input of other required information

and text, spectral and graphical output resulting from molecular mechanics and quantum chemical calculations, and on use of databases of previously calculated structures, energies, properties and spectra accessible from ***Spartan***. This section is intended as a general reference to ***Spartan'26***.

What this guide ***does not do*** is thoroughly document the performance and cost (in computation time) of the different molecular mechanics and quantum chemical models available in ***Spartan'26*** or explicitly recommend specific models or combinations of models for use on chemical problems. Nor does it show the utility of graphical models in presenting and interpreting the results of the calculations. However, the choice of models used in the tutorials can be taken as implicit recommendation as to what we feel as suitable both in terms of accuracy and practicality, and limited information about the relative cost of these models is supplied, specifically. performance and cost in **Appendix A**. They are covered in greater depth in ***A Guide to Molecular Mechanics and Quantum Chemical Calculations**** included as a PDF under the **Help** menu. The guide also provides a collection of illustrative examples.

Additional appendices provide an overview of the program's overall architecture as well as its present capabilities and limitations (**B**), a directory of functions under its menus (**C**), a listing of commonly-used options (**D**), a listing of units (**E**), the proper journal citation for ***Spartan'26*** and the **SSPD** (**F**), instructions for installing the Cambridge Structural Database (**G**), directions for making databases from ***Spartan*** calculations (**H**), examples of pharmacophore input (**I**), input of experimental infrared, UV/visible and NMR spectra (**J**), setting up and accessing remote ***Spartan*** servers (**K**) and providing chemical shift standards for NMR calculations with different theoretical models (**L**). Additional materials relating to several of these appendices may be found as PDFs under the **Help** menu. Finally, a practical glossary is included defining many specialized terms and acronyms unique to computational chemistry, quantum chemical calculations, and methodology nomenclature.

An up-to-date version of this ***Tutorial and User's Guide*** is available via online documentation at: https://www.wavefun.com/spartan-documentation.

* This 2003 reference is somewhat dated as some models cited have certainly been overtaken (replaced) by modern models (in particular, missing are assessment of the new Head-Gordon and Truhlar functionals). An updated version is this guide is a work in progress.

Table of Contents

Introduction		9
Section I	**New Feature Tutorials**	17
Section II	**Operations and Overview**	37
Chapter 1	Operating Spartan	38
Chapter 2	Walking through Spartan	49
Section III	**Introductory Tutorials**	67
Chapter 3	Building Organic Molecules in 3D	69
Chapter 4	Sketching Organic Molecules in 2D	78
Chapter 5	Spectra & Properties of Organic Molecules from SSPD	87
Chapter 6	Graphical Models for Organic Molecules from SSPD	91
Chapter 7	Spectra, Properties, & Graphical Models for Organic Molecules from Quantum Chemical Calculations	96
Section IV	**Advanced Tutorials**	108
Chapter 8	Groups of Organic Molecules	109
Chapter 9	Spectra of Organic Molecules	128
Chapter 10	Inorganic and Organometallic Molecules	144
Chapter 11	Chemical Reactions	152
Chapter 12	Medicinal Chemistry	173
Chapter 13	Flexible Molecules	188
Chapter 14	Dry Labs: Using the SSPD	209
Chapter 15	Using the Cambridge Structural Database	220
Section V	**Features and Functions**	224
Chapter 16	File Menu	225
Chapter 17	Edit Menu	231
Chapter 18	Model Menu	233
Chapter 19	Geometry Menu	245
Chapter 20	Build Menu	260
Chapter 21	Setup Menu	292
Chapter 22	Display Menu	341
Chapter 23	Search Menu	399
Chapter 24	NMR Menu	426
Chapter 25	Options Menu	429
Chapter 26	Activities Menu	450
Chapter 27	Help Menu	452

Appendices

Appendix A	Computational Topics	454
Appendix B	Capabilities & Limitations	556
Appendix C	Menus	563
Appendix D	Commonly-used Calculation Options	574
Appendix E	Units	576
Appendix F	Citation	578
Appendix G	Configuring the CSD	579
Appendix H	Constructing Custom Databases	581
Appendix I	Pharmacophore Input	582
Appendix J	Input of Experimental Spectra	583
Appendix K	Platform Dependencies	584
Appendix L	NMR ^{13}C Chemical Shift Standards	590

Glossary 592

Introduction

Every few decades chemists are presented with a new tool so powerful that it changes the very landscape of the subject. NMR spectroscopy is an excellent example, for it fundamentally altered the way that the structures of molecules are revealed, in many cases replacing difficult and time-demanding synthesis with a *simple* spectral measurement. We believe that computational methodology ("molecular modeling") especially in light of advances in artificial intelligence and machine learning models, is poised to have an equal or perhaps greater impact on the way chemistry is done.

Molecular mechanics and quantum chemical calculations provide structures, relative stabilities, properties and spectra of isolated molecules. Because of their inherent simplicity, molecular mechanics calculations have long been used throughout multiple sub-disciplines of the mainstream chemical community. On the other hand, quantum chemical calculations are more time demanding, and only in the last ten to twenty years have fast enough computers become widely available to make their application routine among mainstream chemists.

Under the umbrella of Artificial Intelligence are Machine Learning (ML) models. These include Neural Network (NN) models" or simply "neural nets" trained to provide specific data. In the chemistry space, this often includes energies, structures, properties, NMR chemical shifts, etc. based (arguably sufficient levels of) experimental data. In the case of ***Spartan*** a series of similarly developed ML approaches (technically neural network models) trained on a large (potentially infinite) source of quantum-chemical data, are included. These models have been formally defined and validated within the last several years in a set of increasingly detailed publications,[1-5] in an area of chemistry that is exploding (application of AI or Machine Learning). The latest ***Spartan*** versions take advantage of more as many as 8 different neural network models (often used in multi-step workflows (flexible NMR prediction, conformational analysis) to supplement and in specific instances *replace* time-consuming quantum chemical calculations.

Quantum chemical calculations are called upon to furnish information about the mechanisms and product distributions of chemical reactions, either directly by calculations on transition states, or indirectly based on the Hammond Postulate, by modeling the steric and electronic demands of the reactants. *Quantitative* calculations leading directly to information about the geometries of transition states, and about reaction mechanisms in general, are becoming more and more common, while *qualitative* models are required for systems that are too large or too complicated to be subjected to the more rigorous treatments. Quantum chemical

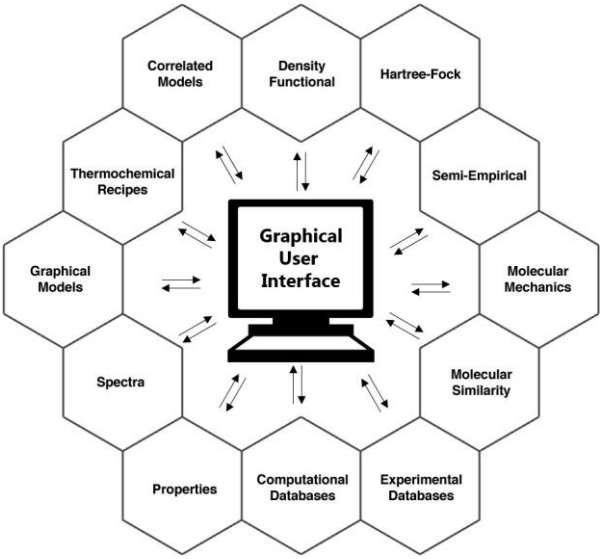
[1] *J. Comp. Chem.* **2025**, 46(1), 70016
[2] *J. Chem. Inf. Model.* **2025**, 65(5), 2314-2321
[3] *J. Comp. Chem.* **2025,** 46(12), 70129
[4] *J. Org. Chem.* **2025**, 90(32), 11478-11485
[5] *ChemRxiv* **2026**, preprint

calculations are also able to furnish infrared, Raman spectra, and *new in this release* VCD spectra. QM also supports excited state calculations, UV/visible spectra, and NMR spectra (shifts and couplings).

The use of calculations to support or challenge structures assigned by NMR or to employ NMR to distinguish between regio and stereoisomers that might result from a specific chemical reaction is of particular value. Finally, quantum chemical calculations can supply information to complement existing experimental data or to replace it altogether, for example, the use of atomic charges for QSAR analyses, and intermolecular potentials for molecular mechanics and molecular dynamics calculations. Machine Learning models have joined the ranks of general "methods" available from ***Spartan***. While currently more restricted in application, they are often incredibly fast, and while not an exact replacement for quantum chemical calculations, they are proving demonstrably reliable. There is every reason to believe that for some tasks, neural network models will become a popular (perhaps the most common) computational approach.

Spartan'26 ("***Spartan***") has been designed to address the ever increasing role that calculations play in chemistry and related fields. It represents a continuing collaboration between Wavefunction Inc., and Q-Chem Inc. Q-Chem code supplements and extends the traditional strengths of ***Spartan***: an easy to learn and use tool for molecular mechanics and quantum chemical and now, machine learning calculations. ***Spartan*** is intended to be ***use by chemists***, not *only* computational chemists who are already familiar with the capabilities of molecular mechanics and quantum chemical calculations (QSAR, Cheminformatics) and ML modeling, but also *experimental* chemists who may have little or no prior experience but want to use calculations much in the same way as experimental techniques like NMR or Mass spectroscopy are employed.

Spartan includes a series of independent modules, tightly connected via a highly functional, but simple and uncluttered graphical user interface. The interface has become less "uncluttered" over the years, but the Spartan philosophy of keeping things as simple as one can is embraced throughout. The interface has been designed not only to greatly reduce the drudgery and possibility for human error associated with the preparation of input, but also to guide and simplify the interpretation of output.



Spartan's interface provides the gateway to a range of modern methods, including molecular mechanics models, semi-empirical and Hartree-Fock molecular orbital models, and correlated models including a wide variety of density functional models, Møller-Plesset models and a selection of high-order wave function-based models such as the popular CCSD(T) approach. Thermochemical recipes such as G3(MP2) and G3 intended to closely mimic the results of CCSD(T) and other high-level correlated models at a small fraction of the computational cost are also supported. A full range of basis sets is supported including the computationally efficient series from Pople-Hehre, cc-pVDZ, cc-pVTZ and cc-pVQZ series from Dunning and the def2 series from Alhrichs and Weigand as well as basis sets customized for individual properties. None of these models is likely to be ideal for every application. While the most sophisticated quantum chemical models may yield excellent results, they will likely be too time consuming for routine application, and it may be necessary to contend with lesser treatments. ***Spartan***'s interface facilitates mixing and matching different molecular mechanics and quantum-chemical models. Results from one model may easily be passed on for further analysis with other (more rigorous) models.

Machine learning (ML) or neural network (NN) approaches offer the possibility of a new class of computational methods that offer the potential of both high reliability and quality as well as low computational cost. Implemented in ***Spartan'26 v.1*** (and later) are a growing set of seven published neural network models. These were developed to (**i**) improve conformer energy differences provided from MMFF94 molecular mechanics, (**ii**) accurately estimate ωB97X-D/6-31G* equilibrium geometries starting from MMFF geometries, (**iii**) accurately estimate total energies from one of the following three:

ωB97X-V/6-311+G(2df,2p)//ωB97X-D/6-31G*
ωB97M-V/6-311+G(2df,2p)//ωB97X-D/6-31G*
ωB97M(2)/6-311+G(2df,2p)//ωB97X-D/6-31G*

(above) models, starting from *either* ωB97X-D/6-31G* equilibrium geometries and energies or from *estimated* ωB97X-D/6-31G* geometries and energies (via neural network), and (**iv**) provide fast, accurate ^{1}H and ^{13}C chemical shifts, trained to reproduce results from the empirically corrected ωB97X-D/6-31G* model. All machine learning models are supported for closed shell uncharged structures comprising the elements: H, C, N, O, F, S, Cl and Br (only). Continued neural network development is underway; additional machine learning models will almost certainly find their way into future ***Spartan*** releases.

Two of the most important quantities to come out of calculation are energy and geometry. While these are unique for a rigid molecule, molecules with one or more rotatable single bond and/or six or higher-member ring comprise a series of conformers, each with its own geometry and energy. ***Spartan*** provides automated (single human step) protocols to furnish either the "best" (lowest-energy) conformer or the set of low-energy conformers that represent the Boltzmann distribution. Molecular mechanics and a variety of quantum chemical models may be used to analyze conformation. In some cases, machine learning (neural network) models may also be employed.

Spartan provides access to several common spectral quantities, in particular infrared spectra (molecular mechanics, semi-empirical, Hartree-Fock, density functional and MP2 models), Raman spectra and VCD spectra (Hartree-Fock and density functional models), NMR spectra (Hartree- Fock, density functional models and a Machine Learning approach for ^{1}H and ^{13}C) and UV/visible spectra (CIS, CIS(D) and time-dependent density functional models). Experimental IR, NMR and UV/visible spectra from freely available on-line databases are easily accessed and superimposed for comparison with calculated spectra.

Spartan provides tools to quantify both the similarity of molecular structures and of chemical environments characteristic of these structures. Also available is the ability to identify molecules that fit into a specific chemical environment (a so-called pharmacophore). Pharmacophores may be extracted from small molecules bound to macromolecules (proteins and nucleotides) from the PDB (Protein Databank).

Spartan includes a variety of graphical tools to assist in interpreting the results of calculations as well as similarity analyses. These include structure models as well as molecular orbitals, electron and spin densities, local ionization potentials and electrostatic potentials that can be displayed as surfaces, slices, and property maps.

Spartan accesses several different databases of calculated and (where available) experimental information. The SSPD (Spartan Spectra and Properties Database version 7) comprises calculated geometries and NMR spectra obtained from the ωB97X-D/6-31G* density functional model for >300k molecules. Most entries also include energies from three additional models (these came primarily as the "training set" data for our energy neural networks): ωB97X-V/6-311+G(2df,2p), ωB97M-V/6-311+G(2df,2p) and the ωB97M(2)/6-311+G(2df,2p) model, allowing more accurate reaction energies to be calculated ("on-the-fly"). About 90% of the SSPD entries have been appended with supplementary IR spectra, calculated using the ωB97X-D/6-31G* density functional model. This allows enthalpies,

entropies and Gibbs energies to be accessed ("on-the-fly"). Finally, heats of formation from the T1 thermochemical recipe are available for about 90% of the SSPD entries.

SSPD also includes ≈8,000 calculated transition-state geometries obtained from the ωB97X-D/6-31G* model for concerted reactions (primarily Diels-Alder, 1,3-dipolar, Cope, Claisen, ene and hydroboration) with improved energies from ωB97X-V/6-311+G(2df,2p).

Spartan'26 allows the user to construct a database to be used by itself or in conjunction with the SSPD v. 7 collection, the minimum requirement being an equilibrium or transition-state geometry obtained from the ωB97X-D/6-31G* model.

Additionally, ***Spartan*** accesses the freely available data collections such as the PDB of >253,000 protein structures, IR, UV/vis, and thermodynamic data from NIST, and chemical shift data from the NMRShiftDB. Finally, and by separate license, **Spartan'26** can search the latest release(s) of the Cambridge Structural Database (the **CSD**) of more than a million X-ray crystal structures.

New features in *Spartan'26*:

Several improvements distinguish ***Spartan'26*** from its predecessor(s), most notably:

Machine Learning (Neural Network) routines: These presently fall into 3 groups, neural networks that provide improved *energies*, a single routine that provides *structure* (equilibrium geometries), and neural net models for *NMR shifts* (proton and ^{13}C). A full description is available in the **Topics** section of the **Activities** menu, or from **Appendix A** (to this reference).

Neural Networks for Conformational Energy Differences:

Spartan'26 includes a neural network termed **Corrected MMFF** or **DLFF03** (the name "corrected" is not intended negatively towards MMFF, as the routine *relies* on MMFF94 structure. It attempts ωB97X-V/6-311+G(2df,2p)[6-311G*] quantum chemical level improvement. From a performance perspective, these energies are essentially "free". In practice, they run on a time scale of less than a second to a few seconds per calculation. While of significant utility alone, this routine is most useful in the context of conformational searching. In ***Spartan'26*** we provide new options for utilizing this neural network in conjunction with) MMFF. Available from the **Equilibrium Geometry** task as well as the multi-step density functional-based routines: **Equilibrium Conformer**, **Conformer Distribution**, and the **NMR Spectrum** task.

Quantum Mechanics Quality Energies

Given input geometries from either: **ωB97X-D/6-31G*** or **DLFFG1** (a neural network trained to estimate ωB97X-D/6-31G* geometry), these are termed:

DLXD2XV or **Est. Density Functional: ωB97X-V/6-311+G(2df,2p)//ωB97X-D/6-31G***
DLXD2MV or **Est. Density Functional: ωB97M-V/6-311+G(2df,2p)//ωB97X-D/6-31G***
DLXD2M(2) or **Est. Density Functional: ωB97M(2)/6-311+G(2df,2p)//ωB97X-D/6-31G***

This represents a significant time savings, over high-quality energy calculations from: ωB97**X-V**/6-311+G(2df,2p), ωB97**M-V**/6-311+G(2df,2p), and ωB97**M(2)**/6-311+G(2df,2p) models lead to significantly improved reaction and conformer energies compared to those from the much *less costly* ωB97X-D/6-31G* model. These routines are *also* offered as either stand-alone calculations, or as options in a number of multi-step protocols accounting for conformational analysis and flexible molecule NMR shift prediction.

A Neural Network model for Structure (Equilibrium Geometry)

Spartan'26 includes a neural network routine that, provided an optimized MMFF structure, will furnish an **Est. Density Functional ωB97X-D/6-31G*** structure (**Equilibrium Geometry**) using **DLFFG1**. This is not an instantaneous calculation; unlike the neural network routines that provide energies, this approach performs an optimization of structure. **DLFFG1** was trained against a large dataset of >6 million energy and force calculations, using organic molecules ranging from ≈200 to 600 amu). This geometry neural network sufficiently reproduces **ωB97X-D/6-31G*** geometries such that 90% of ^{13}C shifts, based on its geometry results, fall within 1 ppm of reference (ωB97X-D/6-31G*//ωB97X-D/6-31G* GIAO calculations with linear scaling and empirical corrections). Note, this routine was not trained for high quality energies, nor does it supply a wave function (necessary for calculation of most graphical models). As such, this neural network is often used in conjunction with either: an **Energy** (single point energy) calculation at **ωB97X-D/6-31G***, or one of the above three QM energy neural networks (in the context of conformational searching).

Estimating accurate Proton and ^{13}C NMR Chemical Shifts

In recent years, the challenging task of assigning structure has been supported by density functional calculations (typically GIAO-based). Calculations have become a useful tool to aid or challenge constitutional and stereochemical assignments. ***Spartan's* 2019 QM-based multi-step protocol** is often successfully applied in support of confirming structural assignment (especially of flexible systems). The major limiting factor to more routine application of the multi-step protocol remains the unfortunately long calculations times (many hours to days). While the original QM-based **NMR Spectrum** task remains an option, beginning in ***Spartan'26***, QM and ML protocols from a convenient **NMR Protocol** task, with four pre-configured and published workflows that furnish ^{1}H or ^{13}C chemical shifts for flexible organic molecules. Use of available neural network models for both geometry and NMR shifts results in performance gains of two to three orders of magnitude over ωB97X-D/6-31G*//ωB97X-D/6-31G* density functional model GIAO calculations. Typical neural network performance is on the order of within .1 ppm for proton and within 1 ppm for ^{13}C shifts of reference shifts from density functional calculations.

Collectively, the **7 neural networks** can be applied to a number of ***Spartan*** tasks including:

Energy is available from five of these routines, however results from equilibrium geometries from Est. Density Functional ωB97X-D/6-31G* task need additional energy calculations performed.

Equilibrium Geometries are available in two ways, from **Corrected MMFF**, and directly from the **Est. Density Functional ωB97X-D/6-31G* (DLFFG1)** approach.

Conformational Searching. Neural network models are also optionally available in the following tasks featuring conformational searching: **Equilibrium Conformer**, **Conformer Distribution**, **NMR Spectrum**, and **NMPR Protocol**. The **Corrected MMFF** routine can be used, in isolation, to improve energy results from MMFF optimized structures, and in conjunction with the above-mentioned machine learning approaches (in a workflow).

NMR shifts both proton and ^{13}C are available from neural network models from the Energy, Equilibrium Geometry, and NMR Spectrum tasks.

Databases

SSPD (Spartan Spectra & Properties Database). SSPD has been extended to include an additional ≈12,000 organic molecules (primarily drug/drug candidates, pesticides, and natural products), as well as another ≈8,000 transition states for concerted reactions (primarily Diels-Alder, dipolar cycloadditions, Cope, Claisen and ene reactions). These include structures from ωB97X-D/6-31G* and energies from ωB97X-V/6-311+G(2df,2p)//ωB97X-D/6-31G* calculations. Transition states are accessed from the **Guess Transition State** mode in ***Spartan*** graphical user interface.

Natural Products. The collection of natural products included in the previous ***SSPD*** (bundled with ***Spartan'26***) has been extended and modified to provide calculated information on only the lowest energy conformer (ωB97X-V/6-311+G(2df,2p)[6-311G*] energies from ωB97X-D/6-31G* geometries). This provides *both* calculated proton and ^{13}C chemical shifts and experimental ^{13}C shifts for ≈ 2800 published natural products molecules.

Custom (user-generated) database formation. New in ***Spartan'26***, a computational task, a one-step submission for **SSPD compatible**, which performs (minimally) an ωB97X-D/6-31G* geometry and optionally includes ωB97X-V/6-311+G(2df,2p) energies, ωB97X-D/6-31G* NMR chemical shifts and thermodynamic/frequency data (and for legacy purposes, EDF2/6-31G* thermodynamic/frequency data). The **SSPD Compatible** task also applies to transition states (absent the NMR option). User entries can be automatically placed alongside the distributed ***SSPD***. This new task provides an easy to specify option for generation of user-databases of spectra & properties, it also offers a head start in the availability of the current ***SSPD*** collection, a curated collection of >309k equilibrium geometries of organic and organometallic molecules, with a full range of pre-calculated data available from high quality density functional models for structure and energy. Also included are roughly 8,000 pre-calculated transition states.

Parallel Processing

The easy access to high end computational hardware (via cloud computing) has enabled performance assessment on several configurations of high-performance computers. Users running the greater than 16-core license are encouraged to reach out to us via **support@wavefun.com** for access to (and assistance with) benchmarking data.

Graphical Interface

Generate Isomers. The isomer generation tools have been refined with improved defaults and provide automatic stereoisomer/regioisomer elaboration.

2D/3D Display. Improvements have been made to the 2D display mode, originally developed as a building device – in general all features available to 3D are also supported in a new 2D viewing mode. This greatly simplifies some tasks (such as assignment of experimental NMR shifts to atoms). Additionally, all ***Spartan*** versions open native **ChemDraw** files. ***Spartan'26*** for Windows, as in previous releases, provides the ability to build in **ChemDraw** (2D), and import/auto-convert into 3D in Spartan (**Note: ChemDraw** interface can be enabled from **Options** menu > **Preferences** > **Settings** tab).

Dual Display. A display window provides 2D structure view when utilizing the 3D user interface, and for 3D display when using the flat 2D sketch display.

NMR menu. The **NMR** menu (previously named Expt. Data) has been refined, offering a new, atom-level spreadsheet with scratch pad for import, organization and assignment of NMR data (also allows assignment of experimental proton and CH coupling data).

Output Summary. The HTML presentation format has been improved resulting in an order or magnitude (or more) speed increase as some data tables are calculated on-the-fly upon opening. All data from the Output Summary is available for export in single or multi-molecule format. A **References** section has been added to select standard tasks and will be extended going forward.

Computational Enhancements

Spartan'26 (by default) takes advantage of up to 16 cores for parallel jobs, with the option to license/utilize >16 cores for high performance multicore systems. In the initial release, *significant* computational enhancements come by way of the new machine learning models available for several routine tasks and included our implementation of the latest Q-Chem 6.3/6.4 version. This release is the result of continued collaboration with Q-Chem, Inc. (first initiated with the release of ***Spartan'02*** in 2002).

Through this collaboration customers benefit from both GUI and significant computational enhancements (the T1 thermochemical recipe, the multi-step protocols for conformational searching in conjunction with Boltzmann weighted NMR data, and the new machine learning models), as well as a growing range of computational options and modern QM development in Q-Chem. ***Spartan'26*** includes an updated version of our quantum engine to the current Q-Chem version.

Spartan users have consistently and diligently made suggestions and reported program bugs via ***support@wavefun.com***. Many reported issues from the previous release have been addressed in this current update, including update to user documentation (this document). Wavefunction would like to thank our customers for their continued help in improving and simplifying ***Spartan***. User feedback is greatly appreciated!

Section I

New Feature Tutorials

Neural Network Calculations

The following tutorials provide a number of examples utilizing the new neural network (machine learning) routines incorporated in ***Spartan'26***. In order to reduce calculation time and also to provide for assessment of results, these tutorials often start from a file containing results of previous density functional (QM) calculations. These include geometries, isomer energies and conformer energies, NMR shifts, etc. Sections are loosely organized around specific neural network routine tasks: **Estimate ωB97X-D/6-31G* Equilibrium Geometries**, **NMR from Estimated ωB97X-D/6-31G* Equilibrium Geometries**, **Multi-step Conformer Distribution Protocols**, **Improving Energy results**, and **NMR Protocols for Flexible Molecules**.

Basic knowledge of ***Spartan*** operation is assumed. New ***Spartan*** users should visit the **Activities** menu and *minimally* review the first three tutorials, as well as the seventh tutorial: **Groups of Organic Molecules**, before attempting the following examples.

While the primary intent is to expose users to the machine learning models, a secondary goal is to quantify the computational time savings when employing (one or more) neural network routines in place of calculation step(s) previously performed with quantum chemical models. Throughout, an attempt is made to estimate the total time required to perform all steps in the tutorial (including calculations). Calculation time estimates are based on ***Spartan'26*** running on either: a 24-core i9 Intel® Core™ i9 Windows 11 machine or 64-core AMD™ Ryzen Threadripper PRO™ Ubuntu 24.04 server. Note, in general Linux (and Macintosh) operating environments show superior parallel performance (over Windows) for most multicore calculations; the Linux implementation includes the most efficient memory allocation.

Estimate ωB97X-D/6-31G* Equilibrium Geometries

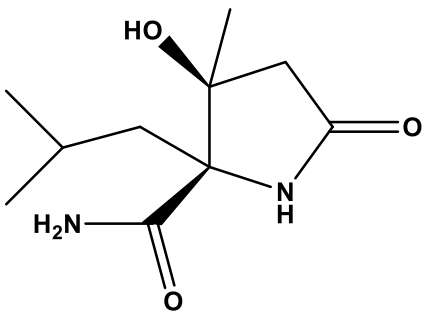
While the ωB97X-D/6-31G* density functional model has proven remarkably successful for calculation of equilibrium geometries (structures), it becomes impractical for routine calculations on rigid or nearly rigid molecules with molecular weights greater than perhaps 1000 amu and for even *smaller* molecules with multiple degrees of conformational freedom (hundreds to thousands of accessible conformers). Routine usage would benefit from an alternate (fast) approach. A good deal of time has been spent exploring post-molecular mechanics options including semi-empirical models such as PM3, PM6, PM7, etc. While these alternatives meet the needs for *faster* results and are applicable to *larger* and *more flexible* systems and generally provide *reasonable* equilibrium geometries, experience shows that these options are no better than molecular mechanics at providing accurate

energies or conformer energy differences, Boltzmann distribution(s).

Spartan'26 employs the **DLFFG1**[2] neural network model trained to reproduce our default workhorse default: ωB97X-D/6-31G* geometries starting from molecular mechanics (MMFF94). This is available for neutral, closed-shell organic molecules including H, C, N, O, F, S, Cl, and Br. Examples follow.

Monacustin structure **[10-30 minutes]**

1. Open the file ***monacustin*** from the **neural network** folder inside the **Tutorials** directory.[3]



monacustin

This file contains the equilibrium geometry for the natural product monacustin, obtained from the ωB97X-D/6-31G* density functional model, and will serve as the reference to compare results from a neural network calculation. The provided calculation results required approximately 4 and a half minutes wall time on an 8-core Intel I7 processor), you can review the **Output** via the **Display** menu. The **Output** tab includes timing data, toward the bottom of text output. **Optional:** To test performance on *your* machine, you can, *click* the **Minimize** entry from the **Build** menu, open the **Calculations** dialog, and submit an **Equilibrium Geometry** using the default **ωB97X-D/6-31G*** density functional model. Note the anticipated calculation on a modern quad-core laptop is on the order of 6-10 minutes.

2. From the **Spreadsheet** (**Display** menu > **Spreadsheet**), make a copy of the monacustin molecule, and paste this into the **Label** column under the existing entry (*right-click* on ***monacustin*** and choose **Copy** then *click* in the empty cell underneath ***monacustin*** and choose **Paste**, and then to **Paste Molecule(s)**. Rename the second copy ***monacustin neural net*** (*double-click* in the new cell and overwrite the auto-generated name). With the copy

[2] Hehre, T.; Klunzinger, P. E.; Deppmeier, B.; Ohlinger, W.; Hehre, W. Accurate Prediction of ωB97X-D/6-31G* Equilibrium Geometries from a Neural Net Starting from Merck Molecular Force Field (MMFF) Molecular Mechanics Geometries. *J. Chem. Info. Mod.* **2025**, *65* (5), 2314–2321. https://doi.org/10.1021/acs.jcim.4c01898

[3] **For Windows**, the ***Tutorials*** directory is found in ***Program Files/ Wavefunction/Spartan26***.
Suggested: copy this to Documents or Desktop prior to opening it in ***Spartan***.
For Macintosh, this is located at the top of the ***Spartan26*** disc image.
Suggested: copy this to another location available to the user (we recommend Documents or Desktop).
For Linux, the ***Tutorials*** directory is found in the install directory.
Suggested: copy the ***Tutorials*** directory to a location that allows write permission, typically the user's home directory.

(*monacustin neural net*) selected, estimate the equilibrium geometry calculation:

3. Open the **Calculations** dialog, select **Equilibrium Geometry** and **Est. Density Functional** to the right of **Calculate**. At present, **ωB97X-D/6-31G*** is the only selection available. Make certain that the **Global Calculations** is *not* checked. **Submit** the job.

4. When completed (a few seconds), return to the spreadsheet, select (*click* on) the first molecule, *monacustin ωB97X-D/6-31G** (the full calculated density functional ωB97X-D/6-31G* equilibrium geometry). Select **Align** from the **Geometry** menu and *click* on the **Align by** button at the lower right of the user interface. (Note: aligning by **Structure**).

5. *Qualify* (visually inspect) the results. From the **Spreadsheet**, *click* inside the boxes to the left of the two monacustin molecules (this enables simultaneous display). From the (default) **Ball and Spoke** display style, it appears that the two structures are nearly identical. Switch to the **Space Filling** (**Model** menu) style; this more accurately displays the overlap of atoms. In the space filling model, visual differences are more difficult to discern, the two appear essentially identical.

6. *Quantify* the results. From the **Spreadsheet**, *click* the **Add** button near the bottom left. *Click* the **Molecule List** tab, and from the bottom center, *click* on **Alignment Scores** to post these to the spreadsheet. In the alignment assessment, a score of 1.0 is a perfect alignment and a score of 0.0 is a terrible alignment. By this metric, the molecules are identical.

7. Explore structural detail. Use the **Measure** feature (**Measure Distance**, **Angle** and **Dihedral Angle** from the **Geometry** menu). *Click* again inside the check boxes to the left of the molecule's names in the spreadsheet, to turn off simultaneous display. With *either* example selected, use the measuring options tools to explore distances, angles, and dihedrals. For example, measure the CO bond lengths in the two carbonyl groups. The distances appear at the bottom right of the screen and can be posted to the spreadsheet by *clicking* on .

8. Explore additional organic molecules. Repeat the above tutorial procedure (steps 2 – 7) for two additional natural products: ***simaroubin A*** and ***bimsic***, sample .spartan files can be found in the **neural network** folder inside the **Tutorials** directory.

9. Do results generally agree between the QM and neural network calculated structures? Results will *very closely* follow the ωB97X-D/6-31G* calculations, see **Appendix A** for assessment. This neural network model's definition, training description, and validation assessment are included in the defining paper: *J. Chem. Inf. Model.* **2025**, 65(5), 2314-2321.

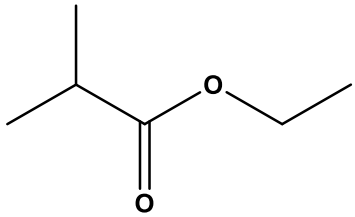
10. When finished, close ***monacustin***, ***simaroubin A***, and ***bimsic***. Note, while the geometries are close to identical, this neural network was not trained on energy and resulting structures typically need an additional calculation to obtain a reasonable energy. See comments below.

NMR from Estimated ωB97X-D/6-31G* Equilibrium Geometries

The neural network routine used to estimate ωB97X-D/6-31G* equilibrium geometries (**Est. Density Functional ωB97X-D/6-31G***) provides an optimized structure, it does not, however, provide an acceptable energy (or the wave function) currently needed to calculate NMR chemical shifts using Density functional GIAO approaches, nor can one generate graphical models such as molecular orbitals or electrostatic potential maps. In order to obtain wave function-based data, an energy (a so called "single point energy") calculation is needed from ωB97X-D/6-31G*. Use of the neural network for structure generation saves ≈ anywhere from 1-3 orders of magnitude in wall/CPU time when compared to the standard ωB97X-D/6-31G* optimization. **Note**: new in ***Spartan'26 v.1*** are proton and ^{13}C neural networks that will accept geometries from *either* **ωB97X-D/6-31G*** or **Est. Density Functional ωB97X-D/6-31G***.

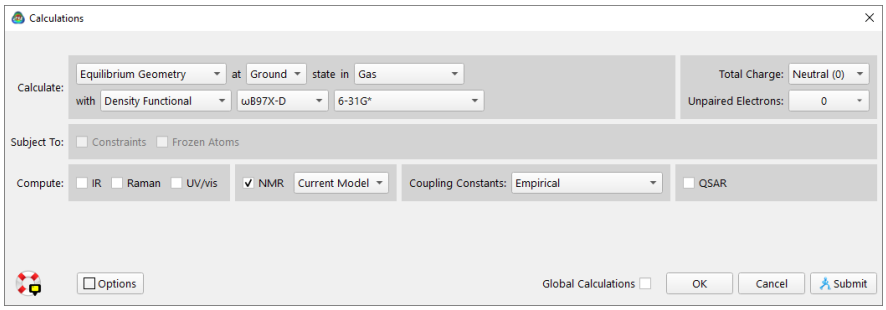
Ethyl 2-methylpropanoate **[10-20 minutes]**

1. Sketch or build ethyl 2-methylpropanoate:



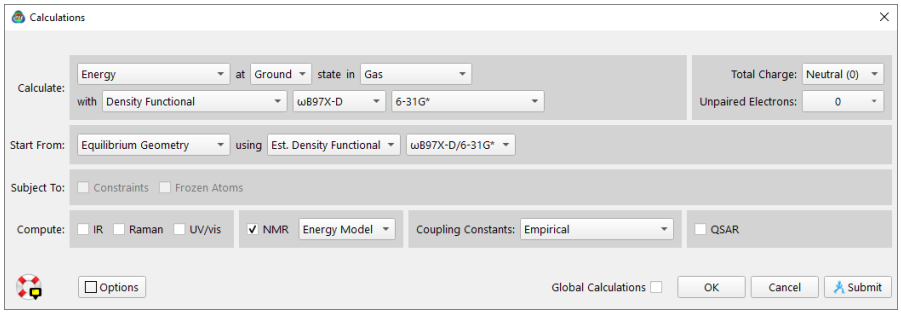
ethyl 2-methylpropanoate

2. Open the **Spreadsheet** and make two copies of the molecule (*right-click*, choose **Copy**, *right-click* cell underneath and choose *paste*, repeat the paste procedure again). For the first molecule, enter the **Calculations** dialog, select **Equilibrium Geometry** to the right of **Calculate**. Select **Density Functional**, **ωB97X-D**, and **6-31G*** from the three menus below and to the right of **with**. *Check* the **NMR** check box. Make sure the **Global Calculations** box at the bottom right of the dialog is ***not checked*** and *click* **OK**.



3. For the second molecule, enter the **Calculations** dialog, select **Energy** to the right of **Calculate**. Select **Density Functional**, **ωB97X-D**, and **6-31G*** from the three menus below and to the right of **with**. From the menu(s) to the right of **Start From:**, select **Equilibrium Geometry** then **using: Est. Density Functional**. *Check* the **NMR** check box. Confirm

Energy Model is selected and that the **Global Calculations** box is ***not checked***.



4. For the third molecule, enter the **Calculations** dialog, select **Equilibrium Geometry** to the right of **Calculate**. Select **Est. Density Functional**, geometry. *Check* the **NMR** check box. Confirm the NMR estimator is selected (not ωB97X-D/6-31G*) and that the **Global Calculations** box is ***not checked***. **Submit** the job.

5. When completed (several minutes), compare the structures and the NMR chemical shifts. Open the **Properties** dialog (**Display** menu), *click* on the carbonyl carbon, and from the **Atom Properties** dialogue, *click* on the **Post** button next to **Chem. Shift**, to post the calculated NMR chemical shifts for this carbon center (for all three examples) into the spreadsheet. One after another, *click* on the two ethoxy carbons, and post their chemical shifts into the spreadsheet. How closely do the NMR results match?

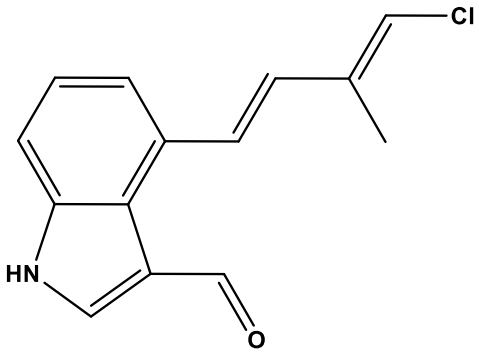
6. From the **Spreadsheet**. Confirm the first entry is selected. From the **Geometry** menu select **Align**. Confirm that **Structure** is selected from the dropdown menu at the lower right of the interface, *click* on the **Align by** button.

7. From the spreadsheet, *click* the **Add** button near the bottom left. *Click* the **Molecule List** tab, and from the bottom center of the **Add** panel, *click* on **Alignment Scores**. How well do the results align? *Check* the box to the left of each entry to turn on display.

8. From the **Model** menu turn *off* the **Coupled** toggle. Use the center mouse-wheel to zoom out if needed. Hold the right mouse-button down to translate the molecules away from each other so that the 3 can clearly be seen separately (not overlapping). From the **Model** menu turn *off* the **Hydrogens** toggle. Return to the **Model** menu and *click* on **Configure**. From the **Labels** tab, *click* the radio button to the left of **Calc. Chem Shift** and *click* **OK**. Compare the shift values on all carbon centers. Differences are very subtle, on the order of tenths of a ppm.

9. Close the ***ethyl 2-methylpropanoate***.

Indiacen B **[10-15 minutes]**



indiacen b

1. From the **neural network** folder inside the **Tutorials** directory, open the file ***indiacen b*** (a slightly larger natural product molecule, this alkaloid is isolated from certain species of myxobacterium or "slime molds"). **Equilibrium Geometry** and NMR chemical shifts are furnished from **ωB97X-D/6-31G***.

2. Open the **Spreadsheet** and make a copy of the molecule **indiacen b DFT**, name it **indiacen b est DFT**.

3. With the *renamed copy selected*, enter the **Calculations** dialog, and specify **Equilibrium Geometry** and **Est. Density Functional**. *Check* the **NMR** check box. Confirm that the **NMR estimator**[4] is selected and that the **Global Calculations** box is ***not checked***. **Submit** the job.

4. When the job completes (just a few minutes), compare the NMR shifts on the carbonyl carbon, the terminal carbon bonded to chlorine, and any other carbon centers you like. Use the **Atom Properties** dialogue and post the calculated shifts (**Chem. Shift**) into the spreadsheet to compare results. While the results are not identical, they are very similar (typically within a few tenths of a ppm). Use of neural network routines (in this case) saves ≈ a factor of 50 or more in computational time.

5. This example explored a relatively small and rigid molecule. Time savings for larger and more flexible systems is even more significant, where select individual calculation steps have the option of utilizing neural network routines rather than QM models.

6. Close ***indiacen b***.

[4] Hehre, T.; Klunzinger, P. E.; Deppmeier, B. J.; Ohlinger, W. S.; Hehre, W. J. Practical Machine Learning Strategies. 4. Using Neural Networks to Replicate Proton and ^{13}C NMR Chemical Shifts Obtained from ωB97X-D/6-31G* Density Functional Calculations. *J. Org. Chem.* **2025**, *90* (32), 11478–11485. https://pubs.acs.org/doi/10.1021/acs.joc.5c00927

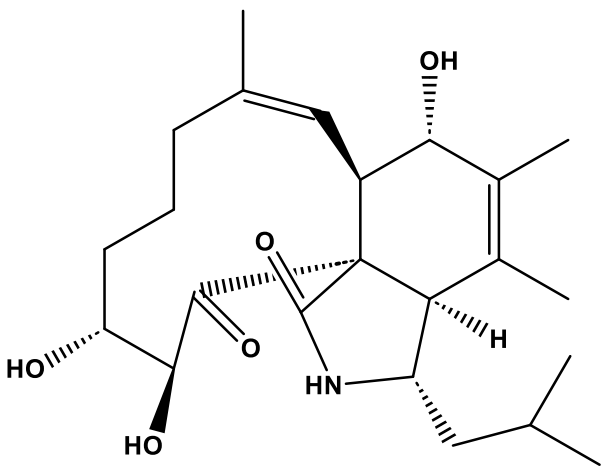
Neural Networks in Multi-step Conformational Protocols I

For the last several ***Spartan*** releases, we have included multi-step workflows combining both molecular mechanics (MMFF94) and quantum mechanics (a number of models) to obtain, what we believe to be, accurate Boltzmann distributions of flexible organic molecules. Both as a stand-alone task (**Conformer Distribution**) and in combination with NMR calculations in the **NMR Spectra**[5] task, these recipes have proven useful in confirming (and in some cases rejecting) structural assignments in the natural products space.

While increasing in importance and utility, the use of predicted NMR (in particular Boltzmann weighted shifts) has been limited by the significant computational demands required (several hours, to *days* for very flexible systems).

Replacing steps in these workflows with neural network calculations *significantly* reduces the required computational requirements, while maintaining quality results.

Aspochalasin U and Boltzmann Weighted NMR [30 minutes]



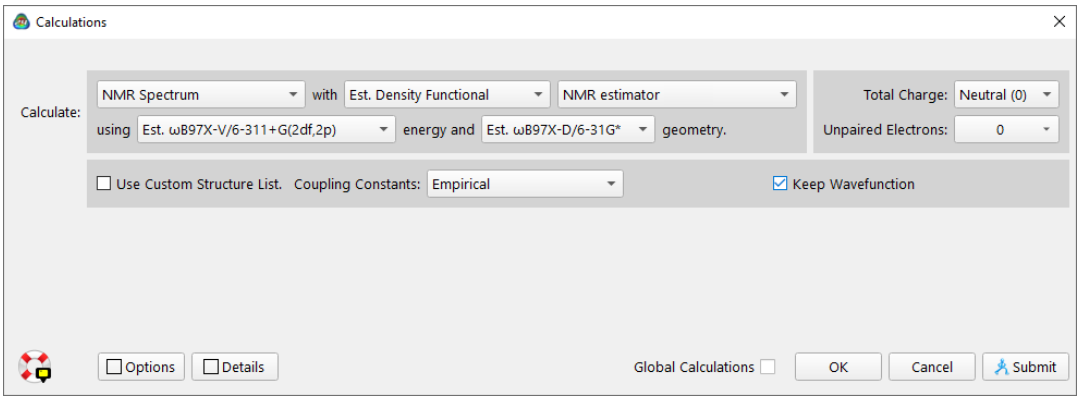
aspochalasin u

1. From the **neural network** folder inside the **Tutorials** directory, open the file ***aspochalasin u***. This file contains **NMR Spectrum** calculation results for the macrocyclic natural product aspochalasin u, one of several alkaloid macrocycles isolated from the desert shrub *ericameria laricifolia* (the "turpentine bush"). Results originate from the default **NMR Spectrum** setup, starting from MMFF conformer searching and including multiple QM model steps to provide both an equilibrium conformation and a set of Boltzmann weighted NMR chemical shifts. This flexible system with more than 100,000 potential conformations,

[5] Hehre, W.; Klunzinger, P.; Deppmeier, B.; Driessen, A.; Uchida, N.; Hashimoto, M.; Fukushi, E.; Takata, Y. Efficient Protocol for Accurately Calculating ^{13}C Chemical Shifts of Conformationally Flexible Natural Products: Scope, Assessment, and Limitations. *J. Nat. Prod.* **2019**, *82* (8), 2299–2306. https://doi.org/10.1021/acs.jnatprod.9b00603

required considerable computational time (20 hours on an 8 core i7 Windows 11 system).

2. Make a *copy* of the molecule named ***aspochalasin u DFT multi-step*** and *paste* it under the existing molecule. Rename the second molecule ***aspochalasin u NN multi-step***. With the second molecule selected, enter the **Calculations** dialog and specify: **NMR Spectrum** with **Est. Density Functional**, and **NMR estimator**.[4] Next to ***using*** confirm the settings updated to: **Est. ωB97X-V/6-311+G(2df,2p)** energy[6] and **Est. ωB97X-D/6-31G*** geometry[2]:



Submit the calculation.

3. When the job completes (perhaps 25 minutes on a 24-core i9 Windows 11 machine, just under *an hour* on the same 8-core i7 Windows 11 machine that the QM job was completed on), compare the returned equilibrium conformer with the example from the longer QM-based NMR Spectrum. **Align** the two molecules. With the first molecule selected, enter the **Geometry** menu and *click* on **Align**. In the lower right of the user interface, *click* on the **Align by** button (note we are aligning by **structure**). From the spreadsheet, *click* the **Add** button near the bottom left. *Click* the **Molecule List** tab, and from the bottom center, *click* on **Alignment Scores**. The neural network example should have good structural alignment (.98 or better) to the standard NMR Spectrum task (from the multistep MM/QM protocol).

4. How well do the calculated (and Boltzmann weighted) NMR shifts compare? *Right-click* on a carbon center; from the resulting contextual menu, choose **Properties**. From the **Atom Properties** dialog, click to post the calculated chemical shift into the spreadsheet (this is the fourth item down named **Chem. Shift**). Note, the molecules are numbered identically and posting the shift value from an atom on one molecule will also post the corresponding shift value from the other copy.

5. Post the calculated **Chem. Shift** values for the remaining unlabeled carbons and return to the spreadsheet for comparison. How similar are the resulting shifts? In most cases you will

[6] Klunzinger, P.; Hehre, T.; Deppmeier, B.; Ohlinger, W.; Hehre, W. Practical Machine Learning Strategies. 2. Accurate Prediction of ωB97X-V/6-311+G(2df,2p), ωB97M-V/6-311+G(2df,2p) and ωB97M(2)/6-311+G(2df,2p) Energies From Neural Networks Trained From ωB97X-D/6-31G* Equilibrium Geometries and Energies. *J. Comp. Chem.* **2025**, *46* (13), 70129. https://doi.org/10.1002/jcc.70129

find the results quite similar, in most cases results vary by only a few tenths to perhaps 1.5 ppm. The alignment and NMR shift results suggest the neural network calculations perform similarly to quantum chemical calculations in the context of the multi-step procedure. In this case (with a rather flexible molecule and using the neural network routine for structure), the calculation run using available neural network routines saves a factor of ≈20 in computational time.

Improving Energy Results

The fact that ωB97**X-V**/6-311+G(2df,2p), ωB97**M-V**/6-311+G(2df,2p) and ωB97**M(2)**/6-311+G(2df,2p) models lead to improved reaction and conformer energies compared to those provided by the simpler and much *less costly* ωB97X-D/6-31G* model is well known. ***Spartan'26*** incorporates three machine learning routines (neural networks) trained to reproduce the above energies starting from either an ωB97X-D/6-31G* geometry, or an est. density functional (neural network) geometry, trained to reproduce ωB97X-D/6-31G* geometries. The better energy calculation is almost instantaneous (seconds).

Structural Isomers of *C15H16O3* **[15-20 minutes]**

1. Open ***C15H16O3 isomers***. The document contains seven $C_{15}H_{16}O_3$ isomers (all natural products) and pre-calculated energies and geometries from ωB97X-V/6-311+G(2df,2p)//ωB97X-D/6-31G* (labeled ***ref*** in the spreadsheet). Isomers are ordered by relative energies (low to high). The underlying QM calculations required several hours of computer time.

2. Inside the **Calculations** dialog, select **Energy** with ***Est. Density Functional***[6] ***ωB97X-V/6-311+G(2df,2p)*** from the right of **Calculate with** and **Equilibrium Geometry** using ***Est. ωB97X-D/6-31G****[2] from the right of **Start From**. Make certain that **Global Calculations** (at the bottom right of the dialog) *is* checked and *click* on **Submit**.

3. When the job completes (a few minutes, 15-45 seconds or so per isomer), return to the spreadsheet. Confirm that the first molecule is selected, *click* on **Add** at the bottom of the spreadsheet, *click* on the **Molecule List** tab and under **Energy from Molecule:** *click* on ***ΔE (kJ/mol)*** near the top left of the panel. Optionally, rename the ***ΔE (kJ/mol)*** column to: **est ωB97X-V:est ωB97X-D**), *double-click* in the ***ΔE (kJ/mol)*** header, backspace to view the full expression, and at the left of the equal sign, rename the previous to: **est ωB97X-V:est ωB97X-D**.

4. Bring up the **Plots** dialog and *click* on **+** to add a new plot. Select **ref** under **X Axis** and select both ***ref*** and ***ΔE (kJ/mol)*** or **est ωB97X-V:est ωB97X-D**, if you renamed the column from the optional step above from the **Y Axes.** *Click* on **Create** at the bottom right of the panel. How well does this combination of neural network routines reproduce the results of the reference calculations?

5. Open the file ***Plot_Est_V_QM***. In this modified plot, the reference energies of the 7 isomers are the larger blue dots, the smaller yellow dots are the energies determined by the neural network. The neural networks perform quite well on the task of reproducing both the

order and the underlying relative energy results from high-level quantum chemical models (in a small fraction of the time it took to run the reference data calculations).

6. Close both ***C15H16O3 isomers*** and ***Plot_Est_V_QM***.

Conformers of methyl 4-pentenoate [10-15 minutes]

1. Open ***methyl 4-pentenoate conformers***. The file contains eleven conformers of the small organic molecule methyl 4-pentenoate, pre-calculated from ωB97X-V/6-311+G(2df,2p) from ωB97X-D/6-31G* geometries (labeled ***ref*** in the spreadsheet) and ordered from lowest to highest energy. This relatively small molecule with relatively few conformers (11) required only 30 minutes of wall time, perhaps 4 hours CPU time to complete.

2. Open the **Calculations** dialog, select **Energy** with ***Est. Density Functional***[6] ***ωB97X-V/6-311+G(2df,2p)*** to the right of **Calculate** and **Equilibrium Geometry** using ***Est. Density Functional*** and ***Geometry Estimator***[2] to the right of **Start From**. Make certain that **Global Calculations** (at the bottom right of the dialog) *is* checked and *click* **Submit**.

3. Once completed (perhaps 5-6 minutes), open the spreadsheet and ensure the first molecule in the list is selected. *Click* on **Add** at the bottom of the spreadsheet, *click* on the **Molecule List** tab and *click* on ***ΔE (kJ/mol)*** at the top left of the dialog (under **Energy from Molecule**). *Rename* the ***ΔE (kJ/mol)*** data column: ***est ωB97X-V:est ωB97X-D***.

4. Bring up the **Plots** dialog and *click* on +. Select **ref** under **X Axis**, both ***ref*** and ***est ωB97X-V:est ωB97X-D*** under **Y Axes** and *click* on **Create** at the bottom right of the panel. How well does this combination of neural network routines reproduce the results of the reference calculations? This is a reasonable representation of performance within a conformer distribution. Note that while the equilibrium conformer (the global minimum) remains the same, there is some change in order based on the energies from the neural network (over this relatively small energy window of 6-7 kJ/mol).

5. For comparison, return to the **Calculations** dialog and specify an **Equilibrium Geometry** using **MMFF**, make sure the **Global Calculations** box is checked, and *click* **Submit**.

6. When the job completes (less than a minute), return to the **Spreadsheet**. The ***ΔE (kJ/mol)*** column has now updated to reflect MMFF energies. In comparison, how well does MMFF reproduce the reference energies? Bring up the **Plots** dialog and *click* on +. Select **ref** under **X Axis**, and select all *three*: ***ref***, ***est ωB97X-V:est ωB97X-D***, and ***ΔE (kJ/mol)*** under **Y Axes**, then *click* on **Create** at the bottom right of the panel. While the results of the combined neural network routines do not perfectly match the underlying QM reference data, unlike MMFF (reported on the new plot as *ΔE (kJ/mol)*), the neural network results obtain the same global minimum, and in general, better agreement with reference data.

7. Close ***methyl 4-pentenoate conformers***.

Neural Network Correction to MMFF Molecular Mechanics

MMFF molecular mechanics is the simplest and perhaps most widely used method to explore the accessible conformers of organic molecules. While it provides a reasonable account of conformer energy differences for molecules with only a single degree of conformational freedom (many of which were used to parameterize MMFF), it has not proven to be reliable for molecules with multiple degrees of conformational freedom. ωB97X-V/6-311+G(2df,2p) calculations using MMFF geometries are much better, although they are much more costly (hours as opposed to seconds). ***Spartan'26*** incorporates a neural network trained to reproduce conformer energy differences from ωB97X-V/6-311+G(2df,2p)//MMFF density functional calculations and intended to improve on MMFF results without greatly adding to its cost.

1,10-oxymyrcene hydroxide [10-15 minutes]

1. Open ***1,10-oxymyrcene hydroxide***. This document contains 40 conformers of the slightly larger and more flexible example (compared to *methyl 2-pentenoate* from the previous tutorial), the organic molecule: 1,10-oxymyrcene hydroxide. Conformers are ordered by energy (low to high) from the ωB97X-V/6-311+G(2df,2p)[6-311G*]//MMFF density functional model (labeled ***ref*** and in units of kJ/mol in the spreadsheet). The reference calculations took just under 3 hours wall time and roughly 17 hours CPU time to complete (on a 24-core i7 Windows 11 machine). Two additional empty columns, one labeled ***MMFF*** and the other ***corr MMFF***[7], will be utilized following calculations (below).

2. From the **Calculations** dialog, select **Equilibrium Geometry** and specify **Molecular Mechanics** and **MMFF** to right of **with**. Make certain that **Global Calculations** (toward the bottom right of the dialog) is checked and *click* on **Submit**.

3. When the job completes (less than a minute), ensure that the first molecule in the list is selected. C*lick* on **Add** at the bottom of the spreadsheet, then *click* on the **Molecule List** tab and finally from under **Energy from Molecule**, *click* on ***ΔE (kJ/mol)*** toward the top left of the dialog. Copy the results from the 40 entries in the ***ΔE (kJ/mol)*** column into the **MMFF** column.

4. Bring up the **Plots** dialog and *click* on +. Select **ref** under **X Axis**, both ***ref*** and ***MMFF*** under **Y Axes** and *click* on **Create** at the bottom right of the panel. The resulting plot shows that MMFF is rather poor at reproducing results from ωB97X-V/6-311+G(2df,2p) [6-311G*]//MMFF (the ***ref*** data).

5. Update the data using the Corrected MMFF model. From the **Calculations** dialog, select **Equilibrium Geometry** with **Molecule Mechanics** and **Corrected MMFF** and *click*

[7] Hehre, T.; Klunzinger, P. E.; Deppmeier, B.; Ohlinger, W.; Hehre, W. Practical Machine Learning Strategies I. Correcting MMFF Molecular Mechanics Model to More Accurately Provide Conformational Energy Differences in Flexible Organic Molecules. *J. Comp. Chem.* **2025**, *46* (1), 70016. https://doi.org/10.1002/jcc.70016

Submit.

6. Once completed (perhaps a minute for all 40 conformers), return to the spreadsheet. The ***ΔE (kJ/mol)*** values have updated. Copy the updated ***ΔE (kJ/mol)*** entries into the **corr MMFF** column. From the **Plots** dialog *click* on +, select **ref** under **X Axis**, both ***ref*** and ***corr MMFF*** under **Y Axes** and *click* on **Create** at the bottom right of the panel. Compare the two plots by moving back and for the between view of **Plot1** and **Plot2** (optionally, *clicking* the ◄ arrow at the top of the plots pane moves the selected plot into the 3D interface for simultaneous display. The results visually confirm the corrected MMFF model does a much better job reproducing the results of the reference calculations).

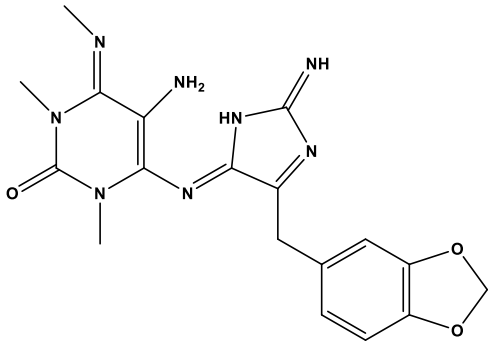
7. Return to the **Spreadsheet**. *Click* on **Add** at the bottom of the spreadsheet, then *click* on the **Linear Regression** tab. Select **ref** for the **Fit**, and choose **MMFF** under **Using**, *click* **Apply**. Create a second regression with **ref** fit to the **corr MMFF** data and *click* **Apply**. Two new entries appear at the bottom of the spreadsheet. *Right-click* on the **Fit1** cell and choose **Properties**. The **Regression Properties** dialog appears. The $\mathbf{R^2}$ value gives an indication of how well the data fit the regression, with 1 being a perfect fit. The fit of MMFF energies is essentially a scatter plot. The $\mathbf{R^2}$ for the corrected MMFF energies is quite good (.9). As (or more) important than the quality of the fit, the RMS value is much better (lower) for the corrected MMFF energies (.46 kJ/mol). From a practical perspective, the computational time savings is *significant*, a minute or less compared to several hours.

8. Close ***1,10-oxymyrcene hydroxide***.

Neural Networks in Multi-step Conformational Protocols II

In a previous example (the flexible macrocycle *aspochalasin u*) we compared the performance of the neural network routines in the Boltzmann weighted **NMR Spectrum** calculation. In the previous example, the NMR results were calculated from structures determined by the **Est. Density Functional** ***ωB97X-D/6-31G****[2] neural network. Another potentially useful strategy would be to utilize the multi-step (NMR Spectrum) routines – but to perform structures (equilibrium geometries) from quantum mechanics and obtain energies (for Boltzmann weighting) from neural network energies.

Leucosolenamine and Boltzmann Weighted NMR **[no calculations]**



leucosolenamine analogue

1. From the **neural network** folder inside the **Tutorials** directory, open the file ***leucosolenamine***. This file contains two copies of a proposed structure for the natural product leucosolenamine a. The first example has completed the default **NMR Spectrum** calculation, using MMFF and multiple QM models to provide both an equilibrium conformation and a set of Boltzmann weighted NMR chemical shifts. The second molecule has replaced the Hartree-Fock geometry step and an initial single point energy ωB97X-D/6-31G* step with a **Corrected MMFF** neural network, and the final energy calculation: ωB97X-V/6-311+G(2df,2p) with an **Est. Density Functional** neural network equivalent. Note that in this case, the final geometry ALSO comes from quantum chemical calculations, as do the NMR shifts.

2. Open the **Spreadsheet** and note that the check boxes to the left of the molecules' names in the far left-hand column are *checked,* specifying display of *both* copies. If your mouse manipulations move ***both*** copies of the leucosolenamine. Visit the **Model** menu and *toggle off* the **Coupled** entry. In this *decoupled* mode, you can move/manipulate the molecules independently. To switch focus between molecules, just *click* on the molecule you wish to select.

3. With both copies displayed simultaneously, an additional **Annotation** box is also displayed. Note, the text applies to the initial position of the examples (if you move the molecules around, the annotations don't follow). To insert an annotation box of your own, *copy* any text (either from inside of ***Spartan***, or any other program) to the clipboard and then click inside the 3D space in the ***Spartan*** UI and *right-click* then choose *paste*. To edit or expand, *right-click* on the annotation and choose **Properties**.). The initial display should include the calculation completed using quantum mechanics on the left, and that from the neural network calculation(s) on the right. Explore the NMR shifts on select carbons.

4. For qualitative results, a few chemical shift labels have been enabled. Confirm these were not "cherry-picked". Compare other carbon centers. *Right-click* on a carbon center without a displayed chemical shift label. From the resulting contextual menu, choose **Properties**. From the **Atom Properties** dialog, click to post the calculated chemical shift into the spreadsheet (this is the fourth item down named **Chem. Shift**). Note, the molecules are numbered identically and posting the value from an atom of one molecule also posts the corresponding shift value from the other copy.

5. Post the calculated **Chem. Shift** values for the remaining unlabeled carbons and return to the spreadsheet for comparison. How similar are the resulting shifts? In most cases you will find the results identical, or to vary by only a few tenths of a ppm.

6. **Align** the two molecules. With the first molecule selected, enter the **Geometry** menu and *click* on **Align**. In the lower right of the user interface, *click* on the **Align by** button (note we are aligning by **structure**). From the spreadsheet, *click* the **Add** button near the bottom left. *Click* the **Molecule List** tab, and from the bottom center of the **Add** panel, *click* on **Alignment Scores**. The alignment and NMR shift results suggest neural network calculations perform similarly to quantum chemical calculations in the context of this multi-step procedure. In this case (with relatively few conformers), the neural network is a factor of ≈3 times faster.

7. The goal (at least one goal), of this tutorial is to demonstrate the option of utilizing *only* the energy related neural networks (retaining the equilibrium geometry at density functional ωB97X-D/6-31G*). This results in NMR shifts that more closely align with those from the standard Density functional (MM and QM based) **NMR Spectrum** task, than utilizing the neural net **Est. Density Functional: ωB97X-D/6-31G*** for structure calculation. There is *still* considerable time savings in utilizing just the energy-based neural networks. Close ***leucosolenamine***.

NMR Neural Networks for proton and ^{13}C Chemical Shifts

Beginning with ***Spartan'24 v 1.3*** **(2025)** and supported in ***Spartan'26 v.1***, proton and ^{13}C shift neural network models, trained to reproduce *empirically corrected* ωB97X-D/6-31G* GIAO results. A performance assessment and discussion are included in **Appendix A**. In general, the neural networks reproduce proton shifts to within .1 ppm and ^{13}C shifts to within less than 1 ppm. The remaining tutorials in this section demonstrate these new NMR Shift options.

Kopsonoline: a rigid organic molecule [5 minutes]

1. From the **neural network** folder inside the **Tutorials** directory, open the file ***kopsonoline***. The file contains 3 copies of the (rigid) natural product kopsonoline.

2. Open the **Spreadsheet** and note, calculations on the first copy named kopsonoline QM//QM, includes NMR shift results from *empirically corrected* ωB97X-D/6-31G*//ωB97X-D/6-31G* GIAO calculations. These took roughly 470 seconds starting from an MMFF geometry. The second member in the list, named kopsonoline NN//NN includes proton and 13C shift results from the neural network models originating from the est. density function neural network (trained to reproduce geometries from ωB97X-D/6-31G*). RMS errors for 13C shifts against experiment is included for quality assessment (recall the neural network is trained to reproduce density functional GIAO shifts, not experiment). The neural network calculations (on the same machine) took less than 3 seconds (a factor of ≈175 improvement in computational time).

3. The third entry, named simply kopsonoline, is set for you to run the same neural network calculations on *your* machine. Open the **Calculations** dialog and click the **Submit** button.

4. When the job returns (seconds), revisit the **Spreadsheet**. How do your time and RMS value compare?

5. Close ***kopsonoline***.

Furoplocamioid A: a flexible organic molecule [5 minutes]

1. From the **neural network** folder inside the **Tutorials** directory, open the file ***furoplocamioid a***. Like the previous example, the file contains 3 copies of the same natural product molecule, in this case, furoplocamioid a. While only a few hundred systematic conformers, this molecule is flexible enough that properly accounting for conformation is likely to improve results. It is also hopefully flexible enough to gain an appreciation for the speed of the neural network models.

2. Open the **Spreadsheet** to review existing data. Included are results from **NMR Spectrum** (the multi-step NMR protocol) calculations using the default density functional approach (ωB97X-D/6-31G*//ωB97X-D/6-31G* GIAO calculations) as well as the est. density functional approach (neural network//neural network) have been performed. Compare calculation times and RMS errors to experiment.

3. The third entry is set up to test neural network model performance on your machine. Click on the third member from the spreadsheet and open the **Calculations** dialog. The details are already specified (**NMR Spectrum** calculation with **Est. Density Functional NMR estimator** using **Est. ωB97X-V/6-311+G(2df,2p)** energy and **Est. ωB97X-D/6-31G*** geometry). Confirm the **Global Calculations** box is not selected and click **Submit**.

4. When the job returns, revisit the **Spreadsheet**. How do your time and RMS value compare to the provided results from the quantum chemical *heavy* option (the first member in the list)?

5. Close ***furoplocamioid a***.

Machine Learning-Based Multistep NMR Protocol Applications

In previous ***Spartan*** releases we have introduced and refined QM-based NMR protocols for flexible molecules. The ***Spartan'26*** release introduces fully integrated, automated protocols for fast, accurate prediction of ^{13}C and ^{1}H NMR chemical shifts utilizing workflows comprised almost exclusively of fast neural network models. Both options, MLHF and MLXD utilize a single QM step to calibrate conformer energies which is subsequently corrected (via neural network) to the quality needed to estimate Boltzmann weights.

Evaluating NMR Assignments I - Constitution [10 minutes]

This tutorial presents general strategies for confirmation/validation of constitutional structure assignment from ^{13}C shift NMR data. We use experimental data from a recent structure confirmation study by Hashimoto and co-workers,[8] where funginocarterphenyl analogs (A-C), whose central phenyl substitution pattern was established, were explored

[8] Kanehira, R.; Maeda, H.; Tanaka, K.; Hashimoto, M. Funginocarterphenyls A–C: Thiazole-Containing *p*-Terphenyls from *Paulkirkia* sp. KT4190. *Biosci. Biotechnol. Biochem.* **2026**, https://doi.org/10.1093/bbb/zbag042

with the 2019 QM-NMR protocol.[5] In the following steps, general procedures for importing experimental ^{13}C shifts and standard **NMR Protocol** workflows are introduced.

1. From the **neural network** folder inside the **Tutorials** directory, open the file ***funginocarterphenyls A-C***. The existing file contains pre-calculated data from the best MMFF conformer, optimized with a density functional model (ωB97X-D/6-31G*), along with corrected GIAO NMR shifts for all three related scaffolds.

2. Open the **Spreadsheet**. This is pre-populated with molecule-level RMS errors (between an existing set of DFT results and experiment). Additionally, reference and solvent meta data are included. The MLHF RMS column is flagged as *pending*; entries await results from calculation. **Close** the **Spreadsheet**.

3. Open the **Enter Expt. $^{1}H/^{13}C$ shifts** tool from the **NMR** menu. This presents an atom-level spreadsheet for organizing and assigning experimental NMR shifts, as well as summary columns including space for experimental (*input*) and calculated (*output*), from ***Spartan***. The **Experimental Chemical Shifts**: will be populated, however the **Scratch Pad**: space is empty.

4. **(OPTIONAL)** Open the **Spreadsheet** and copy the reference doi URL from the **reference** column. Outside of Spartan, open a web browser (Chrome, Safari, Edge, Firefox), visit the URL. Click the PDF link to access the article. This will open a copy in your web browser; ***download*** the article to your local machine and open with Acrobat (or similar PDF reader) in order to perform the remaining steps.

5. With the article open, ***scroll down*** to Table 1; this includes *per atom* shift assignments for both ^{13}C and ^{1}H centers. Position your mouse cursor inside the table and **Copy** (capture) the shift data for the entire table: starting with **C-1** in the upper left and ending with **3.51, s** in the **lower right**. With this data on your clipboard, return to ***Spartan*** and position your mouse cursor inside the **Scratch Pad** area. ***Paste*** the copied data (via *right-click* **Paste** or **Edit** menu **Paste** or **CTRL+V**) into the **Scratch Pad**.

6. If done properly, you will immediately be presented with the **Select Character-Separated-Values-Delimiter** prompt. Choose the **Delimiter Character** that provides the most accessible parsing of the data from the table (typically, but not always, **Space** delimited) and *click* **OK**.

7. Review the overall format of the data: atom labels, carbon types, shift data, and column structure. Shifts can be individually dragged and dropped, but the recommended procedure is to edit the scratch pad data to arrange an entire column of assigned shifts in proper order and then to **copy/paste** to the appropriate ^{13}C column under **Experimental Chemical Shifts**. In general, experimental shifts can be imported and copied as a group. **Note:** As the data format (in this case) combines cells for equivalent carbons (example C-2′/C-6′), some individual hand-editing may still be required.

8. A remaining challenge for proper assignment is confirming that a molecule's atom labels are either identical to or compatible with the atom labeling in the paper (and/or from the instrument or other formatted source). No relabeling is needed in this case, however: should

atom relabeling be necessary, this is *most easily* accomplished from the **2D** view and with **Enter Expt. ^{1}H/^{13}C shifts** selected from the **NMR** menu. With the atom table moved to one side, NMR Atom Labels appear in ***blue*** and can be overwritten either form the **Experimental Chemical Shifts** spreadsheet, or by *clicking* on the atom label itself and then clicking on the corresponding **Edit** dialog that appears.

9. **(END OPTIONAL)** Once you've reviewed the experimental shifts and familiarized yourself with the scratch pad and both shift and atom label editing, open the **Calculations…** dialog. Choose: **NMR Protocol** as the task and leave the default **2026 MLHF** option selected. This specifies the MLHF-NMR protocol.[9] Make sure the **Global Calculations** box is checked and click **Submit**.

10. The **NMR protocol** is specifically designed to assess flexible systems. While the sample molecules in *this* tutorial have only modest single-bond rotation flexibility, they serve as introduction to the mechanics of experimental NMR shift import and general NMR protocol workflow options. The calculation should require no more than a minute or so per analog, even on slower machines. Based on presented structure and default search definitions, ***Spartan*** will attempt a different number of systematic conformers for each funginocarterphenyl analog. As noted, these systems have minimal flexibility, none has more than 36 systematic conformers.

11. Once completed, you'll be prompted with the option to **Open**, **Save**, or **Discard** the corresponding conformer files. In this case, confirm the **Apply To All** check box is selected, and choose **Discard**. In the case of remote submission to an external ***Spartan*** resource, an initial prompt to **Retrieve** is presented. Ensure the **Apply To All** box is checked, and select **No**. If you select **Yes**, a second prompt gives you the option to **Open** or **Save**.

12. Compare the results from the original DFT calculations with the newly populated column (MLHF RMS). Optionally choose **Configure** from the **Model** menu and enable the Δ (Calc. – Expt.) for atom labels. Also from the **Model** menu, optionally turn off **Hydrogens** and change model style to **Tube** (this provides a less cluttered presentation for ease of view). Considering that the initial DFT calculations required more than 3 hours on a reasonably configured laptop, the MLHF performs a similar assessment and provides comparable (potentially "slightly" better results) in less than a minute per molecule.

13. Close ***funginocarterphenyls A-C***.

---

[9] Ohlinger, W.; Klunzinger, P.; Hehre, T.; Driessen, A.; Deppmeier, B.; Johnson, J.; Ohlinger, C.; Hehre, W.; Schnitker, J.; Uchida, N.; Watanabe, Y.; Cohen, R. D.; Bruno, I. J.; Hashimoto, M.; Kanehara, R.; Herzon, S. B.; Riseman, H. J.; López-Pérez, J. L. Practical Machine Learning Strategies. 5. Rapid, Accurate Prediction of ^{1}H and ^{13}C NMR Shifts in Conformationally Flexible Organic Molecules. *ChemRxiv* **2026**, *Preprint.* (Manuscript submitted for publication) https://doi.org/10.26434/chemrxiv.15001441/v1

NMR protocol options are detailed in the essay ***NMR Protocols for Flexible Organic Molecules***, available in **the User's Guide** (**Help** menu) and from the **Topics** collection (**Activities** menu).

Timing comparisons between default NMR Protocols [10 minutes]

When discussing analysis or processing speeds, in general, *faster* is better. It follows, *considerably faster*–as long as accuracy is maintained–can be transformative. This tutorial demonstrates timing benchmarks leveraging the same molecules used to explore timing across multiple hardware and OS configurations in the defining MLHF and MLXD paper.[9] The 4 standard NMR protocol option times are provided for a similar OS/hardware range for comparison to *your* computational environment.

1. From the **neural network** folder inside the **Tutorials** directory, open the file ***NMR Protocol Timings***. Open the **Spreadsheet**. The spreadsheet is pre-populated with molecule-level RMS error and average timing data from: **2019 QM**, **2024 QM**, and **2025 MLXD** approaches across the range of hardware below, a recent sample from a . Additional columns for the **2026 MLHF** protocol are marked *pending*, awaiting calculation results from *your* machine.

Benchmarking Hardware	
a	3.2 GHz Apple M1™ chip 2020
b	4.5 GHz Apple M4™ chip 2024
c	4.5 GHz Apple M4 Pro™ chip 2024
d	4.5 GHz Apple M5 Pro™ chip 2025
e	4.4 GHz Intel® Xeon® W chip 2019
f	4.8 GHz Intel® i7-10700 chip 2020
g	4.9 GHz Intel® i7-12700 chip 2022
h	5.4 GHz Intel® i9-13900H chip 2023
i	5.4 GHz Intel® i9-13900HX chip 2024
j	5.7 GHz AMD™ Ryzen 9™ 9950X chip 2025
k	5.2 GHz Intel® i9-10850K chip 2020
l	4.5 GHz AMD™ Ryzen 9™ 7950X chip 2022
m	4.5 GHz AMD™ Ryzen Threadripper™ 5995WX chip 2022

2. The spreadsheet includes a timing analysis and experimental data from two natural products, the terpenoid, abundisporin G[10] and the alkaloid, uncarialine D.[11] Experimental ^{13}C shifts have *already* been imported and assigned, allowing for statistical comparisons between predicted and observed shifts. Close the **Spreadsheet**.

3. Open the **Calculations…** dialog. Confirm the task is set to **NMR Protocol** with **2026 MLHF**. Ensure the **Global Calculations** check box is checked and *click* **Submit**.

4. When the calculations complete (a minute or less per natural product on a well-configured machine) return to the **Spreadsheet**. Compare your machine's MLHF protocol performance against the Sample MLHF Time column. How well does your timing compare? NMR accuracy is typically reproducible down to the +/- 1–2 hundredths ppm level, across hardware. How well do your MLHF results (use RMS error) compare with the Sample MLHF RMS results? Review results from the other NMR protocol options (both QM and

[10] Sum, W. C.; Ebada, S. S.; Kirchenwitz, M. et al. Neurite Outgrowth-Inducing Drimane-Type Sesquiterpenoids Isolated from Cultures of the Polypore Abundisporus violaceus MUCL 56355. *J. Nat. Prod.* **2023**, *86*, (11), 2457–2467. https://doi.org/10.1021/acs.jnatprod.3c00525

[11] Huang, KP.; Xu, L. L.; Li, S. et al. Uncarialines A-E, new alkaloids from Uncaria rhynchophylla and their anticoagulant activity. *Nat. Prod. Bioprospect.* **2023**, *13*, 13. https://doi.org/10.1007/s13659-023-00377-0

ML-based). Timings provided for the remaining NMR protocols are the ***averages*** from the above **Benchmarking Hardware** table (14 hardware-configurations, a–m).

5. When finished, close both the **Spreadsheet** and the file ***NMR Protocol Timings***.

6. For those interested in additional benchmarking, reach out to support@wavefun.com. An additional file is accessible through Help content, supporting a parallel performance analysis for QM calculations, up to 16-cores (larger configuration test files available from our **Support Team**).

Evaluating NMR Assignments II – Stereochemistry [15 minutes]

For practical purposes, stereochemical assignment is a *special case* of structural assignment. It is assumed that the overall agreement between predicted NMR shifts (typically ^{13}C) and experiment is sufficient for confirmation of constitution (with RMSE typically 2.2 ppm or less). A discussion on predicted NMR performance and practical confidence ranges are deferred to the **Topics** section (validation of this approach was recently established in work carried out by Haniff and co-workers (insert IMNP paper reference).

1. From the **neural network** folder inside the **Tutorials** directory, open the file ***bipoladisin F***. This file contains a set of four diastereomers from one of a series of recently isolated sesterterpenoids.[12] While ***Spartan'26*** includes the ability (given suer-specification of stereo-center) to generate a list possible diastereomers, this step has already been completed. Demonstration of the **Generate Isomer(s)** tool is given in **Chapters 9**, **11**, and **13**.

2. Open the **Spreadsheet**. The diastereomers are named including chiral variations at carbons 4, 5, and 11 (the entire group is systematically labeled based on the original isomer from which it is generated). Reference and solvent meta data appear in the spreadsheet along with a **Spartan pre-calc** column. This column has been formatted with lower precision purposefully; we will revisit this following the calculation.

3. Open the **Calculations…** dialog, confirm this is set to complete the **NMR protocol** task with the **2026 MLHF** workflow. Confirm the **Global Calculations** checkbox is checked. **Submit** the job. It should take no more than a minute or so on a reasonably configured machine. Upon completion, you may either **Save** or **Discard** the auxiliary conformer files (these files consist of the conformers retained from each potential diastereomer, often worth review but not specifically used in this tutorial).

4. Return to the **Spreadsheet**. The first two columns will populate with appropriate RMS error and DP4 probability (%). Which diastereomer best matches experiment? Reminder, RMSE, MAE and MaxAE along with DP4 % (where applicable) are available from the **NMR** tab in the **Molecule Properties** dialogue (**Display** menu or *right-click* the background

[12] Shen, Y.; Li, Q.; Ding, N.; Yu, M. et al. Bipoladisins A–G, Uncommon Sesterterpenoids with Three Types of Carbon Skeletons from Bipolaris maydis and Their Reversal Activity on Paclitaxel Resistance. *J. Nat. Prod.* **2025**, *88*, (6), 1286–1297. https://doi.org/10.1021/acs.jnatprod.5c00024

to access), and can be posted to the spreadsheet via the post button (down arrow icon) to the left of each metric in the **NMR** tab.

5. From the **Spreadsheet** *click* on the previously mentioned **Spartan pre-calc** column. *Right-click* on the header cell and choose **Format Selected…** In the resulting dialog, set the formatting to **Decimal** and choose **2** for **Decimal places** (rather than zero). Now compare these results to your own. Are they consistent?

6. Close ***bipoladisin F***.

Section II
Operations and Overview

The two chapters that make up this section provide an overview of the components which control access to ***Spartan***'s graphical user interface (**Operating Spartan**) and following that, a *guided tour* through some of the more fundamental operations of the interface (**Walking Through Spartan**). Together, these offer sufficient information to begin use of ***Spartan***. Description of the full range of ***Spartan***'s capabilities is deferred to **Section V** which is intended to serve as a reference to the current ***Spartan'26*** feature set. A definitive ***Spartan*** application article is in revision and planned for publication in *late* 2026.

Chapter 1
Operating *Spartan*

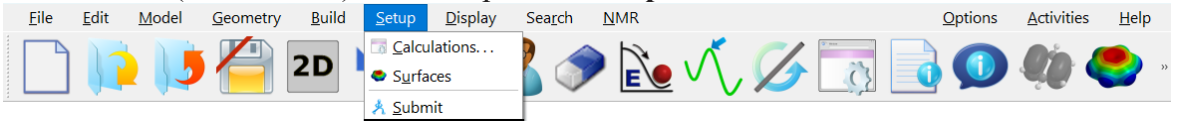
*This chapter describes the general operating features of **Spartan'26**. It should be perused prior to starting the tutorials.*

Starting and Quitting *Spartan*

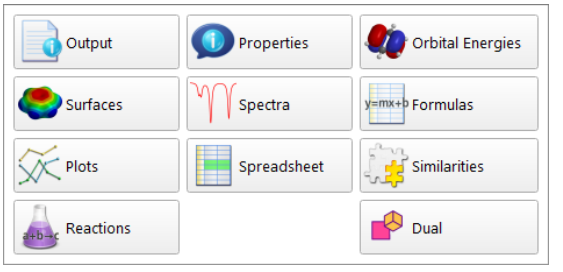
To start ***Spartan*** under Windows, *click* on the **Windows** button, then *click* on **All Programs** and locate and open ***Spartan'26*** (or *double click* on the ***Spartan*** shortcut icon on the desktop). To start under Macintosh, *double click* on the ***Spartan'26*** icon in the Applications list. To open under Linux, bring up a terminal and *type* **spartan26**. To quit ***Spartan***, select **Exit** from the **File** menu (Windows) or **Quit Spartan'26** from the **Spartan'26** menu (Macintosh), or *click* the **Close** button toward the top right of the ***Spartan*** interface, Windows and Linux, (top left for Mac).

Menus and Icons

Program functions may be accessed either from the menu bar or from icons in the toolbar which is directly underneath the menu bar. The menu bar may either be accessed as pull-down menus (**Classic List**), for example, the **Setup** menu:



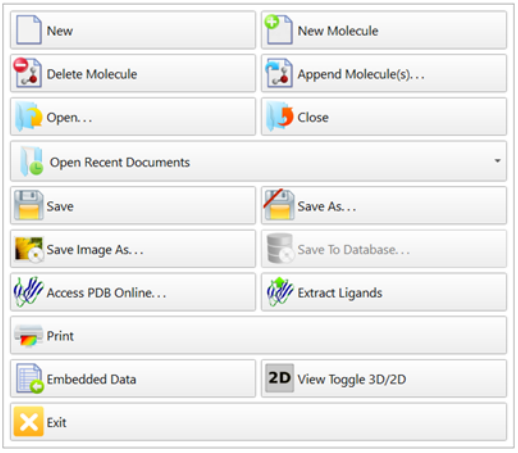
or from a list of icons presented in a palette (**Button Pad**), for example, the **Display** menu:



Selection is made in the **Settings** tab of the **Preferences** dialog (**Preferences...** under the **Options** menu; **Chapter 25**).

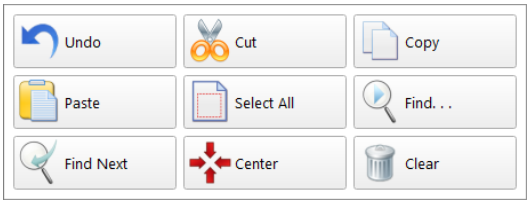
Icons for all menu functions (as shown alongside text in both **Classic List** and **Button Pad** styles) are available on screen below the menus. There are too many icons to be simultaneously displayed and the choice (beyond the default initially configured) is made in the **Icons** tab (**Preferences...** under the **Options** menu; **Chapter 25**). Icon size is also user selectable in the **Settings** tab (**Preferences...** under the **Options** menu; **Chapter 25**).

File



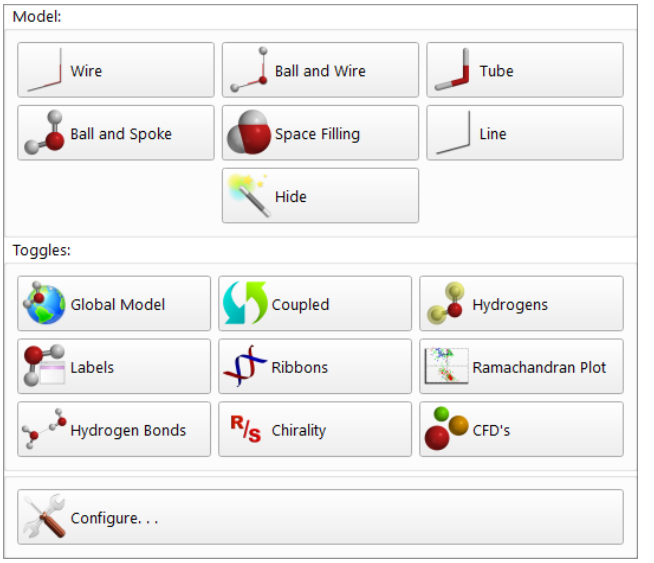
Allows you to build or sketch a new molecule, add or delete a molecule from an existing ***Spartan*** document, read in a molecule that you have previously saved, or a file created by another program. It also allows you to print text and graphics, embed external files (for example, Word and Excel files) into ***Spartan*** documents, to access PDB files online and to extract ligands from these files. A Save to Database option is a one stop save to database (SSPD) operation. And new in ***Spartan'26*** 2D/3D View Toggle.

Edit



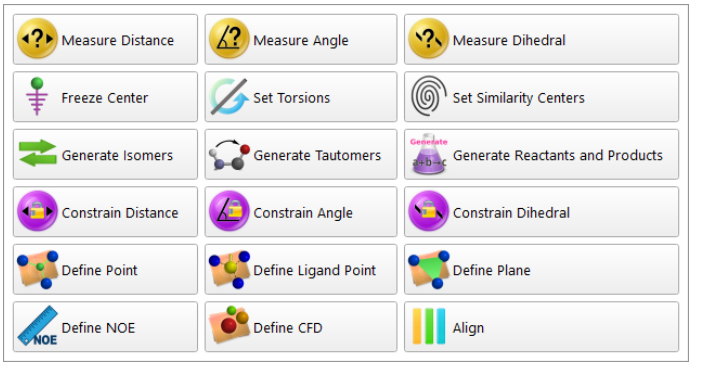
Allows you to transfer information to and from the clipboard, to undo the last operation, to find text strings and molecule fragments, to center molecules on screen, and to clear the active molecule by deleting it.

Model



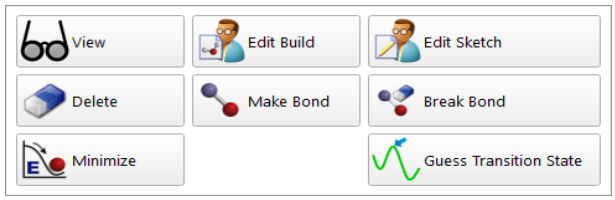
Provides control of model style of your model, display of (R/S) chirality labels, hydrogen bonds and chemical function descriptors, as well as coupling or decoupling mouse function for molecules in a multi-molecule document. Allows display of Ramachandran plot for a protein structures brought in from PDB. Enables labels, including atom labels by atomic charges or experimental or calculated NMR chemical shifts (or the difference between them) among other quantities (custom label displays also available).

Geometry



Allows you to measure and constrain bond lengths, angles and dihedrals, make lists of stereo and regioisomers and tautomers, define points, ligand points and planes, define NOEs, specify frozen atoms, alter default settings for conformational degrees of freedom, select centers for similarity analysis and align molecules.

Build

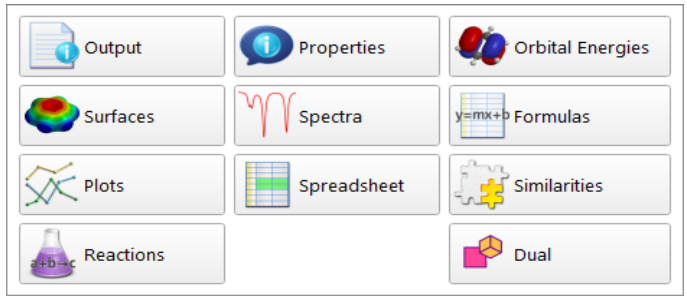


Allows you to sketch molecules in 2D or build them in 3D, and to edit molecules that you have already sketched or built. Provides tools for making and breaking bonds, deleting atoms or groups of atoms in 3D builder as well as both 2D and 3D minimization tools (MMFF) for "cleaning up" 2D drawings and preparing 3D structures for calculation. Allows seamless access to ChemDraw™ (Windows only). Allows you to define transition states in terms of curved (reaction) arrows attached to reactants (or products) in either 2D or 3D, and to guess a transition-state geometry based on a library of reactions. Permits experimental proton and ^{13}C chemical shifts as well as observed HH and CH couplings to be associated with 2D drawings or 3D structures.

Setup

Allows you to specify and submit calculations and graphical models, either locally or to a remote resource via ***Spartan's*** submission tools.

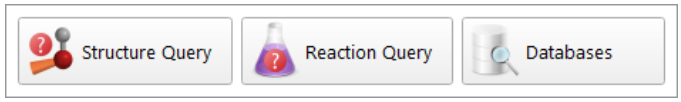
Display



Allows you to display text output, molecular and atomic properties, QSAR descriptors, thermodynamic quantities, orbital energy diagrams, surfaces and property maps and infrared, Raman, NMR and UV/visible spectra, as well as to access experimental IR, NMR and UV/visible spectra from freely available databases on the internet or from local files. ***Spartan'26*** further enhances the utility of the **Summary Output** providing a series of tabbed tables containing spectroscopic data that may be copied or printed. Allows you to present data in a spreadsheet and make plots from and perform regression analysis on these data, to access the results of a similarity analysis and to compute reaction energies based

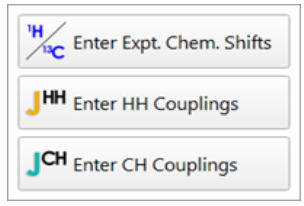
either on user data or from entries in the SSPD (Spartan Spectra and Properties Database). New in ***Spartan'26*** is the dual display option providing a manipulable copy of the user's molecule in 2D (while the UI is in 3D view mode) or 3D (when in the 2D Sketch mode) to complement the existing visual mode.

Search



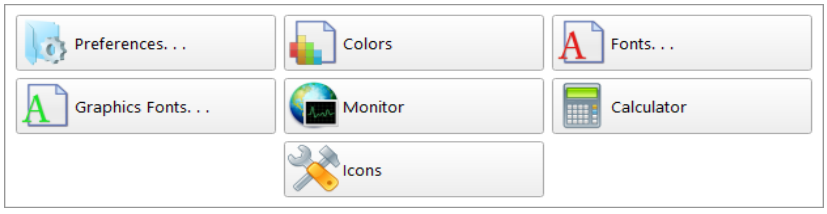
Allows database queries, searching and mining of the SSPD for calculated structures, properties and spectra, and searching the SRD for calculated transition-state structures. Allows you to search the CSD (Cambridge Structural Database)* and to extract ligands from the PDB (Protein Data Bank). Allows you to match an unknown infrared spectrum to calculated spectra in the SIRD (Spartan Infrared Database) which is derived from the EDF2/6-31G* entries in SSPD or to experimental spectra from the XIRD (NIST database). Allows you to guess a transition-state geometry based on a library of reactions.

NMR



Allows you to attach experimental proton and ^{13}C chemical shifts to a 3D structure, and to specify observed HH and CH couplings.

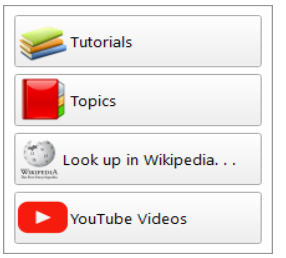
Options



Allows one to set display standards, specify databases, monitor executing jobs, and customize icons and menu display and other aspects of the graphical interface. Allows one to access external ***Spartan*** computational servers or configure the local machine to accept remotely submitted jobs.

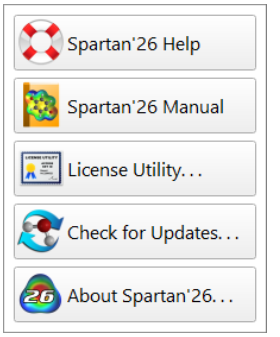
* Requires CSD to be licensed. Visit https://www.ccdc.cam.ac.uk/solutions/csd-licence/ for info.

Activities



Allows you to display tutorials and topics inside of ***Spartan*** and to access our YouTube videos and search Wikipedia.

Help



Provides access to information on ***Spartan's*** general operation, the ***Spartan'26 Tutorial and User's Guide*** (current version of this document), and a number of computational FAQ's. Also provides access to the program's **License Utility**. This allows for modification to the license, activation, extending maintenance, transferring to a new machine, and accessing the license server to check if license update is available to apply. New in ***Spartan'26*** is a utility allowing users to **Check for Updates…** This will compare the user's version with the current release available and offer a link to update if a newer version exists (note: while this alerts of a new version, the user must choose to download/update, we do not do this automatically).

A complete listing of menu functions is provided in **Appendix C**.

Additional Icons

A variety of other icons appear in ***Spartan***, both in individual dialogs and in the message bar at the bottom of the screen.

Post to Spreadsheet

Search Transition State Library

Lock/Unlock Constraints

Move Up/Down Dialog

Server Shortcut

Help/Tool Tips

Surface clipping planes

Play

Pause

Revert to Fullscreen

Step

Stop

Extend to Fullscreen

Restore Default Settings

Look Up in Wikipedia

Add or Remove Tables

Make List

Fit

Tabs

Spartan'26 assigns a tab to each open document. These appear in a single row along the bottom of the screen in the order that the documents were created or opened. Backward and forward step keys (and) at the far right provide access to tabs outside of those displayed. On the left of each tab is a check box. If checked, this indicates that the document will be (permanently) displayed on screen, even if it is not the selected document.

Mouse/Keyboard Operations

The following functions are associated with a standard mouse and keyboard. Acknowledged, depending on platform there is reasonable variability in mice and mice configuration. In general, we try to take advantage of the full range of modern mouse features but accommodate older mice (or no mice – in the case of touch pad or touch-enabled devices). The table below is also available from within the ***Spartan*** GUI from the **Help** menu > ***Spartan'26*** **Help** entry.

Using the Mouse

The following functions are associated with the two-button mouse.

keyboard	left button[a]	right button[a]
-	selection, X/Y rotate	X/Y translate
Shift	range selection, Z rotate	scaling
Control	multiple selection, global X/Y rotate[b]	global X/Y translate
Control + Shift	global Z rotate[b]	scaling
Alt	group picking	...
Alt ("View" mode)	group selection [c]	...
Alt ("Build" mode)	bond rotate	bond stretching
Control ("Build" mode)	fragment X/Y rotate	fragment X/Y translate
Control + Shift ("Build" mode)	fragment Z rotate	scaling[b]
	Double-click left button	**Mouse Wheel**
-		Zooming, scroll up/down
-Bonding mode	substitution, insertion	...
-Bonding mode	select atom [d]	...

a) Left and right buttons together with no modifier keys are used for defining a section box.
b) In this context, 'Global' means all visible molecules.
c) With both left and right mouse buttons selected a selection box for multiple object selection is available.
d) Select atom as opposed to selecting an open valence.

Mouse/keyboard operations may be broadly separated into two categories: selection (picking) and manipulation (translation/rotation).

These broadly fall into two categories: selection (picking) and manipulation (translation/rotation/zooming).

Selection. *Clicking* (left button) selects objects on screen and/or of menu items. Center wheel or both left and right buttons together are used to define a selection box for copying to the clipboard, as well as for group selection. Together with the **Shift** key, the left button allows for selection over a range. Together with the **Ctrl** (**Control**) key, the left button allows for multiple selection. Both range and multiple item selection apply not only to text items in lists, but to atoms and bonds in molecules as well. Together with the **Alt**/**Option** key), the left button allows for selection of an entire group (detached molecular fragment).

In (3D) **Build** mode, *double clicking* (left button) on an atom exchanges it with the atom or atomic fragment selected in the model kit. *Double clicking* on an atom while holding down the **Ctrl** (Win or Linux) or the **Command** key (Mac) leads to inversion in chirality of the atom and *double clicking* on the molecule while holding down both the **Ctrl** (Win or Linux) /**Command** key (Mac) and **Shift** key, will invert the absolute configuration of the molecule. In (2D) sketch mode, *double clicking* on an atom with another atom selected in the sketch pane replaces it with the selected atom. *Double clicking* on an atom with charge/radical selected in the sketch pad adds charge or radical designation to the atom and adjusts hydrogen count. Unless a ring icon is selected from the sketch palette, *double clicking* on a bond increases the bond order by 1 (single to double or double to triple). *Double clicking* on a triple bond replaces it with a single bond. Once an initial fragment, group or ring has been drawn, *double clicking* on the

background in either 2D or 3D mode will insert the selected item alongside the existing structure.

Manipulation. The left button is used for rotation, and the right button is used for translation and scaling in ***Spartan's*** main screen. With no keys depressed, moving the mouse while holding down the left button gives rise to rotation about the X and y (screen) axes, while moving the mouse while holding down the right button gives rise to translation in the x and y (screen) directions. Together with the **Shift** key, moving the mouse while holding down the left button gives rise to rotation about the z direction, while moving the mouse while holding down the right button gives rise to scaling. The center (scroll) wheel on the mouse may also be used for scaling.

In **Edit Build** mode (only), the default is focus on the full set of fragments that make up the molecule being constructed, and rotations and translations refer to this set of fragments as a whole. Use of the **Ctrl** key changes focus to a single fragment (the selected fragment), and rotations and translations now refer only to this fragment.

In **Edit Build** mode (only), moving the mouse while holding down the **Alt/Option** key and left mouse button rotates about the selected bond. Bond rotation may also be accomplished by moving the mouse up and down inside the marked area at the left of the screen while holding down the left button.

Additional keys control various ***Spartan*** functions.

Keyboard Operations

Page Up, Page Down Home, End	Moves up (**Page Up**), down (**Page Down**), to the top (**Home**) and to the bottom (**End**) of the set of open molecules. Also, moves up and down pages in the **Output** dialog.
Insert (**option** for Mac)	In **Edit Build**/**Edit Sketch** fragment mode only, inserts a new fragment on screen. This is accomplished by selecting the fragment from the model kit, holding down the **Insert** key and *clicking* on screen. Insertion may also be accomplished by *double clicking* on the background following selection of a fragment.
Delete	Deletes a fragment, free valence, CFD, reaction arrow, the contents of a selection box, spectrum, curve, or plot. This is accomplished by holding down the **Delete** key and *clicking* on the fragment, etc. Required following text or data entry into spreadsheet or dialogs.

Touch-Pad/Track-Pad Operations

Standard Pad	
Rotate	One-finger press and rotate
Translate	Two-finger press and translate
Zoom (translate in the Z pane)	Two-finger rest (move fingers up and down)
Rotate in the Z plane	Hold **Shift** key down, One-finger press up or down.
Rotate around bond	In **New Build** or **Edit Build**, press on a bond to select it. A red arrow indicates selected bond. Move the cursor over vertical strip on the left of the interface and then *one-finger* press and move the cursor up or down.
Two Button Pad	
Rotate	One-finger left-press and rotate
Translate	One-finger right-press and translate
Zoom (translate in the Z pane)	Two-finger rest (move fingers up and down)
Rotate in the Z plane	Hold **Shift** key down, one-finger left press (up or down)
Rotate around bond	In **New Build** or **Edit Build**, left-press on a bond to select it. Ared arrow will indicate selected bond. Move the cursor over vertical strip on the left-hand side of the interface and then left-press and move the cursor up or down

Touch-Screen Operations

Tapping = clicking and *double tapping = double clicking*. Range and multiple selection have not yet been implemented. One finger motions on screen are equivalent to left button motions. Two finger motions are equivalent to right button motions. Pinching is equivalent to scroll wheel operations.

Selecting Molecules

While two or more molecules may be simultaneously displayed, only one molecule may be selected. Only the selected molecule has access to all capabilities. Molecule selection occurs by *clicking* on its structure model or on any of its associated graphical surfaces.

Where the molecule belongs to a document with more than a single molecule, selection from among the different molecules may be made using the buttons or the scroll bar at the bottom left of the screen. *Clicking* on step keys at the bottom left of the screen animates the display of molecules in the document, that is , steps through them sequentially. Animation speed is controlled from the **Settings** tab (**Preferences…** under **Options** menu; **Chapter 25**).

Clicking on ⏸ stops the animation. If the spreadsheet associated with the document is open (**Spreadsheet** under the **Display** menu; **Chapter 22**), selection can also be made by *clicking* on the molecule label at the left of the spreadsheet.

Two or more molecules from the same document may be displayed at once (although only one may be selected). Molecules are marked for display by *checking* the box immediately to the left of the molecule label in the spreadsheet.

Stereo Displays

Spartan supports red-cyan stereo. Red/blue glasses must be worn. To enter stereo mode, *access* the Preferences dialog, **Options** menu. The **Settings** tab contains an **On/Off** selection for **Stereo** mode. Does not apply to the 2D sketcher.

Changing Colors and Setting Preferences

Colors and **Preferences...** under the **Options** menu (**Spartan'26** menu, **Macintosh**) allows for changing default background and graphical object colors, and for setting (and resetting) program preference defaults, respectively; **Chapter 25**. Further control of color is available in the various **Style** dialogs accessible from **Properties** under the **Display** menu; **Chapter 22**.

Monitoring and Terminating Jobs

The **Monitor** lists jobs that are either queued or executing on the local machine or on a remote server (**Monitor**). Queued jobs may be started and executing jobs may be terminated. The current text output and molecular structure of executing jobs may be examined. The **Monitor** is accessible from the **Options** menu (**Chapter 25**).

Chapter 2

Walking Through ***Spartan***

This chapter, in the form of a tutorial, introduces a number of basic operations in ***Spartan*** *required for molecule manipulation, property query and spectra and graphics display. This chapter is requisite for new* ***Spartan*** *users. It shows how to: i) open molecules, ii) view different model styles and manipulate molecules on screen, iii) measure bond distances, angles and dihedral angles, iv) display energies, dipole moments, atomic charges and infrared, VCD, NMR and UV/visible spectra, and v) display graphical surfaces and property maps, and via the spreadsheet, display multiple molecules from the same document. Other spreadsheet operations are not illustrated, no molecules are sketched or built, and no quantum chemical calculations are performed.* The final example illustrates one of the neural networks incorporated in* ***Spartan'26****.*

1. Start ***Spartan***. *Click* (left mouse button) on **File** from the menu bar that appears at the top of ***Spartan's*** main window. *Click* on **Open...** from the **File** menu that appears. Alternatively, *click* on the icon at the top of the screen. A file browser appears.

Move to the ***Tutorials*** directory** and *open* ***Walking Through Spartan*** (**File** menu **Open** ***Walking Through Spartan***).

> *Tap* on ***Walking Through Spartan*** and *tap* on **Open** (or *double tap* on ***Walking Through Spartan***). Note that *click* (left mouse button) and *tap* (one finger) are equivalent as are *double click* and *double tap*. We shall only indicate *click*/*double click* throughout the text that follows.

* The results of quantum chemical calculations in this chapter have been obtained using the ωB97X-D/6-31G* density functional model.

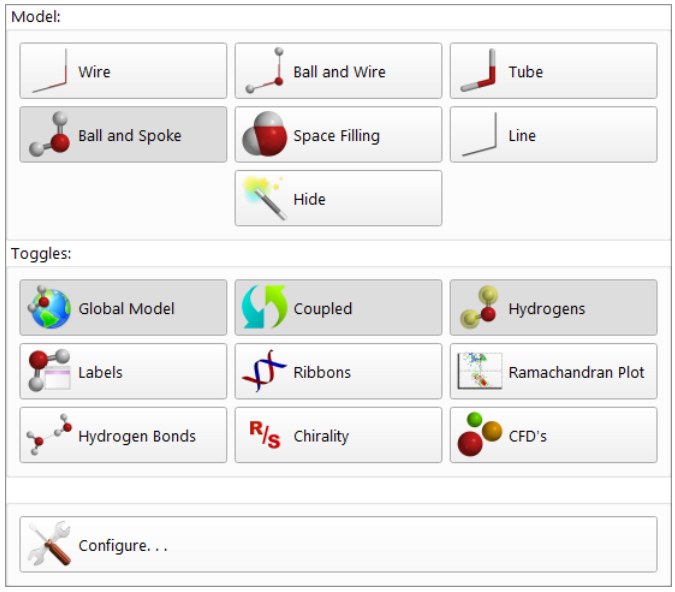
** **For Windows**, the ***Tutorials*** directory is found in ***Program Files/Wavefunction/Spartan26***. It needs to be copied to another location available to the user (we recommend Documents or Desktop) prior to opening it in ***Spartan***.
For Macintosh, this is located at the top of the ***Spartan26*** disc image. It needs to be copied to another location available to the user (we recommend Documents or Desktop).
For Linux, the ***Tutorials*** directory is found in the install directory. Copy the ***Tutorials*** directory to a location that allows write permission, typically the user's home directory.

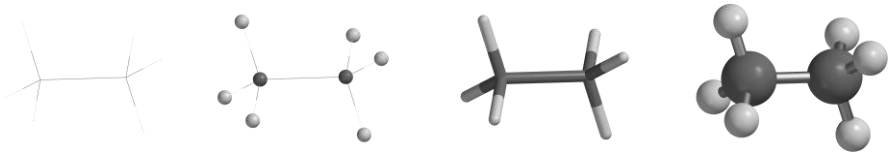
A single file containing ***ethane***, ***acetic acid dimer***, ***propene***, ***ammonia***, ***hydrogen peroxide***, ***acetic acid***, ***water***, ***cyclohexanone***, ***camphor (RR)***, ***epicamphor (SS)***, ***camphor***, ***3-benzoylaniline***, ***ethylene***, ***benzene***, ***aniline***, ***cyclohexanone***, and two copies of the natural product ***camalexin*** (3-(1,3-thiazol-2-yl)-1H-indole) will be opened. A ball-and-spoke model for the first molecule (***ethane***) will be displayed, and its name appears at the bottom right of the screen. The appearance of the name indicates that the molecule is included in the Spartan Spectra and Properties Database (SSPD).

2. Practice rotating (*move* the mouse while holding down the left button) and translating (*move* the mouse while holding down the right button). Use the scroll wheel to zoom in and out, alternately move the mouse (up or down) while holding down both the right-mouse button and the **Shift** key.

> To rotate, move one finger across the screen. To translate, move two fingers across the screen. To zoom out, pinch two fingers together. To zoom in, move two fingers apart.

Click on **Model** from the menu bar.



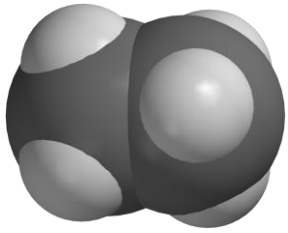


Wire Ball-and-Wire Tube Ball-and-Spoke

One after another, select **Wire**, **Ball and Wire**, **Tube** and finally **Ball and Spoke** from the **Model** menu. All four models for ***ethane*** show essentially the same information. The wire model looks the most like a conventional line formula. It uses color to distinguish different atoms, and one, two and three lines between atoms to indicate single, double and triple bonds, respectively.

The ball-and-wire model is identical to the wire model, except that atom positions are represented by small colored spheres, making it easy to identify atom locations. The tube model is identical to the wire model, except that bonds are represented by solid cylinders. The tube model is better than the wire model in conveying three-dimensional shape. The ball-and-spoke model is a variation on the tube model; atom positions are represented by colored spheres, making it easy to see atom locations.

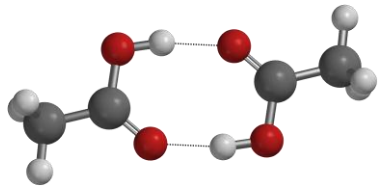
Select **Space Filling** from the **Model** menu.



Space-Filling

The space-filling model is different from the other model styles in that bonds are not shown. Rather, each atom is displayed as a colored sphere that represents its approximate relative size. The space-filling model provides a measure of a molecule's size. While lines between atoms are not drawn, the existence (or absence) of bonds can be inferred from the extent to which spheres on neighboring atoms overlap. If two spheres substantially overlap, then the atoms are almost certainly bonded, and conversely, if two spheres barely overlap, then the atoms are not bonded. Intermediate overlaps suggest weak bonding, for example, hydrogen bonding.

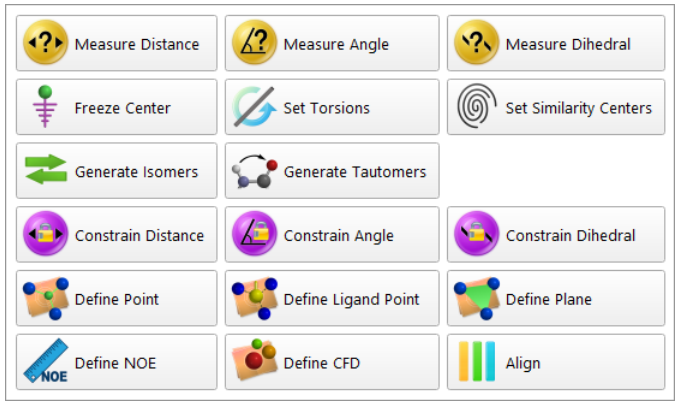
3. *Click* once on the right arrow key at the bottom left of the screen. This will move to the next molecule in the document, ***acetic acid dimer***. Its name will appear at the bottom of the screen. use the backward or forward step keys to get to ***acetic acid dimer*** in the document. From the space-filling model, look for overlap between the (OH) hydrogen on one acetic acid molecule and the (carbonyl) oxygen on the other. Return to a ball-and-spoke model. *Click* on the **Model** menu and select **Hydrogen Bonds**.



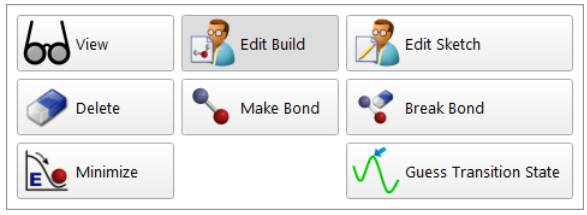
Ball-and-Spoke model for acetic acid dimer with hydrogen bonds displayed

The two hydrogen bonds, that are responsible for the *pose* or *orientation* of the most energetically favorable dimer complex, will be drawn.

4. Distances, angles, and dihedral angles can easily be measured with ***Spartan*** using **Measure Distance**, **Measure Angle**, and **Measure Dihedral**, respectively, from the **Geometry** menu.



a) **Measure Distance**: This measures the distance between two atoms. *Click* once on ►to move to the next molecule, ***propene***. *Click* on the **Geometry** menu and select **Measure Distance** (or *click* on the icon if it appears at the top of the screen). *Click* on a bond or on two atoms (the atoms do not need to be bonded). The distance (in Ångstroms) will be displayed at the bottom of the screen. Repeat the process for different bonds or pairs of atoms. When you are finished, select **View** from the **Build** menu (or *click* on the icon at the top of the screen).

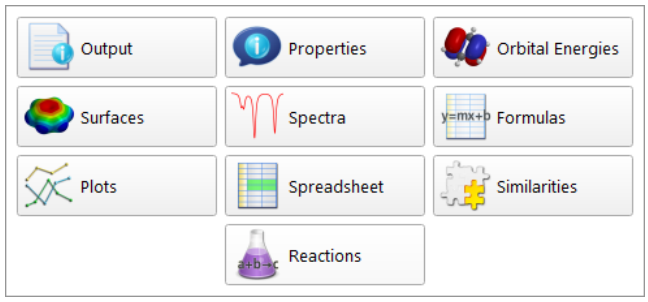


b) **Measure Angle**: This measures the angle around a central atom. *Click* once on ►to move to the next molecule, ***ammonia***. *Click* on the **Geometry** menu and select **Measure Angle** (or *click* on the icon if it appears at the top of the screen). *Click* first on H, then on N, then on another H. Alternatively, *click* on two NH bonds. The HNH angle (in degrees) will be displayed at the bottom of the screen. *Click* on when you are finished.

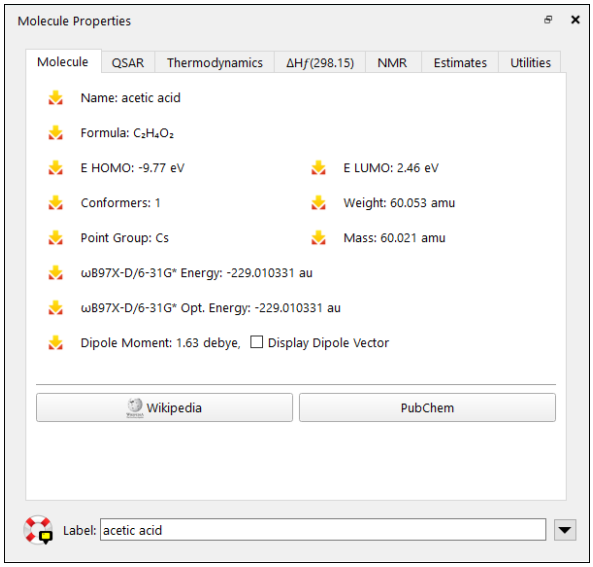
c) **Measure Dihedral**: This measures the angle formed by two intersecting planes,

one containing the first three atoms selected and the other containing the last three atoms selected. *Click* once on ►to move to the next molecule, ***hydrogen peroxide***. *Click* on the **Geometry** menu and select **Measure Dihedral** (or *click* on the icon if it appears at the top of the screen) and then *click* in turn on the four atoms (HOOH) that make up hydrogen peroxide. The HOOH dihedral angle will be displayed at the bottom of the screen. *Click* on when you are finished.

5. Energies, dipole moments and atomic charges (among other calculated properties) are available from **Properties** under the **Display** menu.



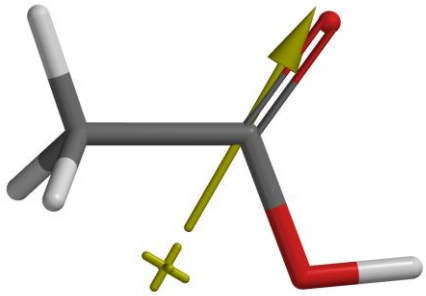
a) **Energy:** *Click* once on ►to move to the next molecule, ***acetic acid***. *Click* on the **Display** menu and select **Properties** (or *click* on icon if it appears at the top of the screen). The **Molecule Properties** dialog appears. It is divided into six parts designated by tabs. **Molecule** provides the energy and other information relating to the isolated molecule, **QSAR** provides quantities that may be used as QSAR descriptors, and **Thermodynamics** provides the entropy, enthalpy, Gibbs energy, zero-point energy and heat capacity. **ΔH*f* (298.15)** provides heat of formation data from NIST (if available) and the SSPD (if available) from the T1 model, and an estimated ΔH*f* from DFT: ωB97X-V/6-311+G(2df,2p)//ωB97X-D/6-31G*, **NMR** relates calculated and experimental NMR chemical shifts and **Utilities** introduces standard atom labels, resets model and atom colors as well as conformational degrees of freedom to standard values, fixes bond typing, changes absolute configuration adds missing hydrogens and adjusts X-H bond lengths. Note that the last three of these require that calculations need to be repeated. Make certain that the **Molecule** tab is selected.



This provides the energy* for acetic acid in atomic units (**Energy** in au).

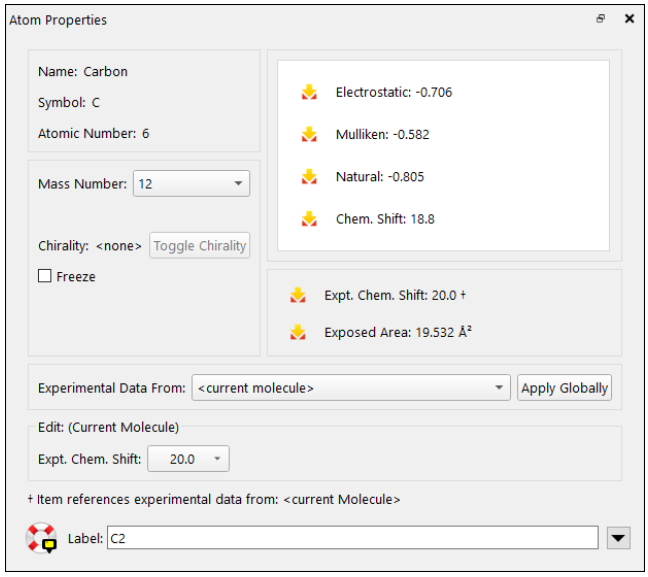
b) **Dipole Moment:** The magnitude of the dipole moment (**Dipole Moment** in Debye) is also provided in the **Molecule Properties** dialog. A large dipole moment indicates large separation of charge. You can attach the dipole moment vector, where the + side refers to the positive end of the dipole, to the model on the screen, by *checking* the box to the left of **Display Dipole Vector** to the right of **Dipole Moment**.

> The vector will not be displayed if the dipole moment is zero. The dipole moment will not be reported if the molecule is charged because in this case it depends on the location and orientation of the molecule in space.



* The calculated energy depends on a number of factors including computational model, basis set (where applicable) and environment. For a discussion, see **Activities** menu Topics: ***Total Energies and Thermodynamic and Kinetic Data***.

c) **Atomic Charges:** To display the charge on an atom, *click* on it with the **Properties** dialog displayed. **Atom Properties** replaces the currently displayed **Properties** dialog.

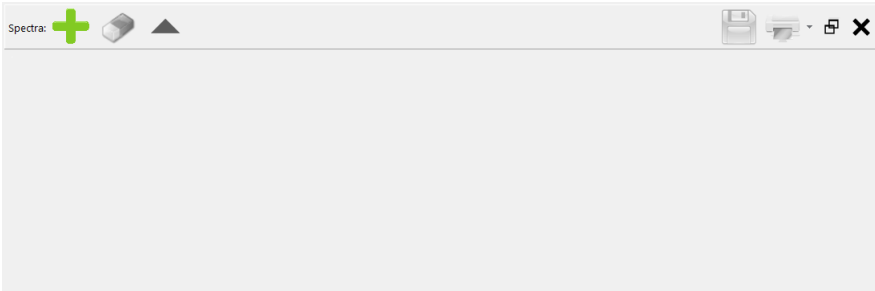


Three different atomic charges, **Electrostatic**, **Mulliken** and **Natural**, are given in units of electrons. A positive charge indicates a deficiency of electrons on an atom and a negative charge, an excess of electrons. Repeat for other atoms. Confirm that the positively-charged atom(s) lie at the positive end of the dipole moment vector. When you are finished, close the dialog by *clicking* on at the top.

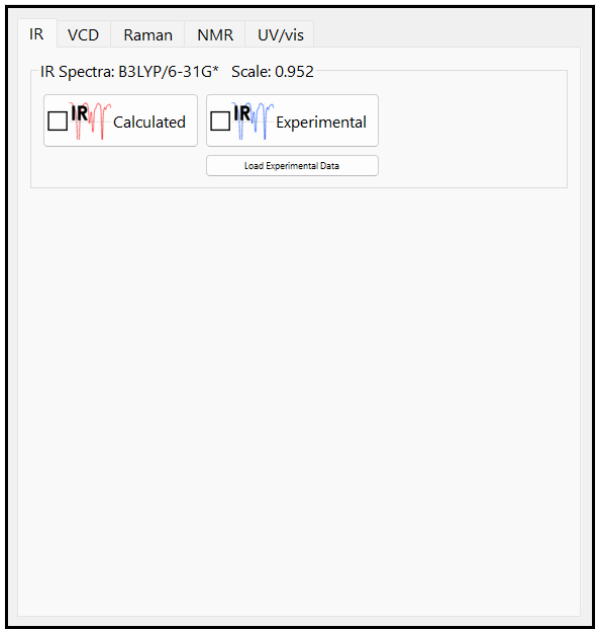
> The three sets of atomic charges for acetic acid are different, sometimes markedly so. Note that the charge on an atom in a molecule cannot be uniquely defined, let alone measured. While the nuclear charge is equal to the atomic number, it is not possible to say how many electrons "belong" to a particular nucleus. The different calculated charges correspond to different ways of counting the number of electrons associated with an atom.

6. **Infrared Spectra:** Molecules vibrate (stretch, bend, twist) even if they are cooled to absolute zero. This is the basis of infrared (and Raman) spectroscopy, where absorption of energy occurs when the frequency of a particular molecular motion matches the frequency of the light. Infrared spectroscopy is important for identifying molecules as different functional groups vibrate at noticeably different and characteristic frequencies.

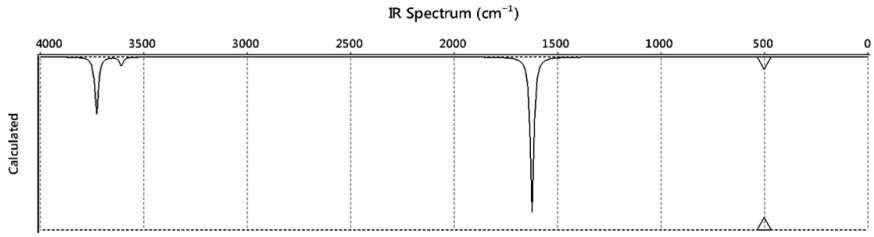
Click once on ▶ to move to the next molecule in the document, ***water***. To animate a vibration, select **Spectra** from the **Display** menu (or *click* on if it appears at the top of the screen). This leads to an empty spectra pane at the bottom of the screen.



Click on at the top left of the pane and select from the available spectra.



The calculated IR spectrum of water from 4000 - 500 cm^{-1} appears in the pane.

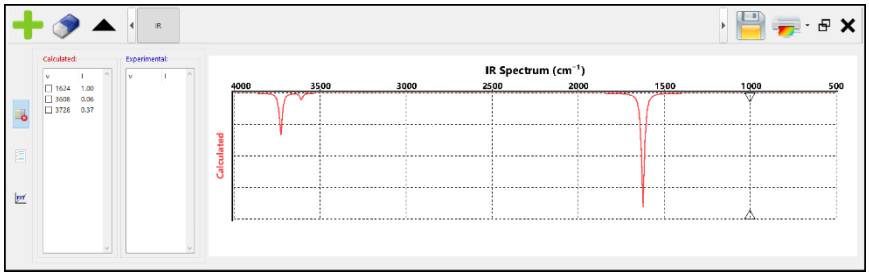


An upward and downward triangle server as a probe indicator, allowing for further inspection of the plot.

There are three lines, one of moderate intensity around 3728 cm^{-1}, one very weak around 3608 cm^{-1} and one very strong around 1624 cm^{-1}. In turn, *click* on each of these peaks. In response, the molecular model will vibrate. The line of moderate intensity corresponds to an asymmetric OH stretching motion, the very weak line corresponds to a symmetric OH stretching motion and the strong line corresponds to the HOH bend.

To translate the plot inside the pane, position the cursor over the spectrum and move the mouse left or right while holding down the right button. To expand or contract the scale of the IR plot from its default range, position the cursor over the spectrum and use the scroll wheel on your mouse (or alternatively move the mouse while holding down both the right button and **Shift** key). To reset the spectra plot to the original values, *click* on in the bar at the top of the spectra pane. To increase (or decrease) the size of the plot, you first need to undock it by *clicking* on at the upper right. You can then size it as you would any other window. To redock the plot, *click* again on .

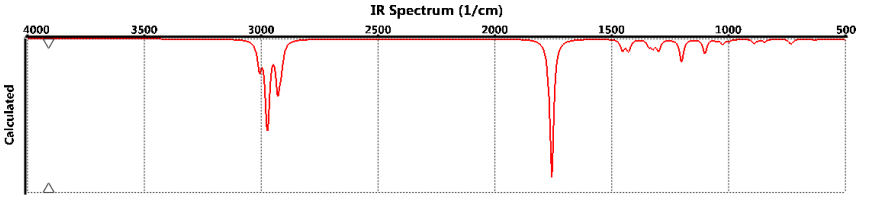
To see a complete listing of frequencies and intensities, *click* on (**Tables**) at the left of the spectra pane.



Click on each entry in the table to highlight the frequency in the spectrum and animate the vibration. *Click* on to dismiss the table and *click* on to remove the spectrum.

> Changing the size of the spectrum dialog as well as translating and altering the visible scale are quite simple with touch screen operations. To resize the spectrum, position one finger inside the menu bar at the top of the spectra pane and move up or down. To translate the scale, move two fingers over the spectrum. To alter the scale, pinch two fingers over the spectrum.

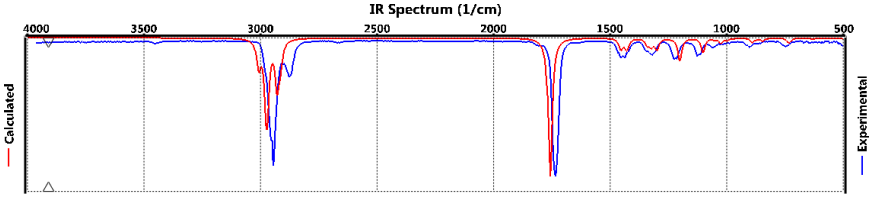
Click once on ▶ to move to ***cyclohexanone***, the next molecule in the list. The spectra pane is still on screen but should be empty. (If it is not on screen, select **Spectra** from the **Display** menu or *click* on if it appears at the top of the screen to restore.) *Click* on in the bar at the top of the spectra pane and on the . The calculated infrared spectrum of cyclohexanone appears.



The calculations actually provide a list of frequencies and an associated list of intensities (some of which may be zero or very close to zero). The spectrum that is displayed uses these data but has been broadened (to account for finite temperature) and scaled (to account for the fact that the underlying energy function is assumed to be quadratic as well as for limitations of different models). The same broadening parameter (temperature) is applied to all molecules and all levels of calculation. The same scaling parameter is used for all molecules, but it depends on the theoretical model.

The largest peak appears at 1759 cm^{-1} and corresponds to a CO stretch. The fact that the line is both intense and isolated from other features in the spectrum makes it a very useful indicator of carbonyl functionality. *Click* on this peak and examine the "vibrating" model for cyclohexanone on screen above the spectrum.

Click again on and this time select . The experimental IR spectrum (from the freely available NIST database) is superimposed on top of the calculated spectrum.



Note that the two spectra are similar but not perfectly matched.

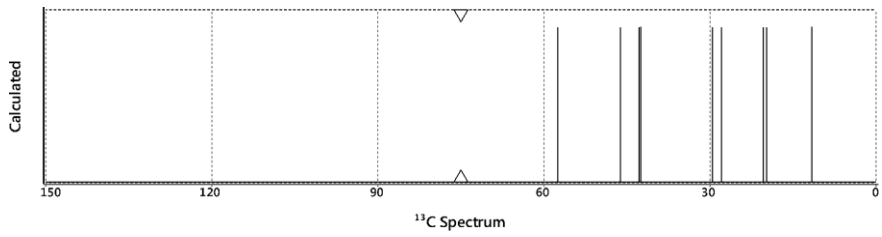
When you are done, select from the bar at the top of the spectra pane.

7. **VCD Spectra:** VCD spectra are often used to distinguish between enantiomers (that is, establish absolute configuration). Along with NMR for structure and diastereomer confirmation, VCD is highly useful in establishing proper stereochemistry.

   Click once on to move to the next molecule in the document, ***camphor (RR)***. With the spectra pane on screen, *click* on in the bar at the top of the spectra pane, click on the VCD tab and select . The calculated VCD and IR spectra for camphor (with RR chirality at bridging carbons). *Click* again on to advance to the next list member, ***epicamphor (SS)***. Note the sign (direction) of the SS enantiomer's spectra are inverted (from the RR epimer). Future updates will include import of experimental VCD spectra for overlaid comparisons. *Click* on to close out of the opened spectra.

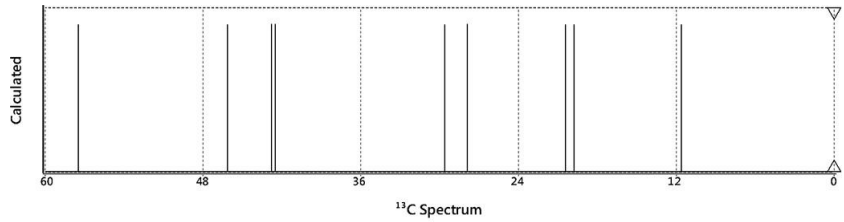
8. **NMR Spectra:** Along with mass spectrometry, NMR spectroscopy is the most powerful tool available with which to assign molecular structure. Many (but not all) nuclei exhibit NMR spectra, but proton and ^{13}C are by far the most important.

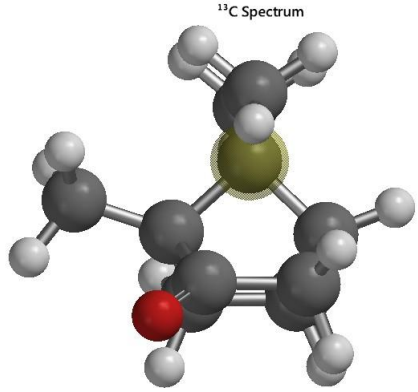
   Click once on to move to the next molecule in the document, ***camphor***. With the spectra pane on screen, *click* on in the bar at the top of the spectra pane, click on the NMR tab and select . The calculated ^{13}C NMR spectrum appears.



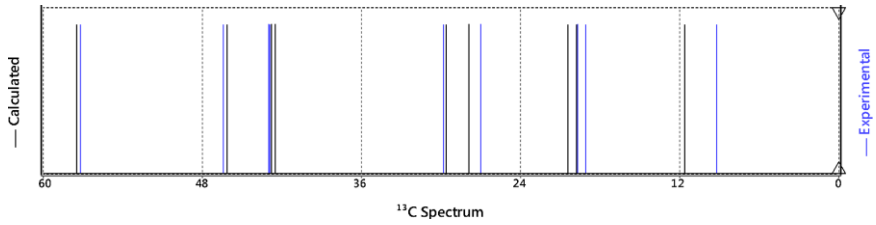
   This comprises nine lines, in the range of 150 to 0 ppm (there is a tenth line corresponding to the carbonyl carbon at 216 ppm). You can zoom out to see this line by using the scroll wheel on your mouse. More instructive is to zoom in on the range from 60 to 0 ppm to get a better look at the other lines.

   Click on one of the probe triangles to select the probe and move the mouse while holding down the left button over the spectrum. When you come to a line, the chemical shift will appear at the top of the spectrum and the atom (or atoms) responsible for this line will be highlighted on the model displayed above the spectrum.



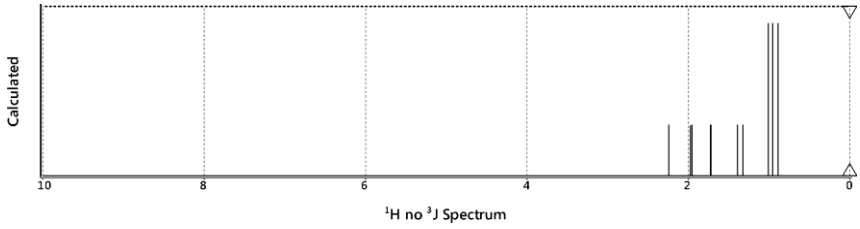


Again, *click* on in the bar at the top of the spectra pane and select . The experimental ^{13}C spectrum obtained from the freely available NMRShiftDB collection will be superimposed on top of the calculated spectrum.



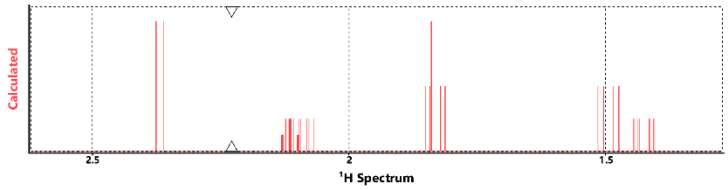
You will see that the overall agreement between calculated and experimental ^{13}C spectra is quite good. As with infrared spectra (see preceding discussion of cyclohexanone), the data resulting from the quantum chemical calculations has been empirically corrected.

Click again on and select . An "idealized" proton spectrum where three-bond HH coupling constants are set to zero appears.



Spectra manipulations are as before and the hydrogens responsible for selected lines are highlighted in the model. No experimental spectrum is available, but the quality of the match would be expected to be similar to that previously observed with comparison of ^{13}C spectra.

Click again on ✚ and select ☐ ^{1}H Calculated from the palette. The spectrum that appears is more complicated and much closer to what would be observed experimentally. In this example, HH coupling constants have not been calculated from quantum mechanics, but they have been estimated based on local environment.

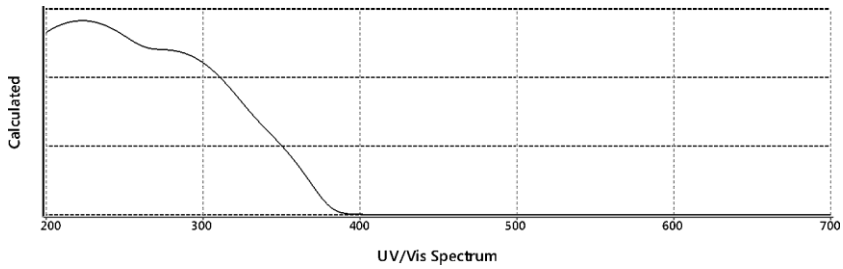


Zoom in on specific lines (scroll wheel) to see the detailed splitting patterns. For example, the two protons at C_3 are both split by the proton at C_4. The doublet at 2.36 ppm shows a much larger splitting than the doublet at 1.85 ppm (you need to zoom in considerably to see that this is a doublet). This reflects the fact that the proton responsible for the line at 2.36 ppm makes a dihedral angle of 43° with the proton at C_4, whereas the proton responsible for the line at 1.85 ppm makes a dihedral angle of 80°.

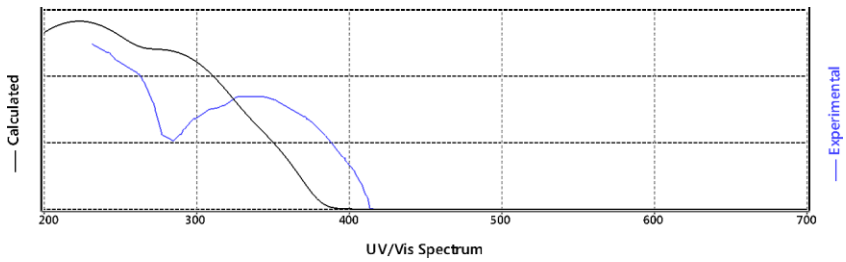
Finally, note that you can switch among the three calculated NMR spectra (as well as the experimental ^{13}C spectrum) for camphor by *clicking* on the associated button in the bar above the spectra pane. When you are done, remove all three spectra. *Click* on three times in succession to remove the spectra.

9. **UV/visible Spectra:** Absorption of light in the visible or ultraviolet range of the electromagnetic spectrum leads to electronic excitation from the ground-state to excited-states and (in the case of absorption in the visible), is responsible for a molecule's color. UV/visible spectroscopy not only offers a "fingerprint" but is also an important screen to identify molecules that may be damaged by exposure to light.

Click once on ► to move to the next molecule, ***3-benzoylaniline***. The spectra pane should still be on screen. *Click* on ✚ and select ☐ UV/vis Calculated . No empirical corrections have been applied to the calculated spectrum that appears.



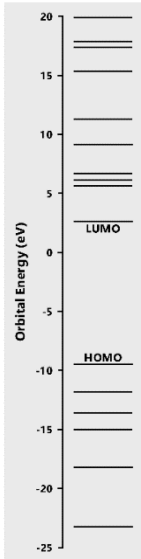
Click again on and select the UV/vis tab and click on . The experimental UV/ visible spectrum from the freely available NIST database will be drawn on top of the calculated spectrum.



The two spectra are visually similar at least qualitatively. However, calculated and experimental UV/visible spectra are likely to be sufficiently different that the theory will not often be able to account for the "color" of the molecule. Where the theory is likely to be more successful is in anticipating changes in color resulting from subtle changes in structure. *Click* on when you are done. Also, remove the spectra pane either by *clicking* on at the top right or by selecting **Spectra** from the **Display** menu or by *clicking* on at the top of the screen.

10. ***Spartan*** permits display, manipulation and query of a number of important graphical quantities resulting from quantum chemical calculations. Most important are the ***electron density*** or simply **density** as it is commonly referred to (that reveals how much space a molecule actually takes up), the ***bond density*** (that reveals chemical bonds), and key ***molecular orbitals*** (that provide insight into both bonding and chemical reactivity). In addition, the ***electrostatic potential map***, an overlay of the electrostatic potential (the attraction or repulsion of a positive charge for a molecule) on the electron density, is valuable for describing overall molecular charge distribution as well as anticipating sites of electrophilic addition. Another indicator of electrophilic addition is provided by the ***local ionization potential map***, an overlay of the energy of electron removal (ionization) on the electron density. Finally, an indicator of nucleophilic addition is provided by the ***|LUMO| map***, an overlay of the absolute value of the lowest-unoccupied molecular orbital (the LUMO) on the electron density.

a) *Click* once on to move to the next molecule in the list, ***ethylene***. *Click* on the **Display** menu and select **Orbital Energies** (or *click* if it appears at the top of the screen). An orbital energy diagram for ethylene will appear at the left of the screen. This provides the energies of all six occupied valence molecular orbitals and the first ten unoccupied molecular orbitals.

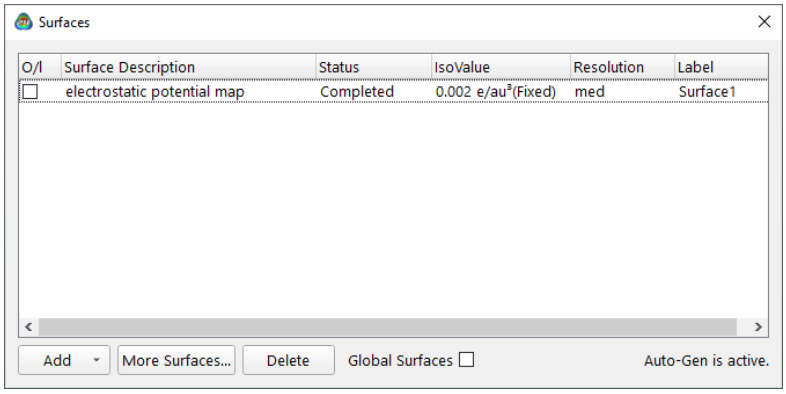


Note: *Right-clicking* on the MO diagram and choosing Properties from the resulting contextual menu provides access to display of up to the first 10 occupied and unoccupied orbitals. Orbitals outside this range can be requested individually from the **Surfaces** dialog (**Setup** and **Display** menu).

Click on the energy level in the diagram labeled ***HOMO***. In a second, the familiar π bond in ethylene will appear. Note that the graphic has "blue and red" regions. These correspond to positive and negative values of the orbital (the absolute sign is arbitrary). Examine the other occupied orbitals (by *clicking* on their respective energy levels in the diagram) as well as the lowest-unoccupied molecular orbital (the ***LUMO***). Note that you can move from one level to the next by moving the mouse up or down while holding down and then releasing the left button. You can also use the up and down arrow keys on your keyboard. *Click* on view when you are done.

> "Swipe" one finger up or down over the orbital energy diagram to move to the next higher or lower energy level.

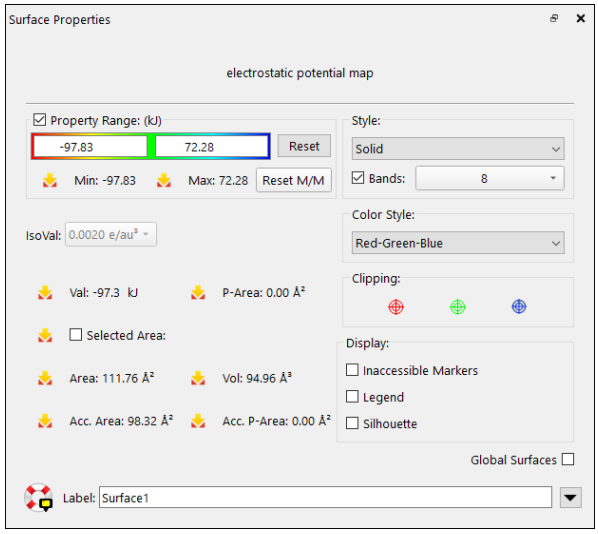
b) *Click* once on ►to move to the next molecule in the list, ***benzene***. *Click* on the **Display** menu and select **Surfaces** from the palette (or *click* on if it appears at the top of the screen). The **Surfaces** dialog appears.



Select ***electrostatic potential map*** inside the **Surfaces** dialog (*check* the box to the left of the name). An electrostatic potential map for benzene will appear.* *Click* on the map. The **Style** menu will appear at the bottom right of the screen. Select **Transparent** from this menu. This makes the map transparent and allows you to see the molecular skeleton underneath. Go back to a **Solid** display (**Style** menu) in order to clearly see color differences. The default color scheme follows the coloring in a rainbow:

red –› orange –› yellow –› green –› blue

By convention, red indicates a negative potential (the benzene π system) attraction to a positive charge while blue indicates a positive potential (the σ system) or repulsion by a positive charge. An alternative scheme with only three colors (red, white and blue) may be selected. *Click* on the **Display** menu and select **Properties** (or *click* on if it appears at the top of the screen) and *click* on the surface.

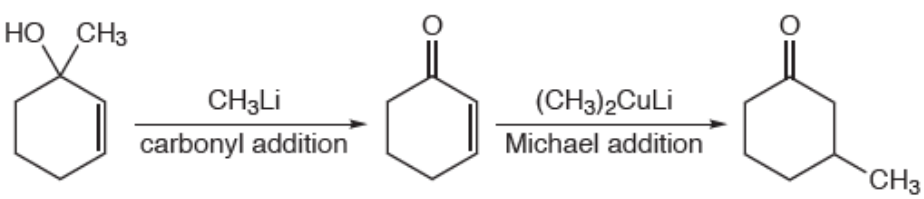


* Discrete displays are the default. You can change the default to continuous displays by turning off the **Bands** checkbox in the **Molecule** tab of the **Preferences** dialog (**Preferences...** under the **Options** menu; **Chapter 25**).

c) *Click* once on ►to move to the next molecule in the list, ***aniline***, and select ***local ionization potential map*** inside the **Surfaces** dialog. By convention, red regions on a local ionization potential map indicate areas from which electron removal (ionization) is relatively easy, meaning that they are subject to electrophilic attack. These are easily distinguished from regions where ionization is relatively difficult (by convention, colored blue). Note that the *ortho* and *para* ring carbons are redder than the *meta* carbons, consistent with the known directing ability of the amino substituent.

d) *Click* once on ►to move to the next molecule in the list, ***cyclohexenone***, and *check* ***LUMO*** inside the **Surfaces** dialog. The resulting graphic portrays the lowest-energy empty molecular orbital (the LUMO) of cyclohexenone. This orbital is delocalized onto several atoms, and it is difficult to tell where exactly a pair of electrons (a nucleophile) will attack the molecule.

A clearer portrayal is provided by a LUMO map that displays the (absolute) value of the LUMO on the electron density surface. By convention, the color blue is used to represent maximum value of the LUMO and the color red, minimum value. First, remove the LUMO from your structure (*uncheck* ***LUMO*** in the **Surfaces** dialog) and then turn on the LUMO map (*check* ***|LUMO|map*** in the dialog). Note that there are two blue regions, one directly over the carbonyl carbon and the other over the β carbon. This is entirely consistent with known chemistry. Enones may either undergo carbonyl addition or conjugate (Michael) addition.



11. New in ***Spartan'26*** is class of models that generally fall under the heading of machine learning models (colloquially, throughout the manual, these are referred to as neural network routines, or more simply neural nets). *Click* once more on ► to move to the next model ***camalexin dft*** (the full SSPD identifies this as 3-(1,3-thiazol-2-yl)-1H-indole. This example includes the results of geometry optimization and NMR chemical shifts from ωB97X-D/6-31G* density functional calculations. The model display has been simplified to omit hydrogens, and the molecule is rendered in the tube display with calculated chemical shift labels turned on for 4 specific carbon centers.

    Click once more on ► to move to the next model ***camalexin est dft nn***. This molecule is a copy of the previous and also has available NMR data from ωB97X-D/6-31G* density functional calculations, the difference in this case is that equilibrium geometry was determined by a neural network routine, rather than quantum mechanics. Move the step keys back and forth to compare chemical shift results.

From the **Display** menu, select the **Spreadsheet**. ***Spartan's*** internal spreadsheet is useful for data organization, analysis, linear regression, and customized functions including multi-molecule display. Resize the spreadsheet so that you can see the bottom two entries. *Click* inside the check box to left of both ***camalexin dft*** and ***camalexin est dft nn***. This will turn on display of *both* molecules. You can manipulate both samples independently (selection is made by *clicking* on the molecule you wish to select). Note, it may be desirable to **Couple** multiple molecule's positions, such that movements are applied to multiple molecules. This is achieved by toggling the **Coupled** (on or off as needed). Position both copies so that the chemical shifts are visible for comparison. How well did the neural network routine perform in comparison to the density functional sample?

12. When you are finished, close the document by selecting **Close** from the **File** menu or alternatively by *clicking* on the icon at the top of the screen.

Section III
Introductory Tutorials

The first step in performing a molecular mechanics or quantum chemical calculation typically involves specifying molecular geometry. ***Spartan*** offers four ways to do this: building in 3D, sketching in 2D, accessing a database and accessing a variety of coordinate formats from the **File** menu. See **Chapter 16** for a list of supported file types.

To build molecules in 3D, ***Spartan*** provides a palette of common atomic fragments, menus of common functional groups and rings, access to the clipboard and tools for making and breaking bonds and for deleting atoms or groups of atoms. The 3D building paradigm is quite general and covers not only organic molecules discussed in chapters which make up this section, but also inorganic and organometallic molecules, as well as polypeptides and polynucleotides. Tutorials illustrating the essential features and capabilities of ***Spartan***'s 3D builder for organic molecules are provided in **Chapter 3**.

Building in 3D may feel unfamiliar or uncomfortable for many chemists, who, after all, have been taught to draw molecules in 2D on paper or on a whiteboard. The advent of tablet computers and touch- screen PC's makes "on-screen" 2D drawing an attractive alternative to 3D building. ***Spartan*** provides a simple but highly functional 2D sketcher directed primarily (but not restricted) to organic molecules. Simple inorganic and organometallic molecules may also be sketched. Common atoms, functional groups and rings are all contained in a single palette. Conversion to 3D structure is automatic, but requires explicit specification of cation, anion, or radical centers (in order to add the proper number of hydrogens to the sketch) as well as designation of stereochemistry. While sketch capability is designed for touch screen devices (and is modeled after the ***iSpartan*** app for iOS devices), it is fully functional using a mouse and keyboard or a touch pad. Tutorials illustrating the essential features and capabilities of ***Spartan***'s 2D sketcher are provided in **Chapter 4**.

Properties relating to geometry, for example, bond lengths and angles, are discussed in these two chapters and a few simple examples are provided. The intention is to touch on only the most commonly-used features, leaving discussion of more advanced functions to tutorials in later chapters.

The tutorials in the remaining chapters in this section may be completed starting from either the 2D sketcher or 3D builder whichever is more comfortable for the user. Tutorials in **Chapter 5** illustrate the range of molecular properties and spectra available from quantum chemical calculations, while those in **Chapter 6** illustrate a variety of common graphical models that can be displayed. No calculations are actually required for these tutorials. Rather, results are taken from the ***Spartan Spectra and Properties Database*** (***SSPD***) and correspond to calculations using either the ωB97X-D/6-31G* density functional method.

The first tutorials that actually requires quantum chemical calculations are provided in **Chapter 7**.

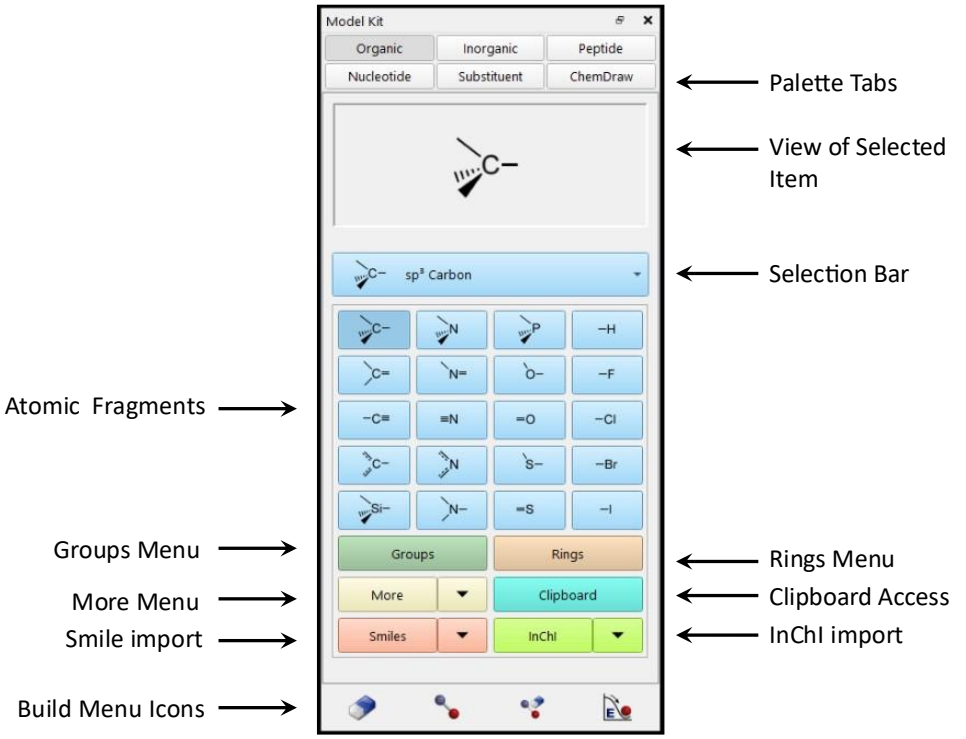
Completing the full set of tutorials in this section will take a few hours and make only a modest demand on computer time (for examples in the last section). This will not make you an "expert" but will expose you to enough that you are comfortable navigating ***Spartan***. "Advanced" tutorials, in the next section, build on this foundation and introduce a wider range of the program's features and capabilities.

Chapter 3
Building Organic Molecules in 3D

The tutorials in this chapter introduce and illustrate tools to build 3D molecular structures. These include atomic fragments, functional groups and rings contained in the organic model kit together with tools for making and breaking bonds, deleting atoms and refining structure.

Organic Model Kit

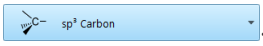
Click on (**New**) to bring up the 3D model kit. The organic model kit contains a selection of atomic fragments corresponding to elements commonly found in organic molecules. The kit may be scrolled if your display is too small to display it.*



* Additionally, you can save space by removing the icons at the bottom of the model kit. *Uncheck* **Show Toolbar** from the **Miscellaneous** tab of the **Preferences** dialog (**Options** menu; **Chapter 25**).

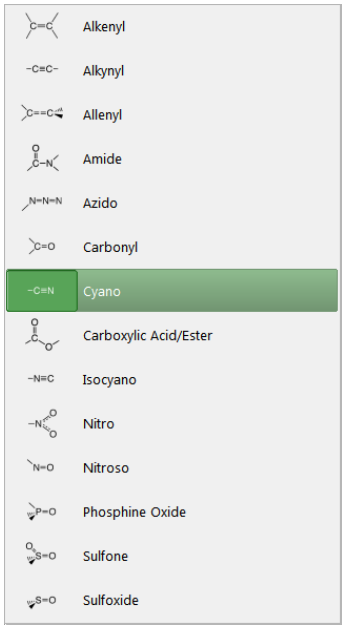
Different hybridization states are provided for some elements (from left to right and then top to bottom):

C(sp^3)	**N(sp^3)**	**P(sp^3)**	**H**
C(sp^2)	**N(sp^2)**	**O(sp^3)**	**F**
C(sp)	**N(sp)**	**O(sp^2)**	**Cl**
C(aromatic)	**N(aromatic)**	**S(sp^3)**	**Br**
Si(sp^3)	**N(planar)**	**S(sp^2)**	**I**

A fragment is chosen by *clicking* on its icon, which is then displayed in a box at the top of the model kit. The name of the fragment is displayed in the selection bar underneath, for example, 

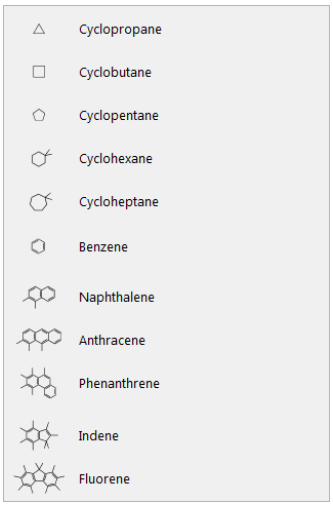
. Once selected, the fragment may be used to initiate building, to add alongside of an existing structure or bond onto an existing structure. To initiate building, *double click* anywhere on screen.* To add alongside of an existing structure, *double click* in a blank area on screen. To bond to an existing structure, *click* on a free valence (***not an atom***). (Free valences are colored yellow on the selected molecule.) Bond type in the case of atomic fragments with multiple bond types, for example, sp^2 carbon, depends on the nature of the free valence selected.

Clicking on the **Groups** button near the bottom of the model kit changes the focus from fragments to groups, one of which will be shown in the box at the top of the model kit and named in the selection bar directly underneath. *Clicking* on the selection bar brings up a menu of available groups.



Once selected from the menu, a group may be used to initiate building, to add alongside of an existing structure on screen, or to add to an existing structure.

* For users of much earlier versions of ***Spartan*** that have grown accustomed to (or new users who prefer) a *single-click* motion to initiate building, this preference can be set from the **Options** menu > **Preferences** dialog > **Settings** tab, by deselecting the **Double-Click Start** check box and *clicking* the **OK** button. The choice to move to a *double-click* start for 3D building was made to maintain consistency with the behavior of the 2D sketch builder.



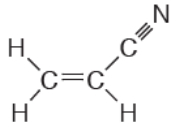
Clicking on the **Rings** button near the bottom of the model kit changes the focus to rings, one of which will be shown in the box at the top of the model kit and named in the selection bar underneath. *Clicking* on the selection bar brings up a menu of available rings.

Once selected from the menu, a ring may be used to initiate building, to add alongside of an existing structure on screen, or to add to an existing structure. Ring-merging is available (that is clicking on a bond to merge rather than attaching via single-bond.

Note that only hydrocarbon rings are available. Heteroatoms may be substituted for carbons, for example, substituting an oxygen for one of the carbons in cyclohexane leading to tetrahydropyran. With cyclohexane on screen, *click* on sp^3 oxygen in the model kit and then *double click* on one of the cyclohexane carbons.

The amide and carboxylic acid/ester groups and the cyclohexane, cycloheptane, naphthalene, phenanthrene, indene and fluorene rings have more than one unique free valence. The free valence that is to be used is marked with a gold • (in the icon shown in the box at the top of the model kit). The marked position circulates among the possible positions with repeated *clicking* on the icon. Selection of an *axial* or *equatorial* free valence in cyclohexane and cycloheptane is indicated by the label **ax** or **eq** appearing alongside the icon.

Acrylonitrile



1. To bring up the organic model kit, *click* (left button) on the **File** menu and select (*click* on) **New**. Alternatively, *click* on if it appears at the top of the screen.* *Click* on trigonal planar sp^2 hybridized carbon from the fragment library.

* Icons are fully customizable from the Options menu > Preferences dialog > Icons tab. So whether or not a particular icon is displayed is at the discretion of the user.

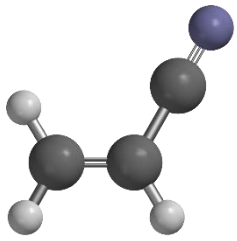
A model of the fragment appears at the top of the model kit. Bring the cursor anywhere on screen and *double click* (left button). Rotate the carbon fragment (*drag* the mouse while holding down the left button) so that you can clearly see both the double free valence (=) and the two single free valences (–).

2. sp^2 carbon is still selected. *Click* on the double free valence. The two fragments are connected by a double bond, leaving you with ethylene, meaning that ethylene is in the Spartan Spectra and Properties Database (SSPD). The name ***ethylene*** will appear at the bottom right of the screen. If you make a mistake and *click* instead on the single free valence, select **Undo** from the **Edit** menu (or *click* on if it appears at the top of the screen). You can also start over by selecting **Clear** from the **Edit** menu (or *click* on if it appears at the top of the screen).

3. *Click* on the **Groups** button at the bottom of the model kit, *click* on the selection bar at the top of the model kit, and select **Cyano** from the groups available from the menu. *Click* on any of the four single free valences on ethylene (they are equivalent). This bonds the cyano group to ethylene, leaving you with acrylonitrile. Its name will appear at the bottom right of the screen.

4. *Click* on at the bottom of the model kit. (You can also select **Minimize** from the **Build** menu or *click* on if the icon appears in the toolbar at the top of the screen.) The molecular mechanics energy (36.2 kJ/mol) and symmetry point group (C_s) are provided at the bottom right of the screen.

5. Select **View** from the **Build** menu (or *click* on the icon in the toolbar). The model kit disappears, leaving only a ball-and- spoke model of acrylonitrile on screen. The name appears at the bottom of the screen as acrylonitrile is also in the SSPD.



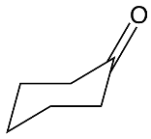
ball-and-spoke model

This model can be rotated, translated and zoomed by using the mouse in conjunction with keyboard functions. To rotate the model, *drag* the mouse while holding down the left button; to rotate in the plane of the screen also hold down the **Shift** key. To translate the model, *drag* the mouse with the right button depressed. To zoom the model, use the center mouse wheel (scroll wheel) if available, or hold down the **Shift** key in addition to the right button while *dragging* the mouse up (zoom in) or down (zoom out).

> Rotate the molecule by moving one finger over the screen. Rotate the molecule in the plane of the screen by twisting two fingers. Translate by moving two fingers. Zoom by pinching two fingers.

6. Close ***acrylonitrile*** (**Close** from the **File** menu or *click* on the icon in the toolbar).

Cyclohexanone



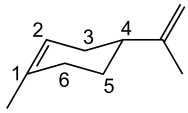
1. *Click* on the **File** menu and select **New Build** from the palette or *click* on at the top of the screen. *Click* on the **Rings** button near the bottom of the model kit, *click* on the selection bar and choose **Cyclohexane** from the menu of rings. *Double-click* anywhere on screen.
2. Select sp^2 carbon from the model kit. *Double click* on any carbon atom (***not a free valence***). The sp^3 hybridized center will be replaced by an sp^2 hybridized carbon.

> Fragment replacement is subject both to the usual valence rules and to the availability of free valences. For example, replacement of sp^3 carbon by sp^2 carbon requires that at least two free valences are available.

Select sp^2 oxygen from the model kit. *Click* on the double free valence on the sp^2 carbon. You have made cyclohexanone. *Click* on at the bottom of the model kit to produce a structure with C_s symmetry. *Click* on . The name ***cyclohexanone*** appears at the bottom of the screen as the molecule is in SSPD.

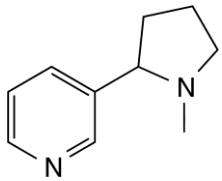
3. Close ***cyclohexanone***.

Limonene



1. Select **New** from the **File** menu (or *click* on the icon in the toolbar) to bring up the model kit. *Click* on the **Rings** button near the bottom of the model kit, *click* on the selection bar and choose **Cyclohexane** from the menu of rings. *Double-click* anywhere on screen.
2. *Click* on the **Groups** button near the bottom of the model kit, *click* on the selection bar and choose the **Alkenyl** group. *Click* on the *equatorial* free valence on C_4 (see figure above for numbering). You have made vinylcyclohexane.
3. *Click* on the **Make Bond** icon ( ) at the bottom of the model kit. One after another *click* on the *axial* free valence on C_1 and then the *axial* free valence on C_2. You have made 4-ethenylcyclohex-1-ene.
4. Select sp3 carbon from the model kit and one after another *click* on the free valence on C1 and on the free valence of the vinylic carbon attached to the ring. You have made limonene. The name will appear at the bottom of the screen as limonene is in SSPD. *Click* on at the bottom of the model kit to give a refined geometry and finally *click* on ( ).
5. Close ***limonene***.

Nicotine



1. Select **New** from the **File** menu. *Click* on the **Rings** button in the model kit. *Click* on selection bar at the top of the model kit and choose **Benzene**. *Double-click* anywhere on screen.

2. *Click* on the selection bar and choose **Cyclopentane**. *Click* on one of the free valences on benzene. You have made phenylcyclopentane.

3. *Click* on the sp^2 nitrogen in the model kit and *double click* on the appropriate (*meta*) carbon (not a free valence) in the benzene ring. You have made 3-cyclopentylpyridine.

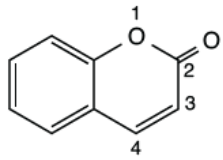
4. *Click* on sp^3 nitrogen in the model kit and *double click* on the appropriate carbon in the cyclopentyl ring. You have made nornicotine.

5. *Click* on sp^3 carbon in the model kit and *click* on the free valence on the nitrogen in the pyrrolidine ring. You have made nicotine. *Click* on at the bottom of the model kit to clean up your structure. The name ***nicotine*** will appear at the bottom of the screen as the molecule is in SSPD.

6. Select **R/S Chirality** from the **Model** menu (). The R/S chirality will be displayed. The S isomer is the naturally occurring isomer of nicotine. If you have built the R isomer and wish to invert the chiral center, hold down the **Ctrl** key on your keyboard and *double-click* on the chiral center.

7. Close ***nicotine***.

Coumarin

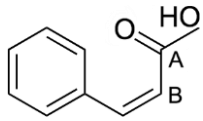


1. Select **New** from the **File** menu . *Click* on the **Rings** button in the model kit. *Click* on the selection bar at the top of the model kit and choose **Benzene**. *Double-click* anywhere on screen.

2. *Click* on the **Groups** button, *click* on the selection bar and choose **Alkenyl**. *Click* on one of the free valences of benzene to make styrene.

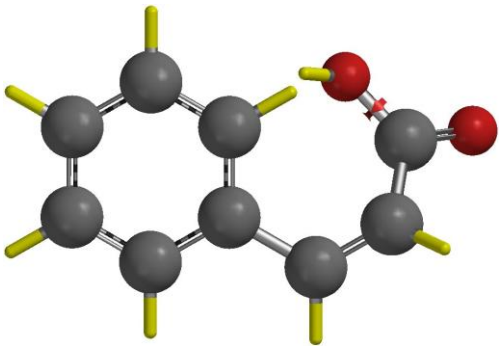
3. *Click* on the selection bar and choose **Carboxylic Acid/Ester**. A yellow dot should mark the free valence ***on carbon*** in the group shown in the window at the top of the model kit. If instead it marks the free valence on oxygen, *click* inside the window to move it. *Click* on the free valence on the methylene end of the vinyl group that is *cis* to benzene to make *cis*-cinnamic acid.

4. It may be necessary to rotate around the bond marked A and B marked below to properly position the molecule for bonding.



In turn, *click* on bonds A and B. A red arrow will curl around the bond and an icon will appear at the top of a narrow shaded area at the far left of the screen. *Click* inside this area and move the mouse up and down. Position the bond such that the OH bond is near the phenyl ring as shown below.

Move one finger up and down inside the area at the far left of the screen to rotate about the selected bond.



When you can clearly see the two free valences, *click* on at the bottom of the model kit and, one after another, *click* on the two free valences. You have made coumarin, its name will appear at the bottom of the screen.

5.Close ***coumarin***.

Spartan is able to fuse rings in 3D. This makes building molecules like coumarin much easier. To build coumarin using the fused-ring approach, select **Benzene** from the **Rings** menu *double-click* to insert benzene on screen. Select **Cyclohexane** from the **Rings** menu, *double-click* on one of the CC bonds in benzene to form benzocyclohexane. Choose sp^3 oxygen and *double-click* to replace C_1 by oxygen (see numbering on previous page). Choose sp^2 carbon and *double-click* on C_2. Choose sp^2 oxygen and add this to the double-bond on the sp^2 carbon. *Click* on **Make Bond** ( ) and *click* on the open *axial* valences the sp^3 ring carbons C_3 and C_4 to complete coumarin.

Note, however, that fusing two rings with both bonds involving sp^3-sp^3 hybridized centers, for example, two cyclohexane rings, can lead to one of two different results, in the case of two cyclohexane rings, either *cis* or *trans*-decalin. In this case, it may be necessary to invert the chirality of one of the resulting ring-fused carbons. This is accomplished by holding down on the **Control** (**Ctrl**) key (**Command** on Mac), positioning the cursor over the carbon and *double clicking*.

Chapter 4
Sketching Organic Molecules in 2D

The tutorials in this chapter introduce and illustrate tools to sketch organic molecules in 2D and then to convert them into realistic 3D structures.

Not only are 2D sketches ("drawings") more familiar to most chemists than 3D structures, they are typically easier to produce especially for complex molecules that may incorporate fused rings or require stereochemistry to be defined. The advent of touch-screen computers makes the argument for sketching as an alternative to building even more compelling. Molecules that require several minutes to build in 3D can be drawn in seconds. The key is automatic and reliable conversion from 2D drawings to 3D structures and from 3D structures to 2D drawings.

Sketch Palette

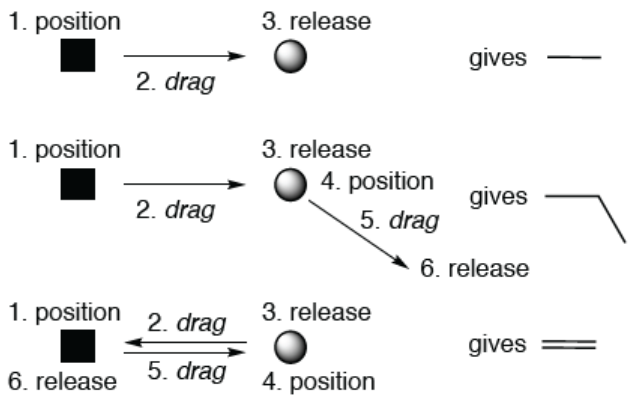
The sketch palette contains tools for making and manipulating 2D drawings, including tools for adding cues to designate stereochemistry.*

These include atoms that are most commonly found in organic molecules (H, B, C, N, O, F, Si, P, S, Cl, Br and I), the phenyl, cyclohexyl and cyclopentyl rings and the carbonyl, acid/ester and amide functional groups. The button immediately below H and B atoms allows for entering additional elements, functional groups, and ligands. The palette also contains stereochemical markers and charge/ radical markers. Complete details are found in **Chapter 20**.

* A row of icons appears immediately above the sketch palette. These are intended to add arrows to designate reaction transition states and to append experimental NMR chemical shifts and observed HH and CH couplings to the sketch palette

Sketching a Molecule

To start a sketch, first select (*click* on) an atom, group, ring or wildcard icon in the palette and then *double click* in the white portion of the screen (the drawing area). To draw a bond, first *click* on an atom, group, ring or wildcard icon in the palette to designate what is at the end of the bond, then position the cursor over the atom in the drawing area where you want the bond to start, move the cursor while holding down the left button ("*drag*" the cursor) to the place in the drawing area where you want the bond to end and release the button. Multiple bonds are made by dragging over existing bonds.*



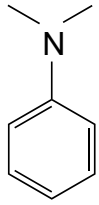
To make a bond touch the screen where you want it to start, move one finger to where you want it to end and lift. Replace position by touch, drag by move and release by lift in the diagram above.

Manipulating a Sketch

To translate the sketch, move the mouse over the screen while holding down the right button. To rotate the sketch (in the plane of the screen), move the mouse up and down while holding down both the left button and **Shift** key. Use the scroll wheel to resize the sketch.

* With ***Spartan'26***, and alternative way of drawing multiple bonds is simply to *double click* on a single bond. *Double clicking* (on a double bond) draws a triple bond and *double clicking* on a triple bond reverts to a single bond. The one exception is *double clicking* on a bond when a ring is selected in the palette as this signifies ring fusing.

N,N-Dimethylaniline



1. Select **New** from the **File** menu (or *click* on at the top of the screen) to bring up the sketch pad. *Click* on in the palette and *double click* on screen.

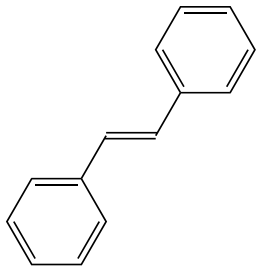
2. *Click* on **N** in the palette. Position the cursor over the "top" carbon on the benzene ring, *drag* it up and release. You have drawn aniline.

3. *Click* on **C** in the palette. Position the cursor over the nitrogen, *drag* it up and to the left and release. **C** is still selected. Again position the cursor over the nitrogen, *drag* it up and to the right and release.

4. *Click* on in the palette to clean up your drawing and *click* on to produce a 3D structure. The name ***N,N-dimethylaniline*** will appear at the bottom of the screen as the molecule is in SSPD.

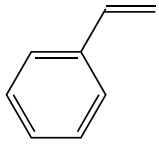
5. Close ***N,N-dimethylaniline***.

***trans*-Stilbene**



1. Select **New** from the **File** menu to bring up the sketch pad. *Click* on C in the palette. Position the cursor on the screen, *drag* it to the right and release. Position the cursor on one end of the line (CC bond) that you have just drawn, *drag* it to the other end and release. You have drawn ethylene.

2. *Click* on . Position the cursor over the left end of the double bond, *drag* it down and to the left and release. You have drawn styrene.

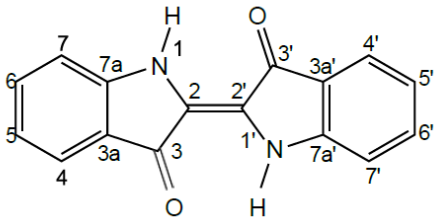


3. is still selected. Position the cursor over the right end of the double bond, *drag* it up and to the right and release. You are done.

4. *Click* on to clean up your drawing. *Click* on to convert it to a 3D structure. The name ***trans-stilbene*** will appear at the bottom of the screen as the molecule is in SSPD.

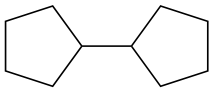
5. Close ***trans-stilbene***.

Indigo

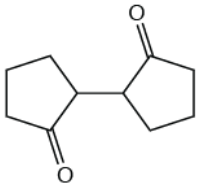


1. Select **New Sketch** from the **File** menu. *Click* on and *double click* on screen.

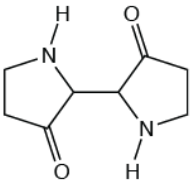
2. is still selected. Position the cursor over C_2 (see numbering in diagram above), *drag* it to the right and release it. Your sketch should appear as below.



3. Select O from the palette. Position the cursor above C_3, *drag* it away from the ring and release. Again position the cursor above C_3, drag it along the CO bond to the oxygen and release it. Alternatively, *double click* on the CO bond that you have just made to turn it into a double bond. Repeat for C_3'. You are left with a drawing.



4. Select N from the palette and one after another *double click* on C_1 and C_1'.



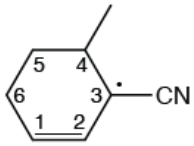
5. Select C from the palette. Position the cursor above C_2, *drag* it to C_2' and release. Alternatively, *double click* on the C_2-C_2' bond.

6. Select from the palette and *double click* on the bond connecting C_{3a} and C_{7a}.* Repeat for $C_{3a'}$ and $C_{7a'}$.

7. *Click* on to clean up your drawing and *click* on to turn it into a 3D structure. The name ***indigo*** will appear at the bottom of the screen as the molecule is in SSPD.

8. Close ***indigo***.

3-Cyano-4-methylcyclohexenyl Radical



This tutorial shows you how to use a wildcard to specify a functional group and how to designate a radical site.

1. Select **New** from the **File** menu . *Click* on in the palette and *double click* on screen.

2. *Click* on C in the palette. Position the cursor over C_4 (see diagram above for numbering), *drag* it up and to the right and release. You have drawn methylcyclohexane.

3. C is still selected. Position the cursor over C_1, *drag* to C_2 and release. Alternatively *double click* on the C_1-C_2 bond. You have drawn 4-methylcyclohexene.

4. *Double-click* on the wildcard icon that is located immediately below H and B in the palette. *Click* on the **Groups** tab and *click* on **CN**. **CN** now appears as the wildcard icon. Position the cursor over C_3 and *drag* it away from the ring and release. You have drawn 3-cyano-4-methylcyclohexene.

5. One of the following icons will appear in the palette directly below the sulfur icon. If it is not the radical marker, click to adjust:

6. With the radical marker selected, *double-click to modify center.*

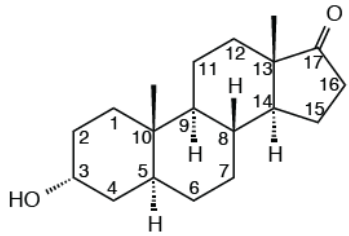
7. *Click* on to clean up your drawing and *click* on to make a 3D structure.

8. Close ***3-cyano-4-methylcyclohexenyl radical***.

---

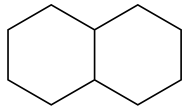
* This is the exception referred to earlier regarding *double clicking* on a bond to change it from single to double (or double to triple).

Androsterone

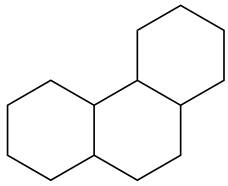


The steroid androsterone is typical example of a molecule that is difficult to build (3D) but quite easy to sketch (2D).

1. Select **New** from the **File** menu to bring up the sketch pad. *Click* on and *double click* on screen. Cyclohexane is still selected. *Double click* on the C_5-C_{10} bond (see diagram above).



2. Cyclohexane is still selected. *Double click* on the C_8-C_9 bond.

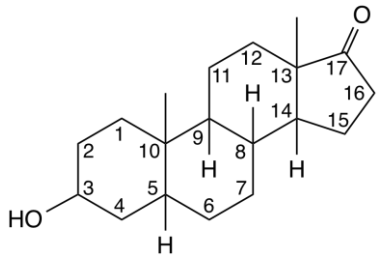


3. *Click* on and *double click* on the C_{13}-C_{14} bond. You have now drawn the complete steroid skeleton.

4. It is necessary to explicitly specify hydrogens at C_5, C_8, C_9 and C_{14} in order to incorporate the necessary stereochemical cues (up and down "wedges") in your drawing. *Click* on H, position the cursor over C_5, *drag* away from the ring and release. Repeat for C_8, C_9 and C_{14}.

5. *Click* on C, position the cursor over C_{10}, *drag* up and release it. Repeat for C_{13}.

6. *Click* on O. Position the cursor over C_3, *drag* down and to the left and release.



7. O is still selected. Position the cursor over C_{17}, *drag* up and to the right and release. Convert the single (CO) bond at C_{17} to a double bond. Again position the cursor over C_{17}, *drag* along the bond to oxygen and release. Alternatively, *double click* on the CO single bond to turn it into a double bond. Your structure should now look as follows.

8. *Click* on —. Position the cursor over C_{10}, *drag* along the bond to the methyl group that you drew in step 5 and release. Repeat for C_{13}. Position the cursor over C_8, *drag* along the CH bond that you drew in step 4 and release. Up wedges will appear for all three centers.

9. *Click* on ▪. Position the cursor over C_3, *drag* along the CO bond that you made in step 6 and release. Position the cursor over C_5, *drag* along the CH bond that you made in step 4 and release. Repeat for the CH bonds at C_9 and C_{14}. Down wedges will appear for all four centers.

Spartan'26 offers an alternative way to specify stereochemistry. Replace steps 4-6 and 8-9 by steps A-D.

A. *Click* on — in the sketch palette. *Click* on H, note — remains selected, position the cursor over C_8, *drag* up and release.

B. *Click* on ▪. H remains selected. One after the other, position the cursor over C_5, C_8, and C_{14}, *drag* down and release.

C. *Click* on — and then on C. One after the other, position the cursor over C_{10}, and C_{13}, *drag* down and release.

D. *Click* on ▪ and then on O. Position the cursor over C_3, *drag* down and to the left and release.

10. *Click* on to clean up your drawing. *Click* on . The name ***androsterone*** should appear at the bottom right of the screen as the molecule is in SSPD. If it does not, you have made an error somewhere. Select **Edit Sketch** from the **Build** menu or *click* on if it appears at the top of the screen to return to the sketch pad.

11. Androsterone incorporates seven chiral centers. To view the R/S assignment, *click* on the **R/S Chirality** entry in the **Model** menu . R/S labels will appear next to each of the chiral centers.

12. Close ***androsterone***.

Chapter 5

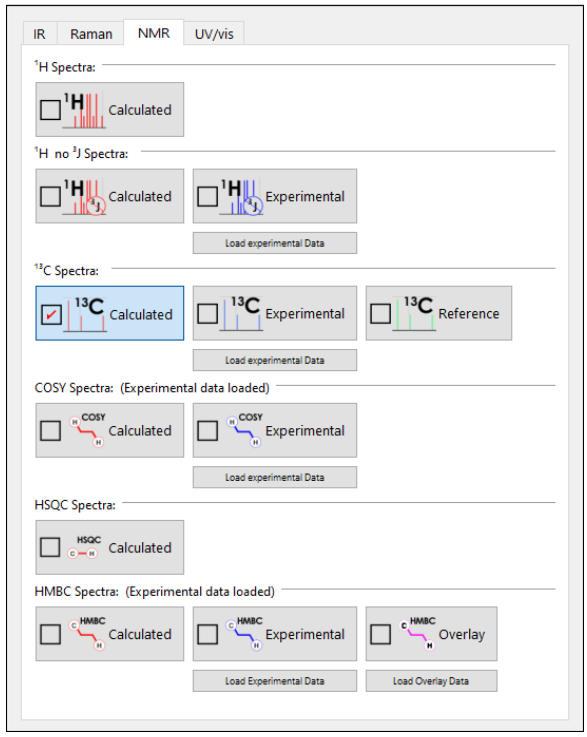
Spectra and Properties of Organic Molecules from SSPD

The tutorials in this and the two chapters that follow involve the molecules that you built or sketched in the preceding two chapters. The emphasis shifts from providing input structures, to analyzing the results of quantum chemical calculations (in this and the following chapter) to actually doing calculations (in the final chapter in this section).

The Spartan Spectra and Properties Database (SSPD) provides atomic and molecular properties and IR, thermodynamic properties, and NMR spectra for >300,000 molecules, calculated using the ωB97X-D/6-31G* density functional model. In addition, the collection contains more than 2,000 transition-metal organometallics and organolanthanides.

^{13}C Spectrum of Limonene

1. Build limonene , minimize and exit the builder . Alternatively, sketch limonene 2D and exit the sketcher .
2. *Click* on the name ***limonene*** at the bottom of the screen. Make sure that ωB97X-D/6-31G* is selected from the dialog that appears and *click* on **Replace**. Properties and spectra for limonene are now available.
3. Select **Spectra** from the **Display** menu or *click* on at the top of the screen. *Click* on at the top of the spectra pane, exposing a tabbed dialog with available calculated spectra (in red) and (possibly) available experimental spectra (in blue).

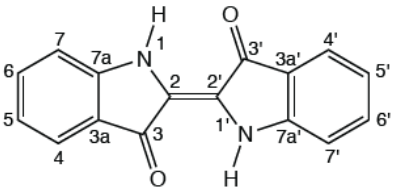


4. Select from the NMR tab. The calculated ^{13}C NMR spectrum appears in the spectra pane. There are ten lines corresponding to ten unique carbons in limonene. Up and down triangles demark a moveable cursor. *Click* on either (or on the dotted line connecting them) to select and with the left mouse button (or your finger), move horizontally over the spectrum. When you intersect an NMR resonance, it will be highlighted in the spectrum, the value of the chemical shift indicated, and the carbon (or carbons) in the structure responsible will be highlighted in the structure model (on the main screen).
5. Again, *click* on at the top of the spectra pane, but this time, select . The experimental ^{13}C spectrum of limonene will be superimposed on top of the calculated spectrum.* Visual comparison will give you an idea of the quality you can expect from NMR calculations. To get an even better idea, shift the range of the scale (initially from 150 to 0 ppm), move the cursor over the spectrum and zoom in with the center mouse wheel. Return to the original setting by *clicking* on in the bar at the top of the screen

Move one finger over the spectrum to select a line. Move two fingers over the spectrum to shift the range of the scale and pinch two fingers to zoom in and out.

6. Close ***limonene***.

Proton Spectrum of Indigo



1. Build indigo, minimize and exit the builder. Alternatively, sketch indigo 2D and exit the sketcher.

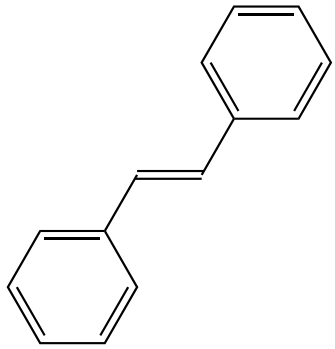
2. *Click* on the name ***indigo*** at the bottom of the screen. *Click* on **Replace**. Properties and spectra for indigo are now available. If the name does not appear, then you have made an error. In this case, re-enter either the builder by selecting **Edit Build** from the **Build** menu, or the sketch pad by selecting **Edit Sketch** from the **Build** menu and correct your model.

3. The calculated proton NMR spectrum of indigo can be displayed in two ways. The simpler "idealized" presentation assumes that three-bond HH coupling constants are zero. Select **Spectra** from the **Display** menu to bring up the spectra pane and *click* on in the bar at the top to show available calculated spectra (in red) and possibly available experimental spectra (in blue). Select from the palette. The spectrum that results shows only lines corresponding to the five unique hydrogens. To see a "familiar" proton spectrum, *click* again on , but this time select . The same four lines appear, but all are split (as in an experimental proton spectrum). The lines at 6.72 and7.88 ppm are doublets due to C_7 and C_4, respectively (and split by C_6 and C_5, respectively). The lines centering at 6.89 and 7.44 are quartets (doublet of doublets), due to C_5 and C_6, respectively (and split by C_4 and C_6 and C_5 and C_7, respectively).

4. Close ***indigo.***

* The experimental spectrum comes from the NMRShiftDB. Requires on-line access, source subject to availability.

Infrared Spectrum of *trans*-Stilbene



1. Build *trans*-stilbene, minimize and exit the builder. Alternatively, sketch *trans*-stilbene 2D and exit the sketcher.

2. *Click* on the name ***trans-stilbene*** at the bottom of the screen and then on **Replace**.

3. Select **Spectra** from the **Display** menu. *Click* on in the bar at the top of the spectra pane and select IR Calculated. The calculated IR spectrum of *trans*-stilbene appears in the spectra pane.

4. As with NMR spectra, up and down triangles demark a moveable cursor. Select and move it horizontally over the spectrum. You will see that as you intersect a line in the spectrum, it will turn yellow and the value of the frequency will appear at the bottom. In addition, the molecular model "vibrates" to reflect the motion that the molecule undergoes. Examine the motions of one or more lines of moderate intensity in the vicinity of 1500 cm^{-1} (at 1446, 1494 and 1609 cm^{-1}). You might find it useful to expand the scale (use the scroll wheel) or to shift it (move the mouse horizontally over the spectrum while holding down the left button). You can return to the original settings by *clicking* on in the bar at the top of the pane.

5. *Click* on from the bar at the top of the spectra pane and then IR Experimental. The experimental IR spectrum of *trans*-stilbene obtained from the freely-available NIST database will be superimposed onto the calculated spectrum. You will need to be online. Note that the two spectra are similar although the experimental spectrum exhibits a number of (small) lines not found in the calculated spectrum.

> To see lists of the 3N-6 calculated IR frequencies as well as the (≤ 3N-6) frequencies assigned to the experimental spectrum together with intensities relative to their respective most intense frequency, *click* on at the left of the spectra dialog. *Checking* individual lines (in the calculated IR) moves the cursor on the spectrum and animates the vibrational motion.

6. Close **trans-*stilbene***.

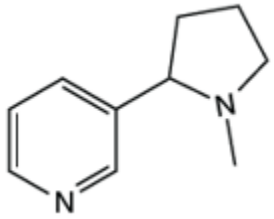
Chapter 6

Graphical Models for Organic Molecules from the Spartan Spectra and Properties Database

The tutorials in this chapter illustrate two of the most commonly-used graphical models, the electrostatic potential map for elucidating molecular charge distributions, and the local ionization potential and LUMO maps for anticipating electrophilic and nucleophilic reactivity, respectively.

Entries in the Spartan Spectra and Properties Database (SSPD) not only include the molecular structure, energy, properties and spectra, but also the wave function. This allows graphical models to be requested and displayed "on-the-fly". You will use the SSPD entries obtained from the ωB97X-D/6-31G* model, rather than calculated results for the required data to complete this chapter.

Electrostatic Potential Map for Nicotine



1. Build nicotine (**File** menu > **New**), minimize the structure (**Build** menu), and return to **3D** view. Alternately, sketch nicotine from the **2D** view and return to **3D**.

2. *Click* on the name ***nicotine*** at the bottom right of the interface. *Click* on **Replace**. The previous structure will be replaced with the entry from the SSPD (including the wave function).

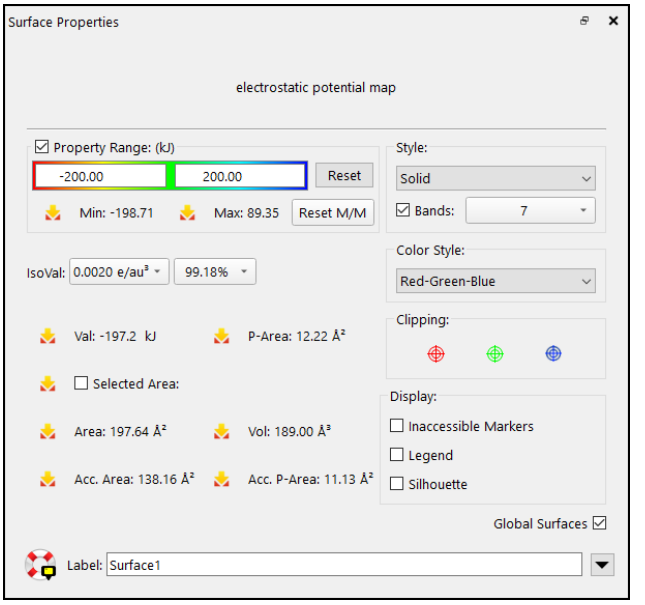
Note: If you see the name ***3-[(2R)-1-methylpyrrolidin-2-yl]pyridine***, instead, go ahead and click replace (you will note in the SSPD preview, next to the model, you also see the word [Enantiomer]. This indicates that you have built 3-[(2R)-1-methylpyrrolidin-2-yl]pyridine, but what exists in the SSPD is this molecule's enantiomer, replacing will *also* result in access to precalculated data for ***nicotine***.

3. Select **Surfaces** from the **Display** menu or *click* on if it appears at the top of the screen. *Click* on **Add** (at the bottom of the **Surfaces** dialog that results) and select **electrostatic potential map** from the menu. This requests an electrostatic potential map (an electron density surface onto which the value of the electrostatic potential is color mapped). The line ***electrostatic potential map*** appears at the top of the dialog.

The graphics calculation will run automatically following your request.* When it completes in a few seconds, *check* the box to the left of ***electrostatic potential map*** in the **Surfaces** dialog. The surface itself corresponds to the electron density and provides a measure of the overall size and shape of nicotine. The colors indicate values of the electrostatic potential on this surface.

By convention, colors toward red correspond to negative potential (stabilizing interaction between the molecule and a positive charge), while colors toward blue correspond to positive potential. The two nitrogen atoms show the largest negative potential (red). You will see that the nitrogen in the pyridine ring is more negative than the nitrogen in the pyrrolidine ring.

4. Quantify your observation. Select **Properties** from the **Display** menu or *click* on if it appears at the top of the screen and *click* anywhere on the electrostatic potential map to bring up the **Surface Properties** dialog.



* If they do not complete automatically, check the box to the left of **Auto-Gen Graphics** under the **Miscellaneous** section in the **Settings** tab in the **Preferences** dialog from the **Options** menu (**Chapter 25**)…

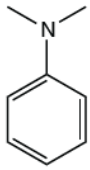
Check the box to the left of **Display Legend** towards the bottom right of the dialog to display the property range on screen. To translate the legend, *click* on the legend to select, then hold down the right mouse button and move the mouse. The legend is useful when making qualitative comparisons of property values. Turn the map such that you can clearly see the pyridine nitrogen and *click* on the area that is *most red*. An arrow marks the point on the surface, and the value of the potential is shown to the right of the legend. Do the same for the pyrrolidine nitrogen.

> Tap on the legend to select. Move two fingers to move it around the screen. Pinch two fingers to make the legend smaller or larger.

5. Nicotine is relatively small and it is easy to associate regions on the map with the underlying molecular skeleton. This becomes more difficult with increasing molecular size. Change the presentation from the **Style** menu located inside the **Surface Properties** dialog or at the bottom right of the screen. Select **Transparent** or **Mesh**. You now can "see through" the map to the underlying molecular structure.

6. Close ***nicotine*** and any open dialogs.

Local Ionization Potential Map for N,N-Dimethylaniline

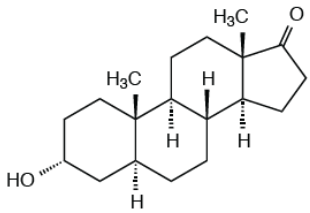


1. Build or sketch N,N-dimethylaniline and *click* on .
2. *Click* on ***N,N-dimethylaniline*** at the bottom of the screen. *Click* on **Replace**. Your structure will be replaced by that in SSPD.

3. Select **Surfaces** from the **Display** menu . *Click* on **Add** at the bottom of the **Surfaces** dialog that results and select **local ionization potential map** from the menu. This requests a map showing the energy required to remove an electron (the "ionization potential") as a function of its location on the electron density surface. Calculation will only take a few seconds. When completed, *check* the box to the left of **local ionization potential map** in the **Surfaces** dialog. The color convention is the same as for the electrostatic potential map, although the scale is completely different. Local ionization potentials are always positive. The default scale (5 to 15 eV) can be changed to highlight the differences. Select **Properties** from the **Display** menu and *click* on the local ionization potential map. Change "5" at the left below **Property** **Range** to "10" and close the dialog. Colors toward red correspond to small ionization potentials (greatest electrophilic reactivity) and colors toward blue correspond to large ionization potentials. Note that the red regions on the map are over the *ortho* and *para* ring positions. This is exactly what is experimentally observed.

4. Close ***N,N-dimethylaniline*** and any open dialogs.

LUMO Map for Androsterone



1. Sketch androsterone 2D and exit the sketcher.

2. *Click* on the name ***androsterone*** at the bottom right of the screen. If the (correct) name is not provided, then you have made a mistake. In this case, re-enter the builder by selecting **Edit Sketch** from the **Build** menu, and make any necessary changes. *Click* on **Replace** to replace your structure by that in the SSPD.

3. Select **Orbital Energies** from the **Display** menu or *click* on if it appears at the top of the screen. An orbital energy diagram will appear at the left of the screen. To examine the molecular orbital corresponding to a line in the diagram, *click* on the line. You will see that the LUMO is a π* orbital localized on the carbonyl group, although it is not at all clear whether it is more or less concentrated on the face away from the two methyl groups. This subtle distinction might be important, for it would be expected that nucleophilic addition would occur from the face on which the LUMO is more concentrated. A |LUMO| map provides a much clearer picture.

4. Select **Surfaces** from the **Display** menu. *Click* on **Add** and select **|LUMO| map** from the menu. The graphics calculation will require only a few seconds. When completed, *check* the box to the left of ***|LUMO| map*** inside the **Surfaces** dialog. The largest (absolute) values of the LUMO are colored blue. Rotate to examine both the faces (with and without the two methyl groups). From the |LUMO| map you can clearly see that the LUMO is more concentrated on the face away from the methyl groups.

> In this example, the LUMO corresponds to the expected π^* orbital primarily localized on the carbonyl bond. This may not always be the case, and it may be necessary to look at higher-energy unoccupied molecular orbitals, LUMO-1, LUMO-2, etc. ***Spartan*** allows any orbital to be superimposed on the electron density surface.

5. Close ***androsterone*** and any open dialogs.

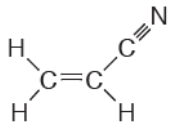
Chapter 7

Spectra, Properties and Graphical Models from Quantum Chemical Calculations

*The tutorials in this chapter provide the earliest and simplest examples of specifying and performing quantum chemical calculations. As in the previous two chapters, they refer to molecules examined in **Chapters 3** and **4**. Calculations are performed utilizing the ωB97X-D/6- 31G* model, and time estimates are based on this model using a quad-core desktop machine (3 GHz Intel I7 processor, 16 GB RAM, 800 GB solid state drive) running in parallel under Windows 10 Pro. Alternate computational models can be used in place of ωB97X-D/6-31G*. Completion times will vary accordingly.*

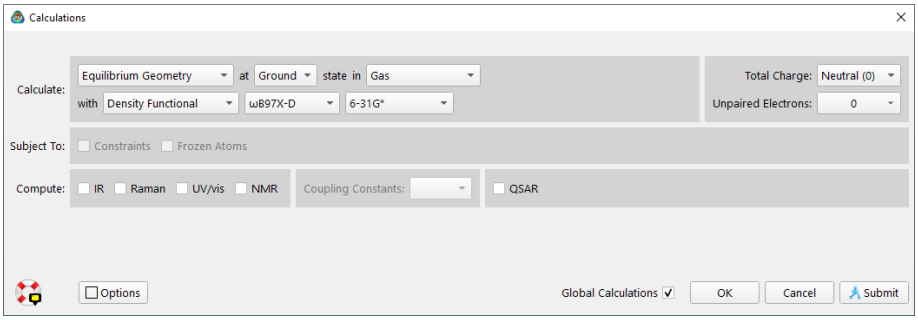
5 min

Acrylonitrile



We return to acrylonitrile for the first tutorial that actually involves quantum chemical calculations.

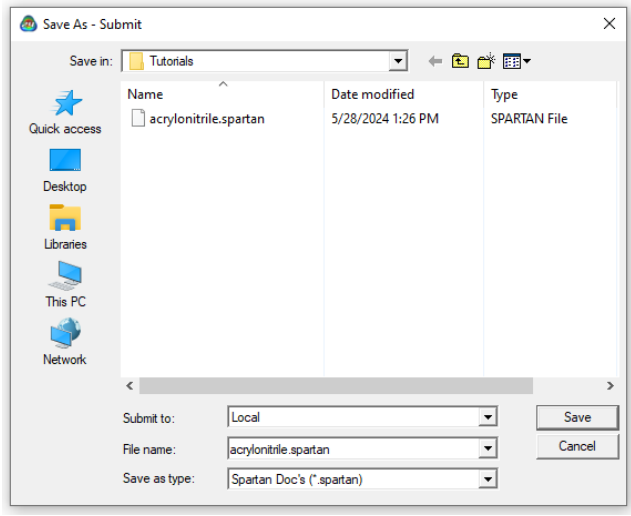
1. Build or sketch acrylonitrile.
2. Select **Calculations...** from the **Setup** menu or *click* on if it appears at the top of the screen and perform the following operations in the **Calculations** dialog that appears.



5. Select **Equilibrium Geometry** from the top menu to the right of **Calculate**. This specifies optimization of equilibrium geometry. At **Ground** state in **Gas** will appear in the menus to the right of **Equilibrium Geometry**.

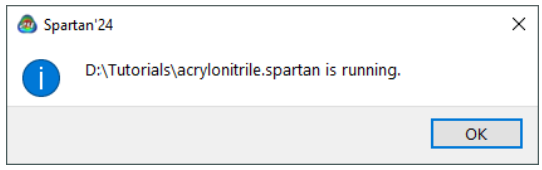
6. Select **Density Functional**, **ωB97X-D** and **6-31G*** from the three bottom menus to the right of **with**. This specifies that the ωB97X-D/6-31G* density functional model will be used for this calculation.

7. *Click* on **Submit** at the bottom of the dialog. A file browser appears.



Because the molecule is in SSPD (even though we will not use the data), the name ***acrylonitrile*** is presented in the box to the right of **File name**.

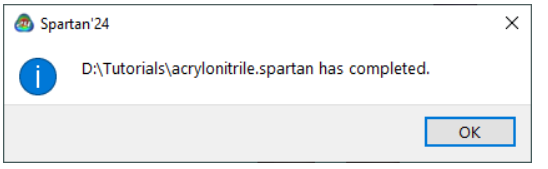
Either use it or type in whatever name you like and then *click* on **Save**. You will be notified that the calculation has been submitted (is running).



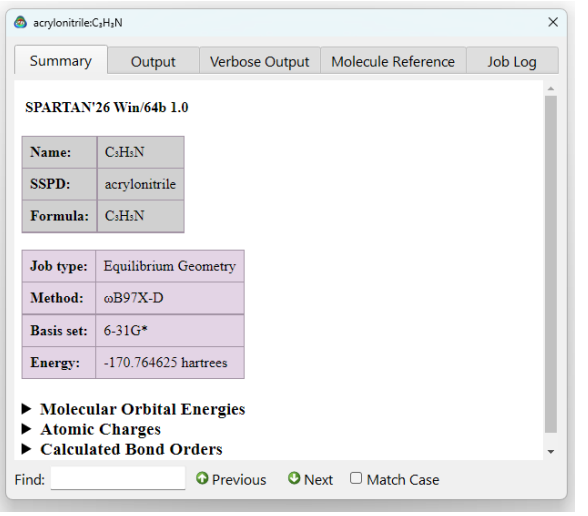
Click on **OK** to remove the message from the screen.

> After a molecule has been submitted, and until the calculation has completed, you will not be permitted to modify any dialogs or other information associated with molecules in the document.

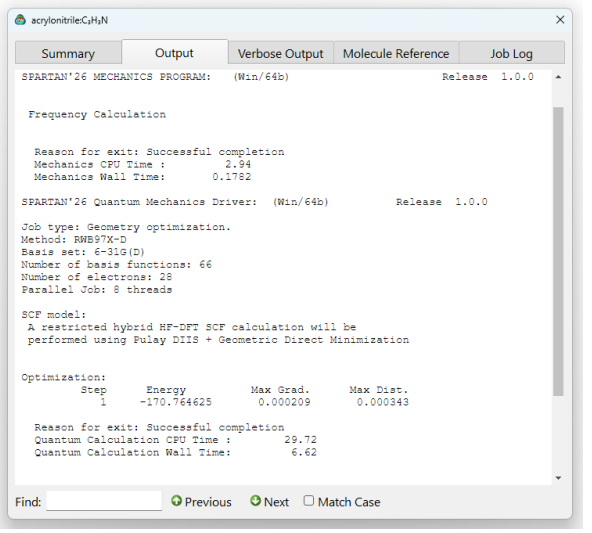
8. You will be notified when the calculation has completed.



Click on **OK** to remove the message from the screen. Select **Output** from the **Display** menu or *click* on if it appears at the top of the screen. Displayed initially is a brief summary (**Summary** tab, describing the task and providing the molecule name and molecular formula together with the total charge (if not neutral) and number of unpaired electrons (if not 0), the calculation model (method and basis set) and the energy.



The summary may be all that you really want (or need) but more detailed information can be obtained by selecting **Output** from the menu at the top of the dialog.



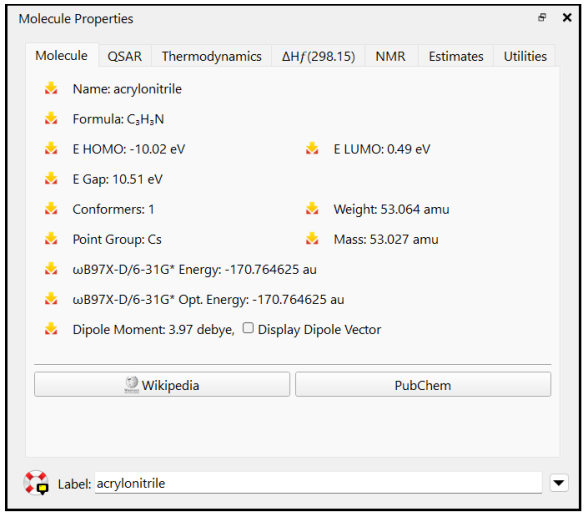
You can scan the output from the calculation by using the scroll bar at the right of the window or by *clicking* (left button) inside the output window and using the scroll wheel on your mouse. Information at the top of the dialog is similar to that reported in the summary, although additional details are provided. Eventually, a series of lines appear, under the heading ***Optimization***. These tell the history of the optimization process. Each line (or ***Step***) provides results for a particular geometry. Ideally, the energy will monotonically approach a minimum value for an optimized geometry. If the geometry was not optimized satisfactorily, an error message, such as: ***OPTIMIZATION has exceeded N steps***, (or similar) will be displayed following the last optimization cycle. Optimization failure is not common, so if it occurs, check to see if you have correctly built the structure.

Near the end of the output is the final total energy* (-170.76462 atomic units for acrylonitrile if you used the ωB97X-D/6-31G* model)**, and the computation time. *Click* on ✕ at the top of the output dialog to close it.

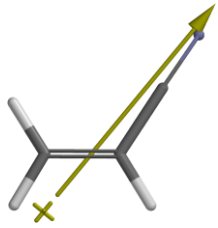
You may examine the total energy and dipole moment among other calculated properties without having to go through the output. Select **Properties** from the **Display** menu to bring up the **Molecule Properties** dialog (make certain that the **Molecule** tab and not the **QSAR**, **Thermodynamics**, **ΔH*f***, **NMR, Estimates** or **Utilities** tab is selected).

* See **Calculations...** (**Setup** menu; **Chapter 21**) for a discussion of how total energy relates to heat of formation and strain energy.

** The number you actually get will likely differ in the last decimal place due to machine precision.

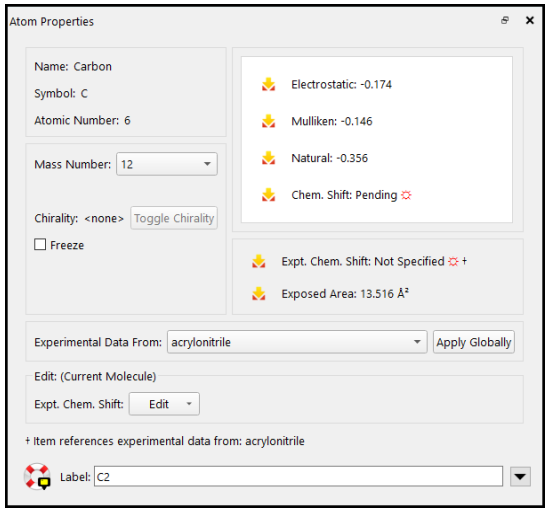


To see the dipole moment vector (indicating the sign and direction of the dipole moment), *check* the box to the left of **Display Dipole Vector**. Wire, ball-and-wire or tube models are best for this display.



Uncheck the box to remove the dipole moment vector.

Click on an atom. The (**Molecule Properties**) dialog will be replaced by the **Atom Properties** dialog.



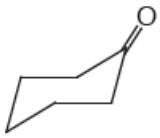
Among other things, this provides three different sets of atomic charges: **Electrostatic**, **Mulliken** and **Natural**. To obtain the charge on another atom, simply *click* on it. Inspect all the atomic charges on acrylonitrile (by *clicking* on the appropriate atoms). When you are finished, *click* on ✕ at the top of the **Atom Properties** dialog to close it.

Select **Surfaces** from the **Display** menu . *Click* on **Add** at the bottom of the **Surfaces** dialog that results and select **electrostatic potential map** from the menu. This requests an electrostatic potential map (an electron density surface on which the value of the electrostatic potential is mapped). The graphics calculation will run without needing to resubmit the job. When it completes in a few seconds, *check* the box to the left of ***electrostatic potential map*** in the **Surfaces** dialog. The surface itself corresponds to the electron density and provides a measure of the overall size and shape of acrylonitrile. The colors indicate values of the electrostatic potential on this surface; by convention, colors toward red correspond to negative potential (stabilizing interaction between the molecule and a positive charge), while colors toward blue correspond to positive potential. The nitrogen (the most electronegative atom) is red; the hydrogens (the most electropositive atoms) are blue.

9. Close ***acrylonitrile*** and any open dialogs.

5 min

Cyclohexanone



This tutorial offers another example of running quantum chemical calculation on a very simple molecule. It also provides an opportunity to further illustrate graphical models for elucidating the stereochemistry of organic reactions.

1. Build or sketch cyclohexanone.

2. Select **Calculations...** from the **Setup** menu . Specify **Equilibrium Geometry**, **Ground** and **Gas** from the top menus to the right of **Calculate**, and **Density Functional**, **ωB97X-D** and **6-31G*** from the menus specifying computational model. *Click* on **Submit** and accept the name ***cyclohexanone***. Wait until the calculation completes before proceeding to the next step.

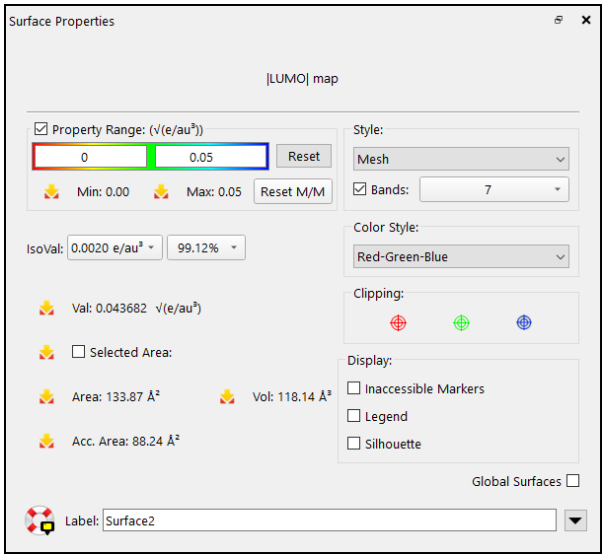
3. Cyclohexanone undergoes nucleophilic attack at the carbonyl carbon, and it is reasonable to expect that the molecule's lowest-unoccupied molecular orbital (the LUMO) will be localized here. To visualize the LUMO, bring up the **Surfaces** dialog (**Surfaces** from the **Display** menu or *click* on . *Click* on **Add** and select **LUMO** from the menu. Also request an electron density surface onto which the (absolute) value of the LUMO has been mapped in color (a so-called LUMO map). *Click* on **Add** and select **|LUMO| map** from the menu. The two graphics calculations will run automatically and will require only a few seconds.

4. *Check* the box to the left of **LUMO** in the **Surfaces** dialog. You will see that the resulting graphic is a π* orbital primarily localized on the carbonyl group, consistent with the fact that nucleophiles (electron pairs) add to the carbonyl carbon. See if you can tell which face of the carbonyl carbon the LUMO is more concentrated on.

5. *Uncheck* the box to the left of **LUMO** in the **Surfaces** dialog (to turn off the display of the LUMO). Then *check* the box to the left of ***|LUMO| map*** to display the electron density surface onto which the (absolute) value of the LUMO has been mapped. By convention, colors toward red indicate small (absolute) values of the LUMO (near zero), while colors toward blue indicate large (absolute) values of the LUMO. We are looking for a "blue spot". Note that it is directly over the carbonyl carbon. This corresponds to the maximum value of the LUMO and is where nucleophilic attack will occur.

6. You will see that the blue spot over the *axial* face of the carbonyl carbon is bigger (and bluer) than that over the *equatorial* face. This indicates preferential attack by nucleophiles onto the *axial* face. Quantify the difference by measuring the (absolute)

value of the LUMO on the two faces. Select **Properties** from the **Display** menu and *click* anywhere on the LUMO map surface to bring up the **Surface Properties** dialog.

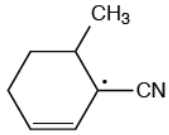


Check the box to the left of **Display Legend** to display the property range on screen. The legend is useful when making qualitative comparisons of property values. Turn the map such that you can clearly see the *axial* face of the carbonyl carbon and *click* on the area of maximum blue. The (absolute) value of the LUMO at the surface point you have selected is provided in the dialog to the right of **Val**. Note the value and then turn the map over such that you can see the *equatorial* face of the carbonyl carbon and *click* on the region of maximum blue on this face. You will find that they support your qualitative conclusions from viewing the image.

7. Close ***cyclohexanone*** and any open dialogs.

5 min

3-Cyano-4-methylcyclohexenyl Radical



3-cyano-4-methylcyclohexene radical will be used to introduce graphical models associated with radicals (molecules with one unpaired electron).

1. Build or sketch 3-cyano-4-methylcyclohexenyl radical.

2. Select **Calculations...** from the **Setup** menu . Specify **Equilibrium Geometry** at **Ground** state in **Gas** from the top menu to the right of **Calculate**, **Density Functional** and ω**B97X-D** and **6-31G*** and from the menus specifying computational model.

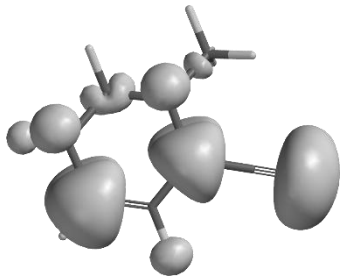
This molecule has one unpaired electron. Change **Unpaired Electrons** from **0** to **1**. *Click* on **Submit** at the bottom of the **Calculations** dialog.* Name it ***3-cyano-4-methylcyclohexenyl radical***. (A name will not be provided as the radical is not in the SSPD.**) Wait (a few minutes) for the calculation to complete before proceeding.

3. Select **Surfaces** from the **Display** menu . *Click* on **Add** and select **spin density** from the menu. *Click* on **Add** and select **spin density map** from the menu. You have requested two different representations of spin distribution. The first presents spin density as a surface of constant value, while the second uses color to map the value of the spin density onto an electron density surface. Finally, request the singly-occupied molecular orbital. *Click* on **Add** one more time and select **aHOMO** (the highest-occupied molecular orbital of α spin that is, the orbital that contains the unpaired electron) from the menu.

4. The three graphics you requested will run without having to resubmit the job and will require only a few seconds to complete. When they are done, *check* the box to the left of ***spin density*** in the **Surfaces** dialog to display the spin density surface. Note that the spin density is delocalized over two of the ring carbons and onto the cyano group.

* ***Spartan*** checks that the structure and the number of electrons specified from **Charge** and **Unpaired Electrons** settings are compatible and if they are not will return an error message.

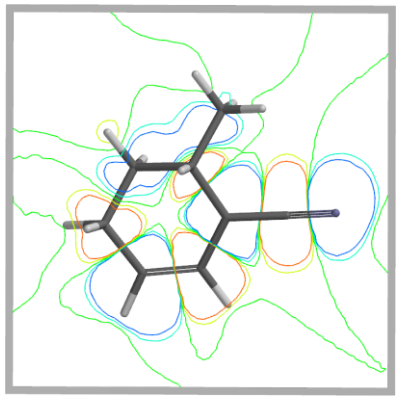
** With very few exceptions, SSPD contains only neutral closed-shell (paired-electron) molecules.



5. Remove the spin density surface (*uncheck* the box to the left of ***spin density*** in the **Surfaces** dialog) and then *check* the box to the left of ***spin density map*** to display a surface on which the spin density is mapped onto the electron density. Note that the areas of maximum spin (colored blue) closely match those where the surface is large in the previous image.

6. Remove the spin density map (*uncheck* the box to the left of ***spin density map***), and then *check* the box to the left of ***aHOMO*** (the molecular orbital which holds the unpaired electron). Aside from the colors (different signs of the orbital), note that this graphic is nearly identical to the previously-displayed image of the spin.

7. *Uncheck* the box to the left of ***aHOMO***. *Click* on **More Surfaces...** at the bottom of the (**Surfaces**) dialog, and select **Slice** from the **Surfaces** menu and ***spin density*** from the **Properties** menu. *Click* on **OK**. A new line ***Slice***, ***spin density*** appears in the window at the top of the dialog.* Select it by *checking* the box at the left. A plane (a slice of spin density) surrounded by a frame appears in the middle of the model on screen. *Click* inside the frame to select. The frame will turn gold. Position the cursor outside the frame, then use the scroll wheel to zoom the plane. You can also translate and rotate the plane independently of the molecule using the usual mouse operations. Alternatively, you can move the molecule and plane together by first *clicking* on the molecule (the frame will now turn white) and then using the mouse. For all operations, be certain to keep the cursor positioned outside of the frame. Size and orient the slice as you wish.

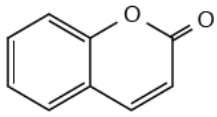


8. Close ***3-cyano-4-methylcyclohexenyl radical*** and any open dialogs.

* You can change the display style from **Contours** to **Solid** or **Transparent** using the **Style** menu at the bottom right of the screen. This will appear only when the slice is selected.

5 min

Coumarin



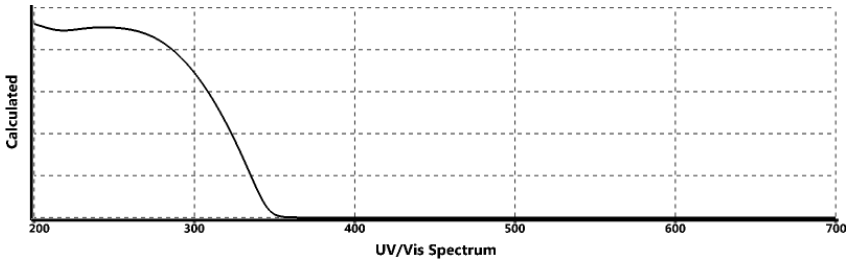
Coumarin illustrates the calculation of a UV/visible spectrum.

1. Build or sketch coumarin.

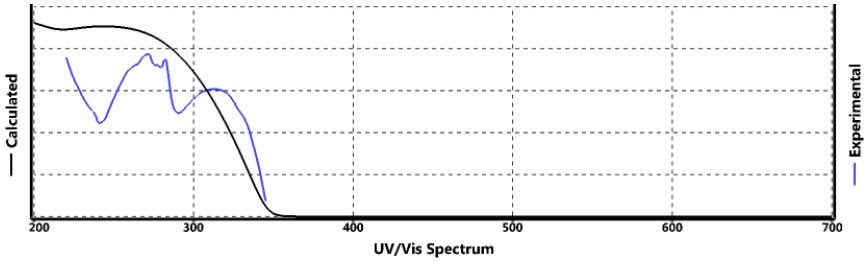
2. The name ***coumarin*** appears at the bottom of the screen. *Click* on it, make sure that **ωB97X-D/6-31G*** is selected from the dialog that appears and *click* on **Replace**.

3. Select **Calculations...** from the **Setup** menu . Select **Energy** from the top menu to the right of **Calculate**. (**Density Functional**, **ωB97X-D** and **6-31G*** should already be selected from three menus immediately below.) You already have the equilibrium structure from SSPD and only need to obtain the energy and wave function as the basis of a UV/visible spectrum calculation. *Check* the box to the left of **UV/vis** (to the right of **Compute**). *Click* on **Submit**. Accept the name ***coumarin***.

4. The calculation will take several minutes. While coumarin is relatively small, calculations on several low-lying excited states in addition to the ground state are required to produce a UV/visible spectrum. When completed, select **Spectra** from the **Display** menu . *Click* on inside the spectra pane and select . The calculated UV/visible spectrum will appear in the spectra pane.



5. *Click* again on and this time select . The experimental UV/visible spectrum from the freely available NIST database will be superimposed onto the calculated spectrum.



6. Close ***coumarin***.

Section IV
Advanced Tutorials

The tutorials in the chapters in this section build upon and extend those in the previous section. **Groups of Organic Molecules** introduces spreadsheets, associated statistics, and plot capabilities. **Spectra of Organic Molecules** provides "real" examples of calculated IR and NMR Spectra. **Inorganic and Organometallic Molecules** illustrates applications to "non-organic" molecules, in particular, molecules incorporating transition metals. **Organic Reactions** describes and illustrates strategies for finding and verifying a reaction transition state. **Medicinal Chemistry** illustrates examples of interactions of small molecules (drugs) in biological systems. **Flexible Molecules** addresses issues involved in identifying the lowest-energy conformer of a flexible molecule and in establishing the relative energies of alternative (higher-energy) conformers, the latter being necessary to "average" properties and spectra. Some of the tutorials in this section require significant computation time, in particular, those involving quantum chemical methods applied to flexible molecules. With this in mind, a number of tutorials have been provided in the last chapter "**Dry Labs". Using the Spartan Spectra and Properties Database** to further illustrate a variety of important features without having to consume computer time (or your time waiting for the calculations to finish).

Quantum chemical calculations utilize the ωB97X-D/6-31G* model. Time estimates are given for completion of the tutorials based on this model using 24-core laptop (5.4 GHz Intel i9 processor, 32 GB RAM, 2 TB solid state drive) running in parallel under Windows 11. Other computational models can be used in place of ωB97X-D/6-31G*. Completion times will vary accordingly.

Icons are provided for all menu entries as they are requested. Icon displays may be customized according to user preference (**Icons** tab under **Preferences** in the **Options** menu; **Chapter 25**).

Chapter 8
Groups of Organic Molecules

The tutorials in this chapter introduce and illustrate a number of basic operations involved in processing groups of molecules, as well as the associated spreadsheet for organizing and fitting data and facilities for making plots.

Computational investigations, like experimental investigations, are rarely restricted to a single molecule, but typically involve a series of related molecules. Here, it may be of interest to compare geometries, energies or other calculated properties, or to compare trends in calculated and measured properties. ***Spartan*** provides facilities for these purposes. In particular, it allows molecules to be grouped, either manually, or automatically as a result of a conformational search, from following a particular vibrational motion, or from a scan of one or more geometric variables. Once grouped, molecules may be aligned to make similarities or differences more apparent based either on their structure, chemical functionality or atom labels. Calculations may be performed either on individual molecules or, just as simply, on the complete group of molecules. On multi-core computers, several molecules can be run at once. The results of calculations on a group can be examined and analyzed individually or altogether to seek out trends. In particular, reaction and activation energies may be obtained and statistical measures, among them the DP4 measure, may be applied to calculated NMR spectra.

Associated with a group of molecules (or an individual molecule) is a spreadsheet. This allows convenient access to virtually any calculated quantity that can be given a numerical value. Additionally, data may be entered manually into the spreadsheet or transferred from another application such as Excel. Data in the spreadsheet may be manipulated, linear regression analyses performed and plots displayed. Alternately, the data in a ***Spartan*** spreadsheet may be transferred to Excel (or other applications) for further analysis. An Excel worksheet can be

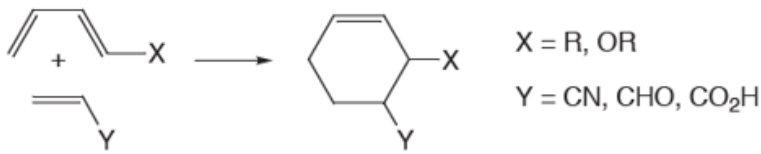
stored inside a ***Spartan*** document if desired (see **Embedded Data**
under the **File** menu; **Chapter 16**).

The tutorials in this chapter introduce and illustrate a number of the basic group operations available in ***Spartan***. These include building a group from scratch and processing groups resulting from both a conformational search and from systematically varying a torsion angle. Also provided is an example of fitting an experimental observable to one or more calculated properties by way of linear regression. Finally, **Reactions** dialog is illustrated both as it operates on molecules in a group and on molecules derived from substitution of grouped molecules.

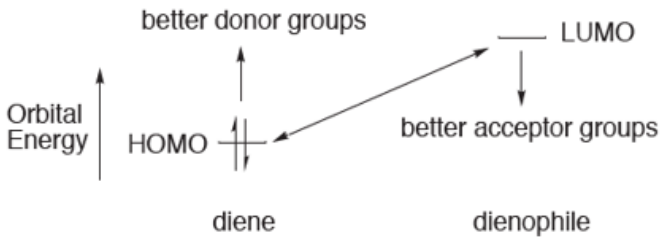
20 mins

Dienophiles in Diels-Alder Cycloadditions

The most common Diels-Alder reactions involve electron-rich dienes and electron-deficient dienophiles.



The rate of these reactions generally increases with increasing π-donor ability of the diene substituent, and with increasing π-acceptor ability of the dienophile substituent. This can be rationalized by noting that donor groups raise the energy of the highest-occupied molecular orbital (the HOMO) on the diene, while acceptor groups lower the energy of the lowest-unoccupied molecular orbital (the LUMO) on the dienophile. Thus, the HOMO-LUMO gap is reduced, leading to increased interaction of diene and dienophile and enhanced reactivity.



To test this hypothesis, you will examine whether or not experimental relative rates of Diels-Alder cycloadditions involving cyclopentadiene and a variety of cyanoethylenes correlate with dienophile LUMO energies.

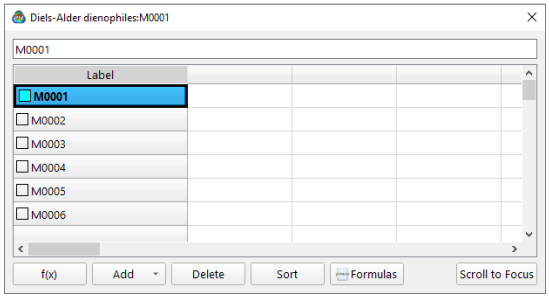
1. Build and minimize acrylonitrile, $H_2C{=}C(H)CN$. Select **View** from the **Build** menu . Copy the structure to the clipboard. Either *right click* on the background and choose **Copy** from the resulting contextual menu or select **Copy** from the **Edit** menu .

2. Select **New Molecule** (not **New**) from the **File** menu . This allows for adding an additional molecule to the current document. The screen will be cleared and the model kit displayed. *Click* on **Clipboard** at the bottom of the model kit and *click* on screen. Acrylonitrile will appear. *Click* on **Groups** in the model kit, select **Cyano** and add to the appropriate free valence on acrylonitrile to make 1,1-dicyanoethylene. *Click* on .

3. Repeat this procedure (**Build New Molecule**, followed by **Clipboard**, followed by **Groups**, followed by four more times to build *cis* and *trans*-1,2-dicyanoethylene, tricyanoethylene and tetracyanoethylene. When you have built all six molecules, *click* on .

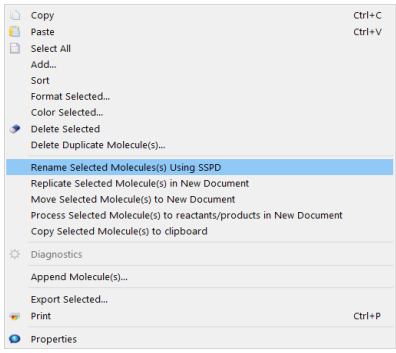
> The 2D sketcher could be used in place of the 3D builder. Sketch acrylonitrile, right *click* on screen and select **Copy** from the menu that appears. Select **Sketch New Molecule** (not **New Sketch**) from the **File** menu, right *click* on screen and select **Paste** from the menu. Add a cyano group to make 1,1-dicyanoethylene. Repeat the sequence to make the remaining molecules and finally *click* on .

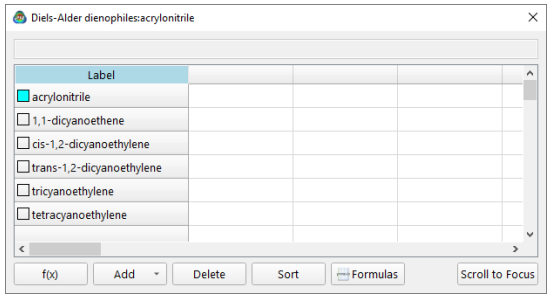
4. The molecules have been grouped together, that is, put into a single document, allowing calculated properties to be accessed via a spreadsheet. Select **Spreadsheet** from the **Display** menu.



To select an individual molecule in the group, *click* on its label (**M0001**, **...**) in the left-hand column, or use the ◄ and ►keys at the bottom left of the screen. *Left-click* on the header cell ("**Label**") to select the entire group, then *right-click* and select **Rename Selected Using SSPD** from the contextual menu that results.

This results in replacement of the default labels by proper names from the Spartan Spectra and Properties Database.





5. Select **Calculations...** from the **Setup** menu and specify **Equilibrium Geometry** at **Ground State** in **Gas** using the ω**B97X-D/6-31G*** model Make certain that **Global Calculations** at the bottom of the dialog is *checked* to ensure that the settings will be applied to all the molecules in the list.

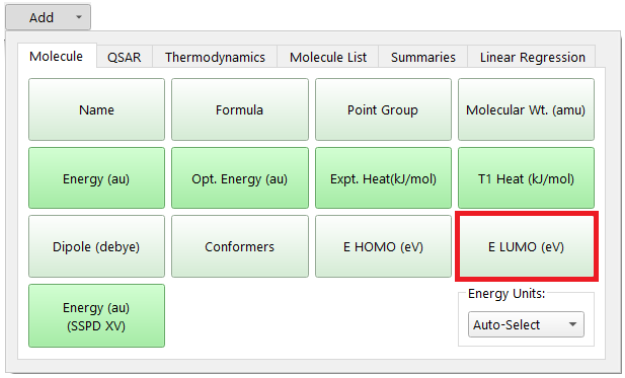
6. *Click* on the **Submit** button at the bottom of the **Calculations** dialog. Name it ***Diels-Alder dienophiles*** it will take several minutes for this calculation to complete. After it completes, enter the following experimental relative rates into the spreadsheet.

Experimental relative rates for Diels-Alder cycloadditions of cyclopentadiene*			
dienophile	**$\log_{10}$ (relative rate)**	**dienophile**	**$\log_{10}$ (relative rate)**
acrylonitrile	0	1,1-dicyanoethylene	4.64
cis-1,2-dicyanoethylene	1.94	*trans*-1,2-dicyanoethylene	1.89
tricyanoethylene	5.66	tetracyanoethylene	7.61

* J. Sauer, H. Weist and A. Mielert, *Chem. Ber.*, **97**, 3183 (1964).

Double click inside the header cell for a blank column in the spreadsheet, type **Log(rate)=** and *press* the **Enter** key (**return** key on Mac). You need to press the **Enter** (**return**) key following each entry (or use the key). Sort by relative rate. *Click* on the column header **Log(rate)**, and then *click* on **Sort** at the bottom of the spreadsheet.

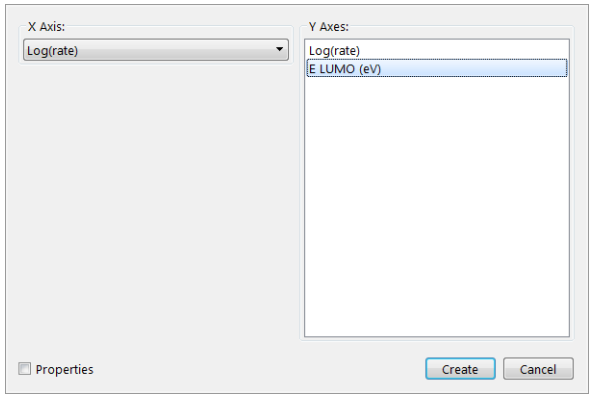
7. *Click* inside a header cell for a blank column and *click* on **Add...** at the bottom of the spreadsheet (alternatively, *right click* inside the header cell for a blank column in the spreadsheet, and select **Add...** from the contextual menu that results). *Click* on the **E LUMO** button under the **Molecule** tab.

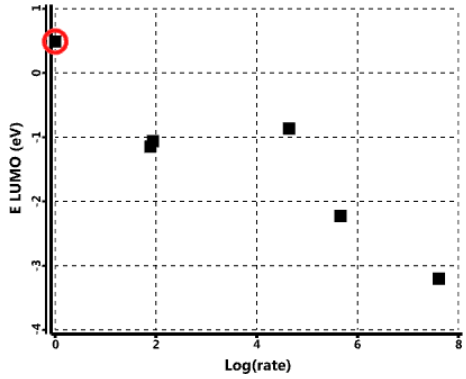


The spreadsheet, which contains both the calculated LUMO energies and experimental relative rates, has served its purpose. Remove it from the screen by *clicking* on at the top).

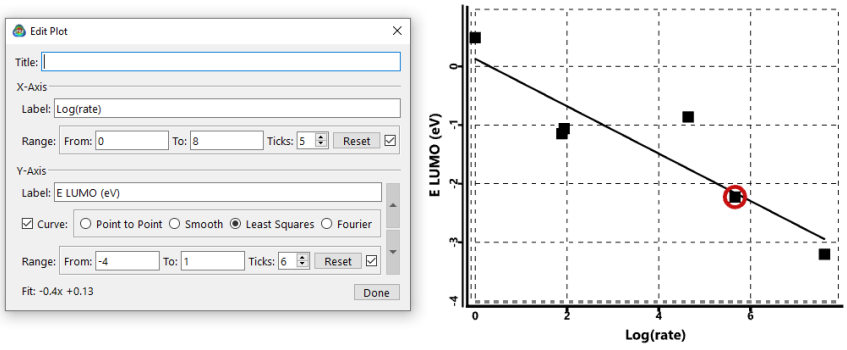
8. Select **Plots...** from the **Display** menu . This leads to an empty plot pane at the right of the screen. *Click* on in the toolbar (top of the pane). Select **Log(rate)** from the list of items in the **X Axis** menu and **E LUMO(eV)** from the **Y Axes*** list. Then, *click* on **Create**.

* The Y axis specification is plural "Axes" as users can generate plots with multiple Y axes displayed. If more than one Y axis is specified, color is used to distinguish data points and plot lines.





9. By default, only data points are displayed. *Click* on in the toolbar at the top of the pane. This leads to display of the **Edit Plot** dialog. *Check* the box next to **Curve**, *click* on the button to the left of **Least Squares** and *click* on **Done** to display a least squares fit of the reaction rates to LUMO energies. The least- squares line will be drawn.

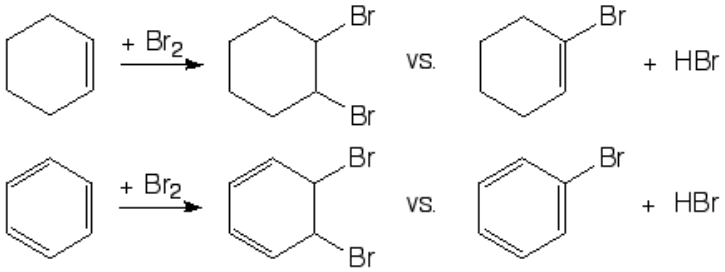


10. Close ***Diels-Alder dienophiles*** and any open dialogs.

no calculations

Addition vs. Substitution

Alkenes normally undergo addition reactions whereas aromatic compounds normally undergo substitution reactions. For example, bromine reacts with cyclohexene to give *trans*-1,2-dibromocyclohexane (the addition product) not 1-bromocyclohexene, whereas it reacts with benzene to give bromobenzene (the substitution product) not *trans*- 5,6-dibromo-1,3-cyclohexadiene.

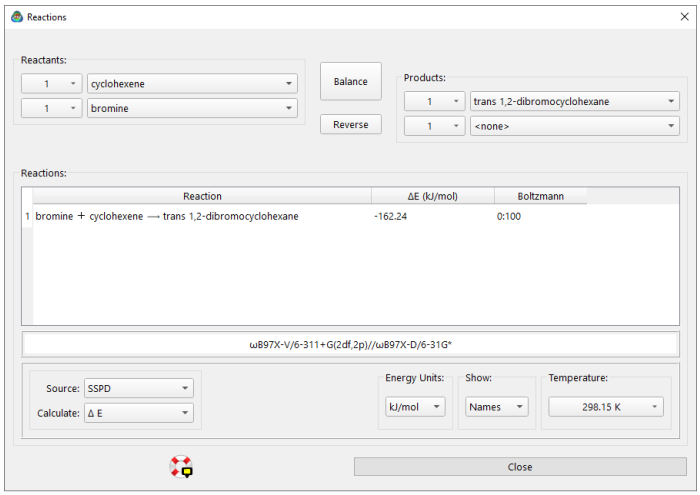


In this tutorial you will use results from the Spartan Spectra and Properties Database (SSPD) as accessed from ***Spartan's*** reaction energy calculator to establish the preferred product for each reaction. Specifically, energies from the ωB97X-V/6-311+G(2df,2p) density functional model stored in SSPD will be employed. No calculations are involved.

1. One after another, build or sketch cyclohexene, *trans*-1,2- dibromocyclohexane, 1-bromocyclohexene, benzene, *trans*- 5,6-dibromo-1,3-cyclohexadiene, bromobenzene, bromine (Br_2) and hydrogen bromide (eight molecules in total). Put all in the same document. Use **New** to or sketch for the first molecule and **New Molecule** to build or sketch each successive molecule.

2. Select **Spreadsheet** from the **Display** menu . *Left click* inside the header cell for the leftmost column, then *right click* and select **Rename Selected Using SSPD** in the menu that appears. *Click* on **OK** in the dialog that results. The molecules can now be referenced using chemical names. Close the spreadsheet.

3. Select **Reactions** from the **Display** menu .



Choose **SSPD Database** from the menu to the right of **Source** and **ΔE** from the menu to the right of **Calculate** at the bottom of the dialog. Compute the energy for Br_2 addition to cyclohexene: select ***cyclohexene*** and ***bromine*** as **Reactants** and ***trans-1,2- dibromocyclohexane*** (leave the second product selection at ***<none>***) as **Products**. Repeat for the corresponding substitution energy (same reactants but the products are 1-bromocyclohexene and hydrogen bromide) and for both addition and substitution reactions of benzene (reactants are ***benzene*** and ***bromine*** and products are ***trans-5,6-dibromo-1-3-cyclohexadiene*** and ***<none>*** for addition and ***bromobenzene*** and ***hydrogen bromide*** for substitution).

Are all reactions thermodynamically favorable (*exothermic*)? Identify any reactions that are not and provide a rationale as to why. Why is there a change in preferred reaction in moving from the alkene to the arene?

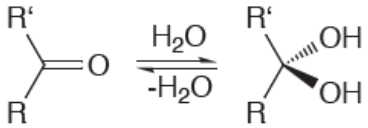
> You can, if you wish, also calculate ΔH and ΔG for these reactions. The first adds zero-point energy and temperature corrections to Energy and the second also adds the entropy. In both cases, ωB97X-D/6-31G* is used to furnish vibrational contributions.

4. Close the document and any open dialogs.

no calculations

Hydration of Carbonyl Compounds

The hydration of carbonyl compounds has been extensively studied primarily because it serves as a model for a number of important reactions, nucleophilic addition to carbonyl compounds among them. In this tutorial, you will use ***Spartan***'s linear regression analysis tool to correlate calculated properties of carbonyl compounds with measured equilibrium constants for their hydration. The molecules are all in SSPD so no calculations are needed.



$$K_{eq} = \frac{[\text{hydrate}]}{[H_2O]\,[\text{carbonyl}]} = \frac{[\text{hydrate}]}{55.5\,[\text{carbonyl}]}$$

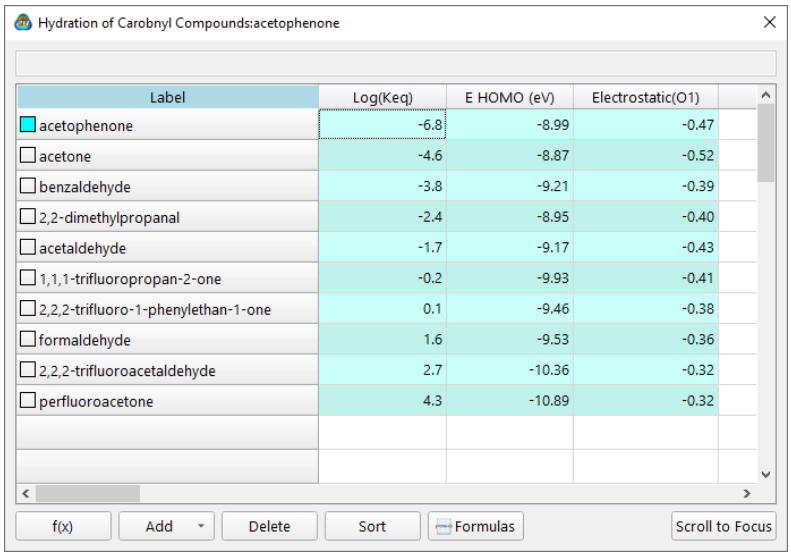
1. Build or sketch all the compounds listed in the next page for which experimental equilibrium constants are available. Build or sketch, start with **New** from the **File** menu for the first molecule and **New Molecule** from the **File** menu for each successive molecule.

2. *Click* on the name of whichever molecule is selected at the bottom of the screen. Confirm that the ω**B97X-D/6-31G*** model is selected and *click* on **Replace** in the dialog that results and then on **All** to replace all molecules with entries from the SSPD.

3. Select **Spreadsheet** from the **Display** menu to bring up the spreadsheet. *Double click* inside the header cell of an empty column, *type* **Log(Keq)** and *press* the **Enter** key (**return** key on Mac). Then enter the experimental equilibrium constants from the table below.

Experimental K_{eq} for hydration of carbonyl compounds			
	log(k_{eq}/55.5)		log(k_{eq}/55.5)
$PhCOCH_3$	-6.8	CF_3COCH_3	-0.2
CH_3COCH_3	-4.6	$PhCOCF_3$	0.1
PhCHO	-3.8	H_2CO	1.6
t-BuCHO	-2.4	CF_3CHO	2.7
CH_3CHO	-1.7	CF_3COCF_3	4.3
* J.P. Guthrie, Can. J. Chem., **53**, 898 (1975); **56**, 962 (1978).			

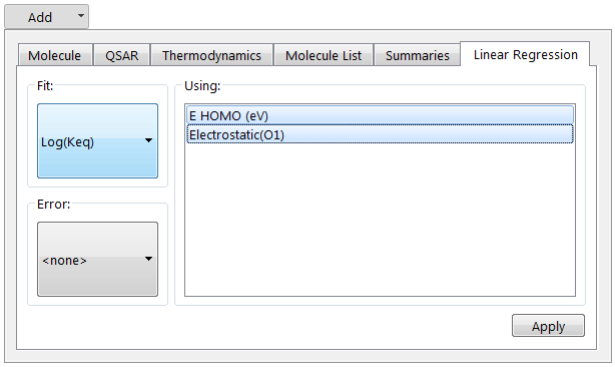
You need to *press* the **Enter** (**return**) key following each data entry. Sort the list according to the value of Log(Keq). *Click* inside the header cell **Log(Keq)** and then *click* on **Sort** at the bottom of the spreadsheet.



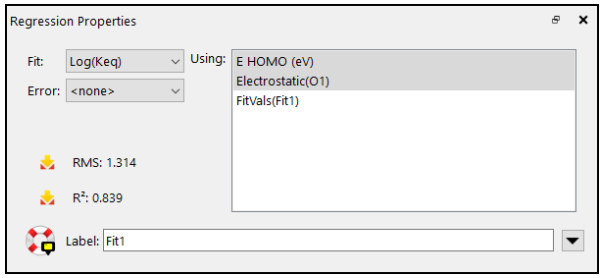
4. Select **Properties** from the **Display** menu, click the icon to the left of **E HOMO**.

5. *Click* on the oxygen atom for whatever compound is displayed. *Click* on to the left of **Electrostatic** (under **Charges**) in the **Atom Properties** dialog (that has replaced the **Molecule Properties** dialog), to place oxygen charges into the spreadsheet.

6. *Click* on **Add...** at the bottom of the spreadsheet, and then *click* on the **Linear Regression** tab at the top of the dialog that results.



Select **Log(Keq)** from the **Fit** menu and both **EHOMO (eV)** and **Electrostatic (O1)*** from the **Using** list. *Click* on **Apply**. A new row (**Fit1**) will be added at the bottom of the spreadsheet. Select **Properties** from the **Display** menu, *click* anywhere on the **Fit1** row. The **Regression Properties** dialog appears.



This provides information about the fit of Log(Keq) to the charge on oxygen, in particular, the value of R^2. The closer to unity, the better the fit.

7. Select **Plots** from the **Display** menu. *Click* on in the bar at the top of the plots pane. Select **Log(Keq)** from the **X Axis** list and **FitVals(Fit1)** from the box below **Y Axes** and *click* on **Add**.

8. To see the correlation (measured and fit rates), *click* on **Edit** from the toolbar at the top of the pane. *Check* the **Curve** box and *click* the **Least Squares** button. Finally, *click* the **Done** button.

9. Close the document and any open dialogs.

* Your numbering may differ as it depends on the order that atoms were introduced during building, but the reference is to the carbonyl oxygen.

no calculations

Acidities of Carboxylic Acids

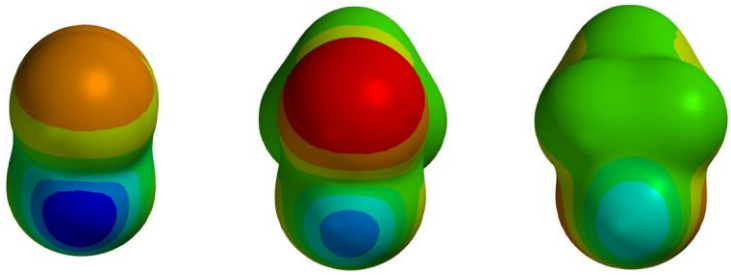
Acid strength is among the most important molecular properties. It is readily available from calculation, either in terms of absolute deprotonation energy, or, more commonly,

$$AH \longrightarrow A^- + H^+$$

as the deprotonation energy relative to that of some standard acid (A^RH).

$$AH + A^{R-} \longrightarrow A^- + A^RH$$

The energy calculations discussed so far apply strictly to gas-phase acidities* and likely do not reflect acidities in solution. This prompts a search for alternative descriptions. One possibility is the value of the electrostatic potential in the vicinity of the acidic hydrogen in the neutral acid. Electrostatic potential maps would certainly be expected to reveal gross trends in acidity, for example, the acidic hydrogen in a strong acid, such as nitric acid, should be more positive than that in a weak acid, such as acetic acid, which in turn should be more positive than that in a very weak acid, such as ethanol.



Nitric acid acetic acid ethanol

In this tutorial, you will use electrostatic potential maps to quantify changes in acid strength due to subtle variations in structure. You will use the Spartan Spectra and Properties Database (SSPD) to eliminate the need for quantum calculations.

* This is not to say that quantum chemical treatments of solvated molecules have not been developed or have not been implemented within ***Spartan***, but rather that in our opinion are not reliable enough to provide quantitative data.

1. One after another, build or sketch trichloroacetic, dichloroacetic, chloroacetic, formic, benzoic, acetic and pivalic acids. Put all molecules into the same document. Use **New** from the **File** menu for the first molecule and use **New Molecule** or from the **File** menu for each successive molecule. *Click* on when you are done.

2. All of the molecules that you have built are available in SSPD. *Click* on the name of whichever molecule is selected at the bottom of the screen, make sure that **ωB97X-D/6-31G*** is selected from the menu in the dialog that results, *click* on **Replace**, and finally on **All**. Structures obtained from ωB97X-D/6-31G* density functional calculations will replace those you have built. The wave function is also provided with the database entry.

3. Select **Spreadsheet** from the **Display** menu . Expand it so that you can see all seven molecules, and that two data columns are available. *Double click* inside the header cell of the first available data column, *type* **pKa** and *press* the **Enter** key (**return** key on Mac). Enter the experimental pK_a's (given in the table below) into the appropriate cells under this column. You need to *press* the **Enter** (**return**) key following each entry.

acid	pK_a	acid	pK_a
trichloroacetic (Cl_3CCO_2H)	0.7	benzoic ($C_6H_5CO_2H$)	4.19
dichloroacetic (Cl_2CHCO_2H)	1.48	acetic (CH_3CO_2H)	4.75
chloroacetic ($ClCH_2CO_2H$)	2.85	pivalic ($(CH_3)_3CCO_2H$)	5.03
formic (HCO_2H)	3.75		

Experimental data from: E.P. Sargeant and B. Dempsey, ***Ionization Constants of Organic Acids in Aqueous Solution***, IUPAC no. 23, Permagon Press, 1979.

4. To display all molecules at once, *check* the box to the left of the molecule name (**Label** column) in the spreadsheet for each entry. To manipulate the molecules independently of one another, *click* to deselect **Coupled** from the **Model** menu . Arrange the seven molecules on screen such that you can clearly see the acidic hydrogen on each.

5. Select **Surfaces** from the **Display** menu and then *click* on **Add** and select **electrostatic potential map**. When the electrostatic potential map calculations complete (they will be marked "completed"), *check* the box at the left of ***electrostatic potential map*** in the **Surfaces** dialog. Select **Properties** from the **Display** menu and *click* on the electrostatic potential map to display the **Surface Properties** dialog. *Click* on (the **Post** icon) to the left of **Max** inside the **Surface Properties** dialog. The maximum value of the electrostatic potential (corresponding to the acidic proton) will be posted to the spreadsheet. Leave the dialog on screen.

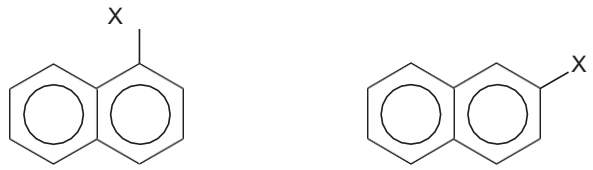
6. Plot experimental pK_a vs. maximum in the electrostatic potential. Select **Plots** from the **Display** menu and then select **pKa** under the **X Axis** menu and **Property Max (Surface)** from the **Y Axes** list. *Click* on **Create**. *Click* on to edit the plot and *check* the box next to **Curve**. Choose the **Least Squares** option and *click* on **Done**. Does there appear to be a correlation between pK_a and the maximum value of the electrostatic potential?

7. Close the document and any open dialogs.

no calculations

Positional Selectivity of Substituted Naphthalenes

Thermochemical stabilities of positional isomers may depend on several factors, including a tendency to minimize unfavorable non- bonded intramolecular interactions (***sterics***) and to minimize overall dipole moment (***electrostatics***). Naphthalene offers a good example with only two different positions.

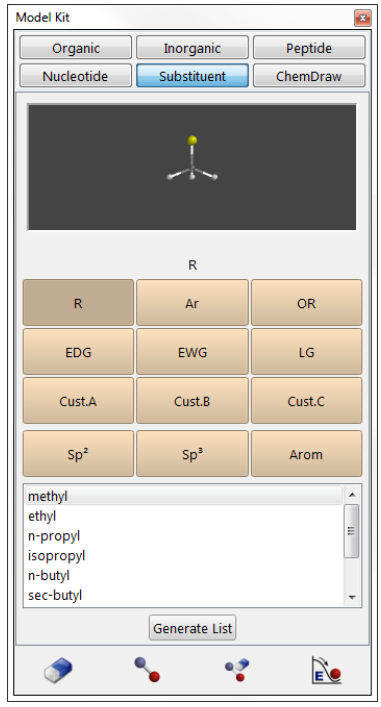


1-substituted naphthalene 2-substituted naphthalene

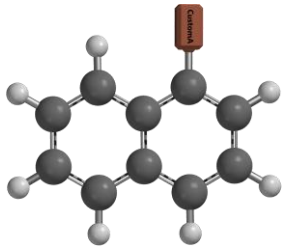
In this tutorial you will use ωB97X-V/6-311+G(2df,2p) energies from the Spartan Spectra and Properties Database (SSPD) to establish differences in heats of formation between 1 and 2-substituted naphthalenes. Instead of building a series of substituted naphthalenes, you will use ***Spartan***'s substituent model kit to construct one "molecule" marked for substitution on the 1-position and the other "molecule" for substitution on the 2-position. No calculations are involved.

1. One after another build methane, ammonia, water, hydrogen fluoride, hydrogen cyanide and formic acid. Put all in the same list (use **New** to enter the builder for the first molecule and **New Molecule** for each successive molecule). Select **spreadsheet** from the **Display** menu, *double click* inside the leftmost cell for each molecule and replace the identifier (***M0001***, …) by proper functional group names (***methyl***, ***amino***, ***hydroxy***, ***fluoro***, ***cyano*** and ***carboxylic acid***). Put the contents on the clipboard. *Click* in the header cell to select all molecules, then *right-click* and select **Copy** from the contextual menu that results. Alternatively, select **Copy** from the **Edit** menu. Close the document (it is not necessary to save the document).
2. Select **New** from the **File** menu and build naphthalene.

Then, select **Substituent** from the menu at the top of the model kit to bring up the substituent builder.



Click on **Cust. A** near the middle of the model kit, then *right click* inside the box near the bottom and select **Paste** from the contextual menu that results. Alternatively, left *click* inside the box and select **Paste** from the **Edit** menu . Names of the six functional groups that you built in the previous step will appear in the box. As you *click* through the list, ball-and-wire models of their structures will appear in the small screen at the top of the model kit. Note that one of the hydrogens (white balls) for each is highlighted in gold. This hydrogen will be removed to make a free valence for attachment. The carboxylic acid group has two different hydrogens, one on carbon (leading to a carboxylic acid) and one on oxygen (leading to a formate). If the hydrogen on the carbon is not already highlighted, *click* on it and it will be highlighted. *Click* on the free valence on naphthalene that will lead to 1-substituted naphthalenes. The molecule on screen will now appear with a marker to indicate substitution at the 1-position.



3. Select **New Molecule** from the **File** menu . Select **Organic** from the tab at the top of the model kit and again build naphthalene. Select **Substituent** from the menu to bring the substituent model kit. Make certain that **Cust. A** is selected (if not, *click* on it). *Click* on the free valence on naphthalene that will lead to 2-substituted naphthalenes. *Click* on to remove the model kit.

4. Select **Spreadsheet** from the **Display** menu . *Double click* inside the leftmost cell for the first molecule and replace the default identifier (**M0001**) with ***1-substituted naphthalenes***. *Double click* inside the corresponding cell for the second molecule and replace the default identifier by ***2-substituted naphthalenes***.

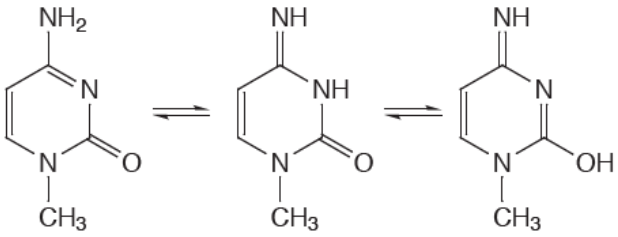
5. Select **Reactions** from the **Display** menu . Choose **SSPD Database** from the menu to the right of **Source** and **ΔE** from the menu to the right of **Calculate**. This signifies that energies obtained from the ωB97X-V/6-311+G(2df,2p) model associated with each of the entries in SSPD will be employed. Select ***1-substituted naphthalenes*** as **Reactants** and ***2-substituted naphthalenes*** as **Products** and *click* on **Compute** at the bottom of the dialog. Reactions will be written (and reaction energies computed) for all six substituents. The **Boltzmann** weights column provides a weighting of positional preferences.

6. Close the document and any open dialogs.

30 mins

Tautomers of Nucleotide Bases

Protons bound to heteroatoms in heterocyclic compounds are likely to be very mobile in solution and, where two or more heteroatoms are present in a structure, different isomers (tautomers) may be in equilibrium. Consider the nucleotide base, like cytosine, but where a methyl group has replaced the sugar-phosphate backbone at the 1-position.



The existence of a low-energy tautomer could have far-reaching consequences, given that the valence structure of cytosine is key to hydrogen bonding in DNA. In this tutorial, you will examine the possible tautomers of 1-methylcytosine for evidence of low-energy structures.*

1. Build or sketch 1-methylcytosine, the **SSPD** identifies this as 4-amino-1-methyl-1,2-dihydropyrimidin-2-one.

2. Note that the word **Tautomer** appears at the bottom right of the screen, indicating that tautomers exist. Select **Generate Tautomers** from the **Geometry** menu. Step through the tautomers using the ◀ and ▶ keys that appear at the bottom right of the screen. To put the tautomers in a list, *click* on ▼ at the bottom right of the screen and *click* on **OK** in the on-screen message "**Generate a Tautomer List**" that results.

3. Select **Calculations...** from the **Setup** menu with the list of tautomers. (To avoid confusion, it is a good idea to close the molecule you used to get the tautomers.) Specify calculation of **Equilibrium Geometry** using the **ωB97X-D/6-31G*** density functional model. Submit the job with the name ***cytosine tautomers***.

* Note, however, that were the energy of an alternative tautomer only 10 kJ/mol higher than that for the normal structure, this would translate into a relative abundance of only about 1% at room temperature. Thus, any alternative tautomers would need to be very close in energy to the lowest-energy tautomer to have noticeable effect.

4. After the calculations have completed (perhaps 10-15 minutes), select **Spreadsheet** from the **Display** menu . *Click* on the cell corresponding to the original structure for 1-methylcytosine. *Click* on **Add** at the bottom of the spreadsheet and then *click* on the **Molecule List** tab; under **Energy from Molecule** select **rel.E** and **Boltzmann Weights** from the available quantities and **kJ/mol** from the **Relative Energy Units** menu. *Click* on the background to release the dialog. Are either of the alternatives close in energy to the normal form of cytosine?

5. If you find that one of the alternative tautomers is within 10 kJ/ mol of 1-methylcytosine, perform ωB97X-V/6-311+G*(2df,2p) energy calculations on this tautomer as well as 1-methylcytosine. These may require upwards of an hour of computer time. What does the X-V functional with the larger basis set report for relative energies?

6. For a more realistic example, examine the RNA nucleotide, close any open files, and return to the **New Build** mode. Click on the **Nucleotide** tab in the **Model Kit**. Make sure that the check box next to **Sequence** is not selected. *Click* on **C** to select cytosine, and from the drop-down menu under this, select **RNA**. *Double-click* on screen to insert cytidine (cytosine riboside). Repeat the procedure of generating the tautomer list (**Geometry** menu, **Generate Tautomers**, and then *click* on to **Generate a Tautomer List**.

7. Submit the following calculation **Energy** with **Est. Density Functional** and **ωB97X-V/6-311+G*(2df,2p)**. To the right of **Start From** specify **Equilibrium Geometry** using **Est. Density Functional** and **ωB97X-6-31G***. Submit the job. Revisit the spreadsheet and again, make relative energy comparisons. What do the neural network calculations suggest with regard to the Boltzmann distribution?

8. Close any open documents and any remaining open dialogs.

Chapter 9
Spectra of Organic Molecules

The tutorials in this chapter illustrate applications involving the calculation of IR and also illustrate the use of calculated IR spectra to identify an experimental unknown. Additionally, calculation of proton and ^{13}C NMR spectra are demonstrated. DP4 analysis is illustrated to identify which stereoisomer best fits an experimental ^{13}C NMR spectrum.

In addition to equilibrium geometries, conformations and conformational energy differences, reaction energies, as well as diverse molecular properties, calculations provide infrared and Raman spectra, NMR spectra and UV/visible spectra. IR and Raman spectra arise from the transitions between ground and excited vibrational states, NMR spectra from transitions between nuclear spin states and the UV/visible spectra from transitions between ground and excited electronic states.

Chemists use spectra, in particular, NMR spectra to provide essential clues needed to assign the structure of an unknown molecule. That is, features in the spectrum may be used to support a proposed structure assignment, whereas their absence suggests that the assignment is likely to be incorrect. On the other hand, a calculated spectrum starts with a known structure. A high degree of similarity with a measured spectrum may be taken as evidence that the calculated molecule is the same (or at least very similar to) that for which the spectrum was measured. Lack of similarity suggests that the two molecules are not the same.

The first two tutorials in this section deal with infrared structure and use the EDF2/6-31G* model. The first details the steps needed to calculate the infrared spectrum of methyl formate and to relate the spectrum to the underlying molecular structure, and the second illustrates use of the Spartan Infrared Database to identify a molecule based on its infrared spectrum.

The remaining tutorials deal with NMR spectroscopy and use the ωB97X-D/6-31G* model and (optionally) the SSPD database. The first of these details the steps involved in calculating and displaying a proton spectrum for 1-methylindole, and the second, the ^{13}C spectrum of caulophyline, both of which are rigid molecules. The third considers the ^{13}C NMR of *cis*-1,2-dimethylcyclohexane, a molecule that rapidly (relative to the time scale of the NMR experiment) moves between two equivalent conformers. Discussion of the NMR spectra of molecules with multiple degrees of conformational freedom and several accessible conformers is provided in a later chapter. The final tutorial examines the dependence of carbon chemical shifts on stereochemistry and illustrates the use of DP4 analysis to decide which best fits the experimental NMR spectrum.

10 mins

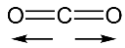
Infrared Spectrum of Methyl Formate

In the harmonic approximation, the frequency at which a diatomic molecule vibrates is proportional to the square root of the ratio of the force constant (the second derivative of the energy with respect to change in bond length) and the reduced mass (the product of the masses of the two atoms divided by their sum). Frequency increases with increasing force constant or stiffness of the bond and decreases with increasing masses of the atoms involved in the bond.

$$\text{frequency} \ \alpha \ \sqrt{\frac{\text{force constant}}{\text{reduced mass}}}$$

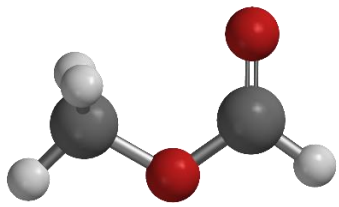
In order to generalize this expression to a polyatomic molecule, it is necessary to use a coordinate system that leads to a diagonal matrix of second energy derivatives (so-called normal coordinates). These differ from internal coordinates (depicting charges in specific bond lengths and angles), and for a polyatomic molecule will typically involve the motions of several (and likely all) atoms.

The intensity of absorption of infrared radiation by a diatomic molecule is proportional to the change in the dipole moment with change in bond length. Since the dipole moment for a homonuclear diatomic molecule does not change with distance, this means that molecules such as N_2 and O_2 are transparent in the infrared. The intensity of each of the individual lines in an infrared spectrum of a polyatomic molecule follows from the change in dipole moment along the associated normal coordinate. Some of the *normal coordinates* may not lead to a change in dipole moment, for example, the symmetric stretch in carbon dioxide where both CO bonds are simultaneously moving, and infrared absorptions will not be observed.

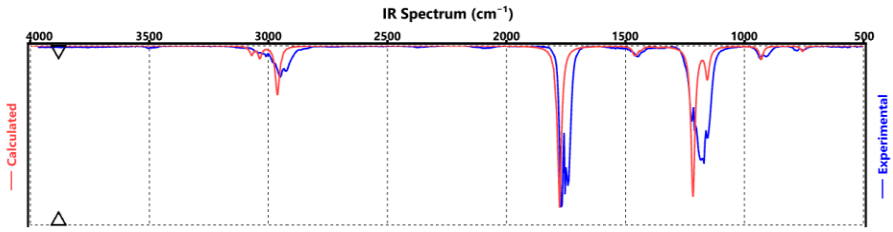


This tutorial illustrates the steps required to calculate and display an infrared spectrum and to compare it with an experimental spectrum. You will explore how the ease or difficulty of molecular motion (the value of the force constant) and changes in atomic masses affect frequency.

1. Build or sketch methyl formate. Select **Calculations...** from the **Setup** menu , and specify an **Equilibrium Geometry** calculation using the **EDF2/6-31G*** density functional model. **Ground** state, in **Gas** and Neutral **Total Charge** with 0 **Unpaired Electrons**. *Check* the **IR** check-box (to the right of **Compute**) and *click* the **Submit** button. Accept the name ***methyl formate***.



2. After the calculation has completed (less than a minute), select **Spectra** from the **Display** menu ( ). *Click* on the in the toolbar at the top of the spectra pane and, from the **IR** tab select from the palette of icons. The experimental IR spectrum of methyl formate is available in the freely available NIST database. *Click* on the again and select from the icon palette. Calculated (in red) and experimental (in blue) infrared spectra are now superimposed.



3. *Click* on either the up or down triangles that define a cursor and slide the mouse while holding down the left button across the spectrum. Position it on the intense line in the (calculated) spectrum near ≈1774-1778 cm^{-1}. Note that the molecular model (on screen above the spectra pane) vibrates. The motion corresponds to stretching of the CO double bond. Position it over the intense line near ≈1211 cm^{-1}. This motion corresponds to a combination of stretching motions of the two CO single bonds.

4. Select **Save As** from the **File** menu to make a copy of methyl formate; name it ***methyl formate d3***. *Click* on to enter **View** mode. Select **Properties** from the **Display** menu and *click* on one of the three hydrogen atoms on the methyl group to bring up the **Atom Properties** dialog. Change **Mass Number** from **1** to **2 Deuterium**. Repeat for the other two methyl group hydrogen atoms. Resubmit the calculation* (it will require only a few seconds) by selecting **Submit** from the **Setup** menu.

* Because force constants (second derivatives) do not depend on mass, the calculation is very simple and will require only a few seconds.

5. Compare the frequencies of the undeuterated and deuterated forms of methyl formate and identify which change the most and which change the least. To do this open the original ***methyl formate*** document and display the calculated IR spectra. To examine the calculated frequencies, *click* on the **Tables** icon . A table of frequencies and intensities is presented. Examine the frequencies associated with CH vibrations on the methyl group. Repeat the procedure examining frequencies associated with CD vibrations on the methyl group. You can move between the documents by *clicking* on the appropriate tab at the bottom of the screen.

6. Close all documents and any open dialogs.

2 mins

Searching *Spartan*'s Infrared Spectral Database

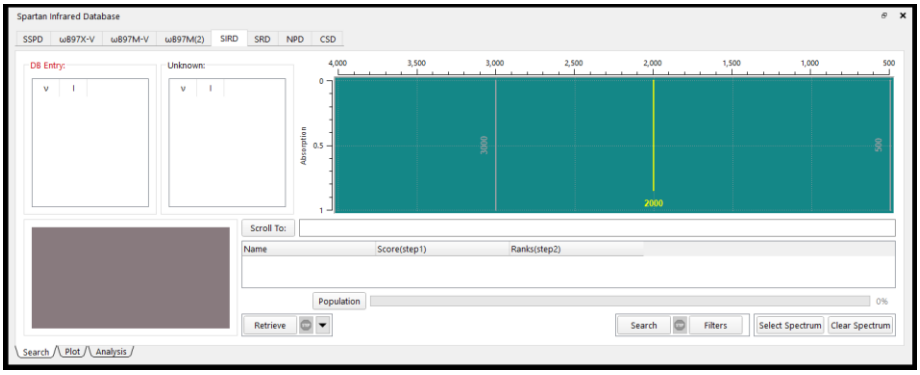
Pattern matching ("fingerprinting") a measured infrared spectrum to a collection of experimental spectra contained in a database is common practice and is appropriate for identifying molecules that have previously been characterized. However, NMR has replaced infrared spectroscopy as the method of choice for structure assignment of "new" (previously uncharacterized) molecules. Pattern matching to databases of calculated infrared spectra is not routine. One reason is that the results of calculations (a set of vibrational frequencies and intensities) do not look like experimental infrared spectra, at least, spectra obtained at normal temperatures. However, a fit of the calculated data to a Lorentzian function in which peak width at half height is treated as a parameter makes the two spectra visually quite similar, not at all surprising as this loosely corresponds to temperature. In fact, an infrared spectrum measured at low temperature comprises a series of sharp lines (rather than bands).

Practical calculations assume that the frequency of an infrared absorption is proportional to the square root of the second derivative of the energy at a minimum in the potential surface for a particular vibrational coordinate, that is, to the curvature of the surface. The so-called quadratic approximation leads to a surface that is too steep, resulting in calculated frequencies that are larger than measured frequencies. For the most part, the error is systematic, with calculated frequencies being between 3% and 15% larger than measured frequencies, depending on the theoretical model. Hartree- Fock models show an error near the top of the range, while density functional and MP2 models show an error near the bottom of the range. Semi-empirical models do not show a consistent error pattern. Such a systematic error can be compensated for, at least in part, by incorporating a single linear scaling parameter into the fitting function.

Spartan fits calculated infrared spectra to the measured infrared spectra using a Lorentzian function that incorporates two parameters, a non-linear parameter accounting for peak width and a linear scaling parameter.

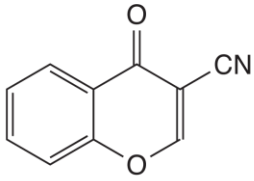
This tutorial illustrates the way in which searches of the Spartan Infrared Database (SIRD) are setup and carried out and the results examined. SIRD is an additional access route to the molecules contained in the Spartan Spectra and Properties Database (SSPD) and derives from EDF2/6-31G* calculations. You may choose from a small selection of measured infrared spectra (or supply your own spectra as a .dx file).

1. You need to have a ***Spartan*** document open in order to access SIRD. Select **New** from the **File** menu, and then select **Databases** from the **Search** menu. *Click* on the **SIRD** tab to bring up the **Spartan Infrared Database** dialog.

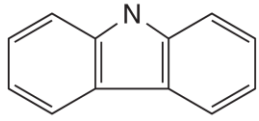


Click on **Select Spectrum** at the bottom right of the dialog. Navigate to the ***spectra of organic molecules*** subdirectory under the ***Tutorials*** directory.*

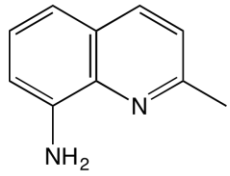
Select one of the following files and *click* on **Open**.



chromone-3-carbonitrile



dibenzopyrrole

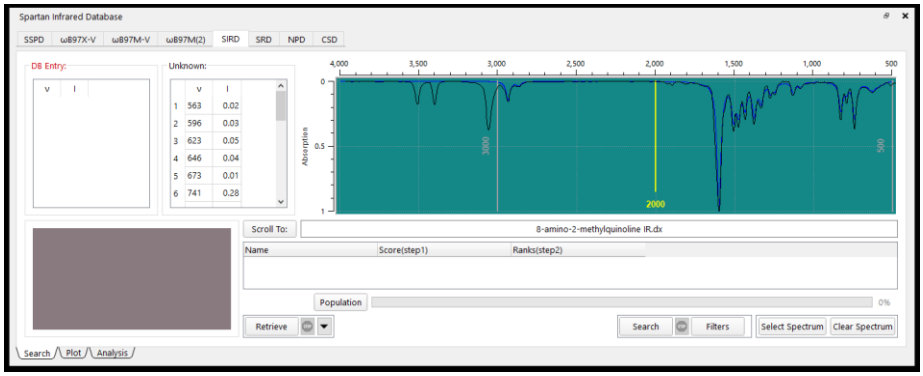


8-amino-2-methylquinoline

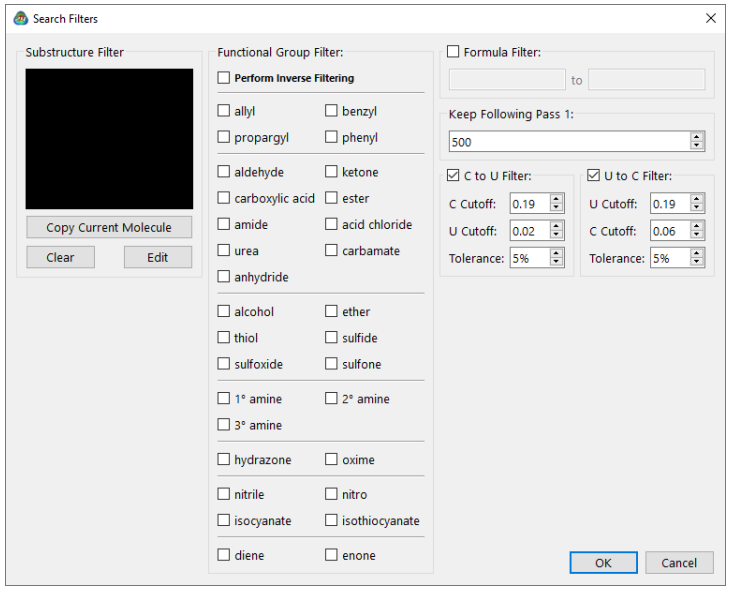
* For Windows, this directory is found in ***Program Files/Wavefunction/Spartan26***. For security reasons, the program file directory is protected. Copy the folder to your desktop or to another location available to the user prior to opening it in ***Spartan***. For Linux, this is found in the directory where ***Spartan*** was installed. For Macintosh, this is located at the top level on the ***Spartan'26*** disc image.

The experimental infrared spectrum of the selected molecule will appear in the

window at the top right of the dialog with the name underneath. Immediately to the left is a scroll box containing the most intense lines in the spectrum (obtained from a fit of the experimental spectrum to a Lorentzian function).



2. *Click* on **Filters** at the bottom right to bring up the **Search Filters** dialog.



This provides for substructure and formula filters as well as functional group filters. You can, if you wish, *check* an appropriate entry. For example, if you have selected chromone-3-carbonitrile as the "unknown", *checking* ***nitrile*** from among the **Functional Group Filters** would limit the search to molecules with nitrile functionality. *Click* on **OK** to exit the dialog.

3. *Click* on **Search** at the bottom of the (**Spartan Infrared Database**) dialog. The search may require a minute or two. When it has completed, scan the list of hits at the bottom of the dialog for the name of your query. It will be at or near the top of the list.* *Click* on the name. The calculated infrared spectrum (red) will be superimposed onto the fit of the experimental spectrum (blue) in the window at the top right of the dialog.

4. Close all documents and any open dialogs.

* This will not always be the case. Errors inherent to the calculations as well as the absence of detail may result in better matches to "incorrect" but closely related molecules. Of course, a query may also not be in the database, so it is possible to find only related molecules.

Proton NMR Spectrum of 1-Methylindole

10 mins

Proton NMR spectroscopy was the first tool available to chemists that allowed definitive assignment of the molecular structures of complex organic molecules. By the 1970's, it had largely replaced infrared spectroscopy and to a large extent chemical proofs of structure. ^{13}C NMR is now more dominant, but proton NMR remains an essential tool in the chemist's arsenal.

An NMR spectrum follows from the fact that nuclei possess spins that can either align parallel or antiparallel to an applied magnetic field, giving rise to different nuclear spin states. The relative energy of these states (ΔE) depends on the nucleus and on the strength of the applied magnetic field, by way of a simple relationship:

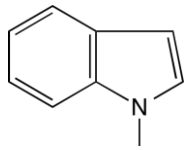
$$\Delta E = \gamma \hbar B_0$$

γ is the gyromagnetic ratio (a constant for a given type of nucleus), $\hbar$ is Planck's constant divided by 2π and B_0 is the strength of the magnetic field ***at the nucleus***. While the two nuclear spin states are normally in equilibrium, this equilibrium can be upset by applying a second magnetic field. The absorption of energy as a function of field strength (a ***resonance***) between the states can then be detected.

The key to the utility of the magnetic resonance experiment is that the energy at which a nucleus resonates depends on its location in the molecule and is measurably different for each (chemically) distinct nucleus. The reason is that the applied magnetic field is weakened by electrons around the nucleus. Nuclei that are well shielded by the electron cloud will feel a lesser magnetic field than those that are poorly shielded and will show a smaller energy splitting. The difference, given relative to a standard, is termed a ***chemical shift***. By convention, both proton and ^{13}C chemical shifts (treated later in this chapter) are reported relative to tetramethylsilane (TMS) as a standard.

While each *unique* proton in a molecule gives rise to a single line (resonance) in the spectrum, the spins on nearby protons add or subtract to the external magnetic field. This leads to a splitting of lines; the splitting pattern depends on the number of neighboring protons and their geometry. In practice, only two-bond ( ) and three-bond ( ) interactions or coupling are important. Discounting splitting, the intensity of the lines is approximately proportional to the number of equivalent protons that contribute.* For example, the proton NMR spectrum of 1-methylindole would be expected to show seven lines, six with unit intensity corresponding to the protons on the indole ring and one line with three times the intensity corresponding to the three equivalent methyl group protons.

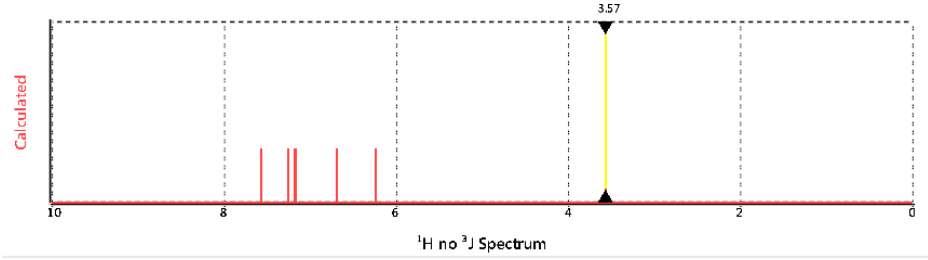
* More generally, the relative intensity of the sum of the lines corresponding to a proton that has been split is approximately proportional to the number of chemically equivalent protons.



In this tutorial, you will use the ωB97X-D/6-31G* model to calculate the proton NMR spectrum of 1-methylindole and compare it with the experimental proton spectrum in the absence of three-bond HH coupling.

1. Build or sketch 1-methylindole. Select **Calculations...** from the **Setup** menu . Specify calculation of equilibrium geometry using the ωB97X-D/6-31G* density functional model. *Check* **NMR** to the right of **Compute** and *click* on **Submit**. Accept the name ***1-methylindole***. The calculation may require a couple minutes.

2. When the calculation has completed (or after you have retrieved results from SSPD), select **Spectra** from the **Display** menu . *Click* on in the bar at the top of the spectra pane and from the **NMR** tab select (proton NMR spectrum in which there is no HH coupling).

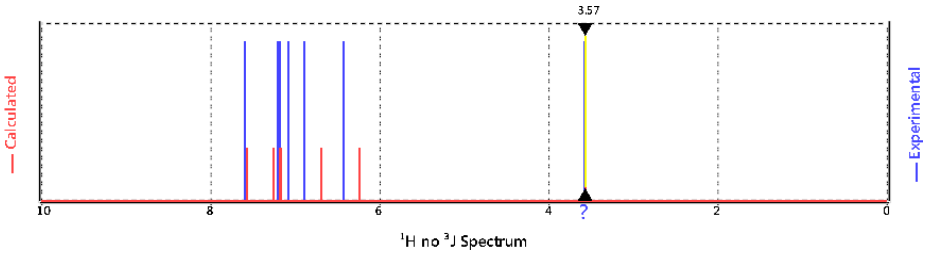


Move the mouse while holding down the left button over the spectrum. When you intersect a line, a numerical value for the chemical shift appears at the top of the spectrum and the proton is highlighted in the structure model.

3. *Click* again on and this time select (experimental proton NMR spectrum with HH coupling constants set to 0). The experimental spectrum will be retrieved from the freely available NMRShiftDB database* and displayed with the calculated spectrum. You will need to be online.

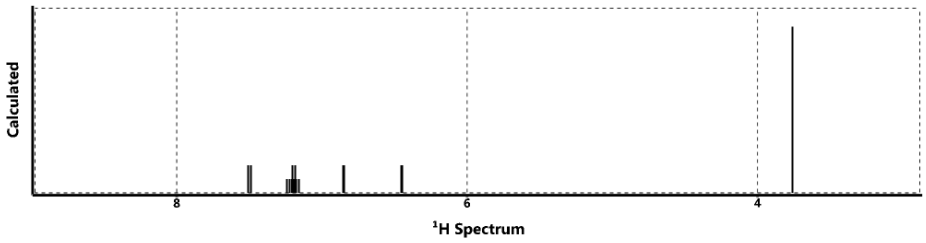
---

* The NMRShiftDB is primarily a collection of ^{13}C spectra, although some proton spectra including that for 1-methylindole are available.



The comparison gives you an idea of the level of agreement that can be expected between calculated and experimental proton spectra.

4. *Click* again on and select (calculated proton NMR spectrum). The calculated proton spectrum that now appears accounts for three-bond HH coupling.



You can focus in on details by a combination of zooming the spectrum (scroll wheel) and shifting the displayed range (move the mouse while holding down the right button). You will see that lines due to protons at C_2, C_3, C_4 and C_7 are doublets, while those due to protons at C_5 and C_6 are quartets (doublet of doublets).

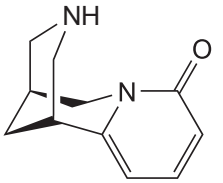
5. Close ***1-methylindole*** and any open dialogs.

20 mins

^{13}C NMR Spectrum of Cytisine

There are several reasons why NMR spectroscopy, in particular, ^{13}C NMR, is one of the most important routine analytical techniques available for characterizing organic molecules. The analysis is straightforward, can be carried out quickly requiring relatively small samples and is non-destructive. The resulting spectrum is quite simple, comprising but a single line for each and every unique carbon.* However, assigning ^{13}C spectra is by no means trivial, even for seemingly simple molecules. The problem is that the positions of the lines in the spectrum (the chemical shifts) are very sensitive to the environment in which the carbons find themselves. Experiments exist to separate ^{13}C resonances based on the number of attached hydrogens, however, where two or more carbons in a molecule reside in "similar environments", it may be very difficult to distinguish them. This is particularly problematic for quaternary carbons.

This tutorial uses the alkaloid cytisine to illustrate the use of calculated ^{13}C spectra to assist in assigning the measured spectrum of the molecule.



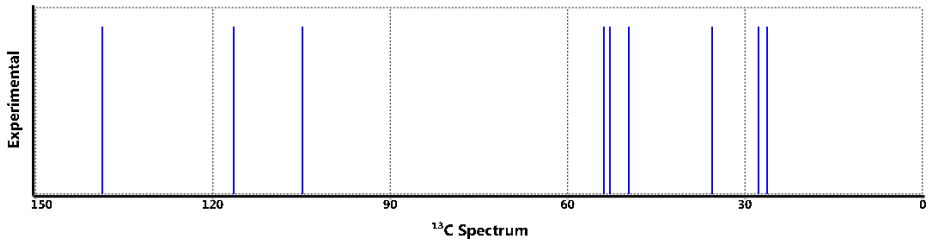
1. Either build or sketch cytisine. Image on previous page. This can pose a challenge, if you end up NOT seeing the name cytisine appear in the lower right of the interface (upon completing your drawing/sketch), you can also build via **InChI** string. From the **New Build** or **Edit Build** mode, in the **Organic** tab of the **Model Kit**, click on the **InChI** button. In the resulting dialog, you can paste the following string:

   InChI=1/C11H14N2O/c14-11-3-1-2-10-9-4-8(5-12-6-9)7-13(10)11/h1-3,8-9,12H,4-7H2/t8-,9+/m0/s1

   and then click **OK**. At the top of the **Model Kit**, a 3-D image of the molecule cytisine will appear. *Double-click* to insert this on screen. Select **Calculations...** from the **Setup** menu and specify calculation of equilibrium geometry with the ωB97X-D/6-31G* density functional model. *Check* **NMR** to the right of **Compute** and leave the NMR setting at **Current Model**. *Click* on **Submit** and accept the name ***cytisine***. The job will require a number minutes to complete (≈ 5-10).

* CH coupling does split carbon resonances but is almost always eliminated by what is termed proton decoupling.

2. When the calculation is done (or after you have retrieved the molecule from SSPD), select **Spectra** from the **Display** menu . *Click* on in the bar at the top of the spectra pane and select (experimental ^{13}C spectrum). The experimental ^{13}C spectrum is drawn.



3. Use the calculated spectrum to associate the individual lines in the experimental ^{13}C spectrum with specific carbons in the structure of cytisine. *Click* on and this time select (calculated ^{13}C spectrum). The calculated spectrum (in red) will be superimposed on top of the experimental spectrum (in blue). This both gives an impression of the performance of the quantum chemical calculations and also allows you to assign the lines in the experimental spectrum to specific carbons. *Click* on one of the triangles that designate the cursor and move the mouse while holding down the left button over the spectrum. A "hit" on a line will display the value of the chemical shift and will highlight the carbons responsible for this line in the structure model.

Optional

4. Open the **Spreadsheet** and *right-click* on the entry (under the Label) column. *Click* on **Rename Selected Molecule(s) Using SSPD**. Again *right-click*, this time on the name ***cytisine*** and choose **Replicate Selected Molecule(s) in New Document**. This will create a new document, enter the **Calculations** dialog, and specify Energy using the ωB97X-D/6-31G* density functional model. To the right of **Start from**, select Equilibrium Geometry using Est. Density Functional ωB97X-D/6-31G*. *Check* the **NMR** to the right of **Compute** and change the NMR setting to **Energy Model**. This specifies determination of structure (Equilibrium Geometry) using one of the NN (neural network) models, followed by a subsequent Energy (single point energy) calculation using the quantum chemical ωB97X-D/6-31G* Density Functional model, and finally an NMR calculation will be performed based on the quantum chemical energy. *Click* on **Submit** and provide the name ***cytisineNN***. The job will require less time to complete (perhaps 2-3 minutes).

5. Compare NMR Shift results. From the original ***cytisine*** document, *right-click* on the carbonyl carbon and choose **Properties** from the resulting contextual menu. From the **Atom Properties** dialog, note the value of the reported Chem. Shift. *Click* on the tab at the bottom of the GUI next to your ***cytisineNN*** document. Repeat the

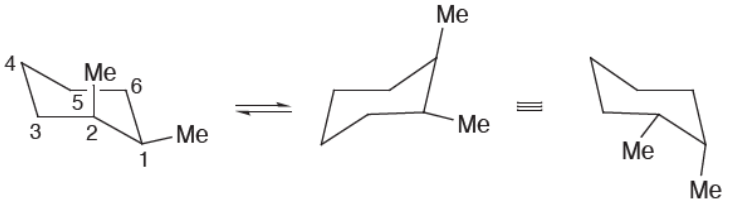
procedure above, how similar are the chemical shifts on the carbonyl carbon in the document where geometry was determined by the neural network (Est. Density Functional > ωB97X-D/6-31G* model)? Repeat the procedure on other centers. Are the shifts generally in good accord between the two calculations (bearing in mind one took 3-4 times longer than the other)?

6. Close both ***cytisine*** and ***cytisineNN*** any open dialogs.

10 mins

^{13}C NMR Spectrum of *cis*-1,2-Dimethylcyclohexane

At normal temperatures, the NMR spectrum of a conformationally- flexible molecule represents an average (Boltzmann-weighted average) of the NMR spectra of all accessible conformers. Only when the temperature is lowered will the spectrum reveal its conformational components. The molecule *cis*- 1,2-dimethylcyclohexane is a special case where the (two) conformers are actually the same (identical).



The room temperature ^{13}C NMR spectrum comprises only four lines at 34.9, 31.9, 24.2 and 16.4 ppm relative to TMS, corresponding to an equal weighting of C_1 and C_2, C_3 and C_6, and C_4 and C_5 and the two methyl carbons, respectively. When the sample is cooled the spectrum reveal eight distinct lines.

1. Either build or sketch *cis*-1,2-dimethylcyclohexane. Select **Calculations...** from the **Setup** menu and specify calculation of equilibrium geometry with the ωB97X-D/ 6-31G* density functional model. *Check* **NMR** to the right of **Compute** and leave the NMR setting at **Current Model**. *Click* on **Submit** and accept the name ***cis-1,2-dimethylcyclohexane***. The calculation will take less than 10 minutes.

2. When the calculation is done, select **Spectra** from the **Display** menu , *click* on in the bar at the top of the spectra pane and from the **NMR** tab select (calculated ^{13}C spectrum). There should be only four lines, the positions of which should correlate fairly well with what is observed experimentally at room temperature (within 1.5 - 2 ppm)*. ***Spartan*** has recognized that there are two equivalent structures and has calculated the average.

3. Close ***cis-1,2-dimethylcyclohexane*** and any open dialogs.

* ^{13}C chemical shifts from ωB97X-D/6-31G* calculations have been empirically corrected.

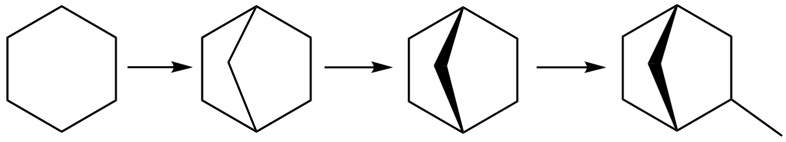
10 minutes

Stereochemical Assignments from ^{13}C Spectra

NMR spectroscopy, in particular ^{13}C spectroscopy, is without doubt the method of choice to establish the three-dimensional structure of organic molecules. Only X-ray diffraction provides more definitive results, although the requirement of a crystalline sample often limits its application. It is now practical to routinely calculate the NMR spectra of moderate size (MW on the order of 500 amu) organic molecules. Accurate calculations provide a "virtual NMR spectrometer" offering organic chemists an entirely new paradigm for structure determination, that is direct comparison of a measured spectrum with calculated spectra for one or more chemically reasonable candidates.

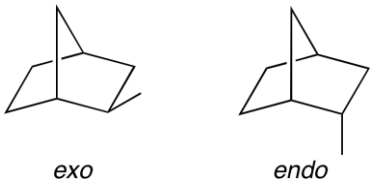
In this tutorial, you will first obtain ^{13}C chemical shifts for *endo* and *exo* stereoisomers of 2-methylnorbornane using the ωB97X-D density functional model, then enter the experimental shifts for one of the two isomers and use the DP4 metric to decide which provides the better match. Note that isomers **MUST** be labeled identically in order to make meaningful comparisons.

1. Sketch 2-methylnorbornane. Start with a six-member ring, draw a one-carbon bridge between two carbons, add "up" stereochemical markers. Finally, add a carbon to the 2 position.



You need to add a stereochemical marker to distinguish between *endo* and *exo* stereoisomers, but before you do so, copy the sketch to the clipboard. *Right click* on screen and select **Copy** from the menu. Add the marker to give the *endo* stereoisomer. Select **Sketch New Molecule** (not **New Sketch**) from the **File** menu, *right click* on screen and select **Paste** from the menu. Add the marker to give the *exo* stereoisomer.

Click on to give 3D structures for both isomers.

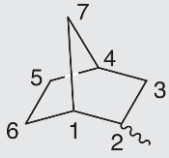


2. Bring up the spreadsheet, *left-click* on the **Label** cell to select all molecules, then *right-*

click and choose **Rename Selected Using SSPD**. Specify calculation of equilibrium geometry using the ωB97X-D/6-31G* density functional model and *check* the **NMR** box. Make sure **Global Calculations** is *checked* and submit the job. The job (two molecules) will require a few minutes to complete.

3. When the calculations are completed, select the first molecule. Add experimental shifts (from the table below). This can be accomplished from either the **Expt. Data** menu, or by bringing up the **Properties** dialog and one after the other *clicking* on the individual carbons and then on the **Edit** drop down menu to the right of **Expt Chem Shift:**. The number pad is used to enter the experimental shift values given in the table below:

Experimental ^{13}C NMR Data



C_1	42.2	C_5	30.6
C_2	34.6	C_6	22.4
C_3	40.7	C_7	38.9
C_4	38.2	CH_3	17.4

4. *Click* on to revert the **Atom Properties** dialog to the **Molecular Properties** dialog. *Click* on the **NMR** tab (in the **Molecule Properties** dialog). Make sure the **Experimental Data From** is set to whichever molecule you have entered in experimental shifts from and *click* the **Apply Globally** button. *Click* on to the left of **DP4** in the dialog. Bring up the spreadsheet to see the DP4 score for the two stereoisomers. The scale is a percentage from 0 to 100 with the larger numbers corresponding to better fits. (You can also examine DP4 scores from the **Properties** dialog if you flip between the stereoisomers).

5. The DP4 calculation is dependent on availability of both experimental shifts and calculated shifts. The specification of **Experimental Data From** determines where this data comes from. In general, it is easiest to specify a single isomer as the source. The **<current molecule>** selection should be reserved for cases where the experimental data is input into an initial structure, but additional (stereo)isomers are subsequently generated from the initial structure as well (**Geometry** menu **Chapter 19**). In this case, the experimental shifts are included for each isomer generated from the **Generate Isomers** tool in the **Geometry** menu. Because the endo and exo structure were built "manually"; we took the additional step of using the clipboard to ensure consistent atom labeling. This is not necessary when stereoisomers are generated from ***Spartan's*** **Generate Isomers** tool.

6. Close the document and any open dialogs.

Chapter 10
Inorganic and Organometallic Molecules

*This chapter illustrates applications to inorganic and organometallic molecules constructed using the inorganic builder in **Spartan**'s model kit.*

Organic molecules typically comprise only a relatively few elements and obey conventional valence rules. The geometry about individual atoms is "tetrahedral" (sp^3), "trigonal planar" (sp^2) or "linear" (sp). As such they are easily constructed using the organic model kit. On the other hand, construction of many inorganic and organometallic compounds in particular those involving transition metals may require the enhanced flexibility provided by the inorganic builder.

Transition-metal inorganic and organometallic compounds may also require different quantum chemical methods from those that have proven to be satisfactory for organic molecules. While there are a large number of experimental structures, almost entirely from X-ray diffraction, accurate experimental thermochemical data on molecules incorporating transition metals is virtually non-existent making assessment of the various models problematic.

The PM3 and PM6 semi-empirical models have been explicitly parameterized for most transition metals and generally (but not always) provide a reasonable account of geometries. However, they would not be expected to properly account for reaction energies as they generally perform poorly for reactions involving molecules without metals.* Hartree-Fock and MP2 models are known to produce poor geometries where transition metals are involved and cannot be recommended.* On the other hand, DFT models appear to be quite successful for calculation of geometries of molecules incorporating both transition metals and lanthanides (although their performance for energy comparisons remains to be clarified).

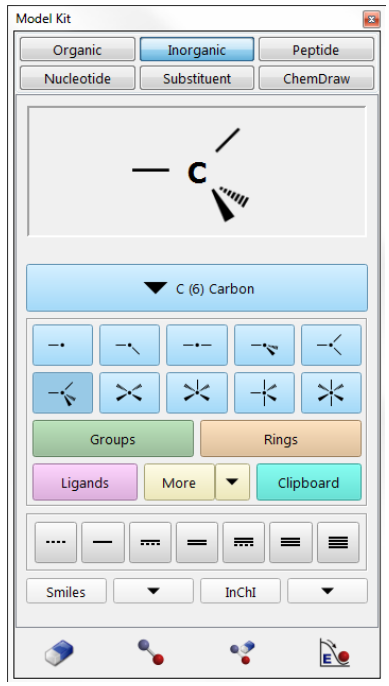
* It should be noted that experimental thermochemical data for inorganic molecule especially for those incorporating transition metals is far less common (and likely less reliable) than that for organic molecules. The poor performance of MP2 is likely a result of its dependence on a Hartree-Fock wave function.

1 min

Sulfur Tetrafluoride

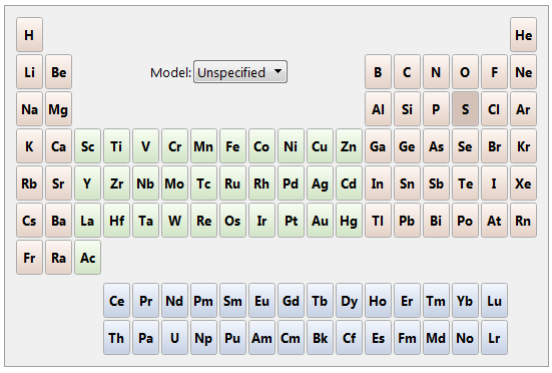
Sulfur tetrafluoride cannot be constructed using ***Spartan's*** organic builder. This is because sulfur is not in its normal bent dicoordinate geometry, but rather in a trigonal bipyramid geometry with one of the *equatorial* positions vacant. However, the molecule can easily be made using the inorganic builder.

1. Bring up the inorganic builder by *clicking* on and then *clicking* on the **Inorganic** tab at the top of the model kit.



The inorganic builder comprises an atom bar (*clicking* on this brings up the *Periodic Table**) followed by a selection of atomic hybrids. Buttons access menus of groups, rings and ligands, additional libraries (**More**) and the clipboard. Finally, a selection of bond types is provided at the bottom of the model kit.

2. *Click* on the atom bar to bring up the *Periodic Table*.



Select (*click* on) **S** in the *Periodic Table* and the five coordinate trigonal bipyramid structure from the list of atomic hybrids. *Double click* on screen. A trigonal bipyramid sulfur will appear at the top of the model kit.

3. Again, *click* on the atom bar, select **F** in the *Periodic Table* and the one-coordinate entry from the list of atomic hybrids. Alternatively switch to the organic builder (*click* on the **Organic** tab) and select F icon. One after another, *click* on both *axial* free valences of sulfur, and two of the three *equatorial* free valences.

4. It is necessary to delete the remaining free valence (on an *equatorial* position); otherwise it will become a hydrogen. *Click* on and then *click* on the remaining *equatorial* free valence.

5. *Click* on . *Click* on . Select **Calculations...** from the **Setup** menu . Specify calculation of equilibrium geometry* using the ω**B97X-D/ 6-31G*** density functional model and *click* on **OK**.

---

* Not all methods are available for all elements listed. Elements for which a specific method (selected in the **Calculations** dialog) are available will be highlighted following selection of a theoretical model or basis set from the **Model** menu that appears in the center of the *Periodic Table*. Note the availability of some heavier elements (>Kr) assumes use of the LANL2DZ basis set which ***Spartan*** will employ automatically when required.

It should be noted that were an incorrect geometry specified at the outset, optimization would lead to the correct structure, as long as the starting geometry possessed no symmetry (C_1 point group). Thus, square planar SF_4 in D_{4h} symmetry would remain square planar, while an "almost" square planar structure (distorted only slightly from D_{4h} symmetry to C_1 symmetry) would collapse to the proper structure upon geometry optimization.

6. Select **Surfaces** from the **Setup** menu, *click* on **Add** and select **HOMO** from the menus the at results and *click* on **OK**.

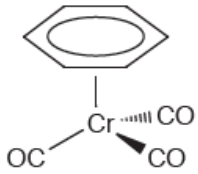
7. Submit the job. Accept the name ***sulfur tetrafluoride***. When completed, select **Properties** from the **Display** menu and *click* on an atom, for example, sulfur. Three different atomic charges will appear in the (**Atom Properties**) dialog (corresponding to different methods for establishing atomic charge). Of these, the procedure based on fitting the electrostatic potential is generally considered to be the best. Are the electrostatic charges consistent with covalent or ionic bonding?

8. *Check* the box to the left of ***HOMO*** inside the **Surfaces** dialog. The HOMO is largely concentrated along the *axial* direction which lacks a fluorine, consistent with the notion that sulfur is surrounded by six electron pairs (five involved in SF bonds and one a "lone pair".

9. Close ***sulfur tetrafluoride*** and any open dialogs.

no calculations

Benzene Chromium Tricarbonyl

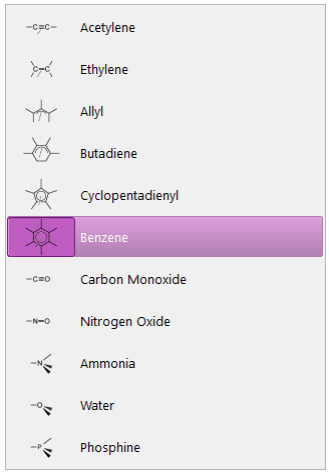


benzene chromium tricarbonyl

Comparison of electrostatic potential maps for benzene chromium tricarbonyl and free benzene will allow you to classify $Cr(CO)_3$ as an electron-donor or an electron-acceptor substituent.

1. Select **New** from the **File** menu and *click* on the **Inorganic** tab at the top of the model kit. *Click* on the atom bar and select **Cr** from the *Periodic Table*. Select the four-coordinate tetrahedral structure from the list of atomic hybrids. *Double-click* anywhere on screen.

2. *Click* on **Ligands** in the model kit and select **Benzene** from the menu of available ligands.



Click on one of the free valences on the four-coordinate chromium center.

3. Select **Carbon Monoxide** from the **Ligands** menu and *click* on the remaining (three) free valences on chromium. *Click* on to produce a refined structure.

> Benzene chromium tricarbonyl among many organometallics may easily be constructed using the 2D sketcher. Inside the sketch pad, *double click* on the wild card icon, select the **Periodic Table** tab and *click* on **Cr**. *Double-click* inside the sketch. *Double-click* a second time on the wild card icon (which will now read **Cr**), select the **Ligands** tab and *click* on benzene ligand. Draw a line upward from **Cr**. *Double click* on the wild card icon a third time (which will now display benzene ligand), select the **Ligands** tab and *click* on carbon monoxide ligand. Draw three lines down from **Cr**. Clean up the sketch and *click* on

4. Select **New Molecule** from the **File** menu . *Click* on **Rings**, select **Benzene** and *click* on the screen. *Click* on and then on . The document now contains both benzene chromium tricarbonyl and benzene.

5. Both benzene and benzene chromium tricarbonyl are available in the Spartan Spectra and Properties Database (SSPD). *Click* on the name of whichever molecule is selected at the bottom of the screen, make certain that **ωB97X-D/6-31G*** is selected and *click* on **Replace**. *Click* on **All** in the dialog that results.

6. Select **Surfaces** from the **Display** menu . *Click* on **Add** and choose **electrostatic potential map** from the menu. Make certain that **Global Surfaces** is checked. Do not close the **Surfaces** dialog.

7. You don't need to submit, as graphics calculations will be performed "on-the-fly". When completed select **Spreadsheet** from the **Display** menu , and *check* the box to the left of the label for both entries. This allows both molecules to be displayed simultaneously on screen. By default, the motions of the two molecules are coupled. To uncouple their motions allowing the two molecules to be moved independently, select (turn off) **Coupled** from the **Model** menu. Orient each molecule so that you can clearly see the benzene face (exposed face in the case of the organometallic).

8. *Check* ***electrostatic potential map*** in the **Surfaces** dialog. Compare electrostatic potential maps for both free and complexed benzene, with attention to the exposed benzene face.* Does the $Cr(CO)_3$ group donate or withdraw electrons from the ring? Would you expect the aromatic ring in benzene chromium tricarbonyl to be more or less susceptible to electrophilic attack than free benzene? More or less susceptible to nucleophilic attack?

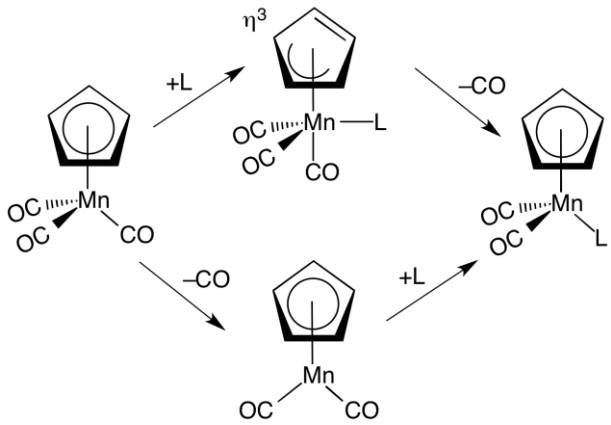
9. Close all documents and any open dialogs.

* Electrostatic potential maps (as well as other maps) for molecules in a single ***Spartan*** document will be put onto the same (color) scale. This allows comparisons to be made among different members.

20 mins

Indenyl Effect

Substitution of one of the carbonyl ligands in cyclopentadienyl manganese tricarbonyl by another two-electron donor ligand may either proceed via an associative or dissociative mechanism. At first glance both mechanisms would seem to be unfavorable, the dissociative pathway because it involves a 16-electron intermediate and the associative pathway because it passes through a 20-electron intermediate. However, the associative mechanism may not be as bad as it first appears as it might be possible to reduce the coordination of the cyclopentadienyl ring from η^5 to η^3 giving rise to an 18-electron association intermediate.



Such a possibility is supported by the observation that substitution of cyclopentadienyl by an indenyl ligand leads to increased association rates. The indenyl effect as it is known, is attributed to the possibility of enhanced aromaticity of η^3 coordinated indenyl relative to η^5 indenyl.

In this tutorial, you will use the PM3 model to calculate geometries for the reactants, intermediates and products for associative ligand exchange by trimethyl phosphine of cyclopentadienyl and indenyl manganese tricarbonyl complexes, and then the ωB97X-D/6-31G* model to calculate energies.

1. Build all eight molecules (the two reactant complexes, the two intermediates and the two product complexes as well as trimethyl phosphine and carbon monoxide) and put into a single ***Spartan*** document. Assume η^5 coordination of cyclopentadienyl and indenyl ligands in reactants and products and η^3 coordination for the two intermediates. η^5 cyclopentadienyl is available from the **Ligands** menu, and η^5 indenyl is in the ligands document accessed by *clicking* on ( ) to the right of **More**. η^3 cyclopentadienyl and η^3 indenyl ligands are not available from menus and need to be built. To build η^3 cyclopentadienyl, first complex an (η^3) allyl ligand to manganese, move to the organic builder, add sp^2 carbons ( ) to both inward pointing allyl free valences, *click* on **Make Bond** ( ) and then on the double free valences on the two fragments that you have just added. To build the η^3 indenyl complex, start with the η^3 cyclopentadienyl complex, select **Benzene** from the **Rings** menu (in the **Organic** builder) and *double click* on the "double bond" in the allyl ligand. Make certain that you minimize the energy of each of the eight structures before proceeding to the next step.

2. Select **Calculations** from the **Setup** menu and specify calculation of equilibrium geometry using the semi-empirical PM3 model. Make certain the **Global Calculations** at the bottom of the dialog is checked. Submit the job with the name ***indenyl effect***.

3. When completed (a few minutes at most), compare the structure of the two reactants (or products) with those of the corresponding intermediates. Verify that the cyclopentadienyl and indenyl ligands have shifted from η^5 to η^3 coordination.

4. Reenter the **Calculations** dialog (**Calculations** under the **Setup** menu) and specify calculation of energy using the ωB97X-D/ 6-31G* density functional model. Resubmit the job. The eight calculations will require several minutes to complete.*

5. Use the reaction energy calculator (**Reactions** from the **Display** menu) to calculate the energy of ligand substitution in both complexes. Is trimethyl phosphine a "better" or "worse" ligand than carbon monoxide? Is the difference diminished or exaggerated in the indenyl complex relative to that in the cyclopentadienyl complex? Next calculate the energy of the two intermediates. Relative to reactants, which is more stable, trimethyl phosphine cyclopentadienyl manganese tricarbonyl or trimethyl phosphine indenyl manganese tricarbonyl?

6. Close ***indenyl effect*** and any open dialogs.

---

* You could do the geometry and energy calculations in a single step. Inside the **Calculations** dialog, specify energy with the ωB97X-D/6-31G* density functional model and set the menus to the right of **Start From** to **Equilibrium Geometry**, **Semi-Empirical**, and **PM3**. Of course, this does not allow you to examine the geometries before doing the density functional calculations.

Chapter 11

Chemical Reactions

This chapter outlines and illustrates strategies for locating and verifying transition states for reactions as well as exploring changes in product distributions as a function of substituents and reactant stereochemistry.

The treatment of chemical reactions adds an entirely new dimension to the application of quantum chemical models. While unique valence structures may generally be written for most molecules and based on these structures, reasonable guesses at bond lengths and angles may be made, it is often difficult to designate appropriate valence structures for transition states, let alone specify detailed geometries. There is a complete absence of experimental data for the structures of transition states. However, transition-state geometries may be calculated. To aid in guessing starting geometries, ***Spartan*** provides both a small library of calculated transition-state structures for common reactions and a facility for automatically matching an entry in this library with the reaction at hand.* The Spartan Reaction Database (SRD) as the library is called, also can be searched by substructure to yield all available transition states of reactions related to the one of interest.

Spartan also provides a procedure for driving user-defined coordinates. Aside from conformational analysis (see discussion in **Chapter 12**), the major application of this is to force reactions, thereby permitting identification of transition states.

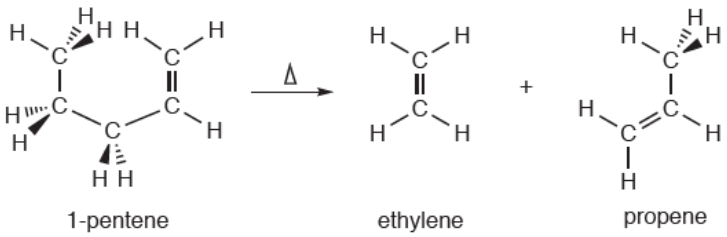
The tutorials in this chapter illustrate ***Spartan's*** automatic procedure for guessing transition-state geometries based on SRD. They also illustrate the use of vibrational analysis (infrared spectroscopy) to verify that a particular structure indeed corresponds to a transition state and to show the motion leading it to reactants and products.

The first four tutorials examine simple organic reactions, while the fifth tutorial provides an example of a transition state calculation for an organometallic reaction. The sixth tutorial illustrates how a reaction may be driven through a transition state. This tutorial, along with the fourth tutorial, draw the connection between relative activation energies and kinetic product distributions. The last tutorial in this chapter illustrates how transition states may be extracted from the Spartan Reaction Database.

*Where a reaction is unknown to ***Spartan's*** library, a fallback technique which averages reactant and product geometries (similar to the so-called linear synchronous transit method) is invoked.

10 mins

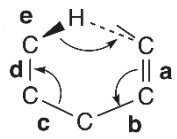
Ene Reaction of 1-Pentene

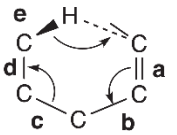


The proposed mechanism of the ene reaction involves simultaneous transfer of a hydrogen atom and CC bond cleavage. In this tutorial, you will obtain the transition state for the ene reaction of 1-pentene from an HF/3-21G calculation, and examine the reaction coordinate for evidence of concerted motion. While the HF/3-21G model is a less rigorous model than many available in ***Spartan***, it is certainly good enough to illustrate the steps involved in obtaining and verifying a transition state.

1. Build 1-pentene in a conformation in which one of the terminal hydrogens on the ethyl group is poised to transfer to the terminal methylene group. To rotate about a (single) bond, first *click* on it to select (the bond will be encircled by a red arrow), and *drag* the mouse up or down in the area below at the left of the screen. Alternatively, hold down the left button and **Alt** key (**option** key on Mac) and *drag* the mouse up and down. *Click* on .

2. Select **Guess Transition State** from the **Build** menu ( ). *Click* on bond **a** in the figure on the following page and then *click* on bond **b**. A curved arrow from double bond **a** to single bond **b** will be drawn. Next, *click* on bond **c** and then on bond **d**. A second curved arrow from bonds **c** to **d** will be drawn. Finally, *click* on bond **e** and then, *click* on the (methyl) hydrogen to be transferred and on the terminal (methylene) carbon to receive this hydrogen. A third curved arrow from bond **e** to the center of a dotted line that has been drawn between the hydrogen and oxygen will appear.

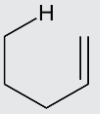




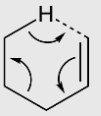
If you make a mistake, you can remove an arrow by selecting **Delete** from the **Build** menu and then *clicking* on the arrow. Alternatively, hold down the **Delete** key as you *click* on an arrow. Select to continue. With all three arrows in place, *click* on at the bottom right of the screen. Your structure will be replaced by a guess at the ene transition state. If the resulting structure is unreasonable, then you have probably made an error in the placement of the arrows. Select **Undo** from the **Edit** menu to return to the model with the arrows and modify (if needed).

Undo is reserved for moving back from the transition-state structure (rather than from undoing reaction arrow placements).

You could also have used the 2D sketcher to provide the guess to the ene transition-state. Enter the sketcher and draw 1-pentene. 2D sketches, do not include hydrogens and you need to draw a CH bond from the terminal sp^3 carbon (C_5).



Click on in the icon bar above the sketch pane. *Click* on the CH bond and then on both H and the terminal sp^2 carbon (C_1). An arrow will be drawn. *Click* on the C_1C_2 bond and then on the C_2C_3 bond. A second arrow will be drawn. *Click* on the C_3C_4 bond and then on the C_4C_5 bond giving the third and final arrow.



Click on to generate a transition state guess; return to .

3. Select **Calculations...** from the **Setup** menu, and specify calculation of transition-state geometry using the HF/3-21G model. Select **Transition State Geometry** from the top menu immediately to the right of **Calculate** and choose **Hartree-Fock** and **3-21G** from the two bottom menus. Finally, *check* **IR** to the right of **Compute**. This will allow you to confirm that you have found a structure that corresponds to a transition state, and that it smoothly connects reactant and product. *Click* on **Submit** and name it ***ene reaction 1-pentene***.

4. When the job completes, animate the motion of atoms along the reaction coordinate. Select **Spectra** from the **Display** menu, *click* on in the bar at the top of the spectra pane and from the **IR** tab, *click* on in the palette that results. *Click* on (Tables) at the left of the spectra pane to bring up a table of frequencies and intensities.

   Click the top entry in the list. It corresponds to an imaginary frequency and will be designated with an **i** in front of the number.

> Within the harmonic approximation, a vibrational frequency is proportional to the square root of the ratio of the force constant (reflecting the curvature of the potential surface along a particular coordinate) divided by a combination of the masses of atoms involved in motion along that coordinate. At a transition state (a maximum in the reaction coordinate), the curvature is negative. Since the mass term is positive the quantity inside the square root is negative and the frequency is an imaginary number.

Calculated:

v	I
i1689	0.01
503	0.01
552	0.25
633	0.10
819	0.16
876	1.00
979	0.03
1071	0.14
1092	0.91
1107	0.16
1133	0.55

Is the vibrational motion consistent with an ene reaction of interest and not with some other process?

5. *Click* on (Make List) at the left of the spectra pane. Controls in the dialog that result allow for changing both the amplitude of vibration (**Amp**) and the number of steps that make up the motion (**Steps**). The default number of steps (11) is fine but change the default amplitude. *Click* inside the box to the right of **Amp:** and enter **0.3** using the number pad that appears. Next, *click* on **Make List** at the bottom of the dialog. This will give rise to a group of 11 structures that follow the reaction coordinate down from the transition state (structure 6) both toward reactant and product. You are done with ***ene reaction 1-pentene***, so close it.

6. *Click* on the list document, select **Calculations...** from the **Setup** menu and specify calculation of **Energy** using the HF/3-21G model (the same level of calculation used to obtain the transition state and calculate the frequencies). Make certain that **Global Calculations** is checked and *click* on **OK**. Select **Surfaces** from the **Setup** menu and specify evaluation of a bond density surface. *Click* on **More Surfaces...**, select **density (bond)** for **Surface** and **none** for **Property**. Before you *click* on **OK**, make certain that **Global Surfaces** is checked.

7. Submit the calculation*. Name it ***ene reaction 1-pentene sequence***. Once the job has completed, enter the **Surfaces** dialog, *check* the box to the left of density and examine the surface that you have calculated. Step through the sequence of structures and keys at the bottom of the screen or animate the reaction . Note, in particular, the changes in bonding revealed by the bond density surface.

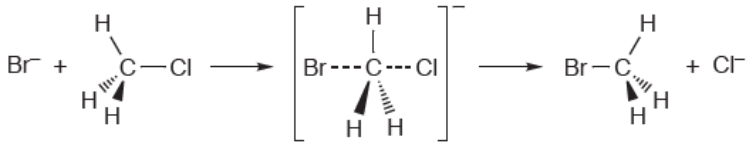
8. Close ***ene reaction 1-pentene sequence*** and any open dialogs.

* In this example, you have requested graphical surfaces prior to submitting the calculation. You could also have requested them to be done "on-the-fly" following the calculation.

5 mins

S_N2 Reaction of Bromide and Methyl Chloride

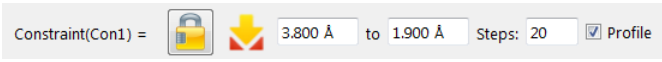
The S_N2 reaction passes through a transition state in which carbon is in a trigonal bipyramid geometry and the entering and leaving groups are colinear. In this tutorial, we will identify it as the "top" of a pathway leading smoothly from reactants to products, employing a semi-empirical molecular orbital model in order to minimize calculation time.



1. Bring up the inorganic builder by selecting **New** from the **File** menu and *clicking* on the **Inorganic** tab *click* on the atom bar, select C from the *Periodic Table* and the five-coordinate trigonal bipyramid from the list of atomic hybrids and *click* on screen. Move to the organic builder (*click* on the **Organic** tab). Select Cl and *click* on one of the *axial* free valences. Select Br and *click* on the other *axial* free valence.

2. Select **Measure Distance** from the **Geometry** menu and then *click* on the C-Br bond. *Click* inside the box to the right of **Length** at the bottom right of the screen and enter **3.8** (3.8Å) using the number pad that appears. You have made a complex representing the reactant.

3. Select **Constrain Distance** from the **Geometry** menu. *Click* on the CBr bond, and then *click* on at the bottom right of the screen. The icon will change to which indicates a constraint is to be applied to this distance. *Check* the box to the left of **Profile** at the bottom of right of the screen. This will result in two additional text boxes.



Leave the value **3.8** (3.8Å) in the leftmost box alone but change the number in the box to the right of **to**, to **1.9** (1.9Å). *Click* inside the box and use the number pad. In the same way, change the number in the box to the right of **Steps** from **10** (the default) to **20**. Twenty calculations with C-Br bond lengths constrained from 3.8Å (the starting point) to 1.9Å (the ending point) will be performed. The transition state should have a C-Br distance in between these values. *Click* on .

4. Select **Calculations...** from the **Setup** menu, and select **Energy Profile**, **Semi-Empirical** and **AM1** from the appropriate menus to the right of **Calculate** in the dialog that results. Make sure to change **Total Charge** to **Anion**.

5. Submit the job. Name it ***bromide+methyl chloride***. When completed, it will give rise to a sequence of calculations placed in ***bromide+methyl chloride.Prof.M0001***. You will be prompted as to whether you want to Open, Save, or Discard this file. *Click* on **Open**. You can close the first document ***bromide+methyl chloride***.

6. Align the molecules that make up the sequence. *Click* on (**Geometry** menu), select **Structure** from the menu at the bottom right of the screen and, one after the other, *click* on the chlorine, the carbon and one of the hydrogens. Finally *click* the **Align by** button at the bottom right of the screen. *Click* on . Use the and keys at the bottom left of the screen to step through the reaction sequence.

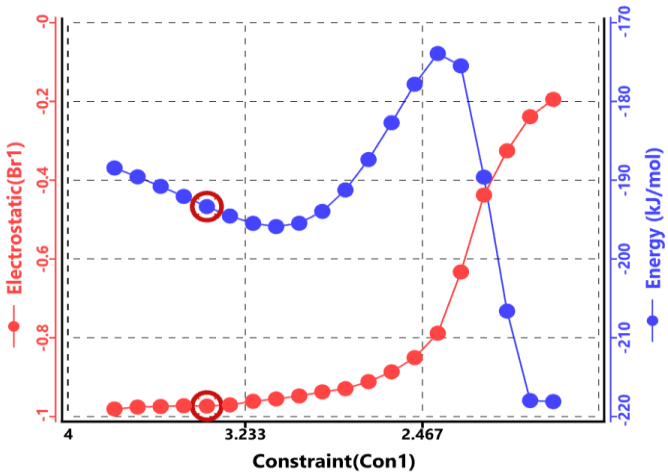
7. Select **Spreadsheet** from the **Display** menu. *Click* on **Add...**. Select **E(kJ/mol)** from among the quantities listed in the **Molecule** tab. Next, enter the (constrained) C-Br distances and bromine charges in the spreadsheet. Select **Constrain Distance** from the **Geometry** menu, *click* on the constraint marker in the model and *click* on at

the bottom right of the screen. *Click* on . Select **Properties** from the **Display** menu to bring up a **Properties** dialog. *Click* on bromine and *click* on to the left of **Electrostatic** under **Charges** in the **Atom Properties** dialog. Close the **Atom Properties** dialog. Close the spreadsheet.

8. Select **Plots** from the **Display** menu ( ) to bring up the **Plots** dialog and *click* on in the bar at the top of the dialog. Select **Constraint (Con1)** (the distance at which the C-Br bond has been constrained) from the **X Axis** menu, and both **E (kJ/mol)** and **Electrostatic (Br1)** from the **YAxes** list in the dialog that results. *Click* on **Create**. By default, only the data points are displayed.

   Click on in the bar at the top of the plots plane, select **Point- to-Point** and then *click* the down arrow to the right of the y-axis field and again, choose **Point-to-Point**, then *click* on **Done**. The visual quality of the plot can be improved by customizing the X-Axis Range, the left arrow icon at the top: can be used to move the plot into the 3D space (with the molecules), and the **Properties** dialog can be employed to enable further customization (*click* on the plot to select either **Plot** or **Curve Properties**.

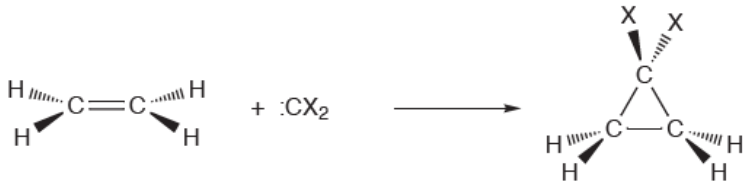
   One plot gives the energy as the reaction proceeds and the other gives charge on bromine. Are the two related? Explain. It might make sense to reverse the X-Axis, and display instead from 4.0 to 1.7 Ångstroms (as this more closely resembles a reaction coordinate:



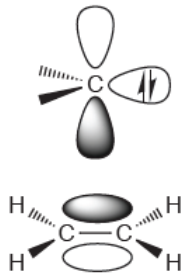
9. Close all documents and any open dialogs.

Carbene Additions to Alkenes

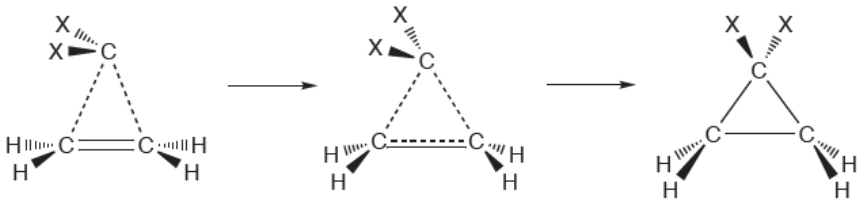
5 mins



Singlet carbenes add to alkenes to yield cyclopropanes. A singlet carbene possesses both a high-energy occupied molecular orbital in the plane of the molecule, and a low-energy, out-of-plane unoccupied molecular orbital, and the reaction presents an interesting dilemma. It is clearly more advantageous for the low-lying vacant orbital on the carbene, and not the high-lying filled orbital, to interact with the olefin π system during its approach.



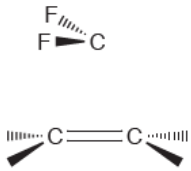
However, this leads to a product with an incorrect geometry, suggesting that the carbene needs to twist by 90° during the course of reaction.



In this tutorial, you will use the Hartree-Fock 3-21G model to obtain the transition state for addition of singlet difluorocarbene, CF_2, to ethylene and then analyze the motion of the fragments.

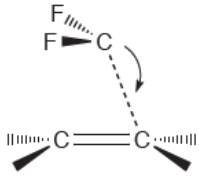
1. Build ethylene using the organic model kit.
2. Select [C–] from the model kit. Hold down the **Insert** key (**option** key on Mac) and then *click* anywhere on screen. Alternatively, *double click* on a blank area of the screen. Next, select [–F] from the model kit and *click* on two of the free valences on the sp^3 carbon. Next, *click* on and *click* on one of the remaining two free valences on the sp^3 carbon.

Click on and *click* on the remaining free valence. You are left with two fragments, ethylene and difluorocarbene. Orient the two as to be poised for reaction.*

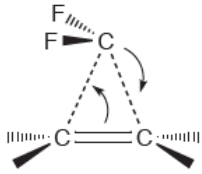


> Translations and rotations normally refer to the complete set of fragments, but if you *click* on a fragment (not on a free valence) to select it and then hold down the **Ctrl** key they will refer to an individual fragment.

3. Select **Guess Transition State** from the **Build** menu ( ). *Click* on the carbon on the CF_2 fragment and then *click* on the closer of the carbons on the ethylene fragment and finally *click* on the CF_2 carbon again. A dotted line is drawn between the two carbons that are to be bonded along with an arrow from the CF_2 carbon to the center of this line.



Click on the CC double bond and then *click* on the other ethylene carbon and on the CF_2 carbon. A second dotted line and arrow will be drawn.



Click on at the bottom right of the screen. Your structure will be replaced by a guess at the transition state.

---

* Proper orientation of the two fragments is not crucial in this case but is primarily to allow you to associate the arrows with the intended reaction. Proper orientation is, however, essential where different stereochemical outcomes are possible.

4. Select **Calculations...** from the **Setup** menu , and request calculation of a **Transition State Geometry** using the **HF/3-21G** model. *Check* **IR** to the right of **Compute** and *click* on **Submit**. Name the job ***difluorocarbene+ethylene***.

5. When the job is complete, examine the geometry of the transition state. In light of the previous discussion, would you describe your structure as corresponding to an early or late transition state? Animate the vibration that corresponds to the reaction coordinate. Select **Spectra** from the **Display** menu . *Click* on the in the bar at the top of the spectra pane and select IR Calculated from the resulting palette of choices. The "infrared spectrum" calculated for the transition state appears. *Click* on at the far left of the spectra pane. This brings up a table of calculated frequencies and infrared intensities. The frequency at the top of the list should be preceded by an "i" designating it as imaginary. It corresponds to the reaction coordinate. Note the value of this frequency as well as that of the largest frequency.

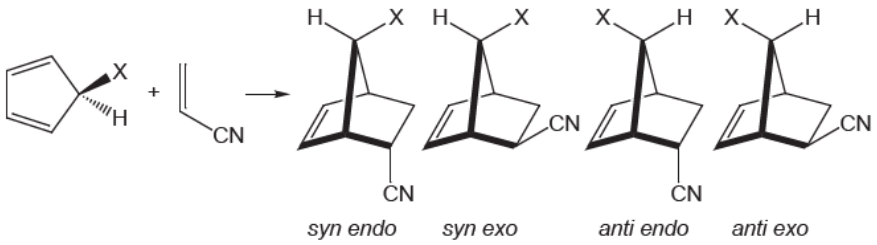
   Click in the check box corresponding to this frequency. Does the animation show that the carbene reorients as it approaches the double bond? Turn the animation off by again *clicking* in the imaginary frequency box. Close the **Spectra** pane and *click* on .

6. Select **Properties** from the **Display** menu and, in turn, *click* on each of the four hydrogens in the transition state. Change the value in the **Mass Number** menu in the **Atom Properties** dialog (**Properties** under the **Display** menu) from **1** to **2 Deuterium**. Resubmit the job. (No additional quantum chemical calculations are involved, but the vibrational analysis needs to be repeated.) When complete, examine the new set of vibrational frequencies, in particular, the imaginary frequency associated with the reaction coordinate. For comparison, examine the largest frequencies which corresponds to one combination of CH stretching motions. You will see that the value of the former has barely changed as hydrogen motions are not significantly involved in the addition reaction, whereas the latter has been markedly reduced.

7. Close ***difluorocarbene+ethylene*** and any open dialogs.

15 mins

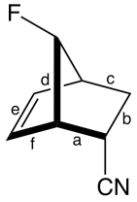
Stereospecific Diels-Alder Reactions



Diels-Alder cycloaddition of 5-substituted cyclopentadienes with acrylonitrile can lead to four stereo products, in which the substituent, X, at the 5 position is *syn* or *anti* to the dienophile, and the nitrile is *endo* or *exo*. *Anti* products are preferred when X is alkyl while *syn* products are favored when X is halogen or alkoxy. In general, *endo* adducts are kinetically favored over *exo* adducts (see following tutorial).

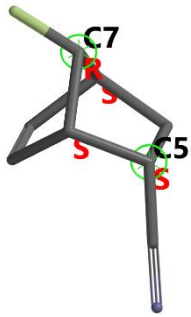
In this tutorial, you will use HF/3-21G molecular orbital calculations to obtain both *syn* and *anti endo* and *exo* transition states for Diels- Alder cycloaddition of 5-fluorocyclopentadiene and acrylonitrile and then use ωB97X-D/6-31G* calculations to estimate relative activation energies. We will use the stereoisomer generation tool first introduced in the previous release (***Spartan'20***) to simplify the steps.

1. Build or sketch the substituted norbornene shown below (the product of *endo* addition of cyclopentadiene and acrylonitrile). *Click* on .



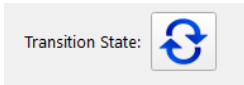
2. Select **Transition State** from the **Build** menu . *Click* on bond **a** (see figure above) and then on bond **b**. A curved arrow will be drawn from **a** to **b**. Next, *click* on bond **c** and then on bond **d**; a second curved arrow from **c** to **d** be drawn. Finally, *click* on bond **e** and then on bond **f**, leading to a third curved arrow. The model on screen should now appear as follows:

3. Select **Generate Isomers** from the **Geometry** menu and *click* on the following two carbons: #1 the carbon atom bonded to the *cyano* group, and #2 the carbon atom bonded to *fluorine* (in the drawing below, these are labeled C_5 and C_7) , this is not necessarily the same numbering as your transition state)in the transition state structure (which will be marked as chiral centers). They will be identified with green circles:



Click on **Generate List** to the right of **Isomers** at the bottom right of the screen, select **New Document** in the dialog that appears and *click* on **OK**. A list of the four stereoisomers will appear in a new document.

4. For each entry in the new list file, visit the **Build** menu **Guess Transition States** entry. The arrows for the Diels-Alder reaction are available for each stereoisomer, as these were part of the model before the isomers were generated. One after the next, enter the **Guess Transition State** mode and click the **Transition State** icon from the lower right of the ***Spartan*** interface:



5. Once you have performed the guess transition state procedure for each potential stereoisomer, Select **Calculations** from the **Setup** menu. Specify **Energy** at **Ground** state in **Gas** from the three top menus to the right of **Calculate** and with **Density Functional**, **ωB97X-D** and **6-31G*** from the three bottom menus. Select **Transition State Geometry** using **Hartree-Fock** and **3-21G** from the three menus to the right of **Start From**. You have requested that transition state geometries for the four stereoisomers be calculated with the **HF/3-21G** model and following that energies calculated using the **ωB97X-D/6-31G*** density functional model. Make certain that **Global Calculations** at the bottom of the dialog is checked and *click* on **Submit**. Name the job ***Diels-Alder stereochemistry***.

6. Once the calculations complete, select **Spreadsheet** from the **Display** menu. If the calculation returns with an error message related to maximum transition state cycles or running out of optimization cycles, you may need to resubmit the job to complete all 4 transition states. You can if you like replace the default labels (M0001, ...) by

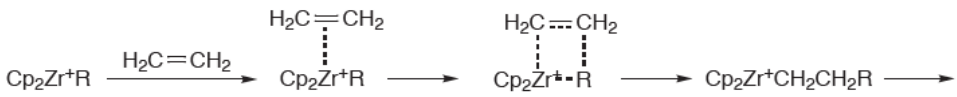
proper identifiers (*syn endo*, *syn exo*, *anti endo*, *anti exo*), to fit the stereoisomers. *Click* on **Add** at the bottom of the spreadsheet, and then on **Energy (au)** in the dialog that results. Identify the isomer with the lowest (most negative) energy and *click* on its label to select it. Again *click* on **Add** and then the **Molecule List** tab at the top of the dialog that results and finally *click* on **ΔE (kJ/mol)** under **Energy from Molecule**. What comments can you make on the relative stability of the transition states (kinetics)?

7. Close all documents and any open dialogs.

30 mins

Ziegler-Natta Polymerization of Ethylene

Ziegler-Natta polymerization involves a metallocene. This complexes an olefin, which in turn inserts into the metal-alkyl bond.



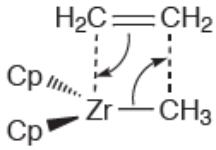
In this tutorial, you will use the semi-empirical PM3 model to obtain a transition state for insertion of ethylene into $Cp_2ZrCH_3^+$ and, estimate the activation energy using the ωB97X-D/6-31G* density functional model.

1. Select **New Build** from the **File** menu and then **Inorganic** from the menu at the top of the model kit. *Click* on the atom bar and select **Zr** from the *Periodic Table*. Select from the list of hybrids (top right of the palette).

   Double-click on screen to insert the planar trigonal **Zr** atom. Select **Cyclopentadienyl** from the **Ligands** menu and *click* on two (of the three) free valences on zirconium.

2. Select **Organic** from the menu at the top of the model kit to move to the organic model kit. Select sp^3 carbon and *click* on the remaining free valence on zirconium. Select **Alkenyl** from the **Groups** menu, hold down the **Insert** key (**option** key on Mac) and *click* anywhere on screen -or- *double-click* in a blank area on screen.

3. Orient the two fragments (Cp_2ZrCH_3 and ethylene) as shown below:



(To move the fragments independently, hold down the **Ctrl** key and use the mouse while in the **Edit Build** mode.)

4. Select **Guess Transition State** from the **Build** menu . *Click* on the ZrC (methyl) bond and one after another, *click* on the methyl carbon and on one of the ethylene carbons. Next, *click* on the ethylene double bond and, one after another, *click* on the other ethylene carbon and on zirconium. *Click* on at the bottom right of the screen. In a few seconds, a guess at the transition state appears.

5. Select **Calculations...** from the **Setup** menu . Specify a transition state geometry calculation using the PM3 semi- empirical model. Change **Total Charge** to **Cation** and *check* **IR** to the right of **Compute**. *Click* on **Submit** and name the job ***Cp2ZrMe cation + ethylene***.

6. When the job completes, select **Spectra** (**Display**) menu . *Click* on the in the bar at the top of the spectra pane and, from the IR tab, select .

7. Calculated infrared spectrum for the transition state appears. *Click* on at the far left of the spectra pane. This brings up a list of calculated frequencies and infrared intensities. The frequency value at the top of the list should be preceded by an "i", designating it as imaginary. *Click* on this frequency and examine the vibrational motion.

8. Perform density functional energy calculations using PM3 geometries to obtain a better estimate for the energy barrier for ethylene insertion. Select **Save As** from the **File** menu to make a copy of ***Cp2ZrMe cation+ethylene***; name it ***Cp2ZrMe cation+ethylene density functional***. Select **Calculations...** from the **Setup** menu and specify calculation of energy using the ωB97X-D/6-31G* density functional model. ***Spartan*** will automatically make use of a pseudopotential for Zr instead of the all-electron 6-31G* basis set. Remove the checkmark on **IR** (to the right of **Compute**). **Total Charge** should still be set to **Cation**. *Click* on **Submit**.

9. While you are waiting for the energy calculation to complete, build both ethylene and $Cp_2ZrCH_3^+$ (name them ***ethylene density functional*** and ***Cp2ZrMe cation***, respectively). For Cp_2ZrCH^+, start with three-coordinate trigonal Zr, and then add two cyclopentadienyl ligands and a tetrahedral carbon. For both, select **Calculations...** from the **Setup** menu , and specify calculation of **Energy** using the **ωB97X-D/6-31G*** density functional model. Select **Equilibrium Geometry**, **Semi-Empirical** and **PM3** from the three menus to the right of **Start From** to designate use of a PM3 geometry. For ***Cp2ZrMe cation density functional*** (only), set **Total Charge** to **Cation**. Make certain that **Global Calculations** at the bottom of the **Calculations** dialog is **not** checked.

10. **Submit** both jobs. When they have completed, calculate an activation energy for the insertion reaction (subtract the sum of the energies of ethylene and Cp_2ZrMe cation from the energy of the calculated transition state). If you copy the reactants to the same file as the transition state, you can use the **Reactions** tool from the **Display** menu, to calculate the activation energy.

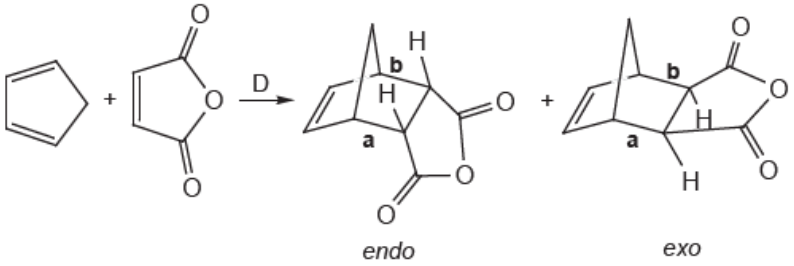
10. Close all documents and any open dialogs.

15 mins

Thermodynamic vs. Kinetic Control

Chemical reactions may yield different products depending on the conditions under which they are carried out. High temperatures and long reaction times favor the most stable (thermodynamic) products, whereas low temperatures and short reaction times favor the most readily formed (kinetic) products. For example, the kinetic product in Diels-Alder cycloaddition of cyclopentadiene and maleic anhydride is the *endo* adduct. Whereas it seems likely that the less-crowded *exo* adduct is the thermodynamic product.

Here you will first obtain pathways for reactions leading to both *endo* and *exo* adducts using the PM3 semi-empirical model and then follow these by ωB97X-D/6-31G* energy calculations to get a better estimate of the difference in activation energies.



1. Build or sketch the *endo* Diels-Alder adduct of cyclopentadiene and maleic anhydride.

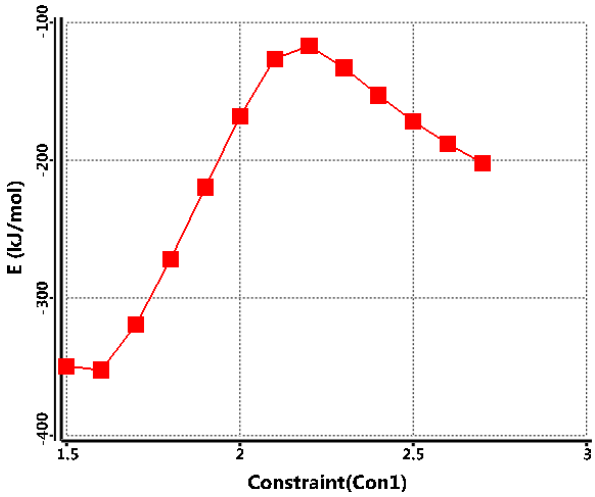
2. Select **Constrain Distance** from the **Geometry** menu, and *click* on bond **a** in the figure from the previous page. *Click* on at the bottom right of the screen (it will then turn to). *Check* the box to the left of **Profile** in the middle of the screen. This will give rise to three additional check boxes. Change the value in the leftmost box to 1.5Å by *clicking* inside the box and entering **1.5** using the number pad. In the same way, enter **2.7** (2.7Å) and **13** (13 steps) into the boxes to the right of **to** and **Steps**, respectively. You have specified that bond **a** will be constrained first to 1.5Å, then to 1.6Å, then to 1.7Å, etc. and finally to 2.7Å. *Click* on .

3. Repeat the process for bond **b**. When you are done, both bonds **a** and **b** will be constrained from 1.5Å to 2.7Å in 13 equal steps. *Click* on .

4. Select **Calculations...** from the **Setup** menu and select **Energy Profile** from the top menu to the right of **Calculate**, and **Semi-Empirical** and **PM3** from the two bottom menus in the dialog that results. *Click* on **Submit** at the bottom of the dialog. Name it ***cyclopentadiene + maleic anhydride endo***.

5. When completed, ***Spartan*** will prompt you that an auxiliary file has been created; chose to **Open** the file. The new document contains the 13 calculations corresponding to the 13 steps in the energy profile.

6. Select **Spreadsheet** from the **Display** menu , and *click* on **Add...** at the bottom of the spreadsheet. Select **ΔE (kJ/mol)** from among the entries in the **Molecule List** tab and *click* on the spreadsheet to release the **Add** menu. Next, *click* on **Constrain Distance** from the **Geometry** menu and *click* on one of the two CC bonds varied in the energy profile. *Click* on at the bottom right of the screen. Close the spreadsheet by again selecting it or *clicking* on or by *clicking* on at the top right. *Click* on . Select **Plots** from the **Display** menu and *click* on to bring up the **Add Plots** dialog. Select **Constraint (Con1)** from among the items in the **X Axis** menu and **ΔE (kJ/ mol)** from the **Y Axes** list and *click* on **Create**. By default, only the data points are displayed. *Click* on in the bar at the top of the plots plane, select **Point-to-Point** in the resulting **Edit Plot** dialog and then *click* on **Done**.



Identify both the reactant and transition state from the plot and estimate the activation energy for the cycloaddition reaction.

7. Repeat steps 1 to 6 for the *exo* adduct. Compare the activation energy for *exo* addition to that for *endo* addition (above). What is the kinetic product?

8. Open ***cyclopentadiene+maleic anhydride endo.Prof.M0001*** . Make a copy . Name it ***cyclopentadiene +maleic anhydride endo DFT***. Select **Calculations** from the **Setup** menu and specify an energy calculation with the **ωB97X-D/ 6-31G*** model in the dialog that results. Make sure that **Global Calculations** at the bottom of the dialog is checked. *Click* on **Submit**. When completed, perform the same spreadsheet and plot operations you did for the PM3 calculations.

9. Repeat the above procedure for the *exo* adduct and compare the two activation energies. What is the kinetic product?

10. Close all documents and any open dialogs.

1 min

Activation Energies of Diels Alder Reactions

In an earlier tutorial, ***Dienophiles in Diels-Alder Reactions*** (**Chapter 8**), you looked for a correlation between LUMO energies for a series of related dienophiles and relative rates of Diels-Alder cycloadditions involving cyclopentadiene. In this tutorial, you will compare calculated activation energies for this same set of reactions with the experimental rates. You will use the previous set dienophiles and obtain transition states from the Spartan Reaction Database (SRD). The only quantum chemical calculation that is required is for cyclopentadiene.

1. Build or sketch cyclopentadiene. Select **Append Molecule(s)...** from the **File** menu, locate the file ***Diels-Alder dienophiles*** (from **Chapter 8**) and *click* on **Open**. If you have not yet completed this tutorial, you can find this file in the ***organic reactions*** folder inside the ***Tutorials**** directory.

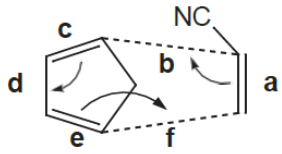
2. Open the **Spreadsheet** from the **Display** menu . *Click* on the **Label** header cell and then *right click* and choose **Rename Selected Using SSPD** from the contextual menu that results. Close the **Spreadsheet** when finished.

3. Select **Calculations...** from the **Setup** menu . Specify calculation of **Equilibrium Geometry** using the **Hartree-Fock 3-21G** model. Make sure the **Global Calculations** box is checked. Submit the job with the name ***Diels-Alder reactants***. It will take a minute or so to complete. While the job is running, you can prepare an additional file with a set of transition states.

4. Select **New** from the **File** menu . **Build** cyclopentadiene, select **Alkenyl** from the **Groups** menu, hold down the **Insert** key (**option** key on Mac) and *click* on screen (or *double click* in a blank area on screen). Cyclopentadiene and ethylene will both appear on screen, although they will not be connected. Select **Cyano** from the **Groups** menu and *click* on one of the free valences on ethylene. Both cyclopentadiene and acrylonitrile will appear on screen.

* For Windows, this directory is found in ***Program Files/Wavefunction/Spartan26***. For security reasons, the program file directory is protected. Copy the folder to your desktop or to another location available to the user prior to opening it in ***Spartan***. For Linux, this is found in the directory where ***Spartan*** was installed. For Macintosh, this is located at the top level on the ***Spartan26*** disc image.

5. Orient the two molecules such that they are poised for a Diels- Alder reaction leading to an *endo* product.



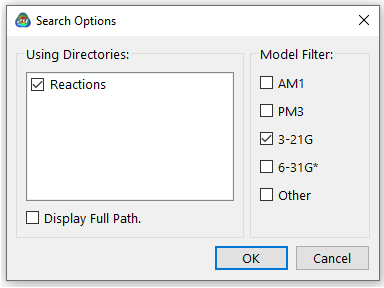
Select **Reaction Query** from the **Search** menu .* *Click* on the bond marked **a** in the above figure and then *click* on the two carbons that when connected will lead to the bond marked **b**. A dotted line will be drawn between these carbons and a curly arrow drawn from the center of bond **a** to the center of bond **b**. Next, *click* on bond **c** and then on bond **d**. A second arrow will be drawn. Finally, *click* on bond **e** and, *click* on the two carbons that when connected will lead to the bond marked **f**. A second dotted line and a third arrow will be drawn.

6. Select **Structure Query** from the **Search** menu . *Click* on all three free valences on the acrylonitrile fragment. They will be replaced by orange cones, meaning that anything (including hydrogen) may be attached to these positions.

7. Select **Databases** from the **Search** menu and *click* on the **SRD** tab at the top of the dialog that results. This leads to the **Spartan Reaction Database (SRD)** dialog.

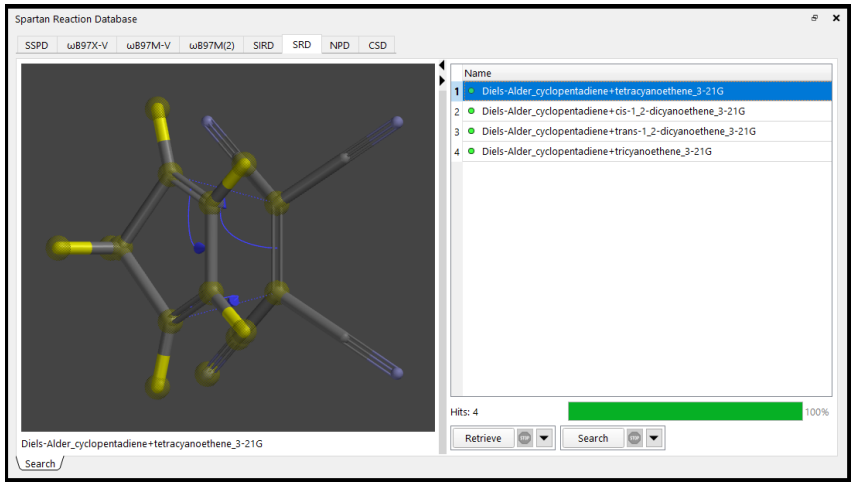
* **Reaction Query** is equivalent to **Transition State** for the purpose of defining curly arrows. It lacks the ability to automatically guess a transition state.

Before you begin the search, *click* on ▼ to the right the **Search** button at the bottom of the dialog. This brings up the **Search Options** dialog.



Check **3-21G** under **Method Filters** and *click* on **OK**. *Click* on the **Search** button to search the Spartan Reaction Database for Diels-Alder transition states between cyclopentadiene and substituted acrylonitriles.

8. When the search completes, a listing of hits appears at the right of the dialog.



This should contain four Diels-Alder transition states for *endo* additions of cyclopentadiene and *trans*-1,2-dicyanoethylene, *cis*- 1,2-dicyanoethylene, tricyanoethylene and tetracyanoethylene. All need to be retrieved into a single document. *Click* on ▼ to the right of the **Retrieve** button at the bottom of the dialog to bring up the **Retrieve Options** dialog. Make sure that **New Document** is selected in this dialog and *click* on **OK**. Hold down the **Ctrl** key, select (*click* on) all four Diels-Alder transition states and *click* on **Retrieve** at the bottom of the dialog. A new (unnamed) document will be created. Close the file, but save it as ***Diels-Alder transition states***.

9. As soon as the job ***Diels-Alder reactants*** completes, select **Append Molecule(s)...** from the **File** menu, navigate to ***Diels-Alder transition states*** and *click* on **Open**.

10. Select **Reactions** from the **Display** menu . One after the other, calculate activation energies for the six Diels-Alder reactions. Use cyclopentadiene and one of the dienophiles for **Reactants** and the appropriate transition state for **Transition States**. Select **Current Document** under **Source** at the bottom of the dialog, and **Activation Energy** under **Calculate**, to enable display of the activation energy. Is the ordering of the calculated activation energies consistent with the experimental relative rates (available in the tutorial ***Diels-Alder dienophiles*** from **Chapter 8**)?

11. Close all documents and any open dialogs.

Chapter 12

Medicinal Chemistry

*This chapter illustrates applications of **Spartan** to areas of relevance to medicinal chemists.*

Medicinal chemistry is a diverse field; the common thread being its concerns with the design, properties and behavior of molecules that are important in biological systems. The tutorials in this chapter illustrate some of the roles that calculations might play.

The first addresses simple models to estimate the rate at which molecules are transported between blood (hydrophilic media) and the tissue in the brain (hydrophobic media); this is not something that can be directly calculated, and instead we will rely first on the so-called polar surface area (PSA) and later refine this using electrostatic potential maps.

The second and third tutorials use similarity analysis to identify "new" molecules (drugs) that are structurally or functionally similar to molecules with known drug action. In the second, involving the antihistamine terfenadine, similarity analysis is not carried out between molecules but between a molecule and an environment. In the third, similarity analysis is used to suggest compounds that "look like" morphine and hence might exhibit similar analgesic action.

The last tutorial accesses the RCSB **P**rotein **D**ata **B**ank (https://www.rcsb.org). While calculations on biopolymers (proteins and RNA/DNA strands) are limited to molecular mechanics models, quantum chemical calculations can be routinely applied to molecules bound to biopolymers. The focus is not the molecule itself, but rather the binding environment. This is turned into a so-called pharmacophore (a simplified representation of the environment); a search for other molecules that may also be consistent with this environment is illustrated.

no calculations

Anticipating Blood-Brain Transport

To be effective, a drug must be capable of transportation to its target. For orally administered drugs, this includes transfer through cell membranes in the intestine. In the case of most neural drugs, this also involves traversing the blood-brain barrier. In this tutorial, you will first examine how well the polar surface area (the area of space- filling model due to nitrogen and oxygen atoms together with any attached hydrogens) correlates with log (C_{brain}/C_{blood})*.

1. Open ***blood brain transport*** from the ***medicinal chemistry*** sub- directory (***Tutorials*** directory).** This comprises a list of drugs/ drug candidates for which experimental data relating to the ratio of concentrations in the brain and in the blood (C_{brain}/C_{blood}) are available. These span a range of nearly 5 log units.

2. Select **Spreadsheet** from the **Display** menu and size the spreadsheet to show all compounds, several extra rows and three columns, in addition to the column of experimental data **log(brain/blood)** that is already displayed. Select **Properties** from the **Display** menu and *click* on the **QSAR** tab in the dialog that results. *Click* on (Post) to the left of **PSA**. Polar surface area values will fill one column in the spreadsheet.

3. *Click* on the **Add** button at the bottom of the spreadsheet and select (*click* on) the **Linear Regression** tab at the top of the dialog that appears. Select **log(brain/blood)** from the **Fit** menu and *click* on **PSA** in the box under **Using**. *Click* on **Apply**. Select **Properties** from the **Display** menu and then *click* on the cell labeled **Fit1** at the bottom of the spreadsheet under the **Label** column. This brings up the **Regression Properties** dialog in which RMSD and R^2 values are reported. The smaller RMSD and the closer to unity R^2,the better the fit. Close the **Regression Properties** dialog.

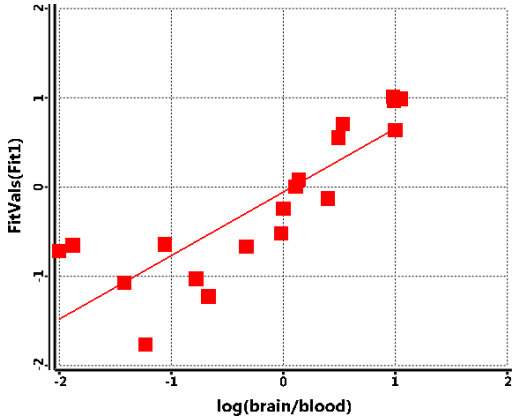
4. Make a plot of log(brain/blood) vs. the regression fit. Select **Plots** from the **Display** menu ( ). *Click* on in the bar at the top of the plots pane to bring up the **Plots** dialog. Select **log (brain/blood)** from the **X Axis** menu, **Fit Vals (Fit 1)** from the **Y Axes** list and *click* on **Create**. *Click* on and check the box to the left of **Curve** and the radio button next to **Least Squares** and finally the **Done** button.

---

* J. Kelder *et al*, Pharmaceutical Res., **16**, 1514 (1999). All 19 named compounds considered in this paper have been included.

** For Windows, this directory is found in ***Program Files/Wavefunction/Spartan26***. It must be copied to another location available to the user prior to opening it in ***Spartan***. For Linux, this is found in the directory where ***Spartan*** was installed. For Macintosh, this is located at the top level of the ***Spartan'26*** disc image.

A least-squares line is drawn through the points.



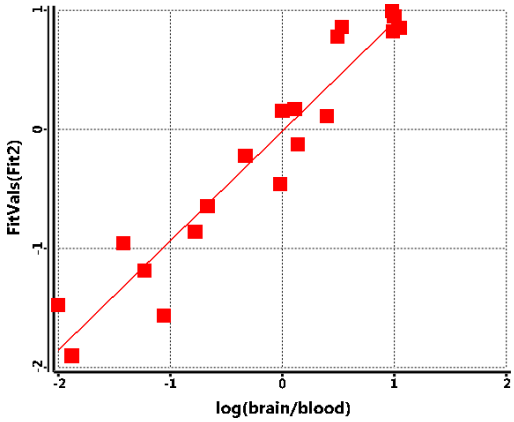
Next, you will consider an alternative definition of polar surface area based on electrostatic potential maps.

5. Select **Surfaces** from the **Display** menu and select ***electrostatic potential map*** in the dialog that results. An electrostatic potential map for the selected drug appears (maps for the remaining drugs may be seen by stepping through the list using the step buttons in the bottom left of the screen). The overall size and shape are that of the electron density and correspond roughly to a conventional space-filling model. The colors indicate the value of the electrostatic potential. Colors toward red designate areas of negative potential (where a positive charge is most likely to be attracted), while colors toward blue designate areas of positive potential (where a positive charge is least likely to be attracted). To see the molecular skeleton underneath the electrostatic potential map, change to a transparent or mesh display. *Click* on the map to select it and select **Transparent** or **Mesh** from the **Style** menu that appears at the bottom right of the screen.

6. Select **Properties** from the **Display** menu . *Click* on the map. *Click* on the post button to the left of **P-Area** the **Surfaces Properties** dialog that results. Polar areas, defined as the area for which the absolute value of the electrostatic potential is > 100 kJ/mol*, will be added to the spreadsheet.

7. *Click* on the **Add** button at the bottom of the spreadsheet and *click* on the **Linear Regression** tab at the top of the dialog that results. Select **log(brain/blood)** from the **Fit** menu and *click* on **Polar Area** from the quantities in the box under **Using**. *Click* on **Apply**. Bring up the **Regression Properties** dialog by *clicking* on the cell labeled **Fit2** under **Label** in the spreadsheet. Note that the R^2 value is better (closer to unity) for this fit than for the previous fit (to PSA defined in the usual manner). Make a plot of log (brain/blood) vs. the new regression fit. Select **Plots** from the **Display** menu ( ) and *click* on in the bar at the top of the plots pane. Choose **log (brain/blood)** from the **X-Axis** menu and **FitVals (Fit 2)** from the **Y Axes** list and *click* on **Create**. *Click* on and again specify **Least Squares** for the **Curve** in the **Edit Plot** dialog

and *click* **Done**. Which property provides better correlation with transport across the blood/brain barrier, PSA or polar area?



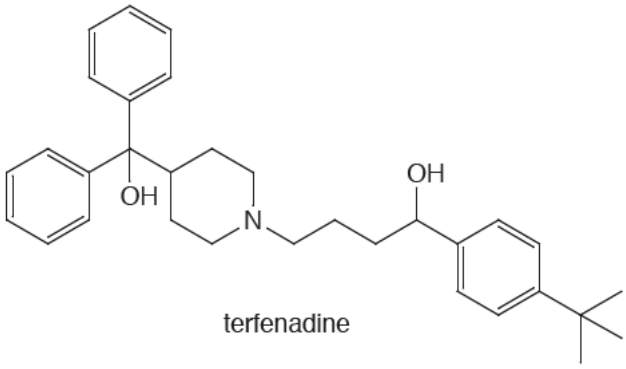
8. Close ***blood brain transport*** and any open dialogs.

* This value may be changed in the **Settings Preferences** dialog (**Preferences** under the **Options** menu; **Chapter 25**).

5 mins

Terfenadine. A Potassium Channel Blocker?

Cardiovascular toxicity due to blocking of potassium ion channels is commonly screened early in the drug development process. One way this is done is to compare drug candidates to a pharmacophore deduced from 3D QSAR studies. This tutorial uses a published pharmacophore* to see whether the antihistamine, terfenadine should be considered a potential channel blocker.



terfenadine

1. Open ***terfenadine similarity library*** from the ***medicinal chemistry*** sub-directory (***Tutorials*** directory).** This file contains the coordinates for a diverse selection of several hundred conformers of terfenadine (the actual number is reported in the **Molecule Properties** dialog) obtained from a similarity library calculation. It has been supplied in order to save computation time. Close ***terfenadine similarity library***.

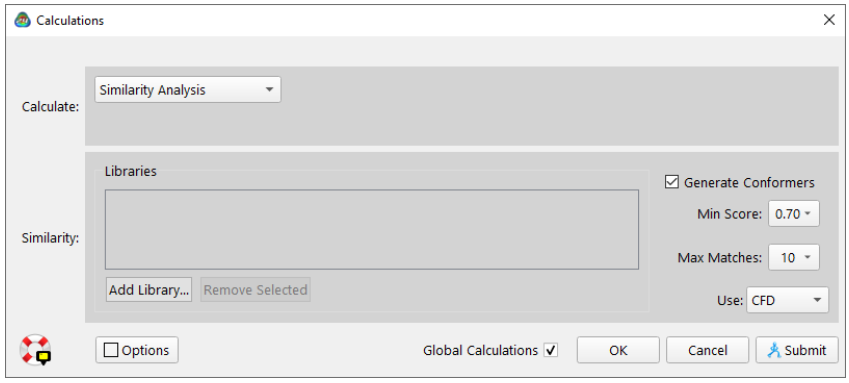
2. Open ***potassium channel blocker pharmacophore*** from the ***medicinal chemistry*** sub-directory (***Tutorials*** directory). This is a five-point pharmacophore comprising four aromatic hydrophobe centers (purple spheres) and one positive ionizable center (red sphere).

* S. Ekins *et al*, J. Pharmacology and Experimental Therapeutics, **301**, 427 (2002).

** For Windows, this directory is found in ***Program Files/Wavefunction/Spartan26***. It needs to be copied to another location available to the user prior to opening it in ***Spartan***. For Linux, this is found in the directory where ***Spartan*** was installed. For Macintosh, this is located at the top level of the ***Spartan'26*** disc image.

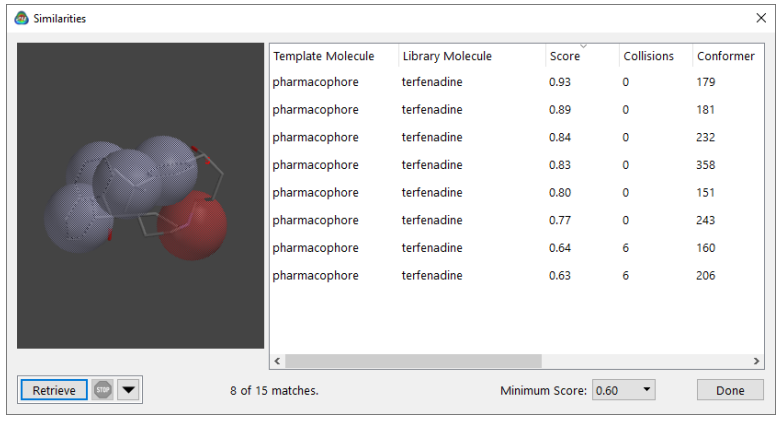
3. Select **Set Similarity Centers** from the **Geometry** menu and then select **CFD** from the menu at the bottom right of the screen next to **Similarity by:**. In turn, *click* on each of the (five) pharmacophore elements. In response, each will be surrounded by a violet circle indicating that it is to be used in the similarity analysis calculation.

4. Select **Calculations...** from the **Setup** menu . Select **Similarity Analysis** from the top left menu to the right of **Calculate** inside the **Calculations** dialog. This leads to a new configuration of the **Calculations** dialog.



Click on **Add Library** to bring up a file browser. Locate ***terfenadine similarity library*** in the ***medicinal chemistry*** sub- directory (***Tutorials*** directory) and *click* on **Open**. This indicates that the library of terfenadine conformers will be searched for a match to the pharmacophore. *Click* on **Submit**.

5. When the analysis has completed (a minute or so), select **Similarities** from the **Display** menu to bring up the **Similarities** dialog, and change the **Minimum Score** to .6 or lower (to expose additional similarity matches).



Multiple hits will appear at the right of the dialog, together with a similarity score (limiting on 1.0 which means a perfect match). *Click* on the hit with the highest score. A composite graphic of terfenadine and the pharmacophore will appear in a window at the left of the dialog. You can manipulate the two (as a single object) in the usual way (you need to position the cursor inside the window). Note that the positive ionizable center is positioned above nitrogen, while the four aromatic hydrophobes are above phenyl rings and the *tert*-butyl group.

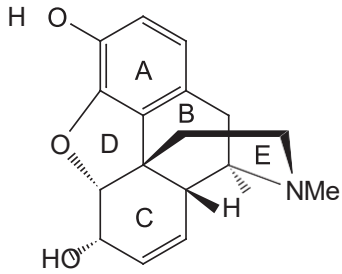
> The scoring algorithm not only accounts for position but also for direction of any nitrogen or oxygen centers that overlap with (non-hydrophobe) pharmacophore elements.

6. You can retrieve a selected hit (without retrieving the pharmacophore), by *clicking* on the **Retrieve** button at the bottom left of the **Similarities** dialog. This can be used for further analysis.

7. Close all open documents and dialogs.

15 mins

Morphine. Structure vs. Pharmacophore



morphine

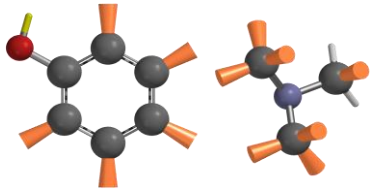
Three "chemical aspects" of morphine appear to be required in order for the molecule to act as an analgesic. These are: i) the nitrogen center (assumed to be protonated in the protein-bound complex), ii) the aromatic ring and iii) the hydroxyl group attached to the aromatic ring. The loss of any of these results in significant reduction of activity. This knowledge may either be used directly to identify other likely analgesics based on structure, or indirectly to construct a simple 3-point pharmacophore that in turn may be employed to find compounds with potentially comparable analgesic activity.

This tutorial comprises three parts. In the first part, you will identify molecules in the Spartan Spectra and Properties Database (SSPD) that incorporate the three required elements. In practice, you will look for molecules that incorporate a nitrogen center substituted by a methyl group and two other (sp^3) carbons and a substituted phenol. You will then select a single hit and generate a conformer library for similarity analysis simply in order to save computer time. This is a superficial investigation. To be comprehensive, one would need to use several (or all) hits.

1. Select **New** from the **File** menu and bring up the organic model kit. Select **Benzene** from the **Rings** menu and *double click* anywhere on screen. Select sp^3 oxygen and *click* on one of the free valences of benzene. You have made phenol. Select sp^3 nitrogen, *double click* on a blank region on screen to insert a non-bonded sp^3 nitrogen.

   Two molecules (phenol and ammonia) appear on screen. *Click* on sp^3 carbon and *click* on all three free valences on the nitrogen. Phenol and trimethylamine now appear on screen. *Click* on .

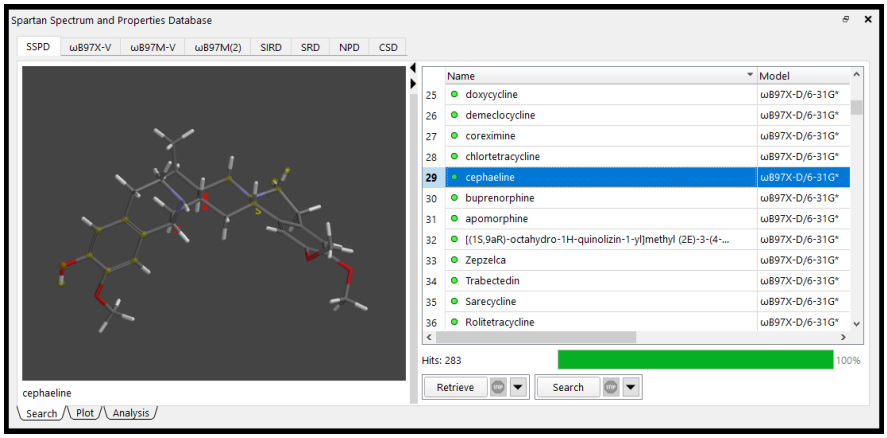
2. Select **Structure Query** from the **Search** menu . *Click* on all five free valences on the carbons in the phenol fragment and on all three free valences on two of the three methyl groups and on one free valence for the remaining methyl group in the trimethylamine fragment. Orange cones will appear at all (twelve) selected positions.



This defines a search in which anything can be "grown" from the selected positions (including hydrogen) but a hit must contain phenol and a nitrogen bonded to three sp^3 carbon centers.

3. Select **Databases** from the **Search** menu . *Click* on the **SSPD** tab in the dialog that results. *Click* on **Search** at the bottom of the **SSPD** dialog.

4. To sort, *click* on **Name** at the top of the list inside in the scroll box at the right of the **SSPD** dialog. Locate and *click* on ***cephaeline***.*



* The search above was done on the full ≈317,000 molecule ***SSPD***. The smaller subset bundled with the ***Spartan'26*** install will find significantly fewer "hits".

Its structure will be displayed in the window at the left of the dialog. You can manipulate the model in the usual way by positioning the cursor inside this window. *Click* on to the right of the **Retrieve** button at the bottom of the **SSPD** dialog to bring up the **Retrieve Options** dialog. *Check* **New Document** under **Retrieve Options** and *click* on **OK**. Finally, *click* on the **Retrieve** button. Dismiss the **SSPD** dialog either by again selecting **Databases** from the **Search** menu or by *clicking* on at the top right-hand corner of the **SSPD** dialog.

Save the molecule you retrieved under the name ***cephaeline***. Close the file used for the search (phenol and trimethylamine) without saving.

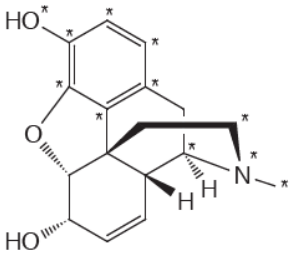
5. For similarity library generation, it is advised to use a more restrictive conformational analysis routine. For this example, the molecule retrieved from the database has been set to do so. To confirm this, select **Set Torsions** from the **Geometry** menu . Your molecule should be augmented with three yellow cylinders (only). If this is *not* the case, select **Preferences** from the **Options** menu. In the **Settings** tab, switch the **Conformer Rules** to **skeletal**, and *click* **OK**. You will still be in the **Set Torsions** mode, in the lower right of the interface, click on the **Reset** button to apply the changes to your molecule.

6. Select **Calculations...** from the **Setup** menu . Select **Similarity Library** from the top left menu to the right of **Calculate**. No further information is required. *Click* on **Submit** at the bottom of the dialog. Name it ***cephaeline library***. You can continue with the tutorial while you are waiting for the job to complete. Note, however, that you need to close ***cephaeline library*** before it is used in **Step 9**.

In the second part of this tutorial, you will specify the key structural components in morphine and perform a similarity analysis based on structure using the conformer library for cephaeline.

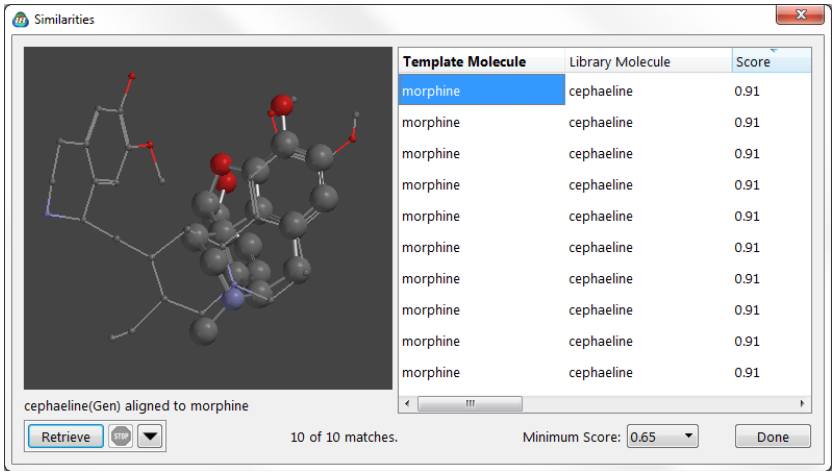
7. You can either sketch morphine (much easier than building it in 3D) or to save time, open it: ***morphine*** from the ***medicinal chemistry*** sub-directory (***Tutorials*** directory).* If you sketch it, *click* on when you are done. The name should appear at the bottom of the screen as morphine is in SSPD. If it does not, you have made a mistake.

8. Select **Set Similarity Centers** from the **Geometry** menu ( ) and select **structure** from the menu at the bottom right of the screen. *Click* on the eleven *'ed atoms in the figure below. These atom centers define the phenol and the primary amine discussed earlier. A violet circle will be drawn around each. If you accidentally select the wrong atom, *click* on the "offending" circle to deselect it.



9. Select **Calculations...** from the **Setup** menu . Select **Similarity Analysis** from the top left menu to the right of **Calculate** inside the **Calculations** dialog. *Click* on **Add Library** to bring up a file browser, locate and select ***cephaeline library***, and *click* on **Open**. Make certain that **Structure** is selected from the **Use** menu at the lower right of the dialog. You have requested that the library of cephaeline conformers be searched for a match to the key structural elements of morphine. *Click* on **Submit**. Supply the name ***morphine***.

10. The similarity analysis will require a few seconds. When it completes, select **Similarities...** from the **Display** menu .



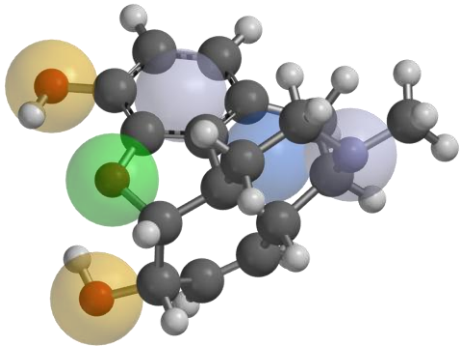
11. Hits appear in the box at the right of the dialog together with a score. Sort according to score (best on top) by *clicking* on **Score** at the top of the box. Select (*click* on) one of the best matches. A composite graphic consisting of morphine and the matching cephaeline conformer from the library will appear in a window at the left. You can manipulate the two structures (as a single object) in the usual way.

---

* For Windows, this directory is in ***Program Files/Wavefunction/Spartan26***. Copy to user-writable location.
For Linux, this is found in the directory where ***Spartan*** was installed.
For Macintosh, this is located at the top level of the ***Spartan'26*** disc image.

Repeat the similarity analysis using CFD's rather than structure.

12. Select **Set Similarity Centers** from the **Geometry** menu , but this time select **CFD** from the menu at the bottom right of the screen. CFD's will augment the morphine structure.



13. Designate three of the CFD's as similarity centers. Select (*click* on) the CFD over the phenolic oxygen. A violet circle will surround this center indicating that it is to be used in the similarity analysis. (If you accidentally select the wrong CFD, *click* on the circle and it will disappear.) Select (*click* on) the CFD at the middle of the phenol ring and the CFD at nitrogen.

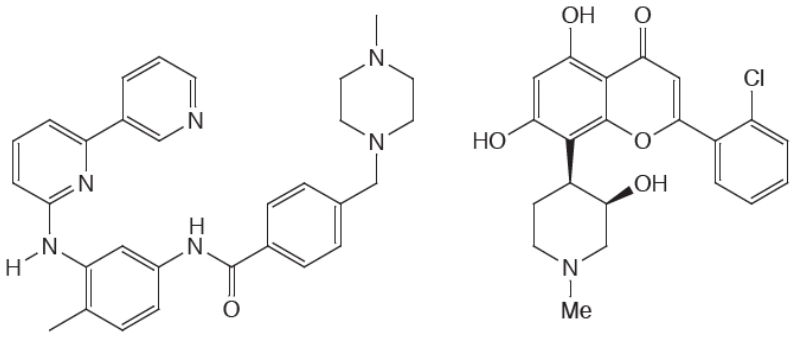
14. Select **Calculations...** from the **Setup** menu . This should already designate **Similarity Analysis** and the appropriate library (***cephaeline***). Change the entry in the **Use** menu to **CFD**. *Click* on **Submit**.

15. When the analysis is complete, select **Similarities...** from the **Display** menu . One or more hits should appear in the box at the right of the dialog. Select a hit to get a composite graphic (CFD and structure from the library). You can easily see the extent to which the two are matched.

16. Close any open documents and dialogs.

5 mins

Gleevec. Making a Pharmacophore from PDB



gleevec flavopiridol

In this tutorial, you will start with a protein structure from the PDB in which the so-called protein kinase inhibitor, gleevec, is found*. Instead of abstracting the molecular structure of gleevec, you will abstract its "binding environment", that is, the locations from which gleevec interacts with the protein host in terms of hydrogen-bond or charge-charge interactions (non-steric contacts). This information, together with knowledge of either the steric requirements of the guest or the space occupied by the protein (excluded volumes), constitutes a structural pharmacophore, that is, a "skeleton" into which other possible "guests" might be accommodated.

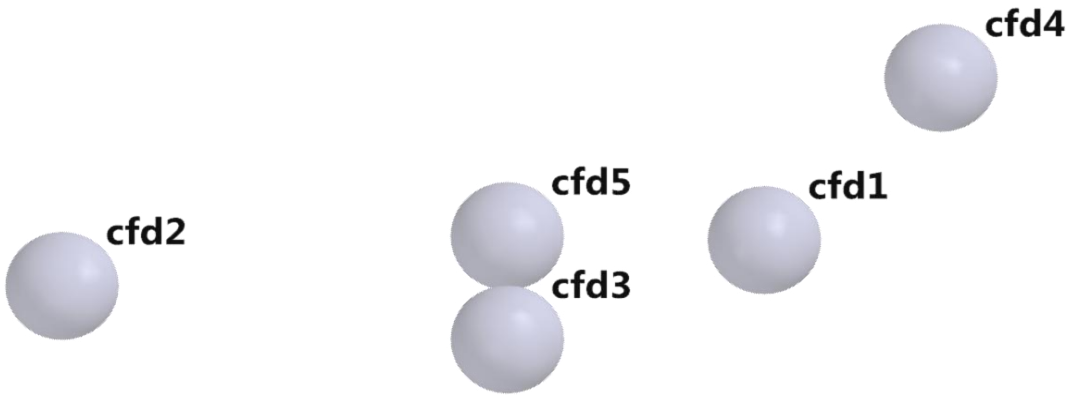
You will then use a subset of the non-steric contacts together with excluded volume elements to see to what extent this pharmacophore fits another protein kinase inhibitor, flavopiridol.

1. Select **Access PDB Online** from the **File** menu to obtain the PDB file ***1opj*** from the PDB or open ***1opj*** from the ***medicinal chemistry*** sub-directory (***tutorials*** directory)*. Select **Extract Ligands** from the **File** menu , *click* on the model for gleevec (it will be designated as Ligand STI 3 or 4) inside the protein structure (displayed as a set of transparent spheres similar to a space-filling model) to select it, and *click* on the **Exact Ligands** button at the lower right of the screen. Inside the **Extract Ligands** dialog, *check* both **HBA/HBD** and **+/- Centers** and **Excluded Volume Centers**. If any other items are selected, make certain to deselect them before you *click* on **OK**.

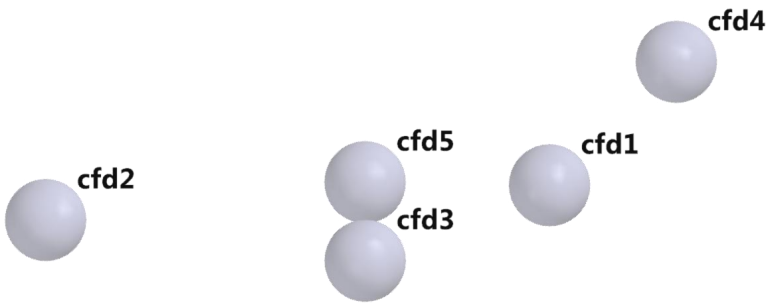
* B. Nager *et. al*., Cell, **112**, 859 (2003): PDB identification 1opj.

2. You are finished with ***1opj*** and may close it. Also, simplify the display of the extracted information by removing the excluded volume elements from view. Select **Configure...** from the **Model** menu and *click* on the **CFDs** tab. *Uncheck* **Excluded Volumes** and *click* on **Apply** (this will hide the excluded volumes and leave the **Configure** dialog on screen).

3. The model that now appears comprises five purple spheres representing non-steric contacts between gleevec and its protein host. Turn on labels. In the **Configure** dialog, *click* on the **Labels** tab, *check* **CFD Labels** and *click* on **OK**.



* For Windows, this directory is found in ***Program Files/Wavefunction/Spartan26***. It must be copied to another location available to the user prior to opening it in ***Spartan***. For Linux, this is found in the directory where ***Spartan*** was installed. For Macintosh, this is located at the top level of the ***Spartan'26*** disc image.



These have been assigned based on the observation that each is within a predefined distance of (one or more) complementary groups from residues in the protein host. Because hydrogen positions are not assigned in the experimental X-ray structure it may not be possible to say whether a particular site on the guest is acting as a hydrogen-bond donor or acceptor. It may not even be possible to say whether a particular site is protonated or deprotonated (although pH may often dictate this). To be "safe", all non-steric contacts have initially been given four CFD definitions: hydrogen-bond acceptor (**HBA**), hydrogen-bond donor (**HBD**), positive ionizable (**+**) and negative ionizable (**–**). Examination of the structure of gleevec suggests more focused assignments: **1** (**HBA**, +), **2** (**HBA**), **3** (**HBD**), **4** (**HBA**) and **5** (**HBA**). (See previous figure for numbering.)

4. You will use only three elements of this pharmacophore (together with the excluded volume elements) to assess whether or not flavopiridol is likely to fit into the same host environment as gleevec. Select **Properties** from the **Display** menu . *Click* on **CFD1**, remove the checks next to **HBD** and **-Ionizable** (leaving **HBA** and **+Ionizable**). Repeat the process, designating **CFD3** as **HBD** only, and **CFD5** as **HBD** only. Next, select **Set Similarity Centers** from the **Geometry** menu . Ensure that the **CFD** (not **Structure**) is selected in the lower right corner of the screen. *Click* **CFDs 1**, **3**, and **5**. A violet circle will appear on each, indicating that it is to be used in the similarity analysis calculation. *Click* on .

Select **Calculations...** from the **Setup** menu and then **Similarity Analysis** from the top left menu to the right of **Calculate**. *Click* on **Add Library...** and then locate and select ***flavopiridol*** in the ***medicinal chemistry*** sub-directory (***Tutorials*** directory) and *click* on **Open**. Make certain that **CFD** is selected next to **Use:** in the lower right of the **Calculations** dialog. *Click* on **Submit**. Name the document ***flavopiridol fit to gleevec pharmacophore***.

5. When the similarity analysis has completed, select **Similarities...** from the **Display** menu . Sort the hits according to similarity score (*click* on **Score** at the top of the dialog). One after the other, *click* on the top scoring entries in the box to the right of the dialog and examine the fits in the window at the left.

6. Close any open documents and dialogs.

Chapter 13

Flexible Molecules

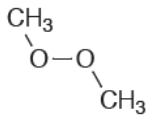
This chapter describes procedures available for calculations on flexible molecules.

Up to this point, we have not paid much attention to the different conformers that molecules might adopt. Many of the molecules dealt with so far are rigid or dominated by a single conformer and, where they are not, we have chosen to downplay the importance of conformation. In fact, the vast majority of molecules are not rigid and knowing the "right" conformer or conformers can be important to properly describing molecular spectra and properties and overall chemical behavior.

The five tutorials in this chapter illustrate a variety of approaches for dealing with flexible molecules. The first tutorial presents a molecule with only a single degree of (conformational) freedom. Here the focus is not only on identifying the "best" conformer, but also on rationalizing why it is the best. The second tutorial explores a molecule with two degrees of freedom. The focus here is on identifying all "reasonable" conformers and comparing the ordering of energies for these conformers with different theoretical models. The third tutorial involves a molecule with several degrees of conformational freedom and hundreds of conformers. Its goal is simply to find the lowest- energy conformer using practical methodology. The last two tutorials have more of a chemical focus. The first of these asks a question about the "price" that needs to be paid in order for a molecule to be properly oriented for an intramolecular reaction. The second asks a similar question, but this time with regard to a proper conformation for binding to a protein.

25 mins

Internal Rotation in Dimethylperoxide



Quantum chemical calculations may be called on to furnish data to parameterize empirical energy functions for use in molecular mechanics and/or molecular dynamics calculations. Perhaps most important are data relating to torsional motions where small changes torsional angles may lead to large changes in overall shape, as well as significant changes in molecular spectra and properties.

Any function chosen to represent the energy of rotation about a single bond needs to reflect the inherent periodicity, that is, it must repeat itself in 360°. The most common choice is a truncated Fourier series.

$$E^{torsion}(\omega) = k^{torsion1}(1 - \cos(\omega - \omega^{eq})) + k^{torsion2}(1 - \cos 2(\omega - \omega^{eq})) + k^{torsion3}(1 - \cos 3(\omega - \omega^{eq}))$$

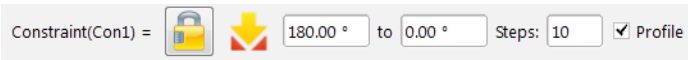
Here, ω^{eq} is the ideal dihedral angle and $k^{torsion1}$, $k^{torsion2}$ and $k^{torsion3}$ are parameters. The first (one-fold) term accounts for the difference in energy between *syn* and *anti* conformers, the second (two-fold) term for the difference in energy between planar and perpendicular conformers, and the third (three-fold) term for the difference in energy between eclipsed and staggered conformers.

In this tutorial, you will examine the energy for rotation about the oxygen-oxygen bond in dimethylperoxide, first using ωB97X-D/6- 31G* density functional calculations to establish geometry and then using ωB97X-V/6-311+G (2df,2p)[6-311G*]calculations to get a more accurate account. Both ωB97X-D and ωB97X-V functionals account for long-range dispersive interactions, the former by an empirical correction and the latter more properly by explicit calculation. The 6-311+G(2df,2p) basis set is much more flexible than 6-31G*, but the calculation can be simplified using the so-called dual-basis set approximation where convergence is first reached with the smaller 6-311G* basis set. You will fit your data to a truncated Fourier series.

1. Build dimethylperoxide. If the molecule is not already in an *anti*-conformation, select **Measure Dihedral** from the **Geometry** menu, *click* on COOC atoms in order, *click* inside the box to the right of **Dihedral...** and enter **180** (180°) using the number pad that appears. ***Do not minimize***.

2. Select **Constrain Dihedral** from the **Geometry** menu. Select the COOC torsion (*click* on COOC atoms in order), and then *click* on at the bottom right of the screen. The icon will change to indicating that a dihedral constraint is to be applied.

3. *Check* the box to the left of **Profile** at the bottom right of the screen. This will result in two additional text boxes.



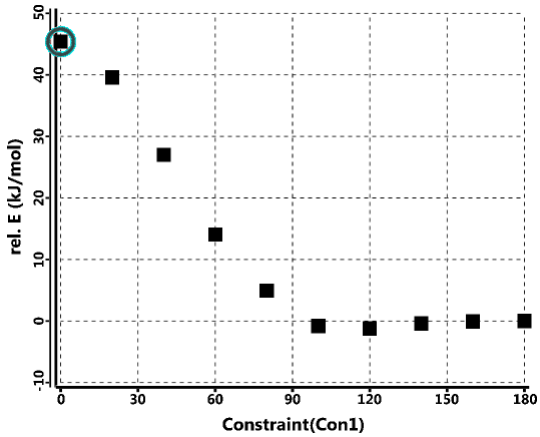
Leave 180° in the original (leftmost) box alone but change the value in the box to the right of "**to**" to 0°. *Click* inside the box and enter **0** using the number pad. **Steps** should be 10. If it is not, *click* inside the box and enter **10** using the number pad. What you have specified is that the dihedral angle will be constrained first to 180°, then to 160°*, etc. and finally to 0°. *Click* on .

4. Select **Calculations...** from the **Setup** menu and specify **Energy Profile** from the top menu to the right of **Calculate**, and **Density-Functional, ωB97X-D** and **6-31G*** from the three bottom menus. *Click* on **Submit** and accept the name ***dimethylperoxide***.

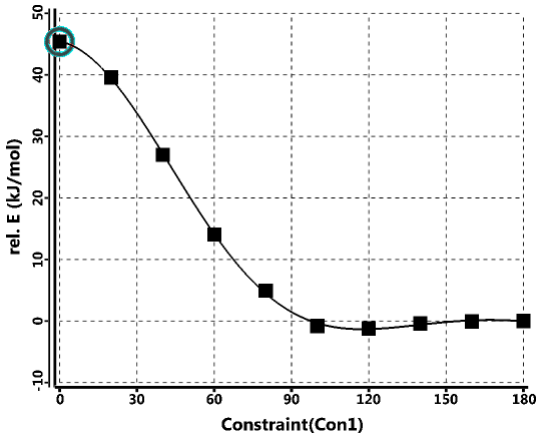
5. When the 10 calculations have completed (perhaps 5 minutes), they return in a new file named ***dimethylperoxide.Prof.M0001***. A prompt will ask you if you want to open this auxiliary file. *Click* on **Open**. Align the conformers. Select **Align** from the **Geometry** menu, select **Structure** from the **Align by** menu at the bottom right of the screen and, one after the other, *click* on both oxygens and on one of the carbons. Then *click* on the **Align by** button at the bottom right of the screen. Select **Spreadsheet** from the **Display** menu, and enter both the energies relative to the 180° conformer, and the COOC dihedral angles. First *click* on the label (**M0001**) for the top entry in the spreadsheet (the 180° conformer), then *click* on the header cell for the leftmost blank column, and finally, *click* on **Add...** at the bottom of the spreadsheet. Select **ΔE (kJ/mol)** from the quantities in the **Molecule List** tab (under **Energy from Molecule**) and *click* on the spreadsheet to release the dialog. Enter the dihedral angle constraints: select **Constrain Dihedral** (**Geometry**) menu, *click* on the constraint marker attached to dimethylperoxide and *click* on at the bottom of the screen (to the right of the value of the dihedral angle). *Click* on .

* The difference between constraint values is given by: (final-initial)/(steps-1).

6. Select **Plots** from the **Display** menu . *Click* on at the top of the (empty) plot pane and select **Constraint(Con1)** from the **X Axis** menu and **ΔE (kJ/mol)** from the **Y Axes** list and then *click* **Create**.



To fit the points to a Fourier series, *click* on at the top of the plot pane, select **Fourier** to the right of **Curve** in the resulting dialog and *click* on **Done**.



7. To get a better account of the energy profile for rotation about the OO bond in dimethylperoxide, perform calculations with the ωB97X-V/6-311+G (2df,2p)[6-311G*] density functional model* (using the equilibrium geometries that you obtained from the ωB97X-D/6-31G* model). First, make a copy of ***dimethylperoxide.Prof.spartan***. Name it ***dimethylperoxide ωB97X-V***. Select **Calculations...** from the **Setup** menu and specify calculation of energy using the ωB97X-V/6-311+G(2df,2p) density functional model. Make certain that **Global Calculations** (at the bottom of the dialog) is *checked* to signify that energy calculations are to be performed on all conformers. *Click* on **Submit**.

8. Energy calculations for all ten conformers will require several minutes to complete. When they are done, draw a new energy plot and compare it to the energy plot produced earlier.

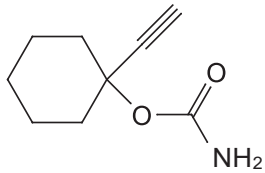
9. For the sake of comparison, return to the ***dimethylperoxide.Prof.spartan*** file, save a copy as ***dimethylperoxide_Est_DFT***. Select **Calculations...** from the Setup menu and specify calculation of **Energy** using **Est. Density Functional** and **ωB97X-V/6-311+G(2df,2p)**. Make sure **Current Geometry** is selected from the **Start From** pull down menu. *Click* on **Submit**.

10. The neural network energy calculations for all ten conformers will require less than 20 seconds to complete. When they are done, draw a new energy plot and compare it to the energy plot produced from the model used in step 7. How well does the neural network reproduce the QM values?

11. Close all documents and dialogs.

30 mins

Ethinamate



Ethinamate is a prescription drug previously used for the treatment of insomnia. It involves two degrees of conformational freedom, rotation of the single bond to the carbonate group (two unique arrangements) and inversion of the six-member ring (two unique chair conformers).

1. Build or sketch ethinamate. *Click* on . Select **Calculations...** from the **Setup** menu and request a **Conformer Distribution** with **Density Functional, ωB97X-V** and the **6-311+G(2df,2p)[6-311G*]** basis set using **ωB97X-D/6-31G*** geometry(ies). This invokes a multistep recipe, starting from a systematic conformer search using MMFF molecular mechanics and proceeding through several immediate steps to gradually reduce the number of conformers for the final energy calculation (*click* on **Details** to explore the steps). *Click* on **Submit** and accept the name ***ethinamate***.

2. This multi-step calculation will require several minutes to complete. When it does, ***ethinamate*** will display the lowest-energy conformer but the total energy, dipole moment and Mulliken atomic charges* will refer to a Boltzmann weighted average over all accessible conformers. A new ***Spartan*** document containing these conformers (***ethinamate_Conf***) has also been returned, choose to Open the auxiliary file. In the new document, enter the total energies into the spreadsheet. You will see that the total energy of the best conformer is in fact lower than that of the Boltzmann averaged value (found in ***ethinamate***).**

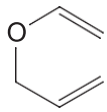
3. Close any open documents and dialogs.

* Natural charges are available only for the 6-311+G(2df,2p) basis sets. Electrostatic charges are too costly to calculate for molecules with large numbers of accessible conformers and are deliberately excluded.

** This may be surprising but is necessarily the case as the Boltzmann average, while dominated by the lowest-energy conformer, includes contributions from higher energy conformers.

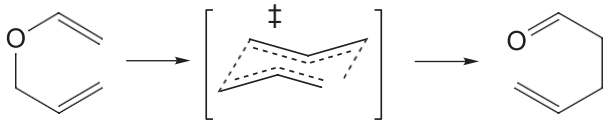
5 mins

Allyl Vinyl Ether



The **Equilibrium Conformer** and **Conformer Distribution** selections under **Calculations** in the **Setup** menu directly access the MMFF molecular mechanics model and either the ωB97X-V or B97M-V density functional models with the 6-311+G(2df,2p) [6-311G*] basis sets. They represent the extremes in terms of cost computation and are intended to be used in two situations with entirely different expectations. There are numerous models and combinations of models in between and the purpose of this tutorial is to explore one of them. Here we seek a semi-quantitative account of the energy required to go from the "best" conformer of a molecule to a conformer poised for chemical reaction. This is different than the expectation of the following tutorial which is to accurately account for the NMR spectrum of a flexible molecule.

Allyl vinyl ether undergoes Claisen rearrangement, the mechanism presumes a chair arrangement of the reactant.



Is this the lowest-energy conformer (global minimum) or is additional energy required to properly orient the molecule for reaction? To find out, you need to locate all the conformers of allyl vinyl ether, identify the best chair structure and evaluate its energy relative to that of the actual global minimum. You will first carry out a conformational search using the MMFF molecular mechanics model and then refine relative conformer energies using single-point ωB97X-D/6-31G* density functional calculations based on 3-21G Hartree-Fock equilibrium geometries. The results are not expected to be as good as those obtained from better geometry and especially better energy calculations, but they should be sufficient for the objective at hand.

1. Select **New** from the **File** menu and either build or sketch allyl vinyl ether.

2. Select **Calculations...** from the **Setup** menu and specify **Conformer Distribution** from the top menu to the right of **Calculate** and **Molecular Mechanics** and **MMFF** from the two bottom menus. *Click* on **Submit** and accept the name ***allyl vinyl ether***.

3. When completed, it will give rise to a series of low-energy conformers* placed in a new file ***allyl vinyl ether_Conf***.** A prompt will ask you if you want to open this file. *Click* on **Open**.*** Select **Spreadsheet** from the **Display** menu . Size the spreadsheet such that all rows (corresponding to different conformers) are visible at one time. *Click* on **Add...** at the bottom of the spreadsheet. *Click* on the **Energy (kJ/mol)** button from the **Molecule** tab (you may need to change the units) and *click* on the spreadsheet (or background) to remove the **Add** dialog. Energies for each of the different conformers will be added to the spreadsheet. Examine the lowest-energy conformer (the top entry). Is it in a chair conformation suitable for Claisen rearrangement? If not, identify the lowest-energy conformer that is suitable. You can keep two or more conformers on screen at the same time by *checking* the boxes immediately to the left of the molecule labels (the leftmost column) in the spreadsheet. To get a clearer idea of structural similarities and differences, align the conformers. Select **Align** from the **Geometry** menu and select **Structure** from the **Align by** menu at the bottom right of the screen. And at bottom left, a message:

Select atoms.

Click on oxygen and on the two carbons of the vinyl group bonded to oxygen to designate them as alignment centers. Each will be marked by a red circle. If you make a mistake, *click* on the circle and it will disappear. When you are done, *click* on the **Align by** button at the bottom right of the screen. *Click* on .

---

* By default, only conformers within 40 kJ/mol of the global minimum will be kept. This can be changed (see **Conformations and Energy Profiles** in the **Spartan'26 Help** menu).

** **M0001** is the default label of the molecule you built. You can change it by altering the **Labels** field in the **Molecule Properties** dialog (**Properties** under the **Display** menu; **Chapter 22**).

*** To avoid confusion, it is a good idea to close the original file ***allyl vinyl ether***.

4. To obtain a better estimate of the energy required to twist allyl vinyl ether into a conformer suitable for Claisen rearrangement, perform ωB97X-D/6-31G* energy calculations using 3-21G geometries. Select **Save As** from the **File** menu to make a copy of ***allyl vinyl ether_Conf***. Name it ***allyl vinyl ether density functional***. Using the copy, delete all conformers except the global minimum and the lowest-energy conformer that appears to be most closely poised for Claisen rearrangement. Select each conformer to be discarded, and then select **Delete Molecule** from the **File** menu . Alternatively, *click* on the cell in the spreadsheet containing the label for the molecule to be deleted, and then *click* on **Delete** at the bottom of the spreadsheet. When you are done, the spreadsheet should contain only two entries.

5. Select **Calculations...** from the **Setup** menu and specify an **Energy** calculation and **ωB97X-D Density Functional** method with the **6-31G*** basis set. Also, specify **Hartree-Fock** and **3-21G** from the **Start from** menu. Make certain that **Global Calculations** at the bottom of the dialog is checked to specify that the dialog settings apply to both conformers. *Click* on **Submit**.

6. When completed, examine the conformer energies. Select **Spreadsheet** from the **Display** menu. *Click* on the **Add** button at the bottom of the spreadsheet and *click* on the **ΔE (kJ/mol)** button from the **Molecule List** tab. *Click* on the spreadsheet to release the **Add** dialog. What does this model report as the energy needed to go from the global minimum to a conformer poised to undergo Claisen rearrangement?

7. Close all open documents and dialogs.

2 hours

^{13}C Chemical Shifts Depend on Conformation

At normal temperatures, the time for nuclear spin relaxation is very long relative to time required for equilibration among conformers. This means that for each of the carbon chemical shifts in the NMR spectrum of a flexible molecule, ^{13}**C**, will be a weighted average of the shifts NMR of the individual conformers, $^{13}C_i$.

$$^{13}\mathbf{C} = \Sigma_i\, \omega_i\, ^{13}C_i$$

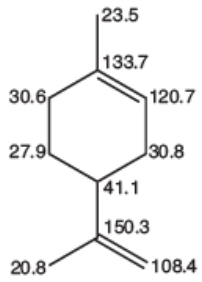
The weight, ω_i, is given by the Boltzmann equation, and depends on its energy, εi (relative to that of the lowest-energy conformer, $\varepsilon 0$) and on temperature, T. Summation is over all conformers (including the lowest-energy conformer), g_i is the number of times that conformer i appears in the overall distribution and k is the Boltzmann constant.

$$\omega_i = g_i \exp - [\varepsilon_i - \varepsilon_0/kT]/\Sigma_j \{\exp - [\varepsilon_j - \varepsilon_0/kT]\}$$

In practice, conformers that are 10 kJ/mol or more above the lowest- energy conformer make a negligible contribution to the total (at room temperature). However, in order to actually know which conformers are important and to construct a proper Boltzmann average, it is necessary to examine all of them.

In this tutorial, you will use the **2019 QM-NMR** protocol, first introduced in the ***Spartan'18*** release to calculate the ^{13}C NMR spectrum of limonene. This comprises two parts, the first a multi-step protocol starting from MMFF molecular mechanics and ending with a series of large basis set density functional calculations to obtain "accurate" Boltzmann weights, and the second a set of small basis set density functional calculations to obtain ^{13}C chemical shifts. The two parts are then combined to produce a Boltzmann averaged ^{13}C spectrum.

1. Build or sketch limonene.



2. Enter experimental ^{13}C chemical shifts (shown above) either on the sketch or 3D model. Select **Expt. Chem. Shifts** [^{1}H/^{13}C] from the **Expt. Data** menu. *Click* in turn on each carbon, *click* inside the box that appears to its right, use the number pad to type in the chemical shift and *click* on **Enter**.

3. Select **Calculations...** from the **Setup** menu, **NMR Spectrum** (**Density Functional**)*, **ωB97X-D** and (**6-31G***)* from the top three menus to the right of **Calculate** and **ωB97X-V/6-311+G(2df,2p)[6-311G*]** energy and **ωB97X-D/ 6-31G* geometry** from the bottom two menus. *Click* **Submit**.

The calculation may require an hour or more.** When completed, bring up the **Properties** dialog (**Display** menu) and *click* on the **NMR** tab. Statistics related to the fit between calculated and experimental ^{13}C shifts are provided on the left at the center of the dialog that results, specifically **RMS**, **Max Absolute** and **Mean Absolute** errors. If the **Boltzmann Average** tab is checked, these correspond to a Boltzmann weighted average over conformers. If the **Best Fit** tab is checked, they instead correspond to the conformer which best fits measured values. You will see that both measures provide good concurrence between calculated and measured ^{13}C chemical shifts.

Spartan provides two different ways to assess the fit between calculated and measured ^{13}C chemical shifts. The first and clearly the more rigorous measure involves constructing a proper Boltzmann-weighted average. This requires accurate conformer energies and associated Boltzmann weights which in turn functionals that take proper account of long-range interactions such as the ωB97X-V functional and large basis sets such as 6-311+G(2df,2p). While such an approach can be applied to medium-size organic molecules with moderate conformational flexibility, it can involve many hours of computer time. The second approach only requires that a set of "good" conformers be identified, and from among these selects the conformer ^{13}C chemical shifts for which best fit measured values. Lacking the need for accurate Boltzmann weights, it can rely on simpler functionals such as ωB97X-D where long range interactions are dealt with empirically and smaller basis sets such as 6-31G*, and as a result, is an order of magnitude faster.

---

* Pre-selected menu items cannot be altered.

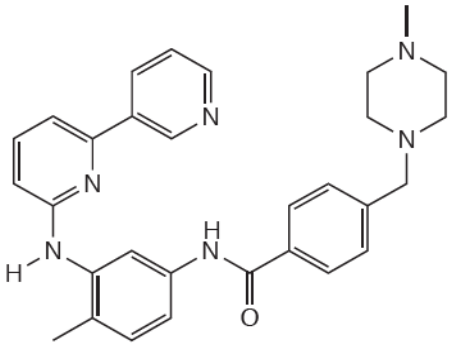
** If you don't want to wait, a version of the file is available from the ***spectra of organic molecules*** directory in the ***Tutorials*** folder. For Windows, see ***Program Files/Wavefunction/Spartan26/Tutorials***. It must be copied to another location available to the user prior to opening it in ***Spartan***. For Linux, the ***Tutorials*** is found in the directory where ***Spartan*** was installed. For Macintosh, ***Tutorials*** is located at the top level of the ***Spartan'26*** disc image.

4. The returned structure corresponds to that with the lowest energy from the ωB97X-V/6-311+G(2df,2p)[6-311G*] model (the leading contributor in the Boltzmann average). A secondary file, **limonene_Conf**, has been created (but not opened). This contains NMR chemical shift calculations for all conformers. Choose to either **Discard** or **Save** the auxiliary file. In the parent file, go to **Model** menu **Configure**… in the **Labels** tab, under **Atom**, *click* the radio button next to (**Calc-Expt.) Chem Shift**. How well does the multi-step protocol (**NMR Spectrum** task) do at reproducing the experimental shifts for limonene?

5. Close any open documents and dialogs.

2 mins

Gleevec. Protein Bound vs. Free Conformer



gleevec

Gleevec (aka Glivec, Imatinib) is a protein kinase inhibitor used in anticancer therapy. It specifically targets a protein kinase coded for in the rogue gene ***bcr-abl***. Several crystal structures with gleevec docked in protein kinases are available in the PDB. In this tutorial, you will examine one* to establish whether or not the conformation of gleevec in the protein is identical (or similar) to that of the free (gas- phase) molecule, and if not, what energy penalty is to be paid to adopt the protein-bound conformation. The next tutorial will return to this same structure and extract a pharmacophore (gleevec's footprint).

1. Retrieve a protein structure from the PDB. Select **Access PDB Online...** from the **File** menu. Type ***1opj*** into the box to the right of PDB ID and *click* on **Open**.

> If you are not online, you may skip this step as the structure is available as ***1opj*** in the ***medicinal chemistry*** sub-directory (***tutorials*** directory).**

The protein will be represented by a ribbon model. Red sections correspond to parts of the chain that are α helices, while blue regions correspond to to parts that are β sheets (the remaining regions are green). A Ramachandran plot (the φ *phi* and ψ *psi* torsional angles connecting the individual amino acids) allows you to see the clustering of α helices and β sheets. To display, select **Ramachandran Plot** from the **Model** menu. The colors of the dots in the resulting plot map one-to-one with the colors of the ribbon model. When you are done, *click* on to remove the plot.

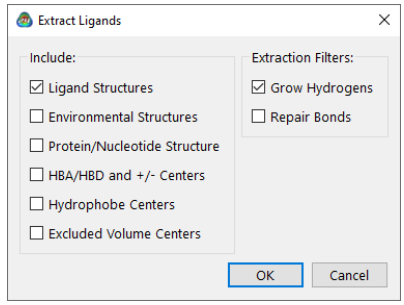
* B. Nager *et al*, Cell, **112**, 859 (2003): PDB identification 1opj.

** For Windows, this directory is found in ***Program Files/Wavefunction/Spartan26***. It must be copied to another location available to the user prior to opening it in ***Spartan***. For Linux, this is found in the directory where ***Spartan*** was installed. For Macintosh, this is located at the top level of the ***Spartan'26*** disc image.

The colors on the ribbon display may be modified, by selecting **Configure...** from the **Model** menu ( ) and *clicking* on the **Ribbons** tab. Coloring **By Residue** will give each amino acid its own unique color.

The protein returned from PDB incorporates two sets of bound molecules, depicted in the model as two sets of translucent spheres. Gleevec is the larger of the two incorporated molecules and is designated by its ***PDB HET*** code ***STI***.

2. Select **Extract Ligands** from the **File** menu . Select (*click* on) one of the gleevec molecules inside the protein structure and *click* on the **Extract Ligands** button at the bottom right of the screen. This leads to the **Extract Ligands** dialog.



Check **Ligand Structures** under **Include** and **Grow Hydrogens** under **Extraction Filters** inside the dialog (if other options are selected, remove them) and *click* on **OK**.

3. A ball-and-spoke model for gleevec will appear on screen. (You no longer need the protein; select and close it.) Put a copy of gleevec on the clipboard. With the molecule selected, choose **Copy** from the **Edit** menu or right *click* on the background and select **Copy** from the menu that appears.

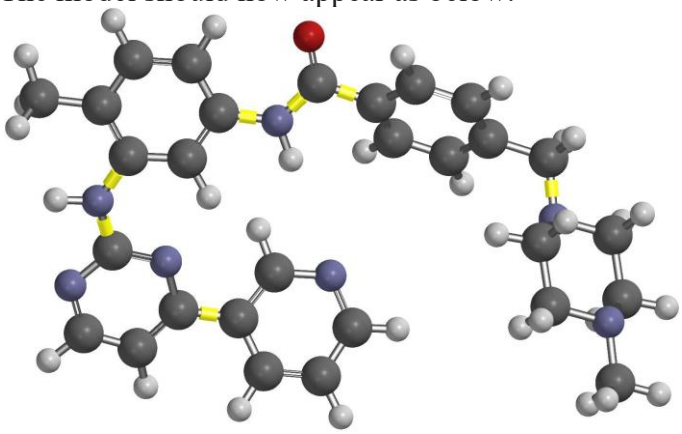
4. Select **New Molecule** (not **New**) from the **File** menu . Select **Clipboard** from the organic model kit and *click* anywhere on screen. *Click* on . You now have two copies of gleevec in a single document. You can step between them using the step buttons at the bottom left of the screen.

5. Select the first copy; select **Calculations...** from the **Setup** menu . **Equilibrium Geometry** from the top left menu to the right of **Calculate** and **Molecular Mechanics** and **MMFF** from the two bottom menus. The conformation will not be altered, but bond lengths and angles will be optimized.

Structure optimization is necessary in order to establish the energy difference between protein-bound and free conformations of gleevec. The resolution of protein X-ray crystallography is not adequate to establish bond lengths and angles to chemical accuracy.

Remove the checkmark from **Global Calculations** at the bottom of the dialog to signify that your choice (geometry optimization) only applies to this copy of gleevec. *Click* on **OK** (not on **Submit**).

6. You will use the second copy of gleevec as the starting point of a conformational search. First, we will adjust some of the default degrees of conformational freedom ***Spartan*** associates with the molecule (in order to reduce the computer time required for the conformational search). With the second copy, select **Set Torsions** from the **Geometry** menu . Your model will be augmented with yellow cylinders designating flexible bonds and yellow circles designating rings, respectively. Remove the markers (circles) on the piperazine ring by *double clicking* on each nitrogen atom in turn. The conformation of the piperazine ring will not change during the search. The model should now appear as below.



7. Select **Calculations...** from the **Setup** menu . Select **Equilibrium Conformer** from the top left menu to the right of **Calculate** and (as before) **Molecular Mechanics** and **MMFF** from the two bottom menus. *Click* on **Submit**, and provide the name ***gleevec protein bound vs. free conformer***.

8. When the calculations have completed (perhaps minutes), select **Spreadsheet** from the **Display** menu . *Click* inside leftmost the cell for the second molecule (the one on which a conformational search was performed) then *click* inside the header cell for a blank data column. *Click* on **Add** at the bottom of the dialog and select **ΔE (kJ/mol)** from the **Molecule List** tab (under **Energy from Molecule**). The relative energy of the protein bound conformer is provided in the spreadsheet. Is it close to zero, meaning that the protein-bound conformer will be present in significant amount in a sample of free gleevec? Or is there an energy barrier, indicating a lower distribution?

9. Close ***gleevec protein bound vs. free conformer*** and any open dialogs.

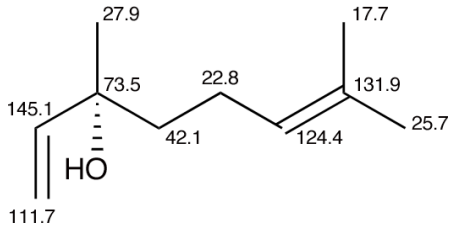
2 hours

Linalool: Calculating the NMR Spectrum of a Flexible Molecule

Spartan provides computational options for calculating proton and ^{13}C spectra as well as 2D COSY and HMBC spectra of flexible molecules. The more rigorous and more costly calculations involve construction of a proper Boltzmann-weighted average. In practice, calculation of accurate Boltzmann weights for all accessible conformers followed by calculation of chemical shifts and (optionally) coupling constants for each of the conformers. Obtaining the Boltzmann weights is the limiting step as it requires density functional models that account explicitly for long-range van der Waals or dispersive interactions coupled with large basis sets. Calculations of moderate size (MW in the range of 500 amu) with 100 or more conformers can easily consume many hours on present generation personal computers.

In this tutorial, you will calculate ^{13}C chemical shifts for the natural product linalool using two different procedures. The first will use MMFF to clean up structure (only) for a randomly built conformer and will then calculate equilibrium geometry and chemical shifts using ωB97X-D/6-31G* (for that specific conformer). The second procedure will involve the multi-step protocol used for the **NMR Spectrum** task that includes MMFF, a neural network model, and multiple QM density functional calculations to provide a Boltzmann averaged set of chemical shifts. Results from both calculations will be compared to reported experimental shifts.

1. Build or sketch linalool. Enter the ^{13}C shifts (see figure below). Select **Expt. Chem. Shifts** from the **Expt. Data** menu, *click* on each chemically unique carbon, *click* on **Edit** to the right of **Expt. Chem. Shifts** at the bottom right of the screen, type in a chemical shift using the number pad and *click* on **Enter**. When you are done, *click* on .



2. Open the **Calculations** dialog (**Calculations** from the **Setup** menu) and select **Equilibrium Conformer** with **Molecular Mechanics** and **MMFF**. *Click* **Submit**. Name the document ***linalool_MMFF.spartan***. When the job completes, return to the **Calculations** dialog. Specify an **Equilibrium Geometry** with **Density Functional ωB97X-D** and **6-31G***. **Submit** the job.

3. When the job completes visit the **Configure…** entry in the **Model** menu. From the **Label** tab *click* the radio button next to **(Calc. – Expt.) Chem Shift**. How well do the ^{13}C shifts for the MMFF equilibrium conformer compare to experiment? In general anything under 2 ppm is considered reasonably good–certainly anything 5 ppm or larger is worth review and may indicate of potential error in assignment.

4. Bring up the **Molecule Properties** dialog and click on the **NMR** tab. Make note of the **Mean Absolute** error. You can, if you like, post this (and the other error values) into the spreadsheet by clicking on the button to the left of the value. In a random conformer, it is not surprising to see a mean absolute error of over 2 ppm.

5. *Right-click* on the background and choose **Copy**. From the **File** menu choose **New Build** and then *right-click* on screen and choose **Paste**. Open the **Calculations** dialog (**Calculations** from the **Setup** menu) and select **NMR Spectrum** (**Density Functional**), **ωB97X-D** and (**6-31G***) from the four top menus to the right of **Calculate** and **ωB97X-V/6-311+G(2df,2p)[6-311G*]** and **ωB97X-D/6-31G*** from the two bottom menus. For additional details on the full multi-step protocol, check the box next to **Details**. The specified calculation will take considerable computational time (1-2 hours). *Click* on **Submit**.

6. The job will take a good deal of time to complete; you can loosely monitor the progress from the **Options** menu **Monitor** dialog. When the job completes, bring up the NMR properties dialog (**Properties** from the **Display** menu and *click* on **NMR**). Make sure the **Boltzmann Average** tab at the center left of the dialog is selected. This will report how well the conformationally searched and **Boltzmann Averaged** calculated shifts fit the experimental data. Review **RMS** error, widely used with values less than or close to 2.0 ppm indicating a good match.

7. The data reported can be configured to report either the conformer that best fits the experimental NMR data (**Best Fit**) or the proper Boltzmann-averaged chemical shifts (**Boltzmann Average**). Compare results from both the Best Fit tab and the Boltzmann Average tab. In general errors these results are typical of the NMR Spectrum calculations on flexible organic molecules. Results from the **Boltzmann Average** tab should be similar to:

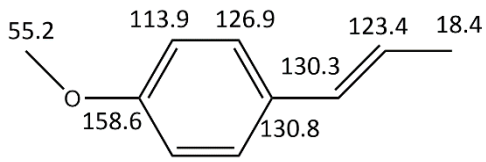
RMS	1.8 – 1.9 ppm
Max Absolute	3.3 – 3.5 ppm
Mean Absolute	1.5 – 1.7 ppm

8. Close any ***Spartan*** documents and open dialogs.

30 minutes

Anethole: Boltzmann Averaging ^{13}C Chemical Shifts

The natural product anethole provides a very simple illustration of obtaining ^{13}C chemical shifts as a Boltzmann-weighted average over different conformers.



The molecule is small and there are only two conformers, so the task can be completed in a relatively short time (perhaps 15 minutes or so).

1. Build or sketch anethole. Enter ^{13}C experimental chemical shifts (provided in the drawing above). Select **Expt. Chem. Shift** from the **Expt. Data** menu and in turn *click* on each unique carbon*, *click* on **Edit** to the right of **Expt. Chem. Shift** at the bottom right of the screen, use the number pad to enter the value of the shift and *click* on **Enter**.

2. From the **Calculations** dialog (**Calculations** from the **Setup** menu), select **NMR Spectrum**, **Density Functional**, **ωB97X-D** and **6-31G*** from the four top menus to the right of **Calculate** and **ωB97X-V/6-311+G(2df,2p)[6-311G*]** energy and **ωB97X-D/6-31G*** geometry from the two bottom menus. You have requested that Boltzmann weights be obtained in a multi-step procedure for the (two) conformers of anethole using the ωB97X-V/6-311+G(2df,2p)[6-311G*]//ωB97X-D/6-31G* model. And, following this, for chemical shifts of the conformers to be calculated with the empirically corrected ωB97X-D/6-31G* density functional model, allowing the Boltzmann average to be obtained. *Click* on **Submit** at the bottom of the dialog.

3. When completed (probably 10-20 minutes), bring up the NMR properties dialog (**Properties** under the **Display** menu and *click* on the **NMR** tab. Make sure that **Boltzmann Average** tab (not **Best Fit**) tab is selected. The reported error statistics indicate an excellent fit of calculated to experimental ^{13}C shifts.

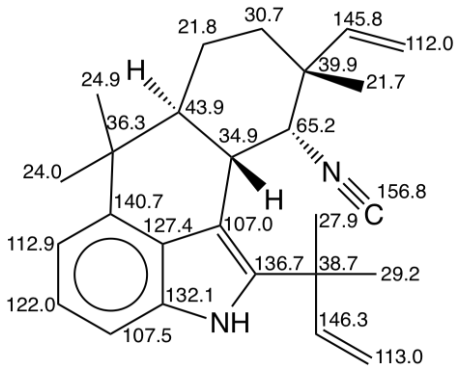
4. Close any open ***Spartan*** documents and dialogs.

* ***Spartan*** detects identical atoms (in this example the two carbons ortho and the two carbons meta) to the methoxy group. You only need to enter experimental data for unique atoms.

4+ hours

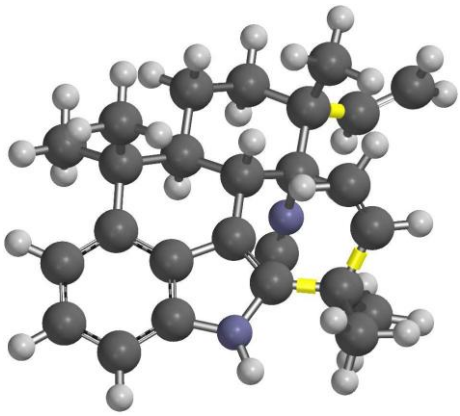
Ambiguine H Isonitrile. DP4 Analysis to Assign Stereoisomer

The structure assigned to the natural product ambiguine H is shown below. It contains four stereocenters, two of which designated R dictate the disposition of substituents attached to the incorporated trans-decalin bicyclic.

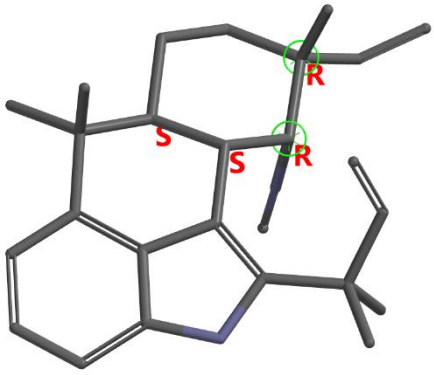


In this tutorial, you will use the DP4 metric to confirm or refute the experimentally assigned stereochemistry about these two centers. This involves chemical shift calculations on the four distinct structures each of which is conformationally flexible.

1. Build or sketch ambiguine H. To enter the experimental ^{13}C chemical shifts, select **Expt. Chem. Shift** from the **Expt. Data** menu. *Click* on each carbon in turn, *click* on **Edit** to the right of **Expt. Chem. Shift** at the bottom right of the screen (or to the right of the atom center), use the number pad to type in a value and *click* on **Enter**. **Note:** acknowledging the tedium of hand-entering (there are 26 carbons), you can, if you prefer, simply start from the copy of ***ambiguine H isonitrile*** from the **flexible molecules** folder in the **Tutorials** directory.
2. Rotatable bonds should be properly marked, but it is a good idea to check. Select **Set Torsions** from the **Geometry** menu. The three bonds shown in the figure below should be marked (encircled with yellow cylinders). If any are not, *double click* on the bond to select. In the interest of reducing computational time, if any other bonds or ring atoms are selected, you may turn these off (*double-click* to deselect torsions, for atoms, *click* to select, and specify *off* for any twist boat or chair flip markers, when finished there should be a total of 3 bonds highlighted:



Make a list of the four stereoisomers. Select **Generate Isomers** from the **Geometry** menu. In turn, *click* on each of the two stereocenters bearing substituents. They will be encircled in green (see figure below). *Click* on **Generate** list to the right of **Isomers** at the bottom right of the screen. Select **New Document** from the dialog that appears and *click* on **OK**. The four isomers (including the RR stereoisomer that you started with) will be placed in a new ***Spartan*** document. *Click* on **Save** (**File** menu) and name it ***ambiguine H isomers***. It's a good time to close the original document.



3. Bring up the **Calculations** dialog (**Calculations** from the **Setup** menu) and select **NMR Spectrum**, (**Density Functional**), **ωB97X-D** and **6-31G*** from the four top menus to the right of **Calculate** and for the bottom 2 menus, to the right of **using**, **Est. ωB97X-V/6-311+G(2df,2p)** and, the **Est. ωB97X-D/6-31G*** geometry from the two bottom menus. You have requested a Boltzmann weighted NMR spectrum be calculated for each stereoisomer (which will require conformational searching for each isomer). Make certain that **Global Calculations** at the bottom of the dialog is checked before you *click* on **Submit**.

4. With the previous release of ***Spartan***, without access to the neural network models (the "Est." models), this calculation would have required many hours (overnight) to complete. It involves calculations on dozens to hundreds of molecules, depending on whether or not the **Set Torsion** step was precisely followed. ***Spartan*** will consider

somewhere between 18 (with only 3 set torsion bonds) and 288 conformers (with additional ring atoms selected) for each of the 4 stereoisomers. When the job completes(total calculation may run 5 or more hours depending on your hardware), bring up the NMR properties dialog. **Properties** under the **Display** menu and *click* on the **NMR** tab). Step through the four isomers and identify the one with the highest DP4 score. Is this the one that has been assigned based on its NMR spectrum? Are any other isomers competitive?

5. Close all open ***Spartan*** documents and dialogs.

Chapter 14

"Dry Labs" Using the Spartan Spectra and Properties Database

The tutorials in this chapter comprise examples where structures, energies, spectra and other properties can be extracted from the Spartan Spectra and Properties Database doing away with the need for new quantum chemical calculations.

The Spartan Spectra and Properties Database (SSPD) is a collection of molecular structures, spectra and diverse properties initially available from multiple computational models. At present the collection has settled on a single density functional model (ωB97X-D/6-31G*) for structure. Previously structures and frequencies were also available from the EDF2/6-31G* model (these are both closer to experimental frequencies and are much faster than the ωB97X-D/6-31G* model). Beginning with release of SSPD v.7, we've included frequencies from the corresponding equilibrium structures run at (ωB97X-D/6-31G*) – only. This allows access to thermodynamic data that is dependent on the frequency calculation – but hopefully alleviates confusion over the previous "mixed-model" data. With the release of ***Spartan'26*** a single standard ωB97X-D/6-31G* model data is used for all structures. At present, something on the order of ***>317,000*** molecules in the collection.

Each SSPD entry includes a wide range of molecular properties including a selection of QSAR descriptors, NMR spectra, IR Spectra, and the wave function, allowing "on-the-fly" generation of molecular orbitals, electron densities and other graphical surfaces as well as the electrostatic potential map and other property maps. In short, SSPD provides a rich source of diverse information for a large selection of molecules.

The emphasis of SSPD on only on primarily one theoretical model (ωB97X-D/6-31G*) is a deliberate attempt to turn attention away from the characteristics of a particular model (and the question "which model is best?") and to focus on the chemistry at hand.

While the choice of the new standard, the ωB97X-D/6-31G* model, may be seen as a continued compromise between accuracy and practicality, in fact, ωB97X-D/6-31G* provides high quality results for a variety of properties that can be accurately measured, including molecular geometry and NMR spectra. Further, both the T1 heat of formation and the ωB97X-V/6-311+G(2df,2p) energy are included in the SSPD entry as a property. We believe that over the wide variety of molecules considered, T1 provides as reliable an indicator of ΔH 298 as the G3(MP2) recipe (which in turn is within measurable experimental error).

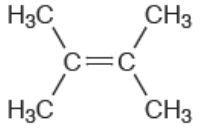
Collections such as the Spartan Spectra and Properties Database may be used to eliminate the need to perform calculations that have already been done. Taking this to the extreme, a database might serve as an exclusive source of information, doing away altogether with the need for quantum chemical calculations. From the instructional perspective (teaching about molecular models and molecular modeling) this allows use of more complex examples than might otherwise be practical due to time constraints. The tutorials in this chapter fall into this category and can be thought as the ultimate in molecular modeling "dry labs".

To complete the tutorials in this chapter, SSPD needs to be available. Either the sample version installed with all copies of ***Spartan*** and which comprises ≈6,000 molecules or the full version which comprises >317,000 molecules which requires a separate download and installation.

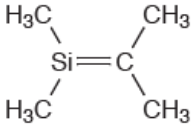
Reactivity of Silicon-Carbon Double Bonds

With the exception of so-called phosphorous ylides, compounds incorporating a double bond between carbon and a second-row element are quite rare. Most curious perhaps is the absence of stable compounds incorporating a carbon-silicon double bond. This can be rationalized by using local ionization potential and LUMO maps to compare the reactivities of olefins and silaolefins.

1. Build both 2,3-dimethyl-2-butene and tetramethylsilaethylene in the same document. Start the second molecule with **Build New Molecule** or **Sketch New Molecule** from the **File** menu instead of **New Build** or **New Sketch**.



2,3-dimethyl-2-butene



tetramethylsilaethylene

2. Both molecules are in SSPD. *Click* on the name of whichever molecule is selected at the bottom of the screen, note the ωB97X-D/6-31G* model is referenced, *click* on the **Replace** button and then *click* on **All**.

3. Select **Surfaces** from the **Setup** menu . First request a local ionization potential map. *Click* on **Add** and select **local ionization potential map** from the menu. Then request a LUMO map. *Click* on **Add** and select **|LUMO| map** from the menu. Surface generation takes less than a second.

4. Select **Spreadsheet** from the **Display** menu and *check* the box at the far left for each entry in the spreadsheet. This allows simultaneous display of both molecules. The molecules should move independently, if mouse movement impacts both molecules, deselect the **Coupled** entry from the **Model** menu to uncouple the

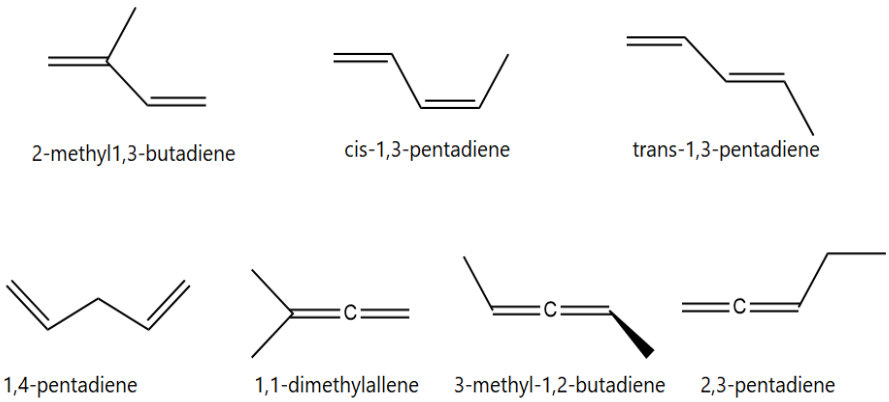
controls. Position the two molecules side by side on screen. Close the spreadsheet.

5. Inside the **Surfaces** dialog, select ***local ionization potential map***. From the **Display** menu, select **Properties** and *click* on one of the displayed surfaces. Towards the top of the **Surface Properties** dialog, you will see the **Property Range** displayed, *click* the button labeled **Reset M/M** (this resets to the minimum and maximum values calculated for the list, rather than the default window. Compare local ionization potential maps for the olefin and silaolefin. Recall that the redder the color, the lower the ionization potential and the more susceptible toward electrophilic attack. Which molecule do you conclude is likely to be more reactive toward an electrophile? Turn off the local ionization and *click* on ***|LUMO| map*** to turn the LUMO maps on. Here, the bluer the color, the greater the concentration of the LUMO and the more susceptible toward nucleophilic attack. Which molecule do you conclude is likely to be more reactive toward a nucleophile? Why are molecules incorporating a silicon-carbon double bond chemically problematic?

6. Close all documents and any open dialogs.

Isomeric C_5H_8 Dienes

In addition to 2-methyl-1,3-butadiene, *cis* and *trans*-1,3-pentadiene (conjugated dienes) and 1,4-pentadiene (a non-conjugated diene), there are three other dienes with the same formula: 3-methyl-1,2- butadiene, 2,3-pentadiene, and 1,2-pentadiene.



1. Build or sketch all seven C_5H_8 dienes (in a single document). *Click* on the name of whichever diene is selected at the bottom of the screen. *Click* on **Replace** and finally *click* on **All** from the dialog that results. This will substitute entries in SSPD for all seven C_5H_8 dienes. Select **Spreadsheet** from the **Display** menu and expand the spreadsheet to see all seven dienes. Select **Properties** from the **Display** menu and make sure that the **ΔH*f*(298.15)** tab is selected. *Click* on the buttons to the left of both **Expt. Heat** and **T1 Heat** in the **Molecule Properties** dialog. This enters

the experimental heat of formation (from the NIST database) as well as the value obtained from the T1 thermochemical recipe into the spreadsheet. The T1 heat of formation will generally provide a more accurate account of relative stability than available from the ωB97X-D/ 6-31G* model itself.

> The T1 model is presently defined only for uncharged, closed-shell molecules comprising H, C, N, O, F, Si, P, S, Cl and Br only. T1 heats of formation are not available for the other entries in the SSPD (but covers 98% of the total).

2. First verify that the T1 model correctly reproduces the ordering of experimental heats. Then use the T1 heats to answer the following questions:

What is the difference in stability (heat) between the best of the three conjugated dienes and 1,4-pentadiene? What is the approximate energy gain due to conjugation? What is the difference in stabilities between the best of the substituted allenes and the best of the conjugated dienes?

3. There are several other C_5H_8 isomers in addition to the dienes discussed above, a number of which will be in SSPD (many more in the full version than in the sample set). Use any one of the seven dienes as a query. Select **Databases** from the **Search** menu and *click* on the **SSPD** tab in the dialog that results. *Click* on to the right of the **Search** button in the **SSPD** dialog and select **Isomer** under **Search By** in the **Search Options** menu that results. *Click* on the **Search** button. Select several (or all) of the ***non-diene*** isomers, by *clicking* on them in the list at the right of the dialog while holding down the **Ctrl** key. *Click* on to the right of the **Retrieve** button, then *click* on **Current Document** in the dialog that results. Close the **Databases** window. Bring up the spreadsheet (**Spreadsheet** from the **Display** menu . Note that not all the experimental heats of formation are available, as the NIST database includes experimental heats of formation for a relatively small number (2-3 thousand) of molecules. According to the T1 model, are any of the non-diene isomers more stable than the best diene?

4. New in ***Spartan'26*** for database entries that include frequencies, a heat of formation from **ωB97X-V/6-311+G(2pf,2p)** starting from **ωB97X-D/6-31G*** geometries. Return to the **ΔH*f*(298.15)** tab in the **Molecule Properties** dialog. *Click* on the buttons to the left of the remaining (DFT-derived) heat of formation. Are the same general trends observed?

5. Close all documents and open dialogs.

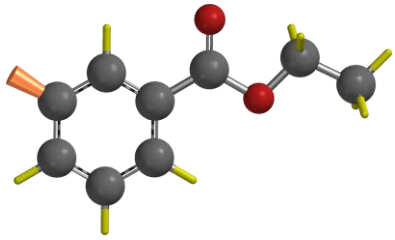
Using Infrared Spectroscopy to Identify an Unknown Ethyl Benzoate Derivative

Unlike NMR spectra (at least ^{13}C spectra) infrared (IR) spectra are rich in detail. Some chemists would view this as a negative, in that the complexity of an infrared spectrum masks its connection with the three-dimensional molecular structure. Others might see complexity as a positive, providing a detailed pattern or fingerprint with which to identify a molecule. This reflects the different ways that chemists actually use NMR and infrared spectroscopy. NMR is most commonly employed to assign structures of new molecules whereas IR is most commonly used to identify previously characterized (known) molecules.

Extensive databases of experimental infrared spectra exist, and sophisticated programs have been developed to search these databases for matches to unknown spectra. ***Spartan*** is able to search the public NIST database of experimental infrared spectra. Infrared spectra calculated from density functional models and adjusted to account for known systematic errors in frequencies and for finite temperature, generally closely match experimental spectra. This suggests that they might be used in lieu of experimental spectra to identify unknowns. This tutorial illustrates the point, attempting to identify which of a series of *meta*-substituted ethyl benzoates contained in the Spartan Infrared Database (SIRD) best matches the infrared spectrum of an unknown molecule.

> SIRD uses the infrared spectra contained in the Spartan Spectra and Properties Database (SSPD) that have been obtained using the ωB97X-D/6-31G* model, these replace the EDF2/6-31G* previously available in SSPD. The fact that there are two different tabs simply reflects the different search paradigms (spectra matching for SIRD vs. substructure matching for SSPD).

1. Build or sketch ethyl benzoate. Select **Structure Query** from the **Search** menu and *click* on one of the free valences on the phenyl ring that is *meta* to the ester group.



2. Select **Databases** from the **Search** menu and *click* on the **SIRD** tab at the top of the dialog that results. Size the dialog to occupy the greater part of the screen. *Click* on **Select Spectrum**, move to the ***using SSPD*** subdirectory under the ***tutorials*** directory*, *click* on ***unknown ethyl benzoate*** and *click* **Open**. In a few seconds, the experimental infrared spectrum of the unknown will appear at the top right of the dialog. A list of the frequencies for all peaks may be found under **Unknown:** immediately to the left is the spectrum. This has been obtained by fitting the experimental spectrum to a Lorentzian function. You can obtain the frequency of a line in the spectrum using the yellow measurement bar. Position the cursor on top of the bar, hold the left button and move the mouse left or right until the bar is over the peak.

3. You can carry out your search with nothing other than the unknown spectrum (see the optional part of this tutorial), or you can restrict it using structural or other information that you have. Assume that you know that the unknown is a *meta*-substituted ethyl benzoate. *Click* on the **Filters** button and then *click* on **Copy Current Molecule** at the top left of the **Search Filters** dialog that results. *Click* on **OK** to exit the dialog. You have restricted the search to *meta*-substituted ethyl benzoates, that is, examine only molecules that contain this particular substructure (ethyl benzoate substituted in the *meta* position). *Click* on the **Search** button and wait for the search to complete. Hits resulting from the search appear at the bottom of the **SIRD** dialog, in numerical order starting from the best (lowest **Score**).

As you *click* on each, the corresponding calculated infrared spectrum will be superimposed onto the experimental spectrum (of the unknown) at the top left of the display. Propose a structure for the unknown.

4. Remove the substructure filter and repeat the search. *Click* on the **Filters** button, *click* on the **Clear** button at the top left of the **Search Filters** dialog and *click* on **OK**. *Click* on the **Search** button. Unlike searches of experimental databases where spectra are directly compared, each of the calculated spectra in the database are adjusted for frequency scale and peak width as it is compared to the unknown. Removing the substructure filter greatly increases the number of spectra that need to be examined. Does the *meta* substituted ethyl benzoate that came in at the top in the previous search, also come in at or very near the top in the full search?

---

* For Windows, this directory is found in ***Program Files/Wavefunction/Spartan26***. It must be copied to another location available to the user prior to opening it in ***Spartan***. For Linux, this is found in the directory where ***Spartan*** was installed. For Macintosh, this is located at the top level of the ***Spartan'26*** disc image.

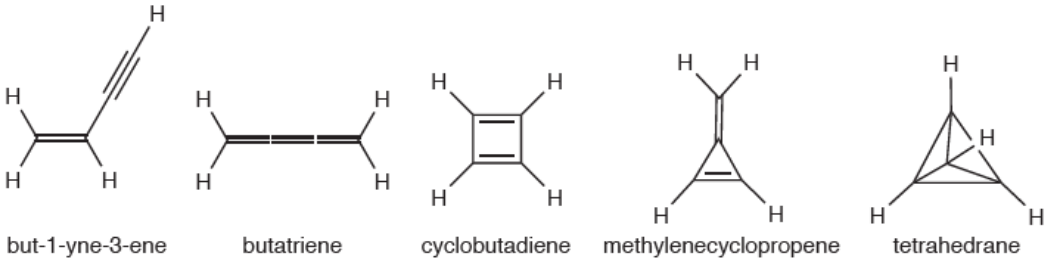
Be patient. A search of the full database will require significantly more computer time than the search of the ≈6,000 molecule sample database supplied with ***Spartan***; it contains >50X as many molecules.

5. Close all documents and any open dialogs.

Infrared Spectra of Short-Lived Molecules

While infrared spectroscopy is by no means as formidable a tool as NMR for elucidating the structures of molecules, it does offer significant advantages which can make it the method of choice in some instances. It is more sensitive, meaning that it can be used to detect transient species, for example, molecules trapped in an inert matrix at very low temperature. On the practical side, an infrared spectrometer is inherently simpler (more robust) and can be put into a much smaller package than an NMR spectrometer. This means that the spectrometer can be brought to the sample rather than the other way around. Imagine, putting an NMR spectrometer on the Mars rover! One downside is that IR spectroscopy, unlike NMR, typically does not provide an unambiguous structure. However, a close match to a calculated spectrum may offer convincing evidence for the veracity of a particular structure (or a mismatch convincing evidence that a structure is incorrect).

The infrared spectrum of an unknown C_4H_4 isomer shows strong absorptions at 854, 1608, 2994 and 3080 cm^{-1}. Possible structures include but-1-yne-3-ene, butatriene, cyclobutadiene, methylenecyclopropene and tetrahedrane.



1. Build or sketch any one of these molecules. Select **Databases** from the **Search** menu , and *click* on the **SSPD** tab. *Click* on to the right of the **Search** button, select **Isomer** under **Search By:** inside the **Search Options** dialog and *click* on **OK**. *Click* on the **Search** button. Structures for all C_4H_4 isomers will appear in a list at the right of the **Databases** dialog. Retrieve all of them to a new document. *Click* on to the left of the **Retrieve** button, select **New Document** and *click* on **OK**. *Click* on each in turn while holding down the **Ctrl** key (or *click* on the first, hold down the **Shift** key and *click* on the last). With all 5 selected, *click* on the **Retrieve** button. Close the **Databases** dialog.

2. Select **Spectra** from the **Display** menu and *click* on at the top of the resulting dialog. From the **IR** tab *click* on (calculated IR spectrum). Move through the list to bring up the spectra for the other isomers. Which, if any, calculated infrared spectrum of the C_4H_4 isomers best fits the unknown infrared spectrum (with peaks at 854, 1608, 2994 and 3080 cm^{-1})?

3. Close all documents and any open dialogs.

Using ^{13}C NMR to Distinguish Structural Isomers

NMR spectroscopy, in particular ^{13}C NMR spectroscopy, is arguably the single most important tool available to an organic chemist to establish molecular structure. Obtaining an NMR spectrum is straightforward and the spectrum itself is very simple, comprising but a single line for each unique carbon the molecule. Because a mass spectrum is normally also available, the molecular formula will be known, and assignment of the NMR supports deciding among possible isomers. Chemical evidence (how the molecule was made) can usually be counted on to eliminate some choices and to strengthen the case for others. Still, pinning an NMR spectrum to a particular molecule can be difficult and fraught with error.

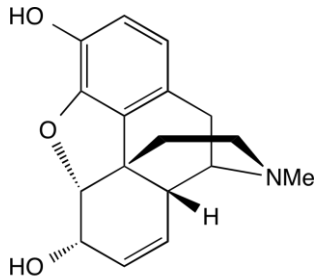
NMR assignment problems can be reduced but not altogether eliminated by requiring that both proton and ^{13}C NMR spectra are consistent with a particular structure. This is routine practice. In principle, ambiguity can be eliminated by cross-correlating the results of the ^{13}C NMR with those of proton NMR with an HMBC spectrum.

The fact that NMR spectra for reasonable size organic molecules may now be routinely calculated in a few minutes to a few hours on a personal computer at a price of $1000 (or less) raises the possibility for another tool to assist with spectral assignments. Whether this becomes common practice ultimately depends on the ability of calculations to obtain NMR spectra that are sufficiently accurate to distinguish among the different isomers.

In this tutorial you will first compare calculated and experimental ^{13}C spectra of morphine. You will then examine calculated spectra for two isomers, norcodeine and hydromorphone. The objective is to provide you with an impression of the ability of the calculations to

reproduce a ^{13}C spectrum for a known compound and more importantly, to challenge structure assignments based on NMR spectroscopy.

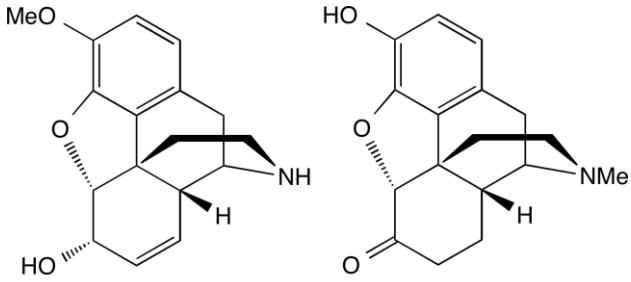
1. Sketch morphine. *Click* on its name at the bottom of the screen and *click* on **Replace**. If the name fails to appear, then you have made a mistake. In this case, select **Edit Sketch** from the **Build** menu and correct your sketch.



morphine

2. Select **Spectra** from the **Display** menu and *click* on in the bar at the top of the plots pane. Select from the palette. *Click* again on and this time select from the palette. Calculated and experimental ^{13}C spectra are now superimposed. Focus your attention on the four "most isolated" resonances, specifically those with calculated chemical shifts of 20.3 ppm, 59.7 ppm, 66.2 ppm and 95.2 ppm.

3. Sketch norcodeine and hydromorphone and retrieve both molecules from SSPD.



norcodeine hydromorphone

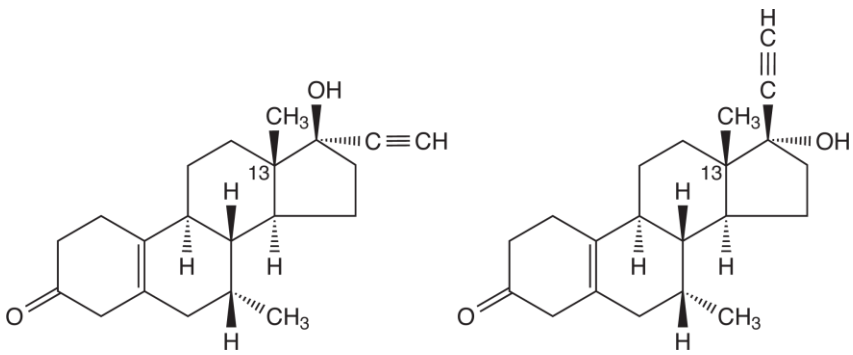
4. Select **Spectra** from the **Display** menu and *click* on and select from the palette. You only need to do this once if you put norcodeine and hydromorphine in the same document. Does either spectrum fit the experimental ^{13}C for morphine (again focus on the four isolated lines)?

5. Close all documents and any open dialogs.

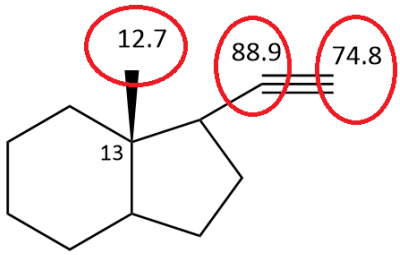
Using ^{13}C NMR Spectra to Distinguish Stereoisomers

As mentioned in the previous tutorial, the combination of 1D and 2D NMR spectra is almost always able to provide unambiguous assignment of molecular structure, including assignment of stereochemistry. However, as noted, practical considerations (time on an expensive instrument) often make 2D experiments (in particular HMBC) the exception rather than the rule. Without the additional experiments, distinguishing stereoisomers based on their ^{13}C NMR spectra may be difficult experimentally, simply because the alternatives are likely to be structurally very similar. For the same reason, it is likely to be an ideal case for calculations, as comparisons between molecules with similar structures should benefit from cancellation of errors. The bottom-line question is whether or not calculated ^{13}C NMR spectra are good enough to be able to clearly distinguish between stereoisomers. This tutorial provides an illustration and a chance to assess.

The hydroxyl CO bond in tibolone, shown below on the left, is *gauche* to the methyl group at C_{13} (the alkyne is *anti*), whereas it is *anti* (and the alkyne is *gauche*) in the stereoisomer shown on the right. You will use just three lines from the experimental ^{13}C spectrum to see if the calculations support or refute its assignment as tibolone.



1. Sketch both stereoisomers and put into the same document. *Click* on the name of whichever molecule is selected at the bottom of the screen, *click* on **Replace** in the dialog that results and *click* on **All**. NMR spectra for both molecules are now available.
2. Focus on tibolone and **Expt. Chem. Shifts** [H/C] from the **Expt. Data** menu. One after the other, enter the experimental chemical shifts for the methyl group at **C13** (12.7 ppm) and for the internal and external alkyne carbons (88.9 and 74.8 ppm, respectively).



3. Select **Properties** from the **Display** menu , and *click* on the NMR tab. Which isomer provides the better agreement in terms of both RMS, mean and maximum absolute error scores as well as the DP4 probability core? Do the calculations support or refute the experimental assignment, or are they ambiguous?

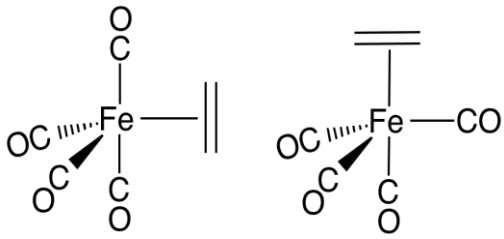
4. Close the document and any open dialogs.

Chapter 15
Using the Cambridge Structural Database

*The Cambridge Structural Database (**CSD**) is a well-organized and easily-accessible depository of well over 1 million experimental structures, almost entirely from X-ray crystallography. It is available on a yearly subscription basis, and when installed and licensed may be seamlessly accessed from **Spartan**. Searches may be carried out based on substructure (as with **SSPD** searches) as well as name, entry RefCode or author. The tutorials in this chapter are intended to illustrate the interface and more importantly provide a sample of the types of questions to which the database may be used in conjunction with computation. These tutorial require access to the **CSD**. See **Appendix G** for configuration instructions.*

Binding to Iron Tetracarbonyl

An olefin bonded to iron tetracarbonyl may occupy either an *equatorial* or *axial* position. Search CSD and review the results to see if one is more common than the other.



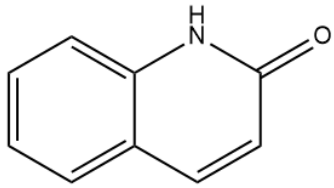
1. Build either axial or equatorial ethylene iron tetracarbonyl. The CSD search does not distinguish between the two. Select **Structure Query** from the **Search** menu and *click* on all the four free valences on ethylene. Select **Databases** from the **Search** menu and *click* on the CSD tab. *Click* on to the right of the **Search** button, in the **Search Options** dialog that results, select **Inorganics** under **Filters** and **Structure** under **Search By**. *Click* on **OK**. *Click* on the **Search** button. More than 50 matches (hits) are returned and displayed at the right-hand side of the dialog. *Click* on each in turn to show the structure on the left-hand side. Does there appear to be a clear preference for *equatorial* or *axial* olefin binding?

2. Build the other conformer and put in the same document as the one that you used to query the CSD. Select **Calculations** from the **Setup** menu and specify calculation of equilibrium geometry with the ωB97X-D/6-31G* density functional model. Make sure that **Global Calculations** is checked. Submit the job. Execution will take several minutes. Which conformer is favored and by how much? Is your result consistent with what you found (or didn't find) in the CSD search?

2-Quinolone

It is generally presumed that energy barriers to interconversion among tautomers are very small and what is observed experimentally is a Boltzmann weighted average.

1. Sketch or build the natural product 2-quinolone.

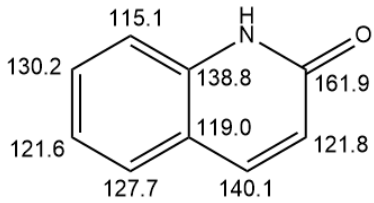


Select **Generate Tautomers** from the **Geometry** menu and *click* on (▼) to the right of **Tautomers** (at the bottom right of the interface) and *click* the **OK** button. ***Spartan*** generates a new document containing both 2-quinolone and it's tautomer 2-quinolinol. Do you note anything special about 2-quinolinol that might lead to it being the dominant tautomer, additional aromaticity, perhaps?

2. Confirm that the OH proton in 2-quinolinol is pointed upward (toward the nitrogen). If it isn't, enter the **Edit Build** mode from the **Build** menu and click on the bond between the ring carbon and OH. Next, position your mouse cursor over the vertical gray shaded section (a bond-rotation tool). *Depress* the left mouse and slide either up or down to rotate the OH group.

3. Enter the **Calculations** dialog and specify an **Equilibrium Geometry** calculation using the **ωB97X-D/6-31G*** model, also *check* the **NMR** check box and make sure the **Global Calculations** box is selected. *Click* on **Submit**. Name the file ***tautomer review***.

4. It should only take a few to complete (perhaps 2-3). Which tautomer is lowest in energy? Compare the ^{13}C results from calculation against the experimental values below.

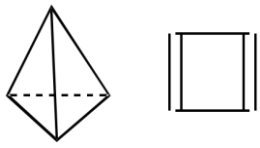


Select **Configure** from the **Model** menu and from the **Labels** tab, select **Calc. Chem Shift** and *click* **OK**. It will be easier to compare if you turn off display of **Hydrogens** from the **Model** menu. You can turn on display of both tautomers from the **Spreadsheet** by checking the box to the left of the molecule **Label**. It may be necessary to turn off the **Coupled** feature in the **Model** menu, in order to manipulate the two molecules independently. Confirm which NMR results best match experiment.

5. **CSD use**: If both tautomers exist independently, there should be experimental evidence in the form of x-ray crystal structures in the **CSD**. With one of the two tautomers selected, *click* on **Structure Query** ( ) from the **Search** menu and then *click* on all free (open yellow) valences (this specifies a search for the specific tautomer as well as any analogues which maintain the same tautomer form). Select **Databases** ( ) from the **Search** menu and *click* on the **CSD** tab. *Click* on ( ) to the right of the **Search** button, make sure **Organics** is selected under **Filters** and **Structure** under **Search By**. *Click* on **OK** to release the **Search Options** dialog. *Click* on the **Search** button. Review the number of hits. Repeat for the other tautomer. For which tautomer(s) do you find crystal structures? Which tautomer is more common in the crystallography literature?

Tetrahedrane and Cyclobutadiene

The C_4H_4 isomers, tetrahedrane and cyclobutadiene, are among the most unlikely of hydrocarbons. It is hard to imagine a molecule that is more strained than the former while the latter is the prototypical antiaromatic molecule. Use the CSD to see if derivatives have actually been characterized.



1. Build cyclobutadiene. Select **Structure Query** ( ) from the **Search** menu and *click* on all open free valences. Select **Databases** from the **Search** menu and click on the CSD tab. *Click* on to the right of the **Search** button, select **Organics** under **Filters** and **Structure** under **Search By** and in the **Search Options** dialog that results and *click* on **OK**. *Click* on the **Search** button.

2. Repeat the process for tetrahedron.

3. Identify the simplest system that is common to both cyclobutadiene and tetrahedrane, that is, one where the substituents are the same. Build the two molecules (or extract the two from the "hits" on CSD. Calculate equilibrium geometries for the two using the ωB97X-D/6-31G* model.

4. Select **Calculations** from the **Setup** menu and inside the dialog select **Density Functional**, **ωB97X-D** and **6-31G*** from the menus to the right of **Calculate**. Submit the job (two molecule). It will require several minutes to complete. Which molecule is more stable?

5. Close any remaining open documents.

Section V
Features and Functions

This section describes the functions available under the menus incorporated into ***Spartan'26*** and is intended to serve as a general reference to the program. The coverage follows the order of the menus presented in ***Spartan***'s user interface: **File**, **Edit**, **Model**, **Geometry**, **Build**, **Setup**, **Display**, **Search**, **Data**, **Options**, **Activities** and **Help**. The functions and usage of each of the menu entries is described. Entries under the **File**, **Model**, **Geometry** and **Build** menus deal primarily with the input and construction of both 2D drawings and 3D structures and with their display and query. Entries under the **Setup** menu specify molecular mechanics or quantum chemical models as well as similarity analysis tools, designate what properties and spectra are to be obtained, request one or more graphical models and submit jobs for calculation in the background. Entries under the **Display** menu access calculated quantities including calculated spectra, under the **Expt. Data** menu - input of experimental IR/Raman, NMR (chemical shifts and observed HH and CH couplings) and UV/visible data is available, and the **Search** menu accesses databases of calculated information. Entries under the **Options** menu set program defaults and user preferences, designate paths to databases of calculated and experimental information, control job queues, designate available servers for remote calculation and control access to ***Spartan*** as a computational server.

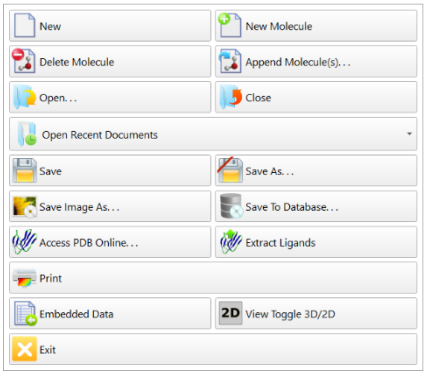
The chapters in this section provide only limited commentary about the performance and requirements of different computational methods and the utility of different graphical models. Additional information is provided under ***Topics*** accessed from the ***Activities*** menu and in ***A Guide to Molecular Mechanics and Quantum Chemical Calculations****, available as a PDF under the **Help** menu.

* This reference was written in 2003, and as such it does not include a full assessment or description of many of the newer computational features included in ***Spartan'26***. In particular many functionals and basis sets are not covered. An updated version is in the works.

Chapter 16

The File Menu

*Operations under the **File** menu access a 2D sketch pad, model kits to build, edit and substitute molecules in 3D, the file system to read and write both native and non-native files and the online PDB database of protein and macromolecular structures. They allow text and on-screen graphics to be printed and text and other files to be embedded into **Spartan** documents.*



New

Brings up a model kit or sketch tools, depending on current viewing mode (2D or 3D) and clears the screen. Model kits are discussed in **Chapter 20**.

New Molecule

Brings up a model kit or sketch tools and clears the screen. **New Molecule** differs from **New**, in that the resulting molecule is appended to the end of the document associated with the molecule (or sketch) that is presently selected.

Delete Molecule

Deletes the selected molecule(s) from a document. Deleting the last molecule leads to an empty document.

Append Molecule(s)...

Appends one or more documents onto the end of the document that contains the selected molecule. **Append Molecule(s)...** leads to a file browser from which one or more documents need to be selected.*

Open...

Opens a file that contains all information associated with a particular molecule or pharmacophore (or list of molecules and/or pharmacophores). In addition to native (.spartan) files (documents), the following are also supported to input: 2D drawings: CambridgeSoft (ChemDraw) CDX (.cdx), and MDL SDF (.sdf)*, SKC (.skc), and TGF (.tgf), files containing 3D structures: MacroModel (.mac), SYBYL Mol (.mol), SYBYL Mol2 (.mol2), and PDB (.pdb), crystallographic Information (.cif), and files containing 1D "strings": SMILES (.smi) and InChi (.inchi). Finally IR, Raman, VCD, NMR and UV/visible spectra may be input (discussion of file formats is provided in **Appendix J**). Experimental spectral data may also be brought into ***Spartan*** from the **Data** menu (**Chapter 24**).

> We use the word "file" to refer to any information contained in the "file system" (the disk on your computer) and the word "document" to refer to files that are native to ***Spartan***. Thus, an MDL SDF ***file*** may be read into ***Spartan*** and then written as a ***document***. Documents may contain information on a single molecule or single pharmacophore or on a collection of molecules and/or pharmacophores. This collection is loosely referred to as a "list" or as a "spreadsheet" (in the context that it is presented in a spreadsheet).

---

*Alternatively, molecules may be appended to an existing document either by copy/paste operations using the clipboard or by *dragging* from an external window. Both require that the spreadsheet associated with the destination document be open on screen.

To copy a molecule open on screen onto the clipboard, first select (*click* on) it, and then select **Copy** from the **Edit** menu. Alternatively, *click* on its label in its spreadsheet (in the leftmost column), and then select **Copy** from the **Edit** menu. The latter permits several molecules to be selected (and copied) at once using the **Shift** and **Ctrl** keys in the usual manner. Once on the clipboard, the molecule or molecules may be moved to the destination list by *clicking* on an empty row header in the spreadsheet (for the destination document) and then selecting **Paste** from the **Edit** menu.

To copy a document from an external window, *drag* it onto the open spreadsheet (associated with the destination document) inside of ***Spartan***. Several documents can be *dragged* at once using the **Shift** and **Ctrl** (**Command** key for Macs) keys in the usual manner.

Cartesian coordinates from any of these files (except spectra files which do not necessarily contain coordinates) may be replaced by coordinates generated based on atomic connectivity using **Replace Coords.** in the **Molecule Utilities** dialog (accessible from **Properties** under the **Display** menu; **Chapter 22**). Additionally, hydrogen atoms that are missing in the input structure may be provided using **Add Hydrogens** from the **Utilities** tab in the **Molecule Properties** dialog. Non-native files are normally hidden from view but may be seen by selecting **All Files** from the **Files of type** menu at the bottom of the dialog.

Close

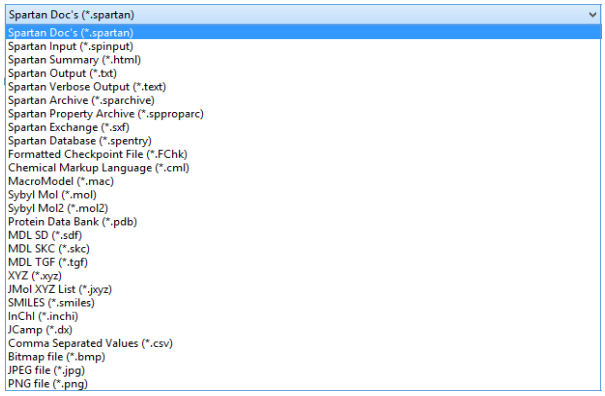
Closes the document containing the selected molecule, as well as any document specific dialogs. If the document has not previously been saved, a name is requested. If a previously-saved document has been altered, verification that the changes are to be saved is requested.

Open Recent Documents

Brings up a list of (at most) the ten recent documents. *Clicking* on one opens the document.

Save **Save As...**

Saves the document containing the selected molecule ***exactly as it appears on screen***. Opening the document will bring it on screen exactly as it was last saved. If the document has not previously been named, **Save** behaves as **Save As...**. Documents may either be saved in native format or in one of the formats listed under **Open**. In addition, bitmap (.bmp), JPEG (.jpg) and PNG (.png) graphics file formats are supported as is the Spartan Database (.spentry) format for creating custom databases (see **Appendix H**). Selection is made under the **Save as type** menu in the **Save As** dialog.



The contents of the spreadsheet (**Spreadsheet** under the **Display** menu; **Chapter 22**) may be printed using **Print** from the contextual menu. The results of a reaction energy calculation (**Reactions...** under the **Display** menu; **Chapter 22**), may be printed using **Print** from the contextual menu.

Save Image As...

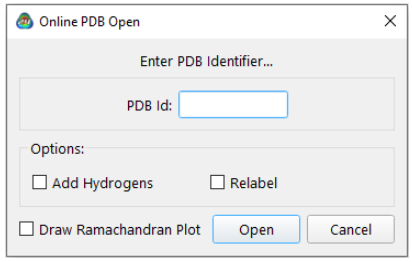
Saves your current molecule system, surface (etc.) as a high resolution PNG (.png), JPEG (.jpg) or bitmap (.bmp) file. Upon save, a new dialog will prompt for dimensions, allowing users to specify height and width (in pixels), PNG format allows saving with a transparent background.

Save to Database...

Saves the current molecule(s) in the SSPD format. This is a new menu item in ***Spartan'26*** and will only be selectable if your system has been submitted as **an SSPD Compatible** (task). See **Chapter 21** (Setup menu).

Access PDB Online...

Provides access to the online **P**rotein **D**ata **B**ank (**PDB**)* comprising more than 253,000 protein and nucleotide structures. Selection results in a dialog.



To access a PDB structure, enter the four-character identification code in the box to the right of **PDB ID** and *click* on **Open**. If the PDB entry contains more than one structure (typical with protein structures obtained from NMR spectroscopy) and/or the PDB ID yields more than one entry, all structures will be returned in a single document. The literature reference from PDB is available from **Molecule Reference** (**Display** menu; **Chapter 22**).

A ribbon model of the PDB file will appear on screen.* A Ramachandran plot associated with a protein structure may either be drawn upon initial retrieval of the PDB file by *checking* the box to the left of **Draw Ramachandran Plot** or later from **Ramachandran Plot** under the **Model** menu (**Chapter 18**).

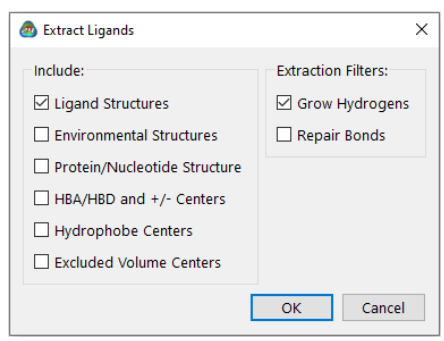
Extract Ligands

This allows "ligands" to be extracted from PDB files** either as molecular structures or as footprints (pharmacophores) that these structures leave behind. The latter is separated into three parts, the first two of which follow from the molecular structure of the ligand and the third which follows from the immediate (protein or nucleotide) environment:

(i) hydrogen-bond donor/acceptor sites and positive/negative ionizable sites
(ii) hydrophobes
(iii) excluded volumes

A pharmacophore can be constructed by requiring (i) to account for hydrogen-bonded and electrostatic interactions and (optionally) ***either*** (ii) or (iii) to account for steric interactions. Selection of (ii) leads to a pharmacophore that is visually simpler, but selection of (iii) may provide a more realistic picture.

Selection of **Extract Ligands** leads to a message at the bottom of left of the screen. Select Ligands. Selection is made by clicking on one of space-filling ligand models. Ligands in PDB files are represented as a transparent space-filling model. *Clicking* on a ligand selects it and displays the PDB **HET name*** at the bottom of the screen (if a HET name is available). Multiple ligands may be selected if desired by holding down on the **Ctrl** key.



All ligands may be selected by *clicking* on the **Select All Ligands** button at the bottom right of the screen. After one or more ligands have been selected, *clicking* on the **Extract Ligands** button at the bottom right of the screen brings up the **Extract Ligands** dialog. This allows selection of **Ligand Structures** of one or more pharmacophore elements (**HBA/HBD and +/- Centers**, **Hydrophobe Centers** including aromatic centers, **Excluded Volume Centers**) or ***both*** structure and one or more pharmacophore elements.

Additionally, amino acids and/or nucleotide residues in close hydrogen-bonding or charge-charge contact with the ligand (**Environmental Structures**) may be identified. This panel also allows the **Protein/Nucleotide Structure** (minus all ligands) to be extracted.

* If the PDB file does not contain a macromodel structure suitable for rendering in ribbon display, choose an alternate **Model** style from the **Model** menu.

** This capability is not restricted to files in PDB but extends to files written in PDB format.

In the case of extraction of **Ligand Structures**, the **Extract Ligands** dialog also provides a set of utilities:

(i) **Grow Hydrogens**

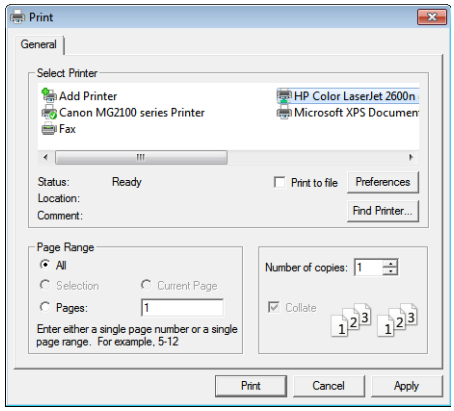
If *checked*, adds hydrogens wherever they appear to be missing, based on the actual geometry and on normal valence rules. This function is also available under **Grow Hydrogens** in the **Molecule Utilities** dialog (**Properties** under the **Display** menu; **Chapter 22**).

(ii) **Repair Bonds**

If *checked*, attempts to fix bonding errors based on the actual geometry in the PDB entry and on normal valence rules. *Clicking* on **OK** at the bottom of the **Extract Ligands** dialog extracts the ligand(s) or pharmacophore(s), placing them in a single new (unnamed) document. *Clicking* on **Cancel** dismisses the dialog without ligand (pharmacophore) extraction.

Print

Selection leads to a dialog in order to designate a printer, specify print layout and number of copies. It also allows printing to a file.



Embedded Data

Is the location where the molecule's thumbnail image and additional data can be stored (within the .spartan format).

View Toggle 2D/3D

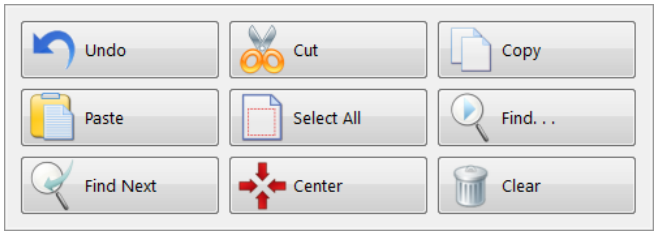
Controls master view mode and toggles between views. When in 3D, only the 2D toggle option is displayed, when in 2D, only the 3D toggle option is displayed.

Exit

Closes ***Spartan***, that is, clears the screen and closes all open documents. A prompt is provided for each document that has not previously been saved.

Chapter 17
The Edit Menu

*Operations under the **Edit** menu provide for undoing commands, copying items to and from the clipboard, finding text and graphics, centering molecules on screen and clearing the selected molecule.*



Undo

Undoes the last operation from the **Build** and **Edit** menus and from the **Molecule Utilities** dialog. Undoes transition-state formation and retrieval from ***Spartan***'s databases.

Cut , Copy , Paste

Cut moves the selected item to the clipboard and removes it from the document. **Copy** copies the item to the clipboard. The item is unaffected. **Paste** transfers the contents of the clipboard to the selected location. The contents of the clipboard are unaffected. Among the important uses of the clipboard within ***Spartan*** are:

(i) Transferring on-screen graphics into other applications such as Microsoft Word® and PowerPoint®.

(ii) Temporary storage of a molecular structure for use in molecule building.

(iii) Transferring data between ***Spartan*** spreadsheets and between a ***Spartan*** spreadsheet and other applications such as Microsoft Excel®.

(iv) Making multi-molecule documents and/or transferring molecules between documents.

Cut operations for (i) and (ii) require drawing a selection box. Position the cursor slightly above and slightly to the left of the item to be transferred, hold down both left and right mouse buttons and *drag* the mouse to a location slightly below and slightly to the right of the item to be transferred and release both buttons. **Copy** operations for (i) and (ii) also refer to the contents of a selection box ***if one has been drawn***, or to the selected molecule if a box has not been drawn. **Copy** operations from a ***Spartan*** spreadsheet refer to all information

associated with a molecule if selection is made on the header cell of the leftmost column, but only to the selected (text) information if selection is made on any other column. Further discussion relating to use of the clipboard in molecule building is provided in **Chapter 20** and for moving data in and out of the spreadsheet is provided in **Chapter 22**.

Select All

Selects all atoms in the selected molecule.

Find..., **Find Next**

Find locates a text string defined in the **Find** dialog if an output window or a spreadsheet is selected, or a structure sequence defined on the clipboard if an on-screen model is selected. **Find Next** locates the next occurrence of a text string or a structure sequence.

Center

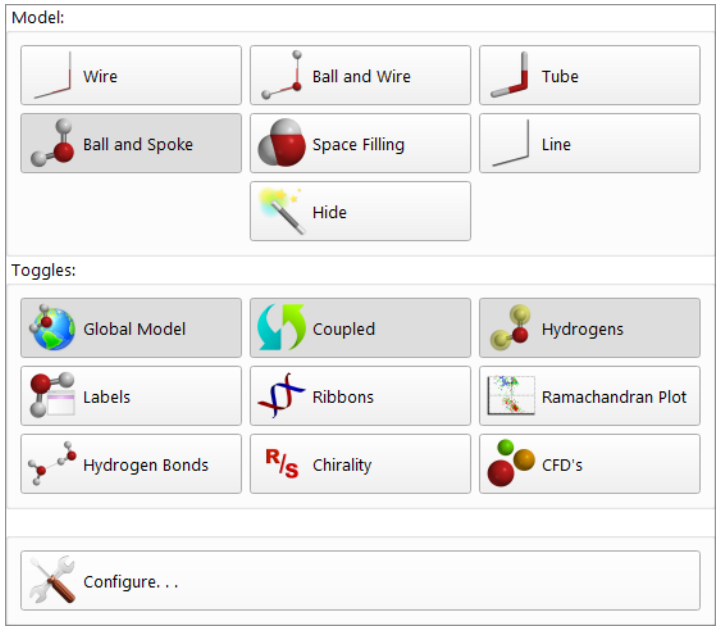
Centers on screen all molecules in the document for which the selected molecule is a member (only the selected molecule is displayed).

Clear

Clears (deletes) the structure and other information for the selected molecule and brings up a model kit. No information is actually removed from the file system until the document is saved.

Chapter 18
The Model Menu

*Structure styles available under the **Model** menu include: Line, Wire, Ball-and-Wire, Tube, Ball-and-Spoke, and Space-Filling (CPK) models. A number of additional items can be toggled on or off, including display of hydrogens, hydrogen bonding, R/S chirality labels, chemical functional descriptors (CFD's), ribbon displays for biopolymers, and Ramachandran plots for proteins from the Protein Data Bank (PDB). Additional options are available for display of Labels (atom and residue), Objects (such as constraint markers, frozen atoms, and user defined planes), Ribbons (custom color and style options) and CFD's. Global Model toggles settings in the **Model** menu between being applied to individual or to all molecules in a document and Coupled toggles between manipulations applied to individual molecules or to all molecules in a document. Configure sets default labels for atoms, ribbons, CFD's and a variety of other objects.*



Only one model style **Wire**, **Ball and Wire**, **Tube**, **Ball and Spoke**, **Space Filling**, **Line** or **Hide**) may be selected. The selected model is designated by being highlighted in the **Model** menu. **Global Model**, **Coupled**, **Hydrogens**, **Labels**, **Ribbons**, **Hydrogen Bonds**, **Chirality**, and **CFD's** operate as toggle switches.

All structure models, and graphics may be displayed either in orthogonal or perspective projections. The latter may be valuable in helping to visualize large molecules. Selection is done in the **Settings** tab (**Preferences...** under the **Options** menu; **Chapter 25**). Both structure models and graphics may be presented in 3D stereo. This may be controlled from the **Settings** tab or by toggling the **3** key on the keyboard. Stereographic displays require perspective projections.

Wire

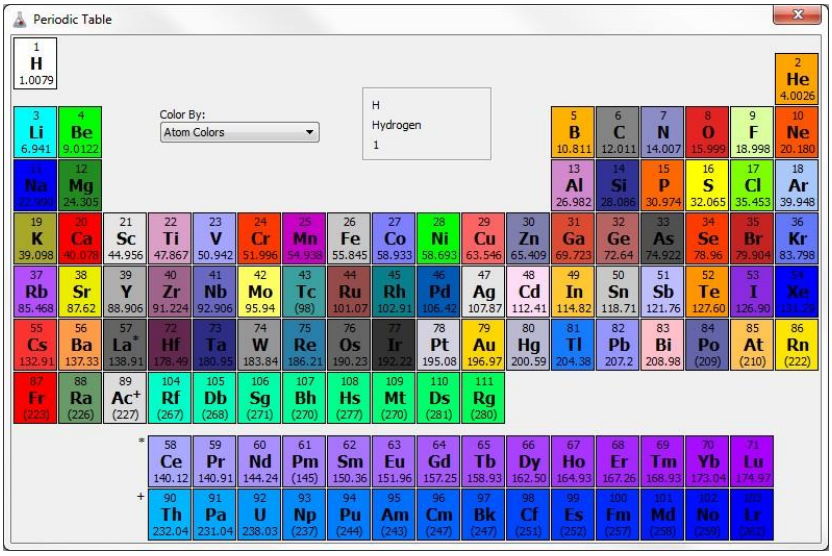
This represents the molecule as a wire model where the vertices represent the atoms.

Wire Model

The bonds are drawn in two colors, one for each of the atoms making up the bond. Default atom colors are given in terms of *Periodic Table* in **Figure 18-1**. These are different than the defaults used in previous versions of ***Spartan***. You can return to the old scheme from the **Miscellaneous** tab (**Preferences...** under the **Options** menu; **Chapter 25**). Select ***Classic*** under **Atom Color Set**.

Atom colors apply globally (to all atoms of given type) and may be changed using the **Set Colors** dialog (**Colors** under the **Options** menu; **Chapter 25**). Colors of individually selected atoms may be set using the **Atom Style** dialog (**Properties** under the **Display** menu; **Chapter 22**). All models use the same color scheme for atoms and provide for the same mechanism of changing colors globally or individually.

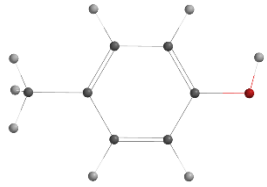
Figure 18-1:Default Atom Colors



Ball and Wire

This represents atoms by small balls and bonds by wires.

Ball-and-Wire Model



The balls are color coded according to atom type, and the wires representing bonds are drawn in two colors (as in wire models).

Tube

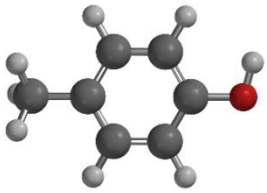
This is similar to the wire model, except that tubes instead of wires are used to represent bonds. As with wire models, bonds are drawn in two colors.

Tube Model

Ball and Spoke

This represents atoms by balls (the size and color of which depend on atom type), and bonds by spokes.

Ball-and-Spoke Model

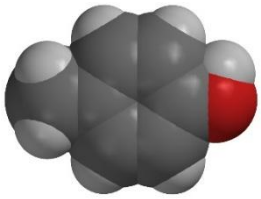


Bond (spoke) color is gray (by default) but it may be changed using the **Set Colors** dialog (**Colors** under the **Options** menu; **Chapter 25**). Colors of individually selected bonds may be set using the **Bond Style** dialog (accessible from **Properties** under the **Display** menu; **Chapter 22**).

Space Filling

This represents the molecule as a composite of spheres, the radii of which have been chosen to approximate van der Waals contact distances.* Also known as CPK models, space-filling models are intended to portray overall molecular size and shape.

Space-Filling Model

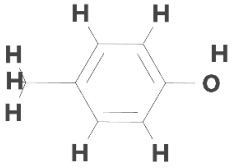


Volume, surface area and polar surface area (PSA)** displayed under the **QSAR** tab in the **Molecule Properties** dialog (accessible from **Properties** under the **Display** menu; **Chapter 22**) correspond to a space-filling model.

Line

A line model is a 2D representation (similar to that provided from drawing programs such as ChemDraw, MarvinSketch or Biovia Draw). It designates carbons as the termini or the intersection of lines and non-carbons by their elemental symbols without use of color.

Line Model



* Default values for van der Waals radii may be changed from the **VDW Radii** dialog accessible from **Preferences** under the **Options** menu (**Chapter 25**). Settings apply to all atoms of given atomic number.

** By default, polar surface area is defined as the area due to nitrogen and oxygen and any hydrogens attached to nitrogen and oxygen. Other combinations of atoms may be employed to define a polar surface area. See **Formulas** under the **Display** menu (**Chapter 22**).

Line models include stereochemical cues (▬ and ⋯) if these have been provided in the 2D drawing or if they are specified from the **Bond Properties** dialog (accessible from **Properties** under the **Display** menu; **Chapter 22**).

Hide

This removes the molecular structure model from the screen. This may be desirable where its display may lead to unnecessary crowding, for example, in proteins where ribbon displays are more appropriate. A structure model may be restored by selecting it from the **Model** menu.

> Different parts of a molecule may be rendered in terms of different model styles and colors. This allows focus on specific regions of a molecule. Regions may be individual atoms and/or bonds or any collection of atoms and/or bonds, contiguous or not. They are selected either by *clicking* on an individual atom or bond or, with the aid of the **Ctrl** (select multiple) or **command** key (Macintosh), **Shift** (select over a range), and **Alt** (select all bonded) or **option** key (Macintosh), by *clicking* on a set of atoms and/or bonds, or by defining a selection box. Discussion provided in **Chapter 1**. Style and color of selected model components is by way of the "**Selected**" **Style** dialog accessible from **Properties** under the **Display** menu (**Chapter 22**).

Global Model

If *highlighted* (turned on), this signifies that all molecules in a document will share attributes. These include presentation of hydrogens, atom and other labels, hydrogen bonds, CFD's and ribbon displays. Global model style is controlled from the **Molecule** tab (**Preferences...** under the **Options** menu; **Chapter 25**). **Global Model** acts in a toggle manner, switching between global and local display. **Global Model** is on by default.

Coupled

If *highlighted* (turned on), this signifies that all molecules in a document selected for simultaneous display will move together. **Coupled** is turned on following molecule alignment (see **Align** under the **Geometry** menu; **Chapter 19**) as well as after either **Generate Isomers** or **Generate Tautomers** has been utilized (see **Geometry** menu). **Coupled** acts in a toggle manner, that is, repeated selection couples or decouples.

Hydrogens

If *highlighted*, this signifies that hydrogens are to be included in the model. Note that 2D SD files, CIF files, as well as structures retrieved from the Cambridge Structural Database

or the Protein Data Bank may lack hydrogens. These need to be added to the model before CFD's (Chemical Function Descriptors) can be assigned (see discussion later in this chapter) and before calculations may be performed. This may be done either upon retrieval (see **Databases** and **Extract Ligands** under the **Search** menu; **Chapter 23**) or using **Grow Hydrogens** in the **Molecule Utilities** dialog (accessible from **Properties** under the **Display** menu; **Chapter 22**). **Hydrogens** acts in a toggle manner, that is, repeated selection turns the display of hydrogens on and off.

Labels

If *highlighted*, this signifies that labels associated with atoms, ribbons and bonds as well as with other attributes specified in **Configure...** (see discussion later in this chapter) are to be displayed in the model. **Labels** acts in a toggle manner, that is, repeated selection turns display of labels on and off. **Labels** is automatically turned on following selection of **Apply** or **OK** in the **Configure** dialog.

Ribbons

If *highlighted*, this signifies that ribbons are to be displayed along with the selected model. (If only ribbons are desired, select **Hide** for the model style.) **Ribbons** acts in a toggle manner, that is, repeated selection turns display of ribbons on and off.

Ramachandran Plot

If *highlighted*, this draws a Ramachandran plot for a protein input from the Protein Data Bank (see **Access PDB Online** under the **File** menu; **Chapter 16**). **Ramachandran Plot** acts in a toggle manner, that is, repeated selection turns the plot on and off. Note that coloring of the points on the plot (red for α-helices, blue for β-sheets, green otherwise) is not based on the actual 3D geometry but rather on assignments in the PDB file. The plot may be removed by again selecting **Ramachandran Plot**.

Hydrogen Bonds

If *selected*, this signifies that hydrogen bonds are to be drawn as part of the model. Hydrogen bonds are defined as non-bonded contacts between a nitrogen or oxygen and a hydrogen attached to nitrogen or oxygen separated by a distance ranging from 1.6 to 2.1Å and making an X–H--Y (X,Y = N, O) angle of >120°. Note that hydrogen bonds are not used in either molecular mechanics or quantum chemical calculations. **Hydrogen Bonds** acts in a toggle manner, that is, repeated selection turns display of hydrogen bonds on and off.

R/S Chirality

If *selected*, this adds R/S chirality labels to the model. **R/S Chirality** acts in a toggle manner, that is, repeated selections turns R/S labels on and off.

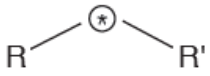
CFD's

If *checked*, this signifies that **C**hemical **F**unction **D**escriptors (**CFD's**) are to be displayed along with the structure model. CFD's are descriptors given to a molecule in order to characterize or anticipate its chemical behavior or to identify commonality among molecules with different structures. They parallel terms in a chemist's vocabulary such as ***lone pair*** (to suggest the role of a hydrogen-bond acceptor) and ***sterically crowded*** (to suggest that getting close may be difficult).

Spartan uses seven different kinds of CFD, the first six of which may be thought of as attributes of a molecule.

CFD	chemical language
1. hydrophobe	sterically-crowded region
2. aromatic	aromatic π system
3. hydrogen-bond donor	acidic hydrogen
4. hydrogen-bond acceptor	lone pair
5. positive ionizable site	basic site
6. negative ionizable site	acidic site

Default CFD assignments follow from atomic connectivity and rules covering common organic functional groups. They may be modified on a case-by-case basis using the **CFD Properties** dialog (accessible from **Properties** under the **Display** menu; **Chapter 22**). CFD's (represented by *'s in the figures below) may either be atom centered, or they may be centered in between atoms. For example, a hydrogen-bond acceptor CFD is placed on an ether oxygen, while an aromatic CFD is placed at the center of a benzene ring.

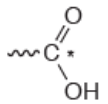


hydrogen-bond acceptor

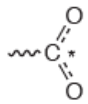
aromatic

CFD's may also be associated with a number of common functional groups. For example, a single CFD is provided for the carboxylic acid group, sited in-plane and equidistant from the carbon and two oxygens.

The CFD's for this and other functional groups may depend on pH. At neutral (or basic) pH this CFD should be designated both as a hydrogen-bond donor and a hydrogen-bond acceptor, whereas at acidic pH it should be designated as a negative ionizable site.



neutral pH: hydrogen-bond donor
hydrogen-bond acceptor



acidic pH: negative ionizable site

Finally, note that default assignments for hydrogen-bond donors and acceptors depend not only on atomic type (N, O, . . .) and (in the case of hydrogen-bond donors) on the availability of hydrogens, but also on known (or presumed) chemical behavior. For example, because an amide nitrogen would generally not be considered to be a hydrogen- bond acceptor, no (hydrogen-bond acceptor) CFD is provided. It is assigned to be a hydrogen-bond donor (if hydrogens are available) and the amide oxygen is assigned as a hydrogen-bond acceptor.

Note that hydrogen-bond acceptor and donor CFD's may be related to the electrostatic potential obtained from quantum chemical calculations. For example, a negative potential associated with an oxygen center suggests that it might serve as a hydrogen-bond acceptor, while a positive potential associated with a hydrogen attached to such a center suggests its role as a hydrogen-bond donor. While a chemist would reach the same conclusions simply by looking at the molecule, the calculations may be able to say which acceptor or donor sites are likely to be strong and which are likely to be weak.

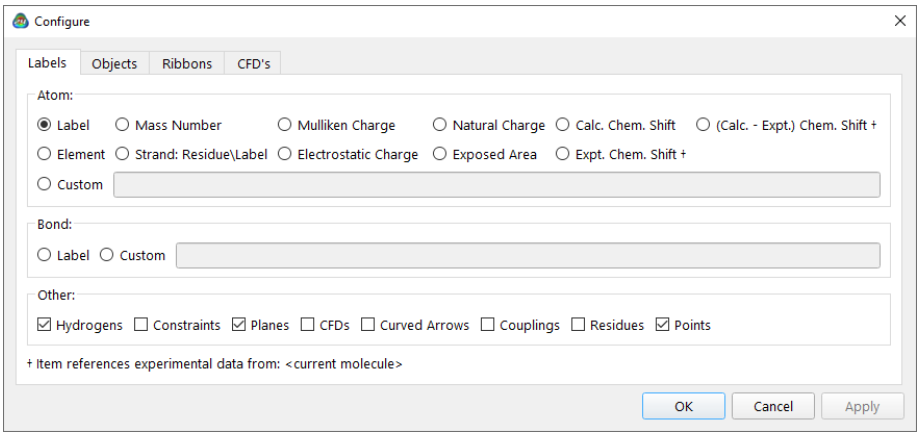
Calculated electrostatic potentials may offer significant advantage over CFD's and other qualitative descriptions when there are no heteroatoms and hydrogen bonding is not possible. The π system of benzene and other aromatics provide the most common examples. The same conditions that make the ring susceptible to electrophilic aromatic substitution mean that it repels other electron-rich regions (and attracts electron-poor regions).

A seventh CFD type, excluded volume, derives from knowledge of a molecule incorporated into a host. Excluded volumes may be obtained for ligands extracted from PDB files using the **Ligands** dialog (**Extract Ligands** under the **File** menu; **Chapter 16**).

Configure...

This selects the types of labels attached to atoms and other molecule attributes.

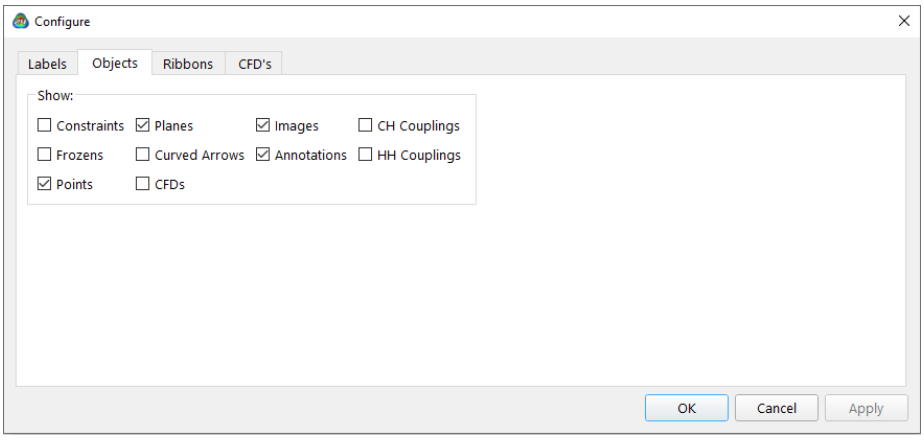
Labels



Under **Atom**, labels may be selected from among the following: **Labels**, a unique element/number combination that may be changed from the **Atom Properties** dialog (accessible from **Properties** under the **Display** menu; **Chapter 22**), **Element**, **Mass Number**, **Mulliken Charge**, **Electrostatic Charge**, **Natural Charge**, **Strand: Residue\Label** (polypeptides and polynucleotides), and **Exposed Area** (of an atom in a space-filling model), **Chem Shift**, **Expt Chem Shift** and **Calc-Expt Chem Shift**. Selection of **Custom** leads to a text box into which a formula defining an atomic label may be entered. Details are provided in **Formulas** under the **Display** menu (**Chapter 22**). Under **Bond**, either **Label** a unique number prefaced by the word "Bond" that may be changed from the **Bond Properties** dialog (accessible from **Properties** under the **Display** menu; **Chapter 22**) or **Custom** entered in a text box. Under **Other**, **Hydrogen Labels**, **Point Labels**, **Plane Labels**, **Constraint Labels**, **Residue Labels**, **Reaction Labels**, **Coupling** (coupling constant) and/or **CFD Labels** may be provided. Default settings (for a new molecule) are made in the **Molecule** tab (**Preferences** under the **Options** menu; **Chapter 25**).

Objects

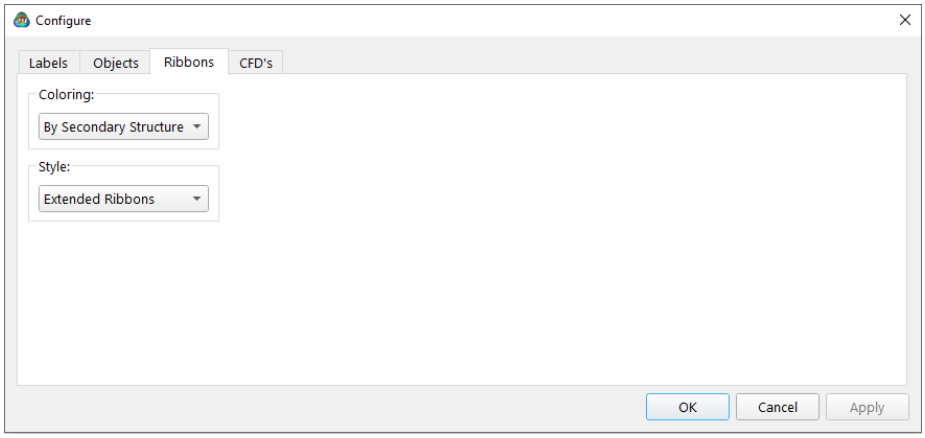
Clicking on the **Objects** tab leads to the **Configure Objects** dialog:



If checked, **Constraint** and **Frozen** markers, **Points** and **Planes**, **Reaction** arrows and **CFD's** attach to the model. If not checked, these are shown only in the respective modes, for example, **Frozen** markers are shown only if **Freeze Center** is selected. **Images** and **Annotations** refer to images and text, respectively, associated with the molecule display. An image or text string may be associated by first putting it on the clipboard and then pasting it on screen. Either may be positioned on screen and sized. **HH** and **CH** refer to experimentally assigned HH and CH couplings.

Ribbons

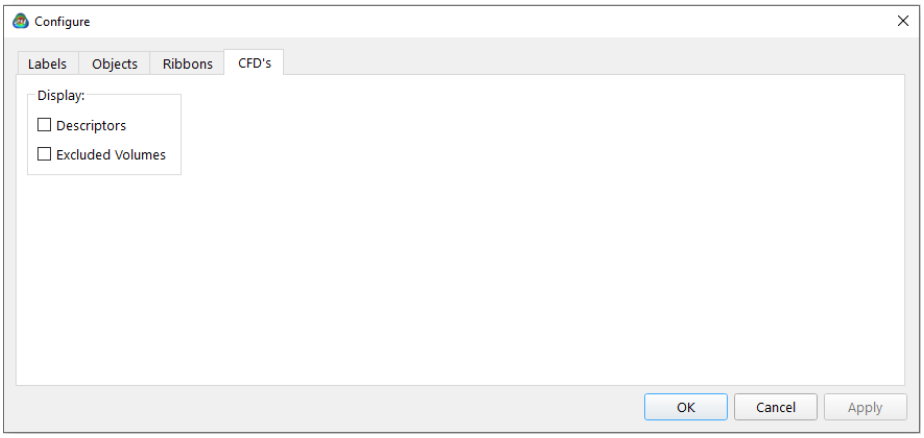
Clicking on the **Ribbons** tab leads to the **Configure Ribbons** dialog:



Ribbon coloring may be selected in a menu from among the following: **Monochrome**, **By Secondary Structure**, **By Strand** or **By Residue**. Ribbon style may be selected in a menu from among the following: **Ribbons**, **Extended Ribbons**, **Beads** or **Lines**.

CFD's

Clicking on the **CFD's** tab leads to the **Configure CFDs** dialog:



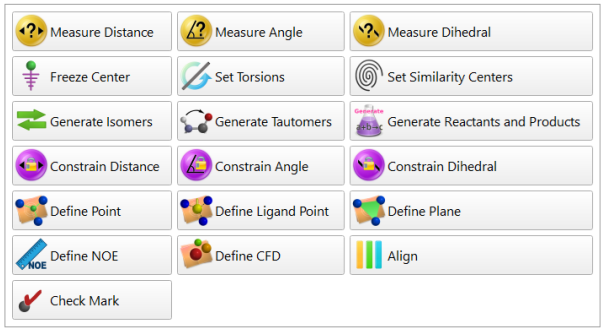
This allows for turning CFD descriptors excluded volumes on or off.

The **Configure** dialog is removed from the screen with all changes maintained by *clicking* on **OK**. *Clicking* on **Cancel** or on removes the dialog but changes are lost. *Clicking* on **Apply** maintains the changes but leaves the **Configure** dialog on screen. Note that **Labels** (from the **Model** menu) will be turned on following either *clicking* on **OK** or **Apply**.

Chapter 19

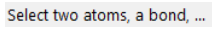
The Geometry Menu

*Functions available under the **Geometry** menu allow querying and changing bond lengths, angles and dihedral angles, defining points and ligand points, planes and chemical function descriptor (CFD) locations, setting geometrical constraints (distances, angles and dihedral angles) or ranges of constraints, freezing atomic centers, altering default bond and ring conformer assignments, introducing NOE's (maximum intramolecular distances) in conformational searching, selecting atoms or CFD's for similarity analysis and aligning molecules. Generation of tautomer list has been moved from the **Search** menu and the ability to generate a list of stereo and/or regioisomers, and set ("check") atoms for BSSE and NICS NMR calculations.*

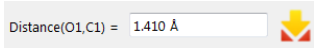


Measure Distance **Angle** **Dihedral**

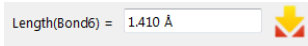
Measure Distance displays the distance (in Ångstroms) between any two atoms. Selection results in a message at the bottom left of the screen:



Clicking on two atoms displays the distance at the bottom right of the screen:



Alternatively, *clicking* on a bond displays its length:



Measure Distance may also be used to alter the distance between atoms by *clicking* inside the box to the right of **Distance (A,B)** = or **Length (A)**=, and either typing or using the number pad. ***Spartan*** includes the ability to alter bond distances inside rings. This is accomplished by temporarily imposing distance constraints followed by molecular mechanization and then removing the constants.

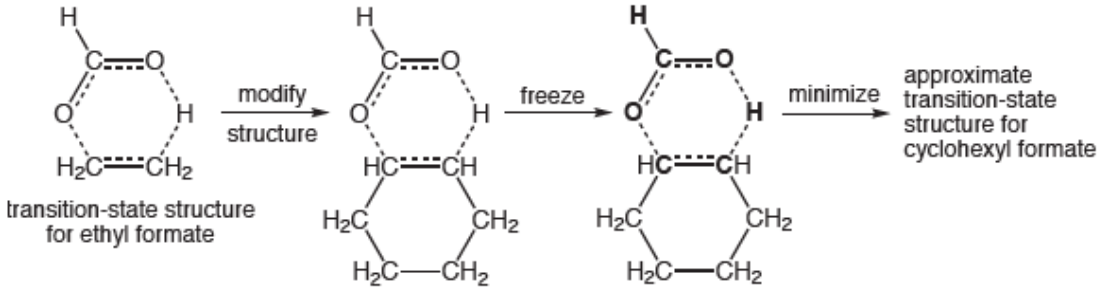
The distance (length) may be entered into the spreadsheet by *clicking* on to the right of its display (see **Spreadsheet** under the **Display** menu; **Chapter 22**). Alternatively, the label "**Distance (A,B)**=" or "**Length (A)**=" may be *dragged* into the spreadsheet, the formula editor (see **Formulas** under the **Display** menu; **Chapter 22**) or into any of the dialogs associated with the Spartan Spectra and Properties Database (see **Databases** under the **Search** menu; **Chapter 23**).

Angle and dihedral angle queries and charges are handled in a similar manner. Angles require that three atoms or two bonds be identified in the proper order while dihedral angles require that four atoms or three bonds be identified in the proper order.

Freeze Center

This allows one or more atoms to be held in place during minimization (in the 3D builder) or (optionally) during equilibrium or transition- state geometry optimization, conformational searching, or energy profile generation using molecular mechanics or quantum chemical methods. In the latter case, use of frozen atoms needs to be explicitly indicated in the **Calculations** dialog (**Chapter 21**).

Atom freezing may be useful in a number of situations, among them guessing a transition-state geometry for a reaction that is closely related to one for which a transition state is available. For example, a good guess at the transition state for pyrolysis of cyclohexyl formate will be obtained by modifying the transition state for pyrolysis of ethyl formate, freezing all but the modified sections (designated in bold in the following figure) and then minimizing.



Selection of **Freeze Center** leads to a message at the bottom left of the screen:

Select atom to freeze.

Clicking on an atom freezes it, *clicking* again thaws it. Buttons at the bottom right of the screen allow for freezing all atoms (**Freeze All**), freezing all non-hydrogen atoms (**Freeze Heavy**) and for thawing all atoms (**Thaw All**).

Another important use of frozen atoms is to adjust XH bond lengths in structures resulting from a search of the Cambridge Structural Database (see **Databases** under the **Search** menu; **Chapter 23**). Hydrogen positions are more often than not poorly located in small molecule X-ray structures, and XH bond lengths are commonly shorter than they should be. Structures incorporating such bond lengths are inappropriate for energy and property calculations and may also be problematic as starting geometries in quantum chemical calculations. Protein crystal structures often lack hydrogens, and hydrogen positions of bound ligands extracted from PDB files may be poorly defined. A reasonable solution to both problems is to freeze all heavy (non-hydrogen) atoms (**Freeze Heavy**) and then to carry out molecular mechanics optimization using either **Minimize** from the **Build** menu (see discussion in **Chapter 20**) or from **Calculations...** under the **Setup** menu (see discussion in **Chapter 21**).

Frozen atoms are indicated by magenta colored markers ( ). Whether or not these are included with the model (outside of **Freeze Center** mode) is controlled by the **Molecule Preferences** dialog under **Preferences...** in the **Options** menu (**Chapter 25**).

Set Torsions

Spartan automatically identifies bonds and rings for conformational searching and specifies default step sizes.

> Three different rule sets have been provided. **Normal** is appropriate to determine lowest-energy conformation or to establish a distribution of low-energy conformers, **Skeletal** is used to build conformer libraries for use in similarity analyses and removes conformers that only involve hydrogen positions. **Normal** and **Skeletal** rule sets consider only chair conformers for saturated six-membered rings, for example, cyclohexane. **Thorough** considers twist-boat conformers in addition to chair conformers. (Different modes of conformational analysis are discussed in **Chapter 21**.) Selection of rule set is made in the **Settings Preferences** dialog (**Preferences...** under the **Options** menu; **Chapter 25**).

Set Torsions allows these defaults to be altered. Selection results in rotatable bonds being marked by a gold cylinder and flexible rings being marked by a gold circle around one or more atoms, and a message at the bottom left of the screen:

Select bonds or ring atoms.

Clicking on a bond or an atom contained in a ring selects it for rotation. In the case of a ring, rotation means that the atom is to be puckered up and down (restricted rotation). The default rotation is provided in a box to the right of **Fold** at the bottom right of the screen. With **Normal** rules, this is typically 2 or 3 for a single bond (step size of 180° and 120°, respectively) and 3 for a flexible ring. A value of either 0 or 1 indicates that the bond (ring) is not to be rotated (rotation by 360°). Other integer values may be entered into the box using the number pad that appears immediately above. The original settings may be retrieved by *clicking* on **Defaults** at the bottom right of the screen. *Double clicking* on a marked bond or ring atom removes the marker and eliminates the conformational degree of freedom.

While conformational degrees of freedom in acyclic molecules are "easy" to define (and default selections usually appropriate), dealing with medium to large rings can be problematic. Experience suggests that torsions applied to ring bonds are likely to be more effective in "catching" the multitude of possible conformers than puckering motions applied to atoms. Experimentation may be necessary.

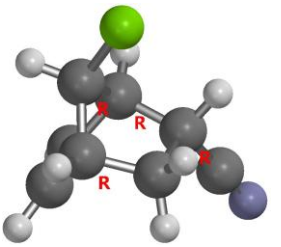
◀ and ▶ buttons at the bottom right of the screen are available to step through the possible single-bond conformers (ring conformers are not provided). Geometries have not been optimized and some conformers may be severely crowded. Any conformer can be selected in lieu of the initial structure. The full set of single-bond conformers may be placed in an unnamed document by *clicking* on ▼ (to the right of the step buttons) and then *clicking* **OK** in the dialog that results. Note that duplicate conformers are not removed.

The total number of conformers (excluding ring conformers) that will be examined in a systematic search (see discussion in **Chapter 21**) is indicated at the bottom right of the screen ***Conformers* = *xx***. This is an upper bound to the number of conformers that will actually be kept as duplicates and high-energy conformers are eliminated. This can be *dragged* into the spreadsheet (see discussion in **Chapter 22**), into the **Formulas** dialog (see discussion in **Chapter 23**), or into any of the dialogs associated with the Spartan Spectra and Properties Database (see discussion in **Chapter 23**). The number of conformers is also available under the **Molecule** tab in the **Molecule Properties** dialog (**Properties** under the **Display** menu; **Chapter 22**). This can be dragged or posted to the spreadsheet.

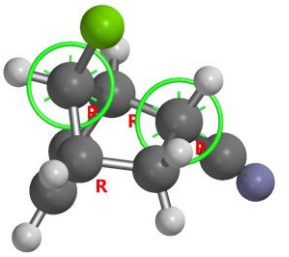
In practice, an efficient systematic search scheme is employed for molecules that incorporate both flexible rings and acyclic free rotors. This divides the search into two parts, the first involving only rings and concluding with the few lowest-energy conformers and the second starting from these conformers to explore the conformations of the acyclic rotors.

Generate Isomers

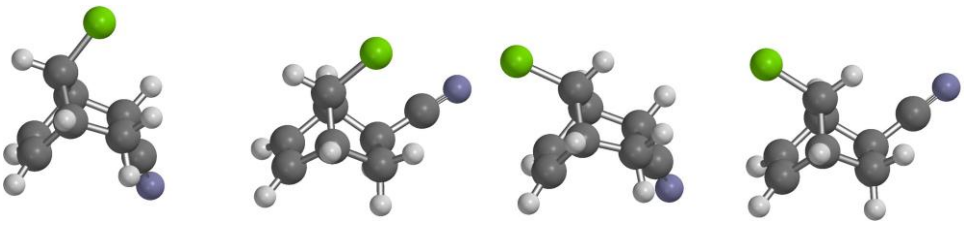
This generates a list of regio and/or stereoisomers starting from a single structure. It's use may simply be a convenient way to examine the relative energies of possible isomeric products of a chemical reaction or to see which of these products best fits on observed NMR spectrum. We first illustrate the more common task of stereoisomer generation. Starting from *endo*-chlorobicyclo[2.2.1]hept-5-ene-2-carbonitrile*, selection of **Generate Isomers** leads to a structure decorated by R/S labels.



Inversion of C_2 and C_7 leads to four possible stereoisomers. To select, *click* on them in turn, following which they will be encircled in green.

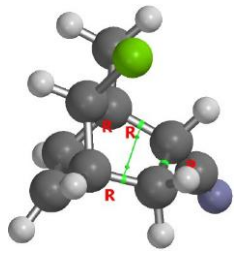


Clicking on **Generate List** to the right of **Isomers** at the bottom right of the screen lead to a dialog requesting if the list is to be placed in a new document or added to the current document. The four isomers are shown below.

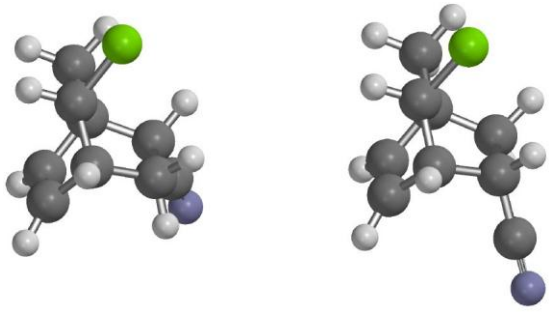


Regioisomer generation is signaled by *clicking* in order on three bonds, for example, in a C_1-C_2, C_2-C_3 and C_3-C_4 in 1-methyl-7- chlorobicyclo[2.2.1] hept-2-ene-2-carbonitrile.

* More precisely this is the *syn* (chloro), *endo* (cyano) stereoisomer.

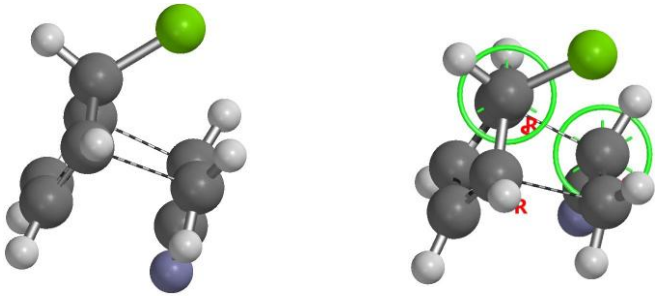


Clicking on **Generate List** gives rise to a pair of regioisomers.



Markers for regio and stereoisomers may be applied together.

Set Isomers can also be used to generate a list of transition states (that would lead to stereoisomeric products) starting from a single transition state. We use the Diels-Alder transition state leading to the four stereoisomers of 7-chlorobicyclo[2.2.1]hept-5-ene-2-carbonitrile. The procedure is the same. Starting from a transition state for one of the isomers, select **Set Isomers** and *click* on C_2 and C_7.

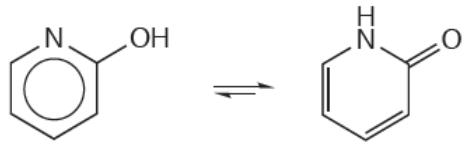


Clicking on **Generate List** then provides the four transition states which will lead to the four stereoisomeric products.

Generate Tautomers

This allows tautomers, that is, isomers arising from rapid transfer of hydrogens among

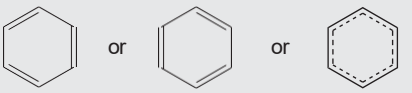
heteroatoms, to be identified based only on the structure. Tautomer identification is especially important in dealing with heterocyclic compounds where two or more different tautomers may exist in equilibrium, and the identity of the dominant species may not be obvious. For example, while 2-hydroxypyridine is an aromatic molecule and would be expected to be unusually stable, the non- aromatic 2-pyridone molecule is actually thermodynamically favored.



2-hydroxypyridine 2-pyridone

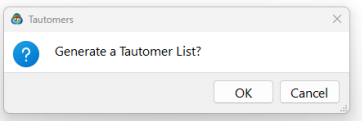
This, of course, has consequences in trying to locate a particular molecule in a database where it may be represented by a different tautomer.

In contrast, ***Spartan*** almost always recognizes different ways to draw the same structure. For example, benzene can be represented as one of two structures with alternating double and single bonds or a single delocalized structure in which all bonds are the same and intermediate in length between single and double bonds.



All of these translate into the same molecule.

The procedure incorporated into ***Spartan*** is limited to tautomers involving nitrogen, oxygen, phosphorous and sulfur. Tautomers involving carbon have intentionally been excluded. There are too many of them, and inclusion would swamp the more likely heteroatom tautomers. Also excluded are zwitterion tautomers, for example, $H_3N^+CH_2CO_2^-$ as a tautomer of glycine. In the gas-phase, these are likely to be much less stable than "neutral" forms. Within these limits, all possible tautomers will be identified.



The existence of (heteroatom) tautomers is signaled by the word **Tautomer** displayed at the bottom right of the screen. If tautomers exist, individual structures may be examined or a complete list of structures generated by selecting **Tautomers** from the **Search** menu (or

clicking on the icon on the top of the screen). Following this, *clicking* on the step buttons at the bottom right of the screen moves through the list of tautomers. Any one of these may be used in place of the original. The full list of tautomers may be generated by *clicking* on (at the bottom right of the screen). This leads to a dialog.

Clicking on **OK** dismisses the dialog and leads to a new document containing the full list of tautomers. (The original document is unaffected.) *Clicking* on either **Cancel** or dismisses the dialog without generating a tautomer list.

> Tautomers are generated without consideration of steric crowding and should be minimized with molecular mechanics (**Minimize** under the **Build** menu; **Chapter 20**) prior to using them for quantum chemical calculations. It may also be important to perform a molecular mechanics search to obtain a guess at the lowest-energy conformer (see **Calculations...** under the **Setup** menu; **Chapter 21**).

Constrain Distance **Angle** **Dihedral**

These introduce one or more geometric constraints during structure minimization (in build mode), and (if requested) during equilibrium or transition-state geometry optimization or conformational searching using methods specified in the **Calculations** dialog (**Calculations...** from the **Setup** menu; **Chapter 21**). They also allow for setting a range of constraints needed for generation of **Energy Profile** calculations. Constraints may be useful in a number of situations, among them:

(i) Constructing conformational energy profiles where one or more dihedral angles need to be fixed while other geometrical variables are optimized.

(ii) Optimizing molecular structures where the values of certain key parameters are known, for example, optimizing the geometry of a molecule with an intramolecular hydrogen bond or a disulfide linkage.

(iii) Building molecules with unusual geometries, for example, molecules with very long bonds, as might be required in the construction of transition states and intermolecular complexes.

Selecting **Constrain Distance** results in a message at the bottom left of the screen:

Select two atoms, a bond, ...

Clicking on two atoms, or a bond results in a message at the bottom right of the screen:

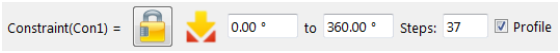
Clicking on changes it to and shows the current distance:

This (constraint) distance can then be changed by *clicking* inside the box and either typing or using the number pad that appears immediately above. Alternatively, the existing distance may be used as the constraint distance. If the selected distance had previously been constrained, the icon would have been initially displayed. *Clicking* on turns the constraint off and returns the icon to . Finally, the value of the constraint (that may be different from the value of the current distance*) may be entered into the spreadsheet by *clicking* on to its right. Alternatively, the label ***Constraint (A,B)=*** may be *dragged* into the spreadsheet, the formula editor (see **Formulas** under the **Display** menu; **Chapter 22**) or into any of the dialogs associated with the Spartan Spectra and Properties Database (see **Databases** under the **Search** menu; **Chapter 23**).

This sequence of operations (bond identification followed by turning the constraint on and off) may be repeated as many times as necessary. Any bonds or non-bonded distances on which constraints are to be imposed are indicated by magenta colored markers. Any constraints introduced are automatically enforced only upon energy Minimization in Build mode , but needs to be requested for methods specified in the **Calculations** dialog (**Calculations...** from the **Setup** menu; **Chapter 21**).

Angle and dihedral angle constraints are handled in a similar manner. Note that ***points and planes may not be used to define constraints***.

Locking in a constraint leads to two additional text boxes at the bottom right of the screen. This allows a sequence of constraints to be defined (from some initial value to some final value in a given number of steps) for the purpose of constructing an energy profile along a predefined set of coordinates (see **Calculations...** under the **Setup** menu; **Chapter 21**):



The leftmost box specifies the initial value of the constraint, the middle box to the right of **to** the final value, and the rightmost box to the right of **Steps:** the number of steps. Values are specified by *clicking* inside the appropriate box and using the number pad. For example, were the initial value set to 0°, the final value to 180° and the number of steps to 10, then a series of ten constraints (0°, 20°, 40°,... 180°) would be specified. This can also be accomplished using the **Constraint Properties** dialog, and the value of the constraint posted to the spreadsheet.

* Note, however, that it may be problematic to carry out a constrained geometry optimization starting from a structure that is very different from that satisfying one or more constraints.

Whether or not constraint markers are included as part of the model (outside of constrain distance, constrain angle or constrain dihedral mode) is controlled from the **Molecule Preferences** dialog (**Preferences...** under the **Options** menu; **Chapter 25**).

Define Point

This defines a point as the geometric (unweighted) center of selected atoms (or points) previously defined. Selection results in display of a message at the bottom left of the screen:

Select atoms. Repeat to terminate.

Clicking on atoms (or points) in any order and *clicking* a second time on any one of the atoms (or points) defines a point (depicted as a small sphere). As many points as desired can be defined and these are treated in the same way as an atom in defining distances, angles, etc. Points move with the molecule as its geometry is altered.

Define Ligand Point

This defines a point of attachment directed perpendicular to the geometric center of the plane defined by three atoms (or best plane in the case of four or more atoms). *Clicking* on atoms (or points) in any order and *clicking* a second time on any one of the atoms (or points) defines a ligand point (depicted as a small sphere). As many ligand points as desired can be defined. A ligand point shares all the characteristics of a normal point, but may also be used to bond to atomic fragments, functional groups, etc. See **Make Bond** under the **Build** menu (**Chapter 20**) for a discussion. Ligand points move with the molecule as geometry is altered.

> **Delete** from the **Build** menu or the **Delete** key may be used to remove a point or ligand point.

Whether or not points and ligand points are shown as part of the model is controlled from the **Molecule Preferences** dialog (**Preferences...** under the **Options** menu; **Chapter 25**).

Define Plane

This defines and displays a reference plane. Selection results in display of a message at the bottom left of the screen:

Select three atoms.

Clicking on three atoms or points defines a plane. As many planes as desired may be defined, and these may be used in defining distances, angles, etc. Planes move with the molecule as its geometry changes.

Delete from the **Build** menu ( ) or the **Delete** key may be used to remove a plane.

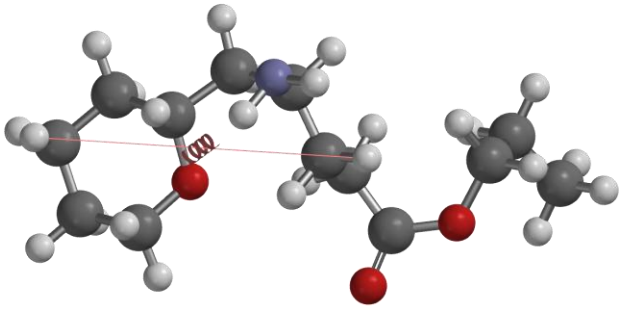
Whether or not planes are included as part of the model for an is controlled from the **Molecule Preferences** dialog (**Preferences...** under the **Options** menu; **Chapter 25**).

Define NOE

Define NOE is used to specify non-bonded distances that need to be kept below a threshold value. These follow from NOE (Nuclear Overhauser Effect) measurements and may be referred to as NOE conditions or simply NOEs. NOEs are specified by *clicking* on two (non-bonded) atoms. In response, a message appears at the bottom of the screen:

NOE(Con1) = 5.000 Å

Clicking on changes it to and enters the default value for the NOE threshold into the box. This value can be changed. A line is drawn between atoms that are to be kept within the threshold value.



Once set, NOEs are used in conformational searching (conformer distribution only; see discussion under **Calculations...** under the **Setup** menu; **Chapter 21**) without further user intervention. They act as a post-processor filter to eliminate conformers that do not satisfy the constraints and do not affect the speed of the search.

Define CFD

This defines the position of a new CFD. The **CFD Properties** dialog (accessible from **Properties** under the **Display** menu; **Chapter 22**) allows CFD size and type to be specified. Selection results in display of a message at the bottom left of the screen:

Select atoms. Repeat to terminate.

Clicking on atoms (or points) in any order, and *clicking* a second time on any one of the atoms (or points) defines the position of a CFD (depicted as a transparent sphere labeled Null). *Clicking* twice on only one atom locates the CFD on the atom. As many CFD's as desired can be defined. These move with the molecule as its geometry is altered.

Whether or not CFD's are included as part of the model is controlled from the **Molecule Preferences** dialog (**Preferences** under the **Options** menu; **Chapter 25**).

Set Similarity Centers

This specifies the set of atomic centers, or **C**hemical **F**unction **D**escriptors (**CFD's**) or pharmacophore elements that are to be used in similarity analysis. Choice of whether the similarity analysis is to be based on structure or CFD's (pharmacophore elements) is made from a menu that appears at the bottom right of the screen following selection of **Set Similarity Centers**. This menu is also accessible from the **Similarity Analysis** task in the **Calculations** dialog (**Calculations...** under the **Setup** menu; **Chapter 21**).

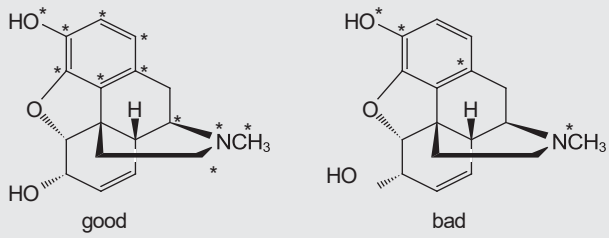
Similarity Based on Structure

Selection of **structure** from the **Similarity by** menu at the bottom right of the screen results in a message at the bottom left of the screen:

Select Atoms.

Clicking on an atom designates it as an alignment center and it marks it with a violet circle. *Clicking* on the circle removes the designation (and the circle). Hydrogens are not used in similarity analysis and may not be selected.

> The computer time required for similarity analysis based on structure depends on the number of molecules in the query, the number of molecules in the library to which comparisons are being made and the choice of similarity centers. Single atoms should never be selected as similarity centers unless they are unique. Rather, contiguous groups of atoms that are likely to be unique should be chosen. For example, analysis carried out on morphine based on selection of its phenol and trialkalkylamine components (left) will be much faster and more informative than that based on selection of the phenol oxygen, two carbons on benzene and the nitrogen (right).



This is because there will be many fewer occurrences of phenol and trialkylamine substructures in the library entries than there will be aromatic carbons and nitrogen atoms and fewer permutations that need to be examined and fewer false positive results.

Similarity Based on CFD's

Similarity analysis may also be based on CFD's. These indicate whether a particular center is likely to act as a hydrogen-bond donor or hydrogen-bond acceptor, is likely to take on positive or negative charge or is a hydrophobe. This is accomplished by selecting **CFD** from the **Similarity by** menu at the bottom right of the screen. In response, the selected molecule is augmented with CFD's, and a message appears at the bottom left of the screen:

Clicking on a CFD designates it as a similarity center and marks it with a violet circle.

Select CFD's.

Clicking on the circle removes the designation (and the circle).

The computer time required for similarity analysis based on CFD's depends on the number of molecules and/or pharmacophores in the query, the number of molecules and/or pharmacophores in the library and the nature of the selected CFD's in the query. For each query/library pair, where the query is represented by i CFD's of type A, j CFD's of type B, k CFD's of type C, etc., the number of different alignment combinations that need to be considered is given by:

$$i!j!k! \ldots$$

(! is the factorial symbol.) This rapidly becomes unmanageable. For example, if there are 6 CFD's of type A, 3 of type B and 2 of type C, the number of combinations is 6!3!2! or 8640. Considerable savings can be achieved by limiting the number of CFD's of a given type in the overall representation.

Similarity Based on a Pharmacophore

Finally, it is possible to base similarity on a pharmacophore. This is set up in exactly the same way as similarity based on CFD's insofar as the elements of a pharmacophore are CFD's (see previous section).

Align

This aligns the selected molecule to all other molecules in the same document based either on atom labels, on structure, or on CFD's. If the selected molecule is a pharmacophore, this aligns the pharmacophore to all molecules in the same document (based on CFD's). Choice of whether alignment is based on labels, structure or CFD's is made from the **Align by** menu that appears at the bottom right of the screen.

> The terms "alignment" and "similarity analysis" have subtly different meanings inside of ***Spartan***. Alignment refers to the operation in which the molecules in a document are reoriented to best coincide with the selected molecule (or pharmacophore). Where a molecule is selected, alignment may either be based on a set of atoms or on a set of CFD's. Where a pharmacophore is selected, alignment is based on a set of CFD's. A single reorientation is provided for each molecule in the document that can be aligned to the selected molecule (pharmacophore).
>
> Similarity analysis refers to the operation in which all molecules in a user-specified library comprising one or more documents are compared one-on-one to all molecules and/or pharmacophores in the selected document (in the graphical user interface). As with alignment, either structural elements or CFD's may be employed. The result is a similarity score from 0 to 1 (1 is perfect) and a manipulable model of the two molecules (or molecule and pharmacophore). It is likely that most comparisons will not be successful, and also that some comparisons may lead to more than one score (corresponding to different orientations of the two components). Similarity analysis both by structure and CFD as well as to a pharmacophore is discussed further in **Chapter 21**.
>
> In short, the role of alignment is to orient molecules and/or pharmacophores in order to highlight their common features, while the role of similarity analysis is to rank pairs of molecules or molecule/pharmacophore pairs according to the extent of their common features.

Align Based on Labels/Align Based on Structure

Selection of either **Labels** or **Structure** from the **Align by** menu results in a message at the bottom left of the screen:

Select Atoms.

Clicking on an atom designates it as an alignment center and marks it with a red circle. *Clicking* on the circle removes the designation (and the circle). Following selection of alignment centers, *clicking* on the **Align by** button at the bottom right of the screen aligns the molecules. If no atoms are selected prior to *clicking* on **Align by**, then alignment is based on all (non-hydrogen) atoms.

Following alignment, two or more molecules may be displayed at once (see **Spreadsheet** under the **Display** menu; **Chapter 22**). Their motions (coordinates) will be coupled following alignment but may be uncoupled allowing the aligned molecules to move independently (see **Coupled** under the **Model** menu; **Chapter 18**). Note that alignment center selections are kept and molecules can be realigned by again selecting **Align** from the **Geometry** menu (or *clicking* on) followed by *clicking* on the **Align by** button.

Align Based on CFD's

It is also possible to base molecule alignment on CFD's. These indicate whether a particular center is likely to act as a hydrogen- bond donor or acceptor, is likely to take on positive or negative charge or is a hydrophobe. This is accomplished by choosing **CFD** from the **Align by** menu at the bottom right of the screen followed by *clicking* on the **Align by** button. In response, the selected molecule is augmented with CFD's, and a message appears at the bottom left of the screen:

Select CFD's.

Clicking on a CFD designates it as an alignment center and marks it with a red circle. *Clicking* on the circle removes the designation (and the circle).

The **CFD Properties** dialog (accessible from **Properties** under the **Display** menu; **Chapter 22**) allows definitions of the individual CFD's to be altered.

Following selection of CFD's, *clicking* on the **Align by** button at the bottom right of the screen aligns the molecules. If no CFD's are selected prior to *clicking* on **Align by**, then alignment is based on all CFD's.

Following alignment, one or more molecules may be displayed at once using spreadsheet functions (see **Spreadsheet** under the **Display** menu; **Chapter 22**). Their motions (coordinates) will be coupled following alignment but may be uncoupled allowing the aligned molecules to move independently (see **Coupled** under the **Model** menu; **Chapter 18**). Note that alignment center selections are kept and molecules can be realigned by again selecting **Align** from the **Geometry** menu (or *clicking* on) followed by *clicking* on the **Align by** button.

Align to Pharmacophore Definitions

Finally, it is possible to align molecules to a pharmacophore contained in the same list. This is set up in exactly the same way as alignment based on CFD's insofar as a pharmacophore element is a CFD (see previous section).

For both structure and CFD alignment (and for alignment to a pharmacophore), an alignment

score from 0 to 1 (perfect alignment), is available in the spreadsheet. This is accessed by *clicking* on the **Add** button at the bottom of the spreadsheet and selecting **Alignment Score** from the **Molecule List** tab of available properties (see **Spreadsheet** under the **Display** menu; **Chapter 22**). A score of 0 is assigned to molecules that cannot be aligned to the selected molecule.

Check Mark

Check Mark is used to designate specific atoms in BSSE (Basis Set Superposition Error) calculations as well as designation of (typically ring) atoms in NICS calculations (Nucleus Independent Chemical Shifts).

Chapter 20

The Build Menu

*The **Build** menu provides access to a sketch palette for drawing organic and organometallic molecules in 2D, model kits and associated tools for building and editing organic, inorganic and organometallic molecules as well as polypeptides and polynucleotides in 3D, for guessing transition-state structures and a facility for generating lists of substituted molecules. The Windows version provides a seamless access to ChemDraw. 2D to 3D conversion and 3D structure refinement is by way of molecular mechanics. With **Spartan'26**, the **Build** menu gives the ability to attach experimental proton and ^{13}C NMR chemical shifts as well as observed HH and CH couplings to either 2D drawings or 3D models.*

Specification of 3D molecular structure is a necessary first step to any molecular mechanics or quantum chemical calculation. ***Spartan*** provides a variety of tools. Organic and organometallic molecules can either be rendered as 2D sketches and later brought into 3D, or directly constructed from 3D fragments. Polypeptides and polynucleotides need to be built in 3D.

2D Sketch Palette

The 2D sketch palette contains tools for making and manipulating drawings. There are also tools for specifying charges and radical sites, for adding cues to designate stereochemistry and for setting limits for a substructure search. Directly above the palette are icons that place curved arrows on 2D drawings thereby defining chemical reactions as well as for attaching experimental proton and ^{13}C chemical shifts and observed HH and CH couplings to 2D drawings.

Palette Atoms. H, **B**, **C**, **N**, **O**, **F**, **Si**, **P**, **Cl**, **Br** and **I** only.

More: Periodic Table/Groups/Ligands/Substituents. *Clicking* on the **More** button that appears below **H** and **B**, leads to a dialog that allows any atom to be specified from the Periodic Table (**Periodic Table** tab), a variety of common functional groups (**Groups** tab) and a number of common organometallic ligands (**Ligands** tab). *Clicking* on an element, group or ligand dismisses the dialog and results in whatever was selected being placed in the icon and available for use.

Common Rings. Three icons facilitate the rapid addition of common rings. The upper icon is for benzene , the middle icon is for cyclohexane, both as a "flat" drawing and as a chair , and the lower icon is for cyclopropane , cyclobutane , cyclopentane and cycloheptane . In the latter two ring icons, *double-tap* **or** *double-click* exposes the additional selection options.

Common Carbonyl Groups. Three icons facilitate the rapid addition of carbonyl , carboxylic acid/ester and amide groups to drawings. In the latter two icons *double-tap*/*double click* exposes alternate attachment options.

Stereochemical Markers. Wedges and dashes, represented by and , can be used to designate in-out stereochemistry. Immediately to the right is a single-bond (unwedged) marker . Once a stereochemical marker has been added to a drawing, it is possible to designate the orientation of hydrogen atoms and/or substituents bonded to six-member rings as **ax(ial)** or **eq(uatorial)** (*ax* and *eq* labels appear only on the drawing, not in the palette).

Open-site Markers. designates an "open-site" in a drawing. Searching a database will return all structures with any atoms (including hydrogen) or groups in open sites.

Charge/Radical Markers. Conventional bonding rules (neutral C makes 4 bonds, neutral N makes 3 bonds, and so on) are enforced when 2D perspective drawings are converted into 3D models. This is accomplished by adding hydrogen atoms to the drawing. For example, a single carbon on screen will give methane, a single line, ethane, and a double line, ethylene. (Hydrogen atoms are added to nitrogen, oxygen, phosphorous and sulfur in the 2D drawings.) When another outcome is desired, for example, for an ion or free radical, charge or radical markers must be added to the drawing.

Two icons and , are used to label atoms that bear formal charges. is used to label atoms that are neutral, open-shell radicals. Each of these markers affects the number of electrons and the number of hydrogen atoms added to the 3D model. For example, **O** will produce a 3D model of water, H_2O. However, adding the appropriate marker will result in 3D models of H_3O^+ , HO^- , or HO radical , respectively. Only one charge/radical marker can be assigned to an atom (radical cations and radical anions cannot be specified in this way) but as many atoms as desired may be provided charge or radical designations. The

sum of charges and radical sites will be used to determine overall molecular charge and number of unpaired electrons needed for a quantum chemical calculation.

Only one charge/radical marker is displayed on the palette, but *clicking* on the icon will cause each marker to appear in turn. Examples of the use of reaction arrows is provided in **Chapter 23**.

Transition State Searching. The Guess Transition States Icon remains available from the tool bar or the Build menu and allows for designation of one or more reaction arrows enabling access to ***Spartan's*** automated procedure for guessing transition states. The tail of the arrow corresponds to the source of the electron pair. If the source is a lone pair, then select the atom that holds the lone pair. If the source is a bond, then select the bond. *Clicking* on an atom or bond highlights (colors gold) the atom or bond. *Clicking* again on the same atom (or same bond) removes the highlighting. The head of the arrow corresponds to the destination of the electron pair. If the destination is an atom (leading to a lone pair), then select the atom that will hold the lone pair by *clicking* on it two times. If the destination is an existing bond (leading to an increase in bond order from single → double or double → triple), then select (*click* on) the bond. If no bond presently exists, select (*click* on) the two atoms that will become bonded upon reaction. These operations result in a curved arrow being drawn on the reactant structure. This extends from an atom, or the center of a bond to an atom, or the center of a bond, or the center of a dotted line that has been drawn between atoms that are to be bonded.

Note that the head and tail do not need to reside on atoms or bonds on the same fragment. Also the tail may involve atoms of two detached fragments.

The process (tail selection followed by head selection) is repeated as necessary to fully define the reaction. Incorrect reaction arrows may be removed by *clicking* on from the palette and *clicking* on the arrow. You then need to select to continue arrow selection.

Once defined, reaction queries can be used to relate the properties of the products of chemical reactions to those of the products for the purpose of mining the Spartan Spectra and Properties Database. Full discussion is provided in **Databases** in **Chapter 23**. Reaction queries can also be used to search the Spartan Reaction Database for transition-state structures associated with the defined reaction (also discussed in **Databases**). In both of these cases, **Open-Site Markers** or structure queries (discussed in **Structure Query** in **Chapter 23**) will also normally be employed. Finally, reaction queries may be employed to automatically provide a guess at a transition state based on similarity to an entry in the Spartan Reaction Database (see discussion in **Transition State** in **Chapter 23**). Here, no additional information is required.

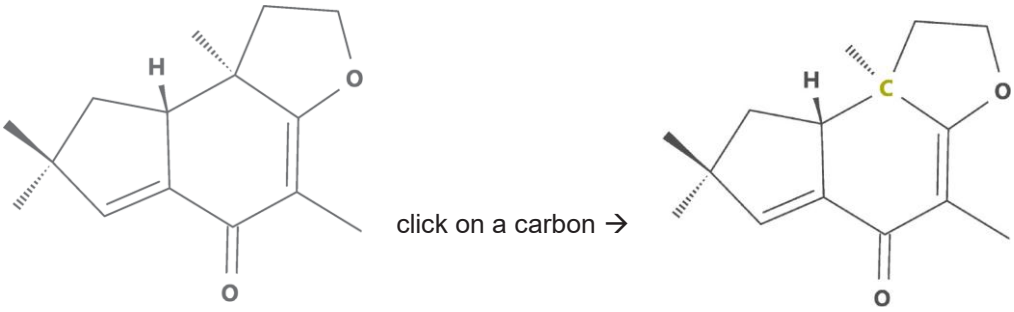
Drawing Tools. undoes the most recent drawing operation. removes or modifies parts of a drawing. deletes an entire drawing (a warning is provided). improves the readability of a drawing by applying a "clean up" procedure. exits sketch mode.

Four icons (in addition to the **Edit** icon) appear on top of the tool palette. These allow "curly arrows" to be added to the sketch to define a transition state, input of experimental proton and ^{13}C chemical shifts and identification of observed HH and CH couplings.

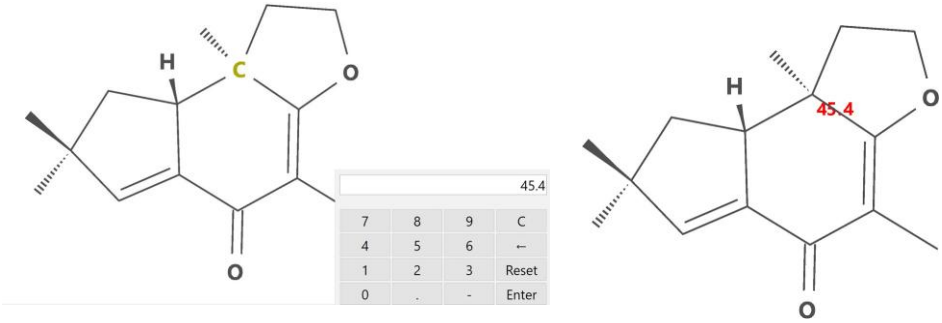
Experimental Data

Proton and ^{13}C Chemical Shifts ( )

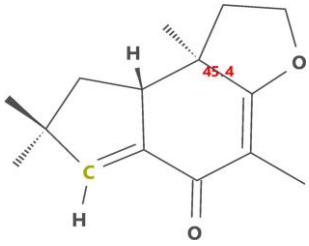
Experimental proton and/or ^{13}C chemical shifts can be added to a 2D drawing. Starting from a 2D sketch (promophyllin A), *click* on above the sketch pad and then on a carbon. If it is a quaternary carbon (as shown below) the only change to the sketch will be that the letter C will appear.



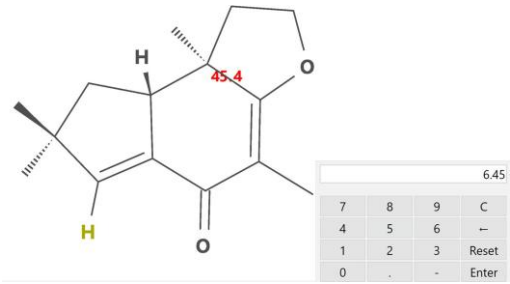
Click on **Edit** to the right of **Expt. Chem. Shift** at the bottom right of the screen and enter a value for the experimental ^{13}C shift using the keypad that results. *Click* on **Enter** and the keypad will be dismissed, and the chemical shift will be attached to the sketch.

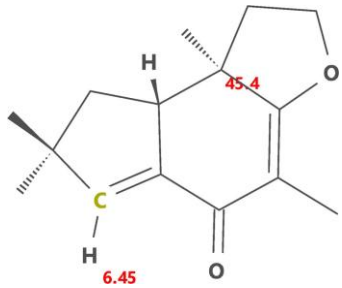


If the selected carbon had one or more attached hydrogens and experimental chemical shifts were to be entered, these will be shown with the selected carbon.



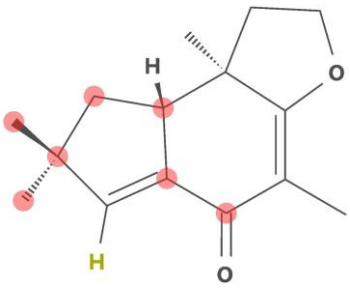
Clicking on one of these hydrogens (only a single hydrogen in the example) results in a keypad into which the experimental proton shift may be entered. *Clicking* on **Enter** again dismisses the keypad and attaches the chemical shift onto the sketch.

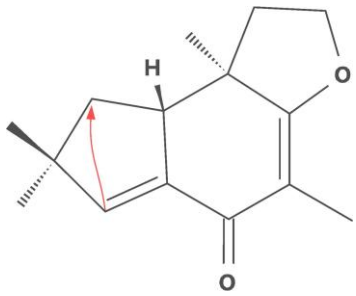




CH Couplings J^{CH}

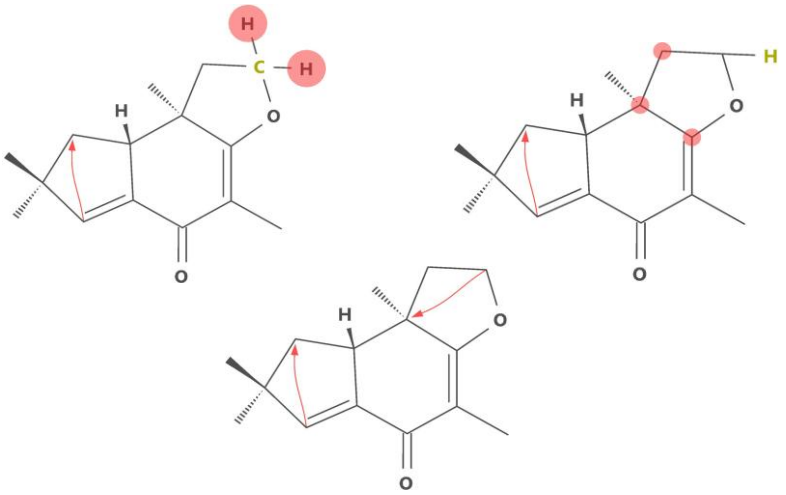
Experimental 2-bond (**H**-C-**C**) and 3-bond (**H**-C-C-**C**) CH "observed" couplings (not numerical values for coupling constants) may be attached to a 2D sketch (with or without experimental chemical shifts attached) by *clicking* on a carbon (H-**C**-C or H-**C**-C-C). If there is only one hydrogen as in this first example, it will be drawn and all carbons which either 2 or 3-bond coupling can occur will be highlighted by translucent red circles. *Clicking* on one of these carbons (in this case three bonds removed)will remove the circles and draw an arrow between the first selected carbon (corresponding to the H in **H**-C-C-C) and the second selected carbon (corresponding to H-C-C-**C**).





The arrow indicating a 3-bond coupling will remain in place.

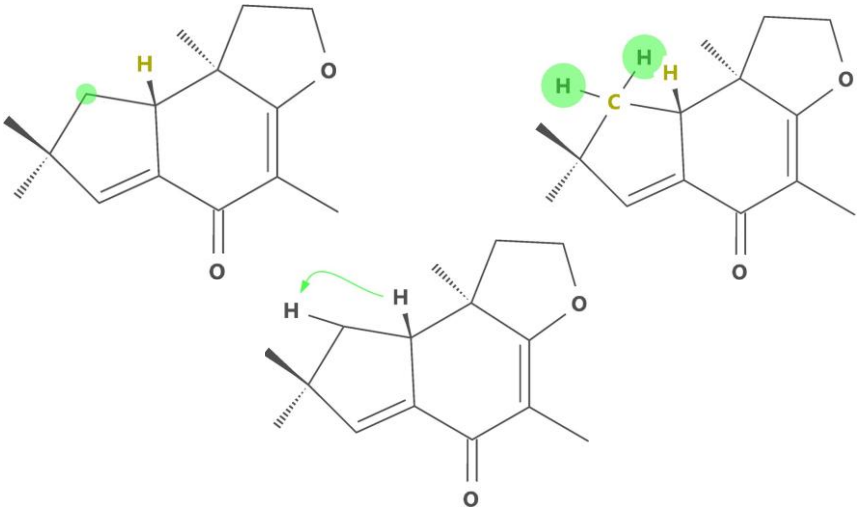
If on the other hand, the initially selected carbon is bonded to more than one hydrogen, an additional click is required to identify which hydrogen before a destination (carbon) is selected.



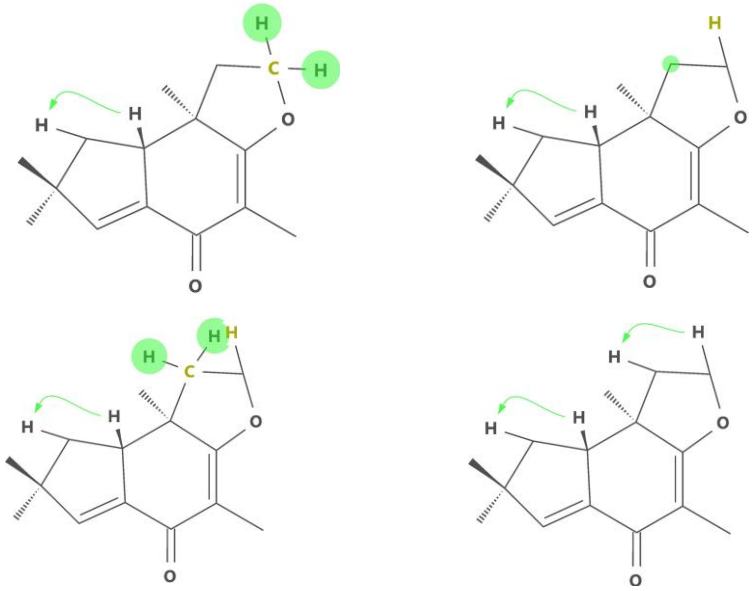
HH Couplings J^{HH}

Experimental 3-bond HH observed couplings (not numerical coupling constant values) (**H**CC**H**) may also be attached to a 2D stretch. There are several cases depending on whether more than one hydrogen is attached to one or both of the carbons and whether or not one or both of the hydrogens are explicit in the 2D sketch.

If one of the hydrogens is explicit, *click* on it following selection of J^{HH}). This results either in one or more carbons being marked with a translucent green circle. Select the relevant carbon (there is only one in this example) and *click* on it to display the hydrogens attached to it (there are two in this example, if there were only one, the selection of the carbon would have eliminated the step). These are also marked by translucent green circles. *Click* on the appropriate hydrogen and the coupling will be displayed by a green arrow.



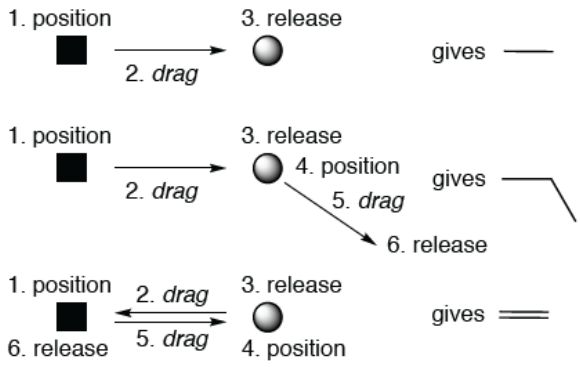
If hydrogens are not explicit, *click* on the carbon to which they are attached. *Click* on the appropriate hydrogen and a coupling arrow will appear.



As before, steps will be eliminated if they are not needed.

Making a Sketch

To start a sketch, first select (*click* on) an atom, group, ring or wildcard icon in the sketch palette and then *double click* in the white portion of the screen (the drawing area). To draw a bond, first *click* on an atom, group, ring or wildcard icon in the sketch palette to designate what is at the end of the bond, then position the cursor over the atom in the drawing area where you want the bond to start, move the cursor while holding down the left button ("drag" the cursor) to the place in the drawing area where you want the bond to end and release the button. Multiple bonds are made by dragging over existing bonds.



To make a bond touch the screen where you want it to start, move one finger to where you want it to end and lift. Replace position by touch, drag by move and release by lift in the previous diagram.

Manipulating a Sketch

To translate the sketch, move the mouse over the screen while holding down the right button. To rotate the sketch (in the plane of the screen), move the mouse up and down while holding down both the left button and **Shift** key. Use the scroll wheel to resize the sketch.

To translate the sketch, move two fingers over the screen. To rotate the sketch in the plane of the screen, "twist" two fingers on the screen. To resize, pinch/spread) two fingers on screen.

Sketch Operations

Add an Atom, Ring, or Carbonyl Group. To add an atom, *click* on that atom's icon. Position the cursor over the atom in the drawing that will connect to the new atom, *drag* it away and release the button. To add a common ring or carbonyl group, *click* on that ring's (group's) icon, position the cursor over the atom in the drawing that will connect to the new ring (group) and *drag* it away and release the button. The carboxylic acid or ester and amide icons contain an arrow that shows which atom in these groups will be connected to the existing drawing. To change the location of this connection point (arrow), *click* on the group's icon until the arrow reaches the desired location.

Add Multiple Bonds. To add a multiple bond, first draw a single bond at the location where the multiple bond is needed, then redraw this line once to make a double bond, and redraw it again to make a triple bond (in other words, position the cursor over one end of the bond, drag to the other end, and release the button).

An alternative way of drawing a multiple bonds is to *double click* (*double tap* on touch-screen devices) on an existing single bond. *Double clicking* a second time gives a triple bond and a third time reverts to a single bond. The only restriction is this cannot be done if a ring has been selected from the palette whereas here *double clicking* results in addition of the ring (ring fusing in the case where the single bond is incorporated into a ring).

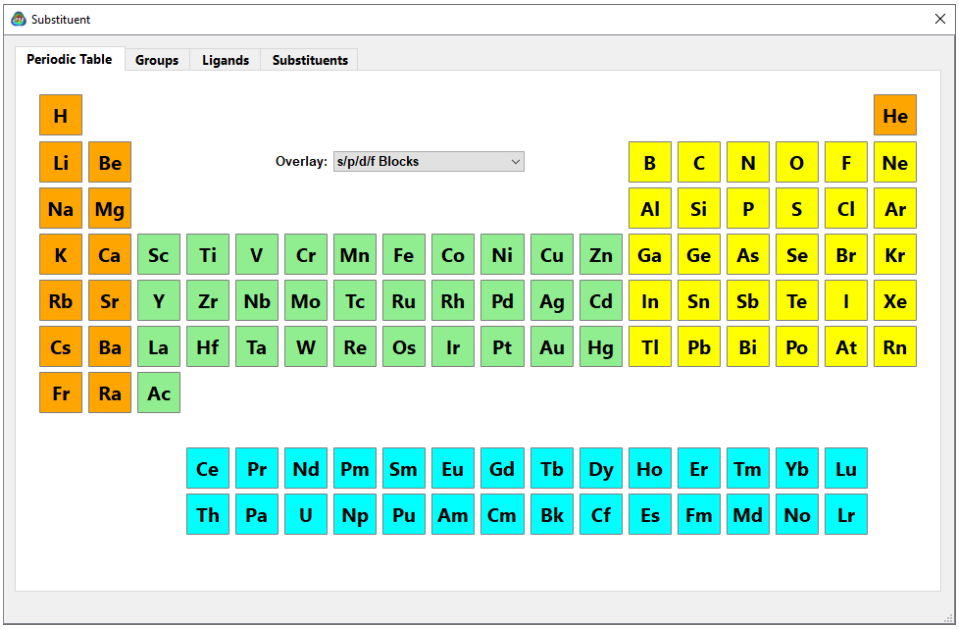
Fuse Rings. *Click* on an icon for the first ring and *double click* (*double tap* for touch-screen devices) on the screen. Next, *click* on the icon for the second ring and double tap the bond that the rings will share. This will create a drawing with a fused bicyclic ring system. Note that the (*cis* or *trans*) stereochemistry of the ring juncture is ill-defined. This is addressed later (**Add Stereochemical Markers**).

This technique can also be used to add rings to existing bonds in any drawing. *Click* on the icon for the ring to be added and *double click* on the bond that will become part of the ring.

Replace an Atom with Another. If a drawing contains atom A where atom B is needed, *click* on the icon for B, then *double click* (*double-tap* for touch-screen devices) on A in the drawing.

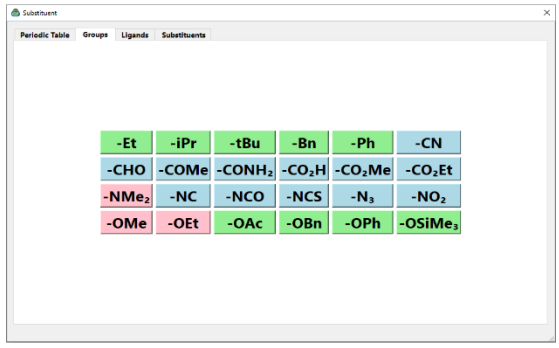
This simplifies the drawing of heterocycles. First, draw an all-carbon ring and then replace specific carbons with heteroatoms.

Access an Element, Functional Group or Ligand from the Library. *Click* on the icon underneath H and B (this icon will initially be labeled with **More**, or the name of the group that was previously selected and will change each time you select a new element or group). A tabbed dialog initially displaying a ***Periodic Table*** *results*.

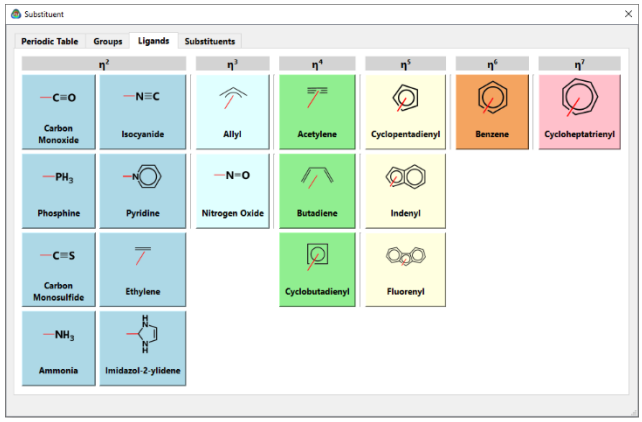


The **Overlay** menu allows the ***Periodic Table*** to be colored based on availability of specific models or basis sets or to separate it into **s**, **p**, **d** and **f** blocks.

If a functional group is desired, *click* on the **Groups** tab and then on a group from the dialog that results. Colors are intended to convey chemical behavior, red represents electron donating, blue represents electron accepting, and green represents relatively neutral groups.



If a ligand is desired, *click* on the **Ligands** tab (above the ***Periodic Table***) and then a ligand from the dialog which results.



Finally, an abridged set of substituents consisting of common R-groups, aromatics, OR (alkoxy), electron-donating, electron-withdrawing, and leaving groups (see **Substituent Builder section** later this chapter):

An element, group, ligand, or substituent defined in this way can be added to a drawing, replaced, or removed using the same drawing techniques used for standard atoms. The element/group can also be redefined. *Double click* on the icon beneath H and B and select the desired element/group.

Add Stereochemical Markers (Dashes and Wedges). 3D information can be added to a drawing by replacing **single** bonds with stereochemical markers: dashes or wedges. The single bond must be drawn before a marker can be added.

To replace a single bond with a marker, *click* either or , then *re-draw* the single bond. To reverse the orientation of the marker, *re- draw* the bond or marker in the opposite direction. One type of marker can be replaced directly by the other. *Click* on the desired marker and then *re-draw* the existing marker. Markers can also be converted back into single bonds. *Click* on , then *double click* on the marker. For best results during 2D-to-3D conversion, all substituent bonds to the ring should be drawn with stereochemical markers.

Spartan'26 provides an alternative way of introducing stereochemical markers either as "new" single bonds or by modifying existing single bonds. Once one of the stereochemical markers (or) has been selected, drawing results in an up or down wedge with the narrow end at the starting point.

Clicking on any other (than or) icon in the tool palette reverts to selection of . Additionally, with a single bond already drawn and with either or selected, the marker may be added by *double clicking* on the bond as an alternative to swiping over the bond. In this case, the "broad" end of the marker (solid or dashed wedge) goes to the terminal atom or group (if one end of the bond is a terminal atom or group). *Double clicking* a second time reverses the direction of the marker and a third time removes it (reverts to an "unmarked" single bond).

Add an *Axial* or *Equatorial* Marker. The orientation of a hydrogen/ substituent on a 6-member ring can be specified by marking **one** ring substituent as either **ax(ial)** or **eq(uatorial)**. If the molecule contains multiple rings, the conformation of each ring can be specified by marking **one** substituent per ring. *Axial* or *equatorial* markers can only be added to stereochemical markers (dashes, wedge) so the bond connecting the substituent to the ring must be drawn with a stereochemical marker first.

To add an *axial* or *equatorial* designation to a stereochemical marker, *click* on either or within the sketch palette, then *hold down* the **SHIFT** key and *double click* on top of the stereochemical marker. This will be designated *axial* and marked with an ***ax***. To specify equatorial, with the **SHIFT** key still depressed, *double-click* again, and the ***ax*** will be replaced with ***eq***. A third *double-click* operation will return to bond to a single bond without stereo-chemical designation.

Note: While it is possible to specify multiple centers as *axial* or *equatorial*, ***Spartan*** only needs one designation to provide the anticipated 3D structure.

Assign Charges and Radical Sites. Formal charges and unpaired electrons can be assigned to individual atoms using charge/unpaired electron markers. To assign a **positive** formal charge to an atom, *click* on and *double click* on the atom in the drawing. To assign a **negative** formal charge or **unpaired electron**, *click* on the charge/ unpaired electron marker until the desired icon appears (tapping the marker rotates it through three possibilities: , , and) and double tap the atom in the drawing.

To replace a charge/unpaired electron marker on an atom with a different marker, *click* on the desired charge/unpaired electron icon and *double click* the marked atom. To remove a marker, *click* on and *double click* on the marker.

Charge/unpaired electron markers play an important role during the conversion of 2D drawings into 3D models in that they determine the number of hydrogens that need to be added to the model (it is usually unnecessary to draw hydrogens unless they are needed to mark stereochemistry).

Hydrogens are not shown in **2D** perspective drawings unless they have been drawn explicitly (exceptions: hydrogens attached to neutral **N**, **O**, **P**, and **S** are shown). When a drawing is converted into a **3D** model, hydrogens are added to the model according to standard bonding rules. A neutral carbon atom is assumed to form four bonds, nitrogen atom three bonds, oxygen atom two bonds, and so on. Analogous logic is used for charged or radical atoms.

Charge and unpaired electron markers are carried over into specification of quantum chemical calculations (**Calculations...** from the **Setup** menu; **Chapter 21**). The total charge is set equal to the sum of the formal charges in the 2D drawing. A model with one unpaired electron is treated as a free radical (drawings that contain more than one unpaired electron may give unanticipated results).

Specify "Open Site" Marker(s). To mark an "open site" for a substructure search on the database, *click* on , locate the atom in the drawing that will carry the open site, and *draw* a bond starting from this location. As many open site markers as desired may be added to the drawing.

To remove an "open site" marker, *click* on , then *double click* on the ? marker in the drawing.

Discussion of database searching is provided in **Chapter 23**.

> Markers can also be placed on a 3D model (**Structure Query** in **Chapter 23**).

Undo the Last Action. *Click* on to return to the drawing as it existed before the last action. Only a single level of undo is presently supported.

Clean Up a Drawing. *Click* on to "clean up" a drawing, that is, to equalize bond lengths, bond angles, and so on. "Clean up" can improve the appearance of a 2D drawing, but not every "clean up" will produce a desirable result. To undo a clean-up operation, *click* on .

Remove an Atom or Bond. *Click* on and then *double click* on the atom or bond. If you *click* an atom, all bonds to that atom will also be removed. Removing a bond, either by *clicking* on an atom or by *clicking* on the bond itself, will also remove terminal atoms, that is, atoms not connected to any other atoms in the drawing will be removed along with the bond.

Remove a Multiple Bond. To remove a multiple bond, *click* on then *double click* on the multiple bond. This reduces the bond order by one. Repeated *double clicks* on a triple bond will reduce the bond order: triple → double → single → no bond.

Remove a Reaction Arrow. To remove a reaction arrow, *click* on and *double click* on the reaction arrow.

Remove a Stereochemical Marker. *Click* on , then double tap the marker. This replaces the marker with a single bond.

Change or Remove an *Axial* or *Equatorial* Label from a Stereochemical Marker.

Tap either or , then *double click* on the stereochemical marker where an *axial* or *equatorial* label appears. This cycles the label (in order) among **ax** (*axial*), **eq** (*equatorial*) and nothing. The stereochemical marker itself will not be modified.

Remove a Charge or Radical Marker. *Click* on and *double click* on the marker.

Remove an "Open Site" Marker. *Click* on and *double click* on the ? marker in the drawing.

Clear the Screen. *Click* on . A warning message will ask you to confirm this operation.

Copy/Paste/Delete Selected. Right *clicking* on screen brings up a menu which allows the sketch to be copied to the clipboard (**Copy**), the contents of the clipboard to be pasted on screen (**Paste**) or a selected portion of the sketch (or the entire sketch) to be deleted (**Delete Selected**).

A 3D structure is obtained from the 2D sketch by *clicking* on .

3D Builders

Spartan provides five different 3D builders for assembling a variety of molecular systems: an organic builder for most organic molecules, an inorganic builder for organic molecules not well represented in terms of an uncharged (non-zwitterionic) valence structure, as well as inorganic and organometallic molecules, builders for polypeptides and polynucleotides and a builder for attaching substituents (lists of chemical groups) to previously-built molecules. The organic and inorganic builders utilize atomic fragments, functional groups and rings (and ligands in the inorganic builder), while the peptide builder uses the set of natural amino acids as building blocks, and the nucleotide builder the set of nucleotide bases.

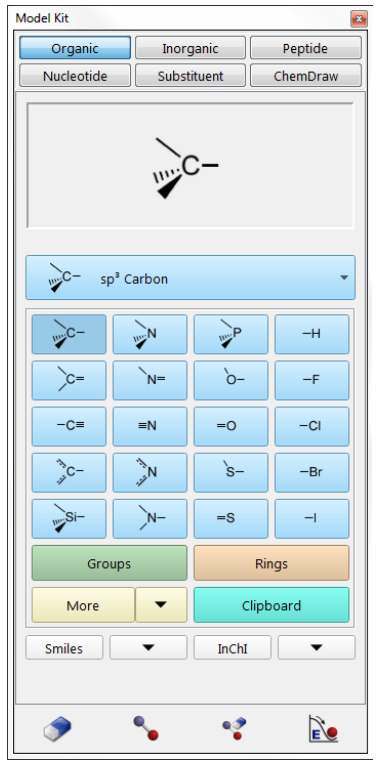
Additionally, two-dimensional molecular structures (drawings) produced using ChemDraw* can also be brought into ***Spartan*** and converted to three- dimensional structures. ***Spartan*** also provides its own 2D sketch capacity (with automatic conversion to 3D structures).

3D molecule construction in ***Spartan*** proceeds much in the same manner as a chemistry student would assemble a structure from a model kit, that is, pieces are taken from the kit one at a time and added sequentially to the molecule under construction. Strategies employed can save time, for example hydrogens are automatically added to all valences left open (there is no need to manually add these). On the other hand, if you wish to use a fragment as a custom building moiety, it is often useful to explicitly add hydrogens to locations you do *not* wish available for attachment.

---

* Seamless access to ChemDraw is available for the Windows version only, turn on the ChemDraw interface from the **Preferences** dialog, **Chapter 25**. All versions are able to read ChemDraw files. ChemDraw is not included with ***Spartan*** and is a commercial 2D drawing program that must be obtained from Revvity https://revvitysignals.com/products/research/chemdraw.

Organic Builder



The organic builder contains a suite of molecule building/editing tools specifically designed to construct the most commonly- encountered organic molecules.

In the center of the builder is a selection of atomic fragments, which from left to right and then top to bottom correspond to:

C(sp^3)	N(sp^3)	P(sp^3)	H
C(sp^2)	N(sp^2)	O(sp^3)	F
C(sp)	N(sp)	O(sp^2)	Cl
C(aromatic)	N(aromatic)	S(sp^3)	Br
Si(sp^3)	N(planar)	S(sp^2)	I

A fragment is chosen by *clicking* on its icon, which is then displayed at the top of the builder. Once selected, the fragment may be used to initiate building, to add alongside of an existing structure or appended onto an existing structure. To initiate building, *double-click** anywhere on screen.

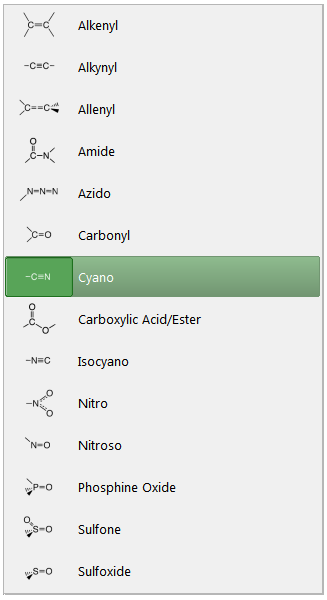
To add alongside of an existing structure, hold down the **Insert** key (**option** key on Mac), and then *click* anywhere on screen or *double click* in a blank area on screen. To bond to an existing structure, *click* on a free valence (***not an atom***). (Free valences are colored yellow on the selected molecule.) Bond type in the case of atomic fragments with multiple bond types, for example, sp^2 carbon, depends on the nature of the free valence selected.

> While only H, C, N, O, F, Si, P, S, Cl, Br and I are available from the organic builder, other elements may be substituted using atom replacement feature available in the inorganic builder (see **General Molecule Building Functionality** later in this chapter). Similarly, bond types may be altered in the inorganic builder. The latter may be necessary in order to alter bonding in structures retrieved from the Cambridge Structural Database or ligands obtained from PDB files. Atom and bond types may also be altered using the **Atom** and **Bond Properties** dialogs, respectively (accessible from **Properties** under the **Display** menu; **Chapter 22**). Note that bond type (or the fact that there is a bond) is not considered in quantum chemical calculations. Only atom identities and coordinates are required. Bond types are, however, necessary for molecular mechanics calculations.

* The default 3D building behavior requires a double-click operation to initiate building. This follows the 2D building behavior. To switch to a single-click behavior, turn off the "Double- Click Start" preference (see **Options** menu, **Chapter 25**).

Menus inside the organic model kit provide access to a number of pre-built fragments corresponding to functional groups (**Groups**) and rings (**Rings**), and to additional libraries of rings (as well as any user-defined structures) stored in ***Spartan***'s file system (**More**). The builder also accesses the clipboard (**Clipboard**).

Groups



Clicking on **Groups** brings up a menu of groups and displays an icon of one group in a box at the top of the builder.

Once selected from the menu, a group may be used to initiate building, to add alongside of an existing structure on screen, or to add to an existing structure.

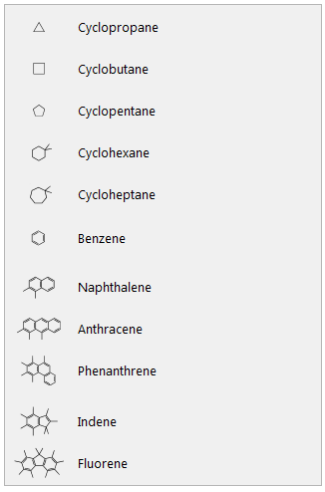
The amide and carboxylic acid/ester groups have more than one different free valence. The free valence that is to be used is marked with a gold • (in the icon at the top of the builder). The marked position circulates among the possible positions with repeated *clicking* on the icon.

Pre-existing and new "custom" groups are available options (if you don't see your favorite groups).

Rings

Clicking on **Rings** brings up a menu of hydrocarbon rings, and displays an icon of one ring in a box at the top of the builder.

Once selected from the menu, a ring may be used to initiate building, to add alongside of an existing structure on screen, or to add to an existing structure.



Cyclohexane, cycloheptane, naphthalene, anthracene, phenanthrene, indene and fluorene have more than one unique free valence. The one that is to be used is marked with a gold • (in the icon). The marked position circulates among the available positions with repeated *clicking* on the icon. Selection of an *axial* or *equatorial* free valence in cyclohexane and cycloheptane is indicated by the label **ax** or **eq** appearing alongside the icon.

All rings in this menu are hydrocarbons, but heteroatoms may be substituted (see **General Molecule Building Functionality** later in this chapter).

More

This provides access to a broader selection of rings as well as to access user-defined entities (rings, groups, ligands, etc.). Upon initial entry, *clicking* on **More** or on ▾ to its right brings up a file browser that has been set to point toward a directory* containing a library of several ***Spartan*** documents of common rings. A list is included below, but note that users can also, very easily build up their own custom libraries of custom rings or more obscure functional groups – the ▾ to the right of **More** will access the file system; the user may point to any user read-writable location to access custom building fragments.

nitrogen heterocycles	**saturated nitrogen rings oxygen**
heterocycles	**saturated oxygen rings sulfur**
heterocycles	**saturated sulfur rings**
mixed heterocycles	**saturated mixed rings**

Clicking on a document followed by *clicking* on **Open** or *double clicking* on a document fills the menu to the right of **More**. Menu entries are selected in the usual way. In response, a ball-and-wire model of the selected ring will appear in a box at the top of the model kit. This may be manipulated (rotated, translated, zoomed) using the usual mouse/keyboard commands (you need to position the cursor inside the box) or with the usual one and two-finger touch commands. The ring may be used to initiate building, to add alongside of an existing structure, or to add to an existing structure. In the latter case, the attachment point (on the ring in the window) needs to be identified by *clicking* on the appropriate free valence.

Three other collections (ligands, chelates and high-coordination fragments) appear in the selected directory. They are intended for use with the inorganic builder as discussed in the next section. User-defined libraries may also be accessed (they are normal ***Spartan*** documents).

Clipboard

Clicking on **Clipboard** accesses the clipboard. A ball-and-wire model of whatever is on the clipboard is displayed in a box at the top of the builder. This may be manipulated using the usual mouse/keyboard commands (you need to position the cursor inside the box or with the usual one and two-finger touch commands). Once selected, the contents of the clipboard may be used to initiate building, to add alongside of an existing structure, or to add to an existing structure. In the latter case, the attachment point needs to be identified by *clicking* on the appropriate free valence.

An empty clipboard is signaled by:

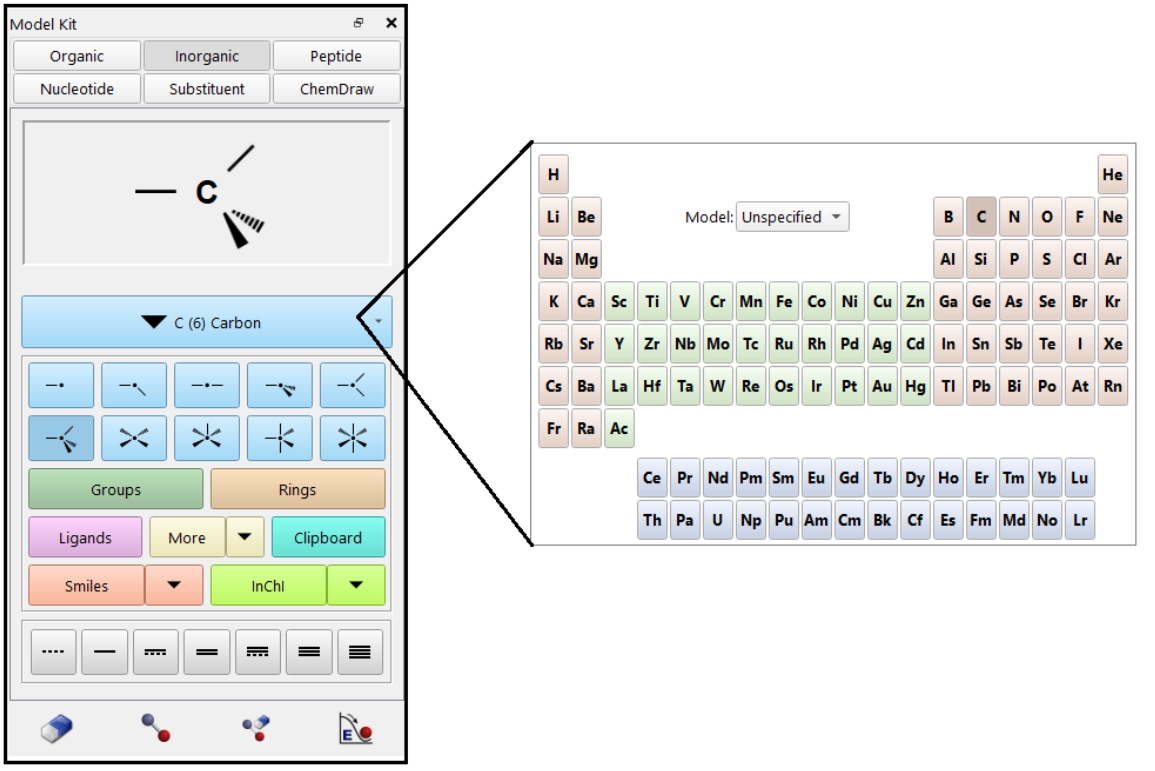
* For Windows, the library directory is located in C:\ProgramFiles\Wavefunction\Spartan. For Linux, the location is /user/Spartan/Library. For Macintosh, the library folder is on the disc image and needs to be copied to a location on the hard drive. Once populated (after first time access) *clicking* on **More** will display the previously selected file in the box at the top of the model kit, while *clicking* on ▾ will bring up a file browser to allow a new selection to be made.

Smiles / InChI

Initially *clicking* on **Smiles** or **InChI** or on ▾ to its right of each will bring up an input dialog where the corresponding string can be entered either manually, or more likely by using Copy/ Paste. Once the string is entered, *click* the **OK** button to release the dialog. A 3D rendering of the molecule (or fragment) will be displayed at the top of the model kit. The attachment point is indicated with a yellow marker. To insert, rather than attach, *double-click* in the blank space of the 3D interface to the left of the model kit. Subsequent *clicking* on the **Smiles** or **InChI** button will return focus to the last entered string in the 3D area at the top of the model kit. To input a new Smiles or InChi string, click the downward triangle immediately to the right of the string type you wish to input. Access to the **Smiles** or **InChI** input dialog is available from both the organic and inorganic builders.

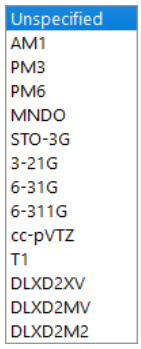
Inorganic Builder

Spartan's inorganic builder allows construction of a much wider class of molecules (including inorganic and organometallic species) than possible with the organic builder. Structures that violate conventional bonding rules may be constructed, as this builder purposefully provides no valence checking. The inorganic builder is reached by selecting **Inorganic** from among the tabs at the top of the builder.*



* Tabs may require too much vertical space on computers or tablets with very small screens. Alternative **Builder Selection Methods** are available in the **Miscellaneous** tab (**Preferences** from the **Options** menu; **Chapter 25**).

Atoms may be selected by *clicking* on the selection bar near the center of the builder. This brings up a full *Periodic Table*. Main- group elements are colored red, transition metals are colored green and lanthanides and actinides are colored blue. The **Overlay** menu inside the *Periodic Table* contains a selection of theoretical models (basis sets, semi-empirical models, and neural network models).



Selecting an entry from this menu leads to re-coloring of the *Periodic Table*. A light green color is used to indicate elements for which the selected model may be used.* Immediately below is a selection of atomic hybrids.

Selection of atom type is affected by *clicking* on the appropriate element in the *Periodic Table*. The entry will be highlighted. Selection of an atomic hybrid follows by *clicking* on the appropriate icon which will then be highlighted.** This combination (atom type + atomic hybrid) may be used to initiate building, to add alongside of an existing structure or to bond to an existing molecular fragment. To initiate building, *double click* anywhere on screen. To add alongside of an existing structure, *double click* in a blank area on screen. To bond to an existing fragment, *click* on the appropriate free valence.

Two of the hybrids (trigonal bipyramidal and square-based pyramidal) may bond either *axially* or *equatorially*. Selection of the appropriate bonding point, marked by a •, is affected by repeatedly *clicking* on the icon; the bonding point alternates between the two sites.

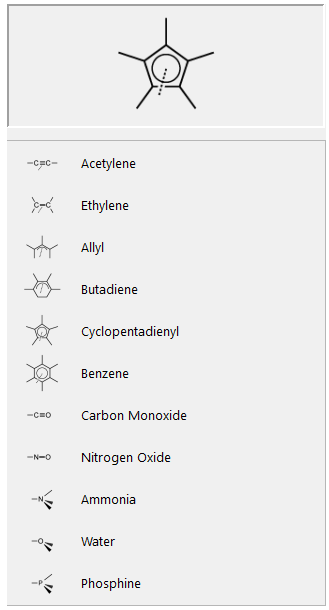
Atoms are connected with whatever bond type (partial single, single, aromatic, double, triple or quadruple) is selected from the bottom of the builder. A bond type may be changed by first selecting a type and then *double clicking* on the bond. Bond types have no impact on quantum chemical calculations but do affect molecular mechanics calculations which reference a Lewis structure (including minimization in the builder; see discussion later in this chapter). No valence checking is performed in the inorganic builder.

---

* The existence of either an all-electron basis set or a pseudopotential is sufficient to classify a model as available for a particular element. Note that while molecular mechanics models are available for all elements, they have been carefully parameterized for only a relatively few of these. Use caution in interpreting molecular mechanics geometries for molecules incorporating heavy main-group elements, transition metals, lanthanides and actinides.

** Hybrids for a number of high-coordination centers are available as a library reachable from **More** (see discussion under **Organic Builder**).

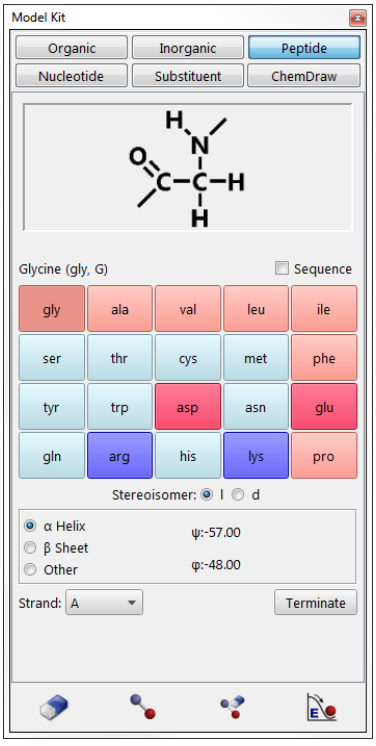
Menus under **Groups**, **Rings** and **More** are the same as those described for the organic builder as is the **Clipboard**. Fragment collection provided:



Ligands

This provides access to a number of pre- built ligands, useful in the construction of inorganic and organometallic molecules. Its operation is analogous to that for the **Groups** and **Rings** menus. *Clicking* on **Ligands** brings up a menu of available ligands, and results in an icon of one ligand from this menu being displayed in a box at the top of the builder. A ligand may be used to initiate building or to add alongside or to an existing structure. Additional ligands are accessible from **More** (see previous discussion). Ligands may also be built with the aid of ligand points (**Define Ligand Point** in the **Geometry** menu; **Chapter 19**).

Peptide Builder



This builder is not intended to be used for constructing proteins (although this is possible). Rather, it is primarily intended to build idealized helix and sheet structures. Protein structures are best entered from the Protein Data Bank (see **Access PDB Online...** under the **File** menu; **Chapter 16**). A builder for construction of polypeptides is reached by selecting **Peptide** from among the tabs at the top of the builder.

At the middle of the peptide builder are icons designating the amino acids (specified by their three-letter codes). An amino acid is selected by *clicking* on its three-letter code, following which either an icon of the amino acid is displayed in the box at the top of the model kit, or the three-letter code for the amino acid is appended to the sequence of codes in the box. Amino acids replace atoms, functional groups, rings and ligands as building blocks. Because the other building blocks are missing, modifications of peptides, aside from modifications in sequence and in overall conformation, should be carried out using the organic or inorganic builders.

There are two different modes of operation: single amino acid mode and polypeptide mode. The former is used to initiate building with a single amino acid, to add a single amino acid alongside of an existing structure or to add a single amino acid to an existing structure, while the latter is used to construct amino acid sequences (polypeptides). **Sequence** off (unchecked) corresponds to single amino acid mode, and on (checked) corresponds to polypeptide mode.

Peptide construction (**Sequence** on) is accomplished in three steps:

Specifying Amino Acid Sequence

This is accomplished by *clicking* in the desired order on the amino acid icons. Building occurs from the N end to the C end of the peptide. In response to each selection, the three-letter code is appended to the sequence of codes in the box at the top of the builder. The stereochemical configuration of the amino acid is by default the l configuration; this may be changed to the d configuration prior to selection of the amino acid, by *checking* **d** to the right of **stereoisomer** in the builder. (It may be changed back to l by *checking* **l**). d amino acids are indicated by **.d** following the code in the box.

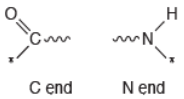
The sequence may be altered by changing the text in the box. Existing amino acid codes may be deleted or changed or new codes can be added. The entire sequence may be specified in this way if desired. Specification of a non-existent code will result in an error message. The sequence can be cleared by *clicking* on the **Clear** button.

Specifying Macroscopic Structure

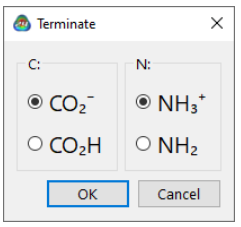
Once sequencing is complete, macroscopic structure (ψ and φ angles), is specified by *clicking* on one of α **Helix**, β **Sheet** or **Other**. In the case of the first two, preset angle values are displayed on the right. In the case of specification of **Other**, boxes appear, into which the desired dihedral angles may be entered.

Terminate

The peptide is not yet terminated, and the two ends are still set up for addition of further amino acids.



where the * indicates a free valence. *Clicking* on **Terminate** at the bottom of the model kit leads to the **Terminate** dialog.



C and N terminating groups may be selected by *clicking* on the desired terminal group. *Clicking* on **OK** removes the dialog and terminates the polypeptide.

Clicking on **Cancel** or removes the dialog but does not terminate the polypeptide.

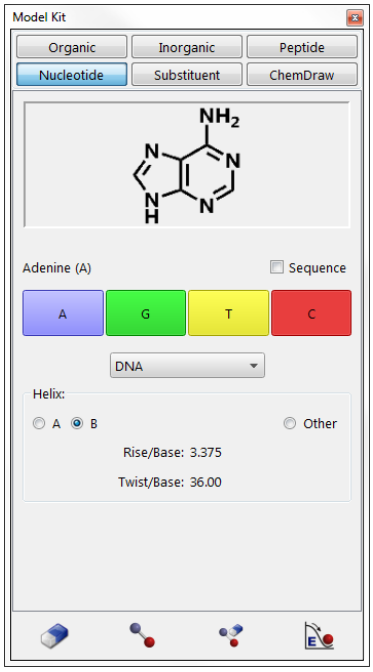
The peptide (or single amino acid) may now be used either to initiate building, by *double clicking* anywhere on screen or added alongside of an existing structure, by holding down the **Insert** key (**option** key on Mac) and *clicking* anywhere on screen, or by *double clicking* in a blank area on screen. If unterminated, it may join an existing structure by *clicking* on a free valence. In the latter case, attachment is made from the N end, unless the free valence corresponds to an unterminated peptide fragment, in which case the appropriate end required to make an amide bond is used.

Nucleotide Builder

Spartan provides a builder for construction of polynucleotides. It is reached by selecting **Nucleotide** from among the tabs at the top of the builder.

At the middle of the builder is a menu designating the type of polynucleotide:

DNA	RNA
DNA (single strand)	RNA (double strand)
DNA-RNA	RNA-DNA



Immediately below this menu are icons, designating the nucleotide bases. Selection of DNA, DNA (single strand) or DNA-RNA from the menu leads to one set of icons.

Selection of RNA, RNA (double strand) or RNA-DNA leads to a second set, the only difference is that uracil (U) has been substituted for thymine (T).

A nucleotide base is selected by *clicking* on its letter, following which either an icon of the base is displayed in the box at the top of the builder, or the letter for the base is appended to the sequence of letters in the box. Nucleotide bases replace atomic fragments, functional groups, rings and ligands as the building blocks in the nucleotide builder. Because these other building blocks are missing, modifications of nucleotides, aside from modifications in sequence and helical structure, need to be carried out using either the organic or inorganic builders.

There are two different modes of operation: single base mode and polynucleotide mode. The former is used to place a single base or base pair on screen, to add a single base or base pair alongside of an existing structure, or to add a single base or base pair to an existing structure, while the latter is used to construct strands of DNA or RNA (or mixed strands). **Sequence** off (unchecked) corresponds to single base (base pair) mode and on (checked) corresponds to polynucleotide mode.

Polynucleotide construction (**Sequence** on) is accomplished in two steps:

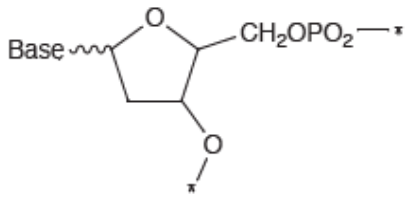
Specifying Base Sequence

This is accomplished by *clicking* in order on the base codes. In response to each selection, the letter code is appended to the sequence of codes in the box at the top of the model kit. The sequence may be altered by editing the contents of the box. Existing base codes may be deleted or changed or new codes added. The entire sequence can be specified in this way if desired. The sequence may be cleared by *clicking* on the **Clear** button.

Specifying Helical Structure

Once sequencing is complete, a helical structure may be specified by *clicking* on **A** or **B**. These correspond to A and B helices, respectively. Selecting **Other** allows user modification of the rise (in Å) per base (**Rise/Base**) and twist (in degrees) per base (**Twist/Base**).

Note that the polynucleotide is not yet terminated, and the two ends are still set up for addition of further bases or base pairs.

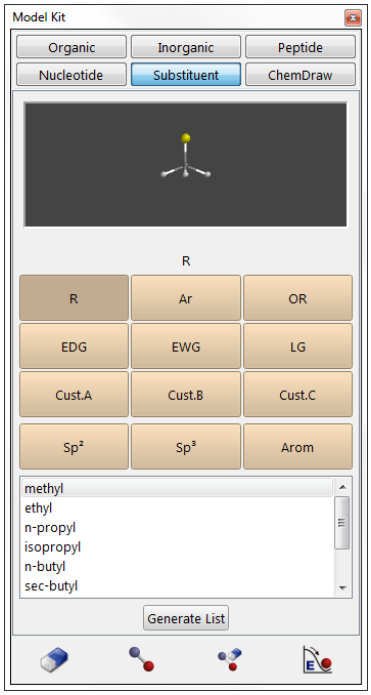


* indicates a free valence. Hydrogens occupy all free valences (except the *'ed positions at the two ends of the chain).

The polynucleotide (or single base pair) may now be used to either initiate building, by *double-clicking* anywhere on screen, added alongside of an existing structure, by first holding down the **Insert** key (**option** key on Mac) and *clicking* anywhere on screen, or *double-clicking* a blank area on screen, or joined onto an existing structure by *clicking* on a free valence. In the latter case, attachment is made from the phosphate end.

Substituent Builder

The substituent builder designates one or more locations for libraries of substituents to be added to a molecule. Additionally, it designates one or more carbon hybrids for replacement by their heteroatom equivalents, for example, sp^3 carbon by sp^3 nitrogen. There are two uses for the **Substituent** builder. The first is to facilitate assembling lists of molecules that differ only in substitution and/or heteroatom. The second application is to use molecules incorporating one or more substituent libraries as queries in searches of the Spartan Spectra and Properties Database (see **Databases** under the **Search** menu; **Chapter 23**) or as elements of reaction energy calculations (see **Reactions** under the **Display** menu; **Chapter 22**). Here, the search or reaction energy calculation is not confined to a single molecule but rather is open to a set of molecules that are related by substitution. Further discussion is provided in the appropriate sections of **Chapters 22** and **23**.



The substituent builder is accessed by selecting **Substituent** from among the tabs at the top of the model kit.

At the center of the builder are four rows of buttons. Buttons in the first two rows provide access to predefined substituent libraries: **R** (alkyl groups), **Ar** (aromatic groups), **OR** (alkoxy groups), **EDG** (electron donor groups), **EWG** (electron withdrawing groups) and **LG** (leaving groups). Buttons in the third row allow access to user-defined substituent libraries (**Cust. A**, **Cust. B**, **Cust. C**), while those in the last row permit substitution of heteroatom hybrids for sp^3, sp^2 and aromatic carbons incorporated in a molecule.

Immediately below the buttons, a dialog lists the entries in the individual libraries (the three custom libraries should be blank upon initial entry). As entries are selected, they will appear in the window at the top of the builder. Predefined libraries may not be altered and individual entries may not be renamed, but existing library entries in custom libraries may be renamed or deleted and new entries added. Modifications to existing substituent libraries need to be performed ***following attachment*** using the **Substituent Properties** dialog (accessible from **Properties** under the **Display** menu; **Chapter 22**).

Creating lists of substituted molecules starting from an unsubstituted molecule occurs in three stages: library preparation, attachment and (optionally) list generation.

Library Preparation

Six substituent libraries have been provided.

> **R** (alkyl group): methyl, ethyl, *n*-propyl, isopropyl, *n*-butyl, *sec*-butyl, isobutyl, *t*-butyl, *t*-amyl
>
> **Ar** (aromatic ring): phenyl, p-tolyl, p-nitrophenyl, 1-naphthyl, 2-naphthyl, 2-pyridyl, 3-pyridyl, 4-pyridyl
>
> **OR** (alkoxy group): methoxy, ethoxy, *n*-propoxy, isopropoxy, *t*-butoxy
>
> **EDG** (electron donor group): hydroxy, methoxy, ethoxy, phenoxy, amino, dimethylamino, mercapto, methylthio, phenylthio
>
> **EWG** (electron withdrawing group): cyano, nitro, trifluoromethyl, carboxy, formyl, methoxycarbonyl, ethoxycarbonyl, methyl sulfonyl, methyl sulfoxyl
>
> **LG** (leaving group): bromo, iodo, tosylate, triflate, mesylate

The contents of an existing library are displayed in the scroll box at the bottom of the builder by *clicking* the appropriate button. A ball-and-wire model of a substituent may be seen at the top of the builder by *clicking* on its name in the scroll box. The model may be manipulated in the usual way. The attachment point is indicated by a gold ball surrounding a free valence.

To create a custom substituent library, first *click* on the **Cust. A**, **Cust. B** or **Cust. C** buttons inside the builder. The scroll box is empty. If a substituent (set of substituents) has previously been copied to the clipboard, it may be pasted into the library by *right clicking* inside the scroll box and selecting **Paste** from the menu that appears. Alternatively, a substituent (set of substituents) may be appended to the document (including an empty document) by *right clicking* inside the scroll box, selecting **Append** from the menu that results and selecting a ***Spartan*** document comprising one or more molecules from the browser. Finally, any ***Spartan*** document may be *dragged* onto the scroll box from the file system. As many substituents (sets of substituents) as desired may be pasted, appended and/or *dragged* into a library. To remove a substituent from the library, *right click* on its name in the scroll box and select **Delete** from the menu that appears. **Ctrl** (**command** for Mac) and **Shift** keys are used for multi-select or selection over a range.

Substituents need to be specified in terms of both their structure and their connection point. One way to specify the latter is to replace all free valences ***except the one that will serve as the connection point*** by hydrogens. For example, to designate a 2-pyridyl substituent, add hydrogens (–H from the organic builder) to four of the five free valences in pyridine, leaving only one free valence in the 2-position available. Alternatively, a connection point can be specified by *clicking* on the appropriate position as the substituent displayed at the top of the builder.

The three buttons in the last row designate heteroatom equivalents of carbon-based hybrids.

sp^2: carbon (C=), nitrogen (N=)

sp^3: carbon (C–), nitrogen (N), oxygen (O–)

Arom: carbon (C–), nitrogen (N)

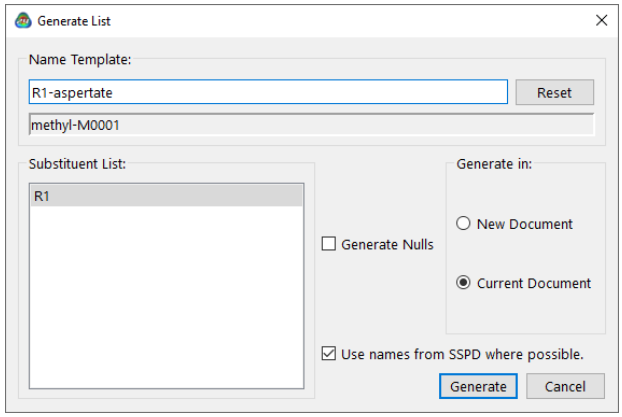
Attachment

Following selection, a substituent library may be used to initiate building, to add alongside of an existing structure or to attach to an existing structure. The model is augmented by an icon containing the library name from the button, for example **EWG**, followed by a number. As many substituent libraries as desired may be attached to a molecule. Each library is given a unique label that can be changed prior to list generation using the **Substituent Properties** dialog. A substituent library may be removed from the molecule by selecting **Delete** from the **Build** menu or *clicking* on ( ) and then *clicking* on the substituent icon.

A document containing one or more molecules with attached substituent libraries may be saved prior to list generation just as any other ***Spartan*** document. Any queries (bond distances, angles, etc.) and calculations refer to the unsubstituted molecule, that is, hydrogens replacing substituent markers. Database searches on the Spartan Spectra and Properties Database (SSPD) and reaction energy calculations based on entries in SSPD pertain to all possible substituted molecules.

List Generation

Generation of a list of substituted molecules follows specification and (optionally) editing of one or more substituent libraries and attachment of each of these libraries at one or more positions in a molecule. Duplicates are eliminated. *Clicking* on **Generate List** at the bottom of the builder, leads to the **Generate List** dialog:



A box at the top of the dialog allows the user to specify (or modify) a **Name Template**, from which the names of substituted molecules will be derived. Upon entry, this takes the form: Ax-By-Cz-molecule name, where A, B, C are the names of the substituent libraries (**R**, **EWG**, etc.), x, y, z are integers and molecule name is the name of the unsubstituted molecule if and as it appears in the SSPD. If the unsubstituted molecule is not in the database, then the box below **Name-Template** will be blank. If the substituted molecule is in SSPD, its name in the database may be used (instead of the name generated from the template) by *checking* the box to the left of **Use names from SSPD where possible**.

Nulls are molecules with less than the full complement of substituents, for example, monosubstituted molecules in the case where two different positions have been substituted. These will not be included in the list of substituted molecules unless **Generate Nulls** is checked.

Note, that the conformations of the substituted molecules have not been subjected to energy minimization to relieve unfavorable crowding interactions. Therefore, the list entries should be examined, conformations adjusted as appropriate and minimized with molecular mechanics prior to using them for quantum chemical calculations.

The list may either be written to a **New Document** or appended to the end of the **Current Document**.

Accessing ChemDraw (Windows Only)*

The ChemDraw program may be seamlessly accessed inside of ***Spartan***** allowing chemical drawings to be produced in a familiar environment and then brought over as 3D structures. The conversion is unambiguous as long as all stereochemical cues are in place in the 2D drawing. Note that the conformation implied by the 2D drawing may be ambiguous and need adjusting.

---

* ChemDraw files (.cdx) may be read with all versions of ***Spartan***.

** ChemDraw is not provided with ***Spartan*** but must be obtained from Revvity, via: https://revvitysignals.com/products/research/chemdraw

To access ***ChemDraw***, select **ChemDraw** from the menu at the top of the Model Kit, and *click* on **New** at the bottom of the panel that results. ***ChemDraw*** will open.

Build as usual in ***ChemDraw***. To return to ***Spartan***, close ***ChemDraw***. A 2D drawing will appear at the center of the panel and a manipulable 3D structure will appear at the top of the panel. *Double-clicking* on screen will move the 2D ChemDraw contents and convert this into a 3D structure in ***Spartan***'s main window.

General Molecule Building Functionality

Additional capabilities are available with **Edit Build** selected:

Multiple Fragments

Multiple fragments may result either from bond breakage (see **Break Bond** later in this chapter) or from use of the **Insert** key (**option** key on Mac), or *double clicking* in a blank area on screen. A fragment is selected by *clicking* on it, following which the associated free valences are colored gold (free valences for any non-selected fragments are colored white). All builder functions apply to the selected fragment only. Rotation and translation also apply to the entire set of fragments, but may be made to apply to the selected fragment by holding down the **Ctrl** key while carrying out these operations.

Fragments may be attached using **Make Bond** (see discussion later in this chapter).

Rotate/Stretch Bonds

In addition to molecule rotation, translation and scaling, the mouse is used to rotate about and stretch bonds not incorporated into rings. This is accomplished via the following sequence of operations:

(i) *Clicking* on the bond which is then marked by a red cylindrical arrow. (The bond connecting the last atom, group, ring or substituent added to the molecule is automatically selected.)

(ii) Simultaneously holding down the **Alt** key (**option** key on Mac) and the left mouse button while *dragging* the mouse up and down, for bond rotation, or the **Alt** (**option**) key and the right mouse button for bond-stretching. Bond rotation (only) also follows from moving the cursor up and down inside the shaded area at the far left of the screen while holding down the left button.

Replace Atom/Fragment

Another function of the mouse is atom replacement. This behaves differently in the organic and inorganic builders. *Double clicking* on an atom (**not a free valence**) while an atomic fragment in the organic builder is highlighted, replaces the atom by the selected fragment. Free valences are adjusted to accommodate the replacement, for example, replacement of sp^3 carbon by sp^3 oxygen results in two free valences being removed. Atom replacements

that violate valence rules or that would disrupt substituents are not permitted. *Double clicking* on an atom (***not a free valence***) while an element in the *Periodic Table* from the inorganic builder is selected, replaces the atom by the selected element, that is, changes the atomic number. No changes in the number or arrangement of free valences is made, and no checking is done. Atom replacement is not available in the peptide, nucleotide and substituent model builders.

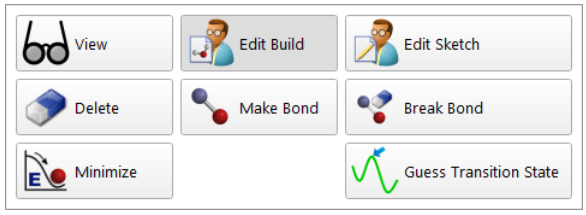
Invert Chirality

In the **Edit Build** mode, *double clicking* on a chiral atom with the **Ctrl** key (**command** key on Mac) depressed inverts the chirality of the atom (R→S or S→R). *Double clicking* on any atom with both **Ctrl** (**command** key on Mac) and **Shift** keys depressed inverts the absolute configuration of the molecule.

> Replace *click* with *tap* for multiple fragment selections. Replace *double click* with *double tap* for atom/fragment replacement and chirality inversion. One finger movement up and down the shaded area at the left of the screen results in rotation about the marked bond.

Building/Editing Menu Functions

Molecule building/editing functions are found under the **Build** menu.



Icons for **Delete**, **Make Bond**, **Break Bond** and **Minimize** are included at the bottom of the model kit. They may also be added to the set of icons below the menus at the top of the interface.

View

This exits build mode, and removes the model kit from the screen.

> Initial entry into the 3D builder is by way of **New Build** or **Build New Molecule** under the **File** menu (**Chapter 16**). **Edit Build**, **Delete**, **Make Bond**, **Break Bond** and **Minimize** are for modifying existing structures.

Edit

In addition to the capabilities discussed under **General Molecule Building Functionality**, this allows access to the libraries of atomic fragments, groups, rings, ligands and substituents, as well as the file system and the clipboard. *Clicking* on any buttons or menus in the organic, inorganic, peptide, nucleotide or substituent builders, leads to **Edit** when in **3D**. (If a builder is not already on screen, selection brings up the last-accessed builder.) A fragment may be used to initiate building by *double clicking* anywhere on screen, to add alongside an existing structure on screen by holding down the **Insert** key (**option** key on Mac) or by *double clicking* in a blank area on screen or be added to an existing structure by *clicking* on the appropriate free valence. Fragment addition can be terminated by selection of any other function.

Edit Sketch

This allows a **2D** sketch to be modified in the 2D sketcher after it has been converted to a 3D structure. Modifications to molecules brought into 3D can be passed back to the 2D sketch builder, essentially allowing for structure modifications in either 2D, 3D, or both.

Delete

This allows atom, free valence and substituent library removal from a structure.
Selection leads to a message at the bottom left of the screen.

Select object to delete.

Subsequent *clicking* on an atom, free valence or substituent library results in its deletion. Deletion of an atom results in deletion of all of its associated free valences. Free valences for any atoms to which the deleted atom was previously connected are restored. Note that atom deletion may result in one or more detached fragments. *Clicking* inside a selection box results in deletion of everything inside the box. Selection of **Delete** does not bring up a model kit nor does it remove a model kit that is present on screen. In 3D, the default settings relegate **Delete** to a one-time operation. Upon deleting, whatever mode was previously selected is restored. This may be overridden by *checking* **Persistent Delete** in the **Settings** tab (**Preferences** under the **Options** menu; **Chapter 24**). Persistent mode is also enabled by *double clicking* on instead of single *clicking*.

> **Delete** is also used to delete points and planes, CFD's and pharmacophore elements as well as spectra and plots (or individual curves that make up plots). Be careful!

Deletion may also be accomplished by holding down on the **Delete** key while *clicking* on the item to be deleted. This mode allows multiple deletions.

Make Bond

This allows bonds to be drawn between free valences and/or atoms. Selection leads to a message at the bottom left of the screen:

Select two free valences.

Clicking on two free valences (on different atoms) will cause these atoms to be linked by a single bond. Alternatively, *double clicking* on each of two atoms will bond them, and *clicking* on a free valence on one atom and *double clicking* on a different atom will bond the two atoms. Note that available free valences are consumed as a result of bond formation, irrespective of whether free valences or atoms are selected.* If the selected atoms are already bonded, this will result in the bond order being increased by one, that is, single → double, double → triple. Selection of **Make Bond** does not bring up a model kit nor does it remove a model kit that is already present on screen. **Make Bond** may be terminated by selection of any other function.

Break Bond

This allows breaking an existing bond resulting in free valences. Selection leads to a message at the bottom left of the screen:

Select bond to break.

Clicking on a bond breaks it and restores free valences. Note that bond breaking may result in detached fragments. Selection of **Break Bond** does not bring up a model kit nor does it remove a model kit that is present on screen. **Break Bond** may be terminated by selection of any other function.

Replace *click* with *tap* and *double click* with *double tap* for delete, make bond and break bond operations.

Minimize

This uses MMFF molecular mechanics to refine the geometry. Selection leads to a message at the bottom left of the screen:

Minimization completed.

* Free valences can be "protected" without altering the molecule by adding hydrogens -H to them from the organic builder).

The molecular mechanics energy* in kJ/mol, displayed at the bottom right of the screen, is continually updated during the minimization process. Minimization may be stopped at any time by *clicking* on the STOP icon at the bottom right of the screen (for most molecules the minimization time is very short). Any geometrical constraints imposed on the structure (see **Constrain Distance**, **Constrain Angle**, **Constrain Dihedral** under the **Geometry** menu; **Chapter 19**) are enforced during minimization. NOE's are not taken into account. Also, any frozen atoms in the structure (see **Freeze Center** under the **Geometry** menu; **Chapter 19**) remain frozen.

Guess Transition State

This allows transition state geometries to be "guessed" based on the similarity with one or more entries in ***Spartan***'s reaction database**. An "exact hit" replaces the guess with transition state for the same reaction with the particular theoretical model entered in the database***. More commonly, the procedure attempts to provide as close a match as possible with a database entry, generally involving a less substituted system or one with different substituents. In this case, those parts of the structure that are common with a transition state in the database will be "frozen" and the remaining parts will be optimized using molecular mechanics. No conformational searching is performed. Note that it may be essential for the reactants to be properly oriented to reflect the desired stereochemical outcome of the reaction.

Where a reaction is completely unknown to the database, a fallback technique (linear synchronous transit) is invoked. This cannot be expected to yield as good a guess as that provided by the database procedure described above.

Input to transition-state guessing procedure has already been described (**Reaction Query** earlier in this chapter). It will be familiar to organic chemists, in that it is based on curved arrows, where each arrow identifies the movement of one electron pair. Curved arrows may either be superimposed onto a three-dimensional structure or onto a two-dimensional drawing.

After all reaction arrows have been properly designated, *clicking* on at the bottom right of the screen replaces the reactant(s) with a guess at the transition state. In the event that the guess seems to be unreasonable, this operation may be undone (select **Undo** from the **Edit** menu). This allows you to review your assignment of arrows and make changes as needed.

---

* The mechanics energy is a combination of the strain energy which is necessarily positive and the non-bonded or intramolecular interaction energy which can be either positive or negative. It will most commonly be a positive quantity, although it can be slightly negative.

** The Spartan Reaction Database (SRD) may also be directly searched by substructure. See discussion earlier in this chapter.

*** At the present time, the majority of database entries derive from very simple (semi-empirical and Hartree-Fock) quantum chemical models.

Chapter 21

The Setup Menu

*This chapter describes functions available under the **Setup** menu. **Calculations...** is used to specify molecular mechanics calculations, semi-empirical calculations, Hartree-Fock molecular orbital calculations and correlated calculations, including density functional calculations, Møller-Plesset calculations, coupled cluster calculations and quadratic configuration interaction calculations for ground-state species, and configuration interaction calculations and (time dependent) density functional calculations for excited-state species. Tasks include calculation of energy, equilibrium structure and conformation, transition-state structure, as well as identifying low-energy conformers for calculating Boltzmann weights and for constructing energy profiles. Not all tasks are available for all methods. Hartree-Fock and density functional calculations may be extended to include a solvation term based on the C-PCM procedure, allowing equilibrium and transition-state structures and molecular properties to be obtained in the presence of an idealized solvent. A wide variety of all-electron Gaussian basis sets are supported for Hartree-Fock and correlated calculations as are pseudopotentials for calculations on molecules incorporating elements for which all- electron basis sets are not available. Also provided are the G3 (MP2), G3, and G4 thermochemical recipes, as well as the T1 recipe that has been parameterized to closely reproduce G3 (MP2) heats of formation with 2-3 orders of magnitude less computation time.*

*New to **Spartan'26** are a series of seven neural nets that have been trained to (1), improve conformer energy differences obtained MMFF molecular mechanics, (2), estimate ωB97X-D/6-31G* equilibrium geometries starting from MMFF geometries, (3), correct HF/3-21G(*) energies to ωB97X-D/6-31G* quality, (4-6), estimate ωB97X-V/6-311+G(2df,2p), ωB97M-V/6-311+G(2df,2p), and ωB97M(2)/6-311+G(2df,2p) energies starting from ωB97X-D/6-31G* equilibrium geometries and energies, and (7), estimate ^{13}C and ^{1}H NMR shifts. These offer the possibility of accurately reproducing the results of high-level quantum chemical calculations at greatly reduced cost.*

*The automated multi-step procedure for identifying the lowest-energy conformer of a flexible molecule and for calculating accurate Boltzmann weights,(first introduced in **Spartan'18**) is now available with the option of utilizing the new (in **Spartan'26**) neural net (machine learning) protocols – which open the procedure up to both larger and more flexible molecules. The scheme can be seamlessly combined with NMR data on the*

individual conformers to yield Boltzmann averaged chemical shifts and coupling constants.

Calculations... *also requests calculation of IR, Raman, VCD, both 1D and 2D NMR and UV/visible spectra. IR spectra may be calculated using the C-PCM solvent model but solvated NMR, Raman and UV/visible spectra are not available. Both 3-Bond HH and 2 and 3-bond CH coupling constants may either be directly calculated or evaluated empirically, whereas longer range HH and CH coupling constants as well as CC coupling constants if needed may be calculated. A variety of molecular properties, QSAR descriptors and thermodynamic quantities are available for calculation and printing.*

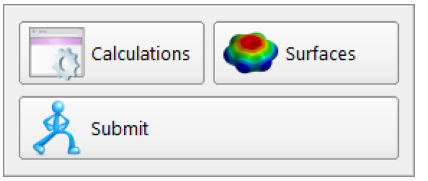
Finally, ***Calculations...*** *specifies conditions for identifying similar molecules, based either on molecular structure or chemical functionality, as well as for identifying molecules that are compatible with a particular molecular environment (a pharmacophore).*

Calculations performed by ***Spartan'26*** *take advantage of 64-bit architecture. Multi-core processors may be used to simultaneously process different jobs or for most tasks involving Hartree-Fock, density functional and RI-MP2 models, may be directed at a single job.*

Surfaces *is used to designate graphical surfaces, including electron and spin densities, electrostatic potentials, local ionization potentials and molecular orbitals, for display as surfaces, property maps and contour plots. Accessible regions on electron density surfaces and property maps based on these surfaces may be demarked.*

Submit *is used to direct jobs for calculation either locally or on a remote resource licensed with* ***Spartan****.*

The **Setup** menu provides access to dialogs for specifying molecular mechanics and quantum chemical calculations including a variety of properties and spectra, for specifying conditions for similarity analysis as well as specifying surfaces and property maps and for submitting jobs for calculation.



Calculations...

Calculations modules perform: molecular mechanics calculations using the MMFF and SYBYL force fields; MNDO (MNDO/d for second-row elements), AM1, RM1, PM3 and PM6 semi-empirical calculations, including PM3 and PM6 calculations on transition- metal inorganic and organometallic systems; Hartree-Fock molecular orbital calculations; a wide variety of density functional calculations including time-dependent density functional calculations (for excited-state species); MP2 and RI-MP2 Møller-Plesset calculations;

configuration interaction (CI) calculations (for excited-state species). All are applicable to energy calculation, equilibrium and transition-state geometry determination, conformational searching and energy profile construction, although some methods may not be practical for some tasks. Available for energy calculations only, are a series of high-order correlated models, including a double-hybrid density functional model, MP3 and MP4 Møller-Plesset models, coupled cluster models and quadratic configuration interaction models. Heats of formation may be calculated from G3 (MP2), G3, G4 (MP2) and G4 procedures "approximations" to coupled-cluster models, and from the much faster T1 recipe that is perhaps best viewed as an approximation to G3(MP2).

New: *Neural networks trained to (1) improve conformer energy differences obtained MMFF molecular mechanics,*[4] *(2) estimate ωB97X-D/6-31G* equilibrium geometries starting from MMFF geometries,*[2] *(3) estimate ωB97X-V, ωB97M-V and ωB97M(2) energies with the 6-311+G(2df,2p) basis set,*[6] *starting from ωB97X-D/6-31G* equilibrium geometries and energies (or est. density functional geometries), are accessible as an option to specific MMFF or QM calculations for select tasks. Two separate neural network models are employed for chemical shift calculation when* ***NMR Estimator***[7] *is chosen (one for* ^{1}H *and one for* ^{13}C*), rather than a QM model for calculation of NMR. Finally, in select tasks a neural network is available to correct/improve HF/3-21G(*) energies to ωB97X-D/6-31G* quality (specifically used in the new* ***MLHF*** *and* ***MLXD*** *protocols).*[9]

A wide selection of basis sets are supported, ranging from the STO-3G minimal basis set (recommended for Hartree-Fock models only), to the 3-21G split-valence basis set (also recommended for Hartree-Fock models only), to 6-31G*, 6-311G*, cc-pVDZ, cc-pVTZ and cc-pVQZ polarization basis sets together with the corresponding augmented representations, the def2 series of double, triple and quadruple σ basis sets and the PCJ basis sets specifically designed for calculation of NMR coupling constants. The dual basis set approximation has been implemented for some (but not all) basis sets, typically providing as much as an order of magnitude reduction in computation cost with very little change in either reaction energies or molecular properties. In addition, pseudopotentials for elements which all-electron basis sets do not exist are available, in particular, second and third-row transition metals and lanthanides. Note, however, that first derivatives of pseudopotentials in the case where the valence incorporates f-type and higher-order functions are only available for lanthanides, and U, Np, and Pu from the actinides. All basis sets, except STO-3G and 3-21G, may be supplemented with polarization functions on hydrogen, additional and/or higher-order polarization functions on heavy atoms and/or diffuse functions on hydrogens and/or heavy atoms.

Quantum chemical calculations also result in a variety of atomic and molecular properties, QSAR descriptors and thermodynamic quantities as well as IR, Raman, NMR and UV/visible spectra. IR spectra are available for molecular mechanics, semi-empirical, Hartree-Fock, density functional (except for double-hybrid functionals) and MP2 and RI-MP2 models. Other spectra are available only for Hartree-Fock and density functional

models. Note, however, that NMR spectra are not presently available where pseudopotentials are employed.

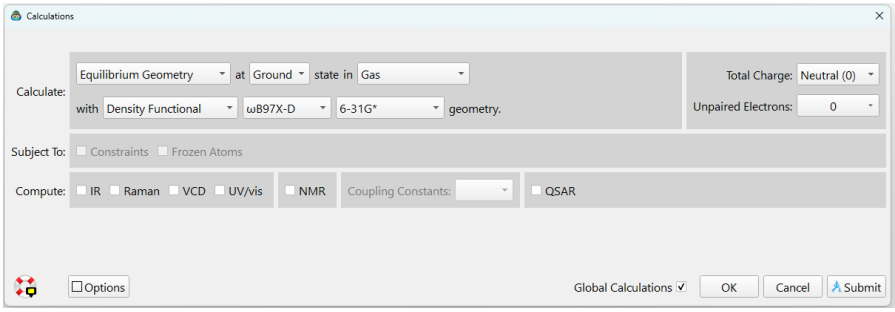
Solvent may be introduced into Hartree-Fock and density functional calculations by way of the C-PCM continuum model. Equilibrium (and transition-state) structures and vibrational frequencies in addition to energies may be calculated in the presence of solvent. Additional solvent models are available (see **Spartan'26 Help** from the **Help** menu).

Discussion of molecular mechanics and quantum chemical methods in general, focusing on the specific methods available in ***Spartan***, is provided in **Topics** under the **Activities** menu.

A more thorough discussion of specific quantum chemical models is provided in the latest Q-Chem User's Manual (≈version 6.3/6.4 at the time of this revision to the ***Spartan*** manual).

Finally, the **Calculations** dialog accesses a procedure to quantify the extent to which two molecules or a molecule and a pharmacophore are similar. This does not involve molecular mechanics or quantum chemical calculations but depends only on the geometry of the system.

Selection of **Calculations...** results in the **Calculations** dialog*.



This contains a number of options, accessed via pull-down menus, buttons and check boxes.

Calculate

This section contains a series of menus and check boxes that specify the task to be accomplished, the electronic state (ground or first excited), the type of calculation, method, basis set and solvent to be employed, as well as details specific to each calculation type.

* The dialog allows for selection of a number of tasks: **Energy**, **Equilibrium Geometry**, **Transition State Geometry**, **Equilibrium Conformer**, **Conformer Distribution**, **Energy Profile**, **SSPD Compatible**, **NMR Spectrum**, **Similarity Library**, or **Similarity Analysis** tasks, as documented in this chapter.

Task

Specification of what is to be done is by way of a pull-down menu:

Energy
Equilibrium Geometry
Transition State Geometry
Equilibrium Conformer
Conformer Distribution
Energy Profile
NMR Protocol
NMR Spectrum
Similarity Library
Similarity Analysis

Energy specifies calculation of energy (and in the case of quantum chemical methods, a wave function) at a single geometry.

By default, ***Spartan*** reports energies from molecular mechanics and heats of formation from semi-empirical calculations, T1, G3(MP2) and other thermochemical recipes in kJ/mol. Energies from Hartree-Fock and electron-correlated models are reported in atomic units (hartrees).

The molecular mechanics energy comprises two parts: the strain energy and the non-bonded energy. The strain energy is the difference in energy between a molecule and its "strain free" analog. It is nearly always positive and less than a few hundred kJ/mol* in magnitude. The non-bonded energy accounts for attraction or repulsion between atomic centers that are not connected due to van der Waals and Coulombic interactions. Because the strain energy of every molecule relates to a different standard, molecular mechanics energies cannot be used to obtain reaction energies (unless there are no changes in bonding between reactants and products).

The heat of formation is the enthalpy at 298K of a balanced chemical reaction in which a molecule is converted to a set of standard products. For example, the heat of formation of ethylene is given by reaction:

$$\mathbf{C_2H_4 \rightarrow 2C\ (graphite) + 2H_2\ (gas)}$$

Here, graphite and hydrogen molecule are the carbon and hydrogen standards, respectively. In practice, the actual measurement is typically carried out for a combustion reaction, for example, for ethylene:

$$\mathbf{C_2H_4 + 3O_2 \longrightarrow 2CO_2 + 2H_2O}$$

Heats of formation may either be positive or negative quantities and generally span a range of a few thousand kJ/mol.

Heats of formation are not suitable for presenting energy data from most quantum chemical calculations (except thermochemical recipes), simply because the standards for several important elements (most notably, carbon) are not well-defined isolated species. In its place is the energy of a reaction that splits a molecule into isolated nuclei and electrons, for

example, for ethylene:

$$\mathbf{C_2H_4 \rightarrow 2C^{+6} + 4H^{+} + 16e^{-}}$$

Total energies, as the energies of such reactions are termed, are always negative and may be very large (tens of thousands of kJ/mol), most commonly given in atomic units (hartrees).

1 atomic unit = 2625 kJ/mol

Aside from a difference in units (see **Appendix E**), it makes no difference whatsoever which standard is employed to investigate thermochemistry.

Note that we will generally treat enthalpy (what is measured experimentally) and energy (what is calculated) as identical quantities. They are not identical but very similar for the purpose of dealing with chemical reactions.

Equilibrium Geometry specifies that the nearest energy minimum will be located, and **Transition State Geometry** that the nearest ***first-order saddle point*** (an energy maximum in one dimension and energy minima in all other dimensions), will be located. Note, however, that not all first-order saddle points will correspond to chemically-meaningful transition states. **Energy Profile** steps along a user-defined range of coordinates.

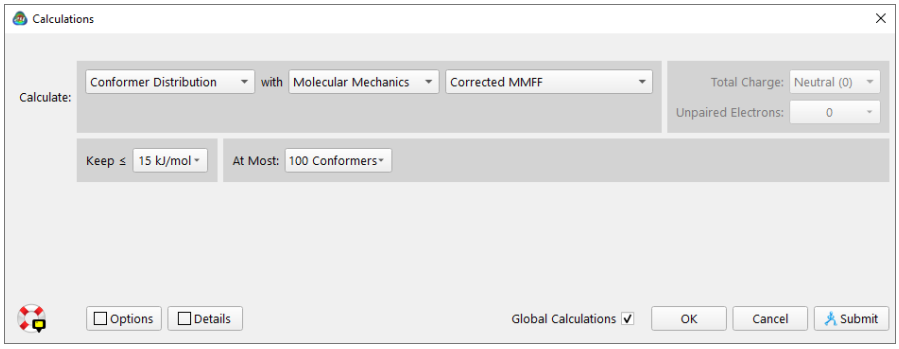
Energy applies to all methods. **Equilibrium Geometry** and **Energy Profile** do not apply to double-hybrid density functional methods, wavefunction-based correlated methods or thermochemical recipes. **Transition State Geometry** does not apply to molecular mechanics methods, double-hybrid density functional methods, wave-function-based correlated methods and thermochemical recipes.

Equilibrium Conformer / Conformer Distribution

Equilibrium Conformer replaces the initial conformer of a molecule with its lowest-energy conformer. **Conformer Distribution** replaces the structure with that of the lowest-energy conformer but substitutes the energy, dipole moment and atomic charges with the corresponding Boltzmann-weighted averages. A list of all conformers that make up the average is also provided as a separate ***Spartan*** document. Both procedures involve examination of all available conformers and may involve MMFF, MMFF corrected by a neural net, or a variety of multi-step protocols starting with MMFF and ending with a large basis set density functional model.

The neural net was trained on conformer energy differences ≈250k flexible organic molecules from several sources including the Spartan Structure and Properties Database (SSPD, collections of natural products and pharmaceuticals and entries from the Cambridge Structural Database (CSD). In it involves a ~2M conformers. It is presently restricted to the elements H, C, N, O, F, S, Cl and Br. MMFF molecular mechanics structures are employed.

Selection of **Molecular Mechanics** from the top menu for both **Equilibrium Conformer** and **Conformer Distribution** leads to nearly identical dialogs (that for **Conformer Distribution** is shown).



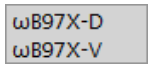
Spartan automatically identifies conformational degrees of freedom (single bonds and flexible rings). **Conformer Rules** are set in the **Settings Preferences** dialog (**Preferences** under the **Options** menu; **Chapter 25**). Finer control may be exercised using **Set Torsions** under the **Geometry** menu (**Chapter 19**).

For molecular mechanics (only) two different search modes are available. A ***systematic search*** looks everywhere and may not be practical for molecules with many degrees of conformational freedom. In practice, that is accomplished in two steps for molecules that incorporate medium to large flexible rings: search only degrees of freedom involving the rings yielding a set of "stable" structures and then for each of these structures, search the acyclic degrees of freedom.

A ***Monte-Carlo search*** follows a path that biases in favor of low- energy conformers (but that does not completely exclude high-energy conformers). While there is no guarantee that the lowest-energy conformer (the ***global minimum***) will actually be located, it can be shown that the set of conformers examined (and kept) approach a Boltzmann distribution. (This is strictly true only for a large number of search steps but appears to be closely approximated for searches with reasonable numbers of steps.) ***Spartan*** chooses which (***systematic*** or ***Monte-Carlo***) search procedure to use based on the number of conformational degrees of freedom. Selection may be forced using the **SEARCH METHOD** (SEARCHMETHOD=) keyword; **Appendix D**.

Both **Equilibrium Conformer** and **Conformer Distribution** may be restricted to "looking at" a user-specified maximum number of conformers (**Maximum Conformers Examined**). If this number is less than the estimated number of actual conformers, this forces a Monte-Carlo search.

Selection of **Density Functional** from the left most of the menus following both **Equilibrium Conformer** and **Conformer Distribution** results in a menu to the right with two entries: either ω**B97X-V** or **B97M-V** functionals both with 6-311+G(2df,2p) [6-311G*] basis set.

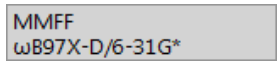


The ω**B97X-V** functional is used for energy calculations only and is coupled with

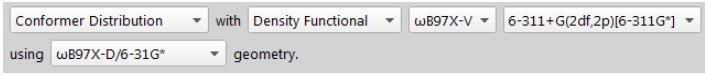
geometries from ω**B97X-D/6-31G***. Similarly, **B97M-V** is coupled with **B3LYP/6-31G*** geometries. Experience suggests that the former combination yields more reliable results albeit at a computation cost that can be close to twice the latter combination. Other combinations of methods may be specified, although their performance has not been assessed.

The second line in the dialog starting from **using** contains a single menu which controls the geometry to be employed for the final step energy calculation. **B97M-V** is initially selected, it contains:

If ω**B97X-V** is initially selected, it contains:



These menus control two of the steps in the multi-step procedure used to determine the lowest-energy conformer or the set of Boltzmann weights. This actually involves intermediate steps to gradually reduce the number of conformers from those produced by the initial MMFF search. The full procedure can be seen by *checking* **Details** at the bottom of the dialog. Shown below are the settings corresponding to menu selections of ωB97X-V (functional), 6-311+G(2df,2p) [6-311G*] (basis set) using ωB97X-D/6-31G* (geometry).



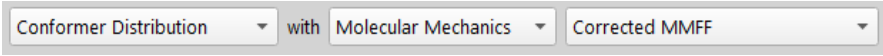
With details exposed, the models used for each of the steps can be altered (or eliminated together) and energy and number thresholds are adjustable.

Selection of MMFF instead of B3LYP/6-31G* or ωB97X-D/6-31G* for geometry calculation eliminates all intermediate steps resulting in a much less costly (and presumably less accurate) protocol.

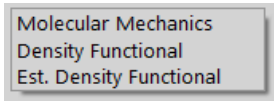
The difference between **Equilibrium Conformer** and **Conformer Distribution** (after the full sequence of steps has been completed), is the number of conformers retained, that is, both tasks have the same computation cost.

Neural Network Models

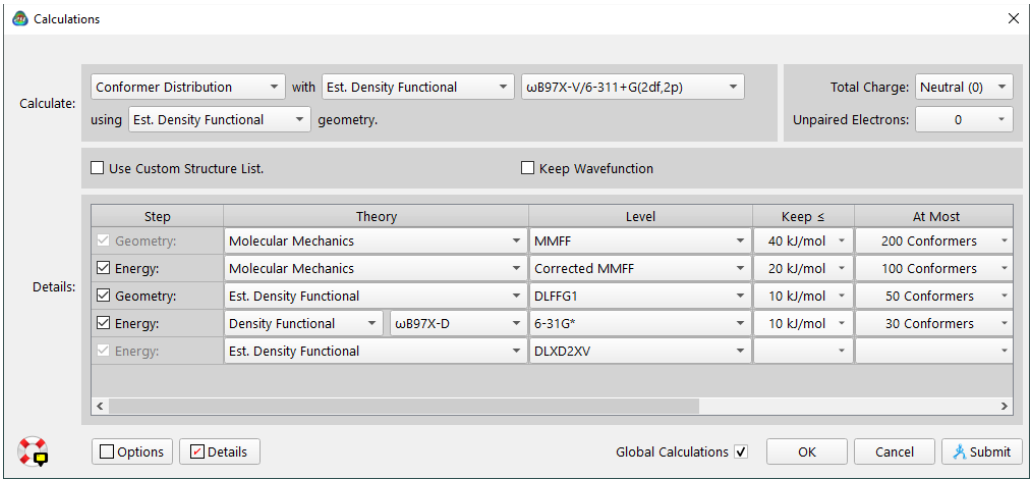
Beginning in ***Spartan'26*** the Molecular Mechanics options included **Corrected MMFF** as an option:



Of significant utility, are the additional machine learning routines accessed from the **Est. Density Functional** option:



as this enables a wide full range of neural networks to be used in tandem in a number steps in the workflow:



NMR Protocol

Contains four predefined and published* workflows for calculation of NMR of flexible molecules. Two primarily QM (2019 and 2024) approaches and two primarily ML (2025 MLXD and 2026 MLHF) approaches. 2026 assessments against experiment suggest these protocols reliable reproduce experimental ^{13}C NMR shifts with an average RMSD of 1.9 ppm for rigid and 2.2 ppm (with additional accuracy improvements of 0.5 ppm in the pipeline).

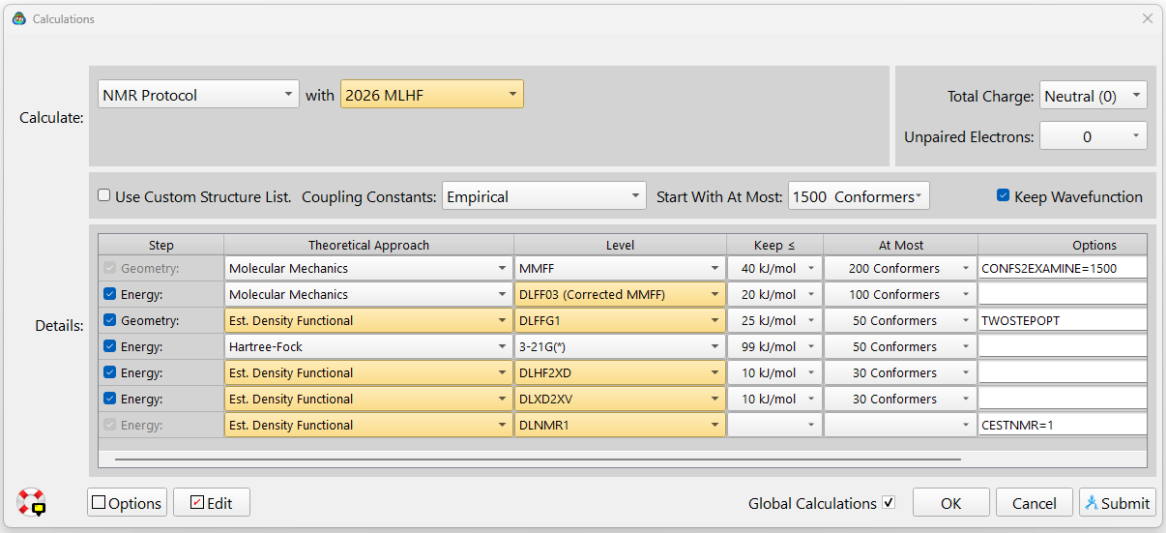
In terms of **QUALITY**: While both QM and ML-based options provide comparable results in terms of ^{13}C and ^{1}H NMR accuracy, it is important to note several key differences.

In terms of **COVERAGE**: The 2019 QM protocol generally supports elements H-Kr, including charged systems. The 2024 QM protocol (including a single ML step) and the remaining ML-based protocols are limited to: H, C, N, O, F, S, Cl, and Br, uncharged and closed shell systems only.

---

* 2019 QM protocols: *J. Nat. Prod.* **2019**, *82* (8), 2299-2306
2024 QM, 2025-2026 ML protocols: *ChemRxiv* **2026**, preprint

In terms of **PERFORMANCE**: The MLHF-NMR protocol is the *fastest* option, with results typically in seconds to a few minutes. The MLXD-NMR protocol is *next*, at a factor of 5-10 slower than MLHF-NMR, requiring minutes to tens of minutes. The QM-based protocols are *last*, at hundreds to thousands of times slower (taking multiple hours to overnight) for flexible natural products.



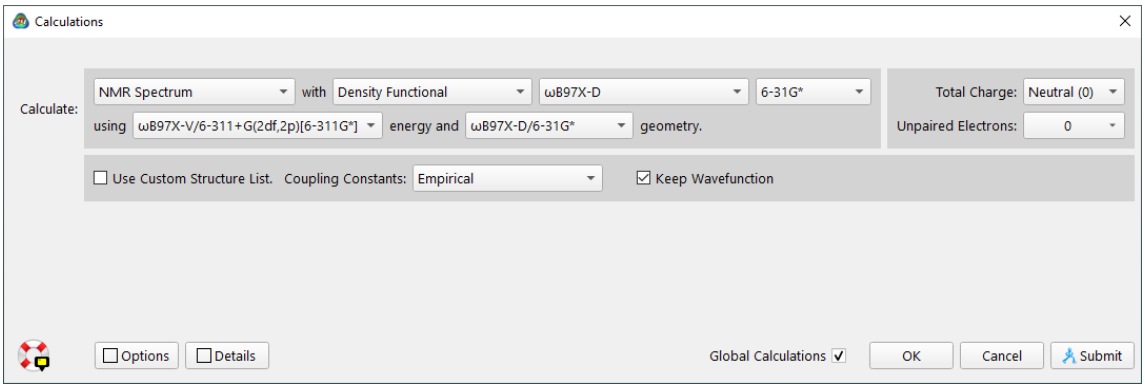
In the NMR Protocol task, details are exposed but initially grayed out. Details can be enabled by clicking the **Edit** check box. These referenced protocols have published definitions, depending on changes, the task name may change to **Custom Protocol**. These protocols are highly tested and have been successfully employed in support (and correction) of multiple published natural products. The protocols largely take the place of the following **NMR Spectrum** task, which we include primarily for backward compatibility. Both options include defaults and the ability to customize calculation details (including the option to use a custom list of conformers, rather than generate this from ***Spartan***).

NMR protocol	total wall time	avg wall time	avg ^{13}C RMSE	avg ^{13}C max E
2019 QM-NMR	26,100 min	261 min	2.12 ppm	4.98 ppm
2026 MLHF-NMR	148 min	1.5 min	2.14 ppm	4.89 ppm

NMR Spectrum

NMR Spectrum calculates chemical shifts and coupling constants* for a flexible molecule or for a set of closely-related (regio and or stereo) isomers of a flexible molecule. A variety of multi-step procedures are available, all starting either from a systematic or sparse-systematic conformer search using MMFF molecular mechanics, or from a user-supplied list of conformers. The best MMFF conformer could then be directly used for NMR

calculations, or the search could be refined using a selection of density functional models for geometry and energy calculations, the latter resulting in a series of Boltzmann weights. Only two models, B3LYP/6-31G* and ωB97X-D/6-31G* are available in menus for NMR calculations. Both have been extensively tested and have been corrected empirically. While not corrected, a wide variety of other calculation models may be specified (see discussion later in this section). Inclusion of the neural network routines associated with conformational searching are also available for the multi-step **NMR Spectrum** calculation. Selection of **NMR Spectrum** leads to:



Checking the **Details** check box will enable additional options at each step in the workflow, including choice of model, energy window and total number of conformers retained:

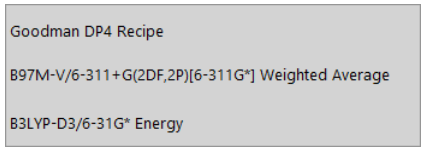
Details:

Step	Theory		Level	Keep ≤	At Most
Geometry:	Molecular Mechanics		MMFF	40 kJ/mol	200 Conformers
Energy:	Molecular Mechanics		Corrected MMFF	20 kJ/mol	100 Conformers
Geometry:	Hartree-Fock		3-21G	30 kJ/mol	100 Conformers
Energy:	Density Functional	ωB97X-D	6-31G*	15 kJ/mol	50 Conformers
Geometry:	Density Functional	ωB97X-D	6-31G*	10 kJ/mol	50 Conformers
Energy:	Density Functional	ωB97X-V	6-311+G(2df,2p)[6-311G*]	10 kJ/mol	30 Conformers
Energy:	Density Functional	ωB97X-D	6-31G*		

Three menus appear to the right of **NMR Spectrum** at the top. The first allows for selection of **Est. Density Functional** or **Density Functional**. When the former is chosen, a default multi-step procedure is enabled that maximizes use of the machine learning (neural network) models.

If **Density Functional** is specified, choice of available functionals ωB97X-D or B3LYP dictate the default of the menu options on the second line, to the right of "using"). When B3LYP is selected, these read:

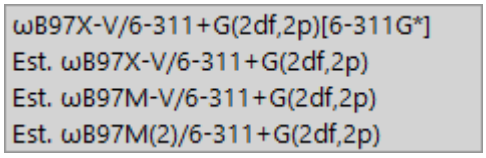
* 3-bond HH and 2 and 3-bond CH coupling constants available from parameterized expressions are replaced if coupling constant calculation is requested. Longer-range CH coupling constants as well as CC coupling constants may also be calculated.



No further information is required for the Goodman DP4 recipe. For the B97M-V/6-311+G(2df,2p)[6-311G*] weighted average calculation, the menu to the right contains two selections:

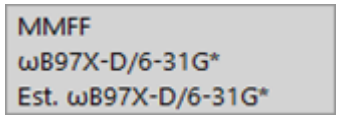
For the B3LYP-D3/6-31G* energy calculation, the menu to the right contains only a single selection.

Where ωB97X-D is selected (top line), the menus underneath reads:

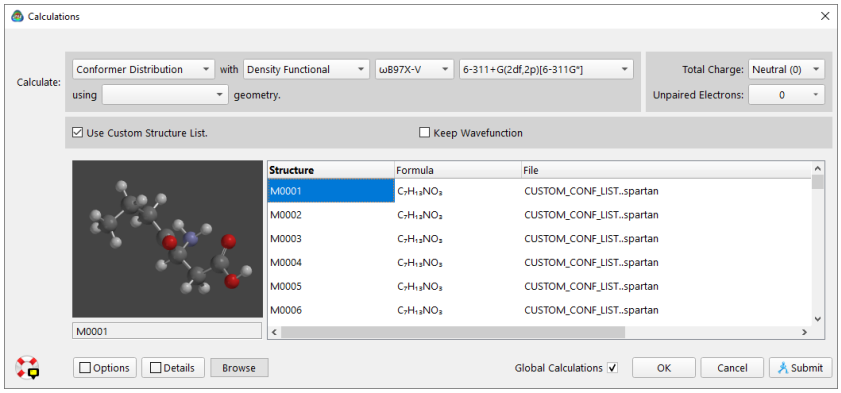


The choices amount using a QM derived energy or a neural network derived energies in the case that one of the Est. options is selected.

In either case, selection of one the above energy options offers to 3 additional choices for geometry:



Checking **Use Custom Structure List** replaces the MMFF conformer search by a user specified list of conformers. This is a ***Spartan*** document and may be pasted or dragged into the dialog that appears.

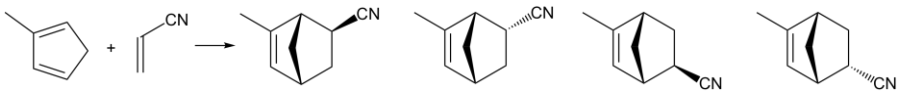


Coupling constants are not calculated by default, although 3-bond HH coupling constants and 2 and 3-bond CH coupling constants are provided from parameterized relationships. These, as well as longer-range HH and CC coupling constants, may be calculated by selecting **Calculated** or **Calculated (Fermi Contact)** in lieu of **Empirical** from the **Coupling Constants** menu to the right at the center of the dialog. In many cases, empirical coupling constants are typically sufficient, and in almost all cases, coupling constants obtained within the Fermi contact approximation which are significantly less costly, are typically very close to calculated values, that is, those obtained without the approximation.

Full details of the multi-step procedures may be seen by *checking* **Details** at the bottom of the dialog. Users are able to alter the default energy and maximum number thresholds between steps in the procedure as well as change the models employed, eliminate specific steps, and via the **Options** check box, can include specific computational items for each step in the customizable protocol). It should be noted that the defaults have been fairly heavily assessed, and in general, we see a high degree of agreement between experimental shifts (in particular ^{13}C shifts) to the results returned from the NMR Spectrum task (that is, the Boltzmann weighted shifts).

There are at least two distinct uses for ***Spartan's*** NMR capability. The first, perhaps of most interest and value to the natural products community, is to provide an accurate NMR spectrum corresponding to a structure proposed for a "new" compound. Serious disagreement between calculated and experimental spectra, that is, outside previously established limits might cast doubt on the validity of the latter, either the specific peak assignments or the proposed structure, while agreement between the two by no means confirms the experiment although it does lend credence.

The second use may be of greatest interest and value to the synthetic organic community, where the identity of the molecule is most likely certain aside from "details" regarding regio and stereochemistry.* For example, comparison of calculated and measured NMR spectra might be of value to decide which of the four distinct products that could arise from a Diels-Alder cycloaddition of 2-methyl cyclopentadiene and acrylonitrile is actually formed.



These two applications have different requirements and may best be met with two different approaches rather than a "one size fits all". Specifically, contesting or supporting the experimental assignment of an "unknown" natural product is quite likely more demanding of the computations than finding the best match among the closely-related isomers than may result in a synthesis.

Similarity Library Similarity Analysis

Spartan provides a procedure to assess and quantify the similarity of one or more molecules in a document to one or more molecules in one or more different documents. Similarity may either be based on molecular structure (geometry) or on chemical function descriptors (CFD's). ***Spartan*** also assesses the similarity of one or more pharmacophores in a document to one or more molecules in one or more different documents.

Three different similarity analysis modes are available, depending on the ***template*** (the selected document in the graphical user interface) and the ***library*** (one or more different documents specified in the **Calculations** dialog).

Three different similarity analysis modes are available, depending on the ***template*** (the selected document in the graphical user interface) and the ***library*** (one or more different documents specified in the **Calculations** dialog).

template	***library***
pharmacophores	molecular structures for which CFD's are assigned on-the-fly
molecular structures	molecular structures
CFD's from molecular structures	molecular structures for which CFD's are assigned on-the-fly

* The results from DP4 analysis are available for posting in the spreadsheet.

Two types of libraries are supported for molecular structures (and CFD's generated on-the-fly from molecular structures). In the first, each entry represents a single molecule. In the second, each entry represents a selection of different conformers of a single molecule and spanning the possible shapes that the molecule may take on. The latter are generated using **Similarity Library** and ***may only be used as library entries in similarity analysis and not as templates***.

Of the three similarity analysis modes, only the second makes direct use of molecular structures (selected atomic centers for the molecules in the template) as a similarity measure. The other two make use of CFD's generated (and optionally edited) for the template and generated on-the-fly in the library. Selection of structure or CFD as the basis of similarity measure must be explicitly stated (see discussion following).

Seven different chemical function descriptors (CFD's) are recognized by the similarity analysis procedure, the first six of which are attributes of a molecule while the seventh derives from knowledge of a molecule bound to a protein or nucleic acid. A list follows, note, in addition to definition, the volume of the spheres are also available for modification.

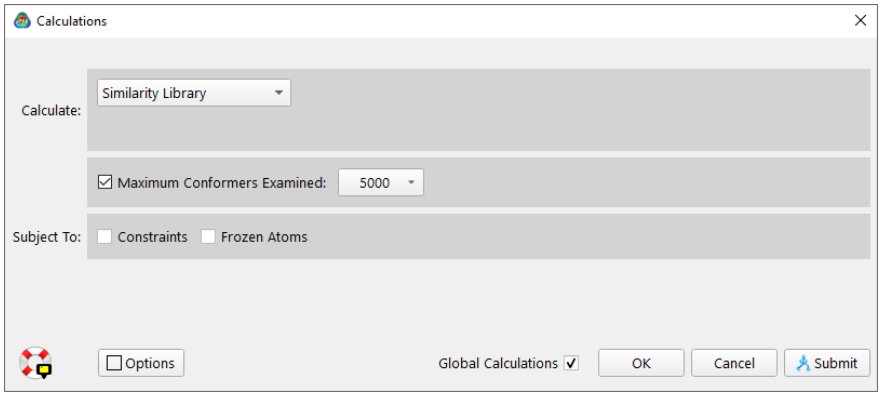
hydrophobe (sterically crowded region)
aromatic hydrophobe (aromatic ring)
hydrogen-bond donor
hydrogen-bond acceptor positive
ionizable site negative ionizable site
excluded volume

The individual CFD's for molecules in the template (only) may either be used or ignored in the similarity procedure. This is specified from **Set Similarity Centers** under the **Geometry** menu (**Chapter 19**). The **CFD Properties** dialog (accessible from **Properties** under the **Display** menu; **Chapter 22**) allows for changing or extending the definition of a CFD For example, the nitrogen in a dimethyl amino group might initially be assigned as a hydrogen bond acceptor CFD but could be changed to a positive ionizable site (as such a center could become positively-charged at biological pH) or could be extended to cover both possibilities.

Spartan assumes that chiral molecules in a library are unresolved (they comprise an equal mixture of enantiomers).* That is to say, similarity comparisons are carried out on both the library molecules as they are stored and (if the molecules are chiral) on their mirror images. This procedure may be overridden (providing comparisons only to the stored molecules) by way of the **Single Enantiomer** keyword entered in the **Options** field. See discussion later in this chapter and also **Appendix D**.

Note that only a single enantiomer of a chiral template molecule is considered. If comparisons involving both enantiomers are desired, the second enantiomer must be added to the query.

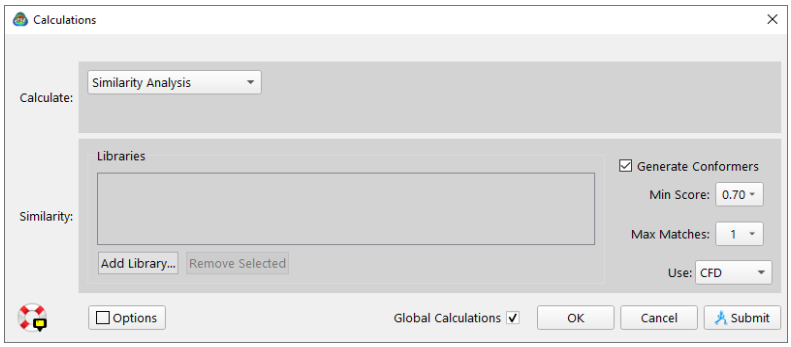
Selection of **Similarity Library** from the **Calculate** menu with either a molecule or pharmacophore (list of molecules/pharmacophores) leads to a dialog for setting up a similarity analysis.



One or more libraries need to be identified. This is accomplished by *clicking* on the **Add Library...** button and gives rise to a file browser. Once selected, *clicking* on **Open** adds the library to the list of templates. A library may be deleted from the list by *clicking* on its name in the list and *clicking* on the **Remove Selected** button at the bottom of the dialog. **Shift** and **Ctrl** keys function in the usual manner in both library addition and removal.

Selection of **Similarity Analysis** from the **Calculate** menu leads to a second dialog.

* Properties from quantum chemical calculations (structures, energies, ...) are independent of absolute configuration..



Prior to starting a similarity analysis, it is necessary to select either **structure** or **CFD** from the **Use** menu near the bottom of the **Calculations** dialog.

Similarity analysis is most valuable where the template molecule is conformationally flexible, and it is not apparent which (if any) of the different shapes it adopts might match a particular target in a library.

Where the target is also flexible and its conformation is not established, the process must be repeated for every template conformer-target conformer pair. However, if the purpose is simply to identify which target molecules a particular template molecule might match, irrespective of conformer, the first good match concludes the analysis. In this case, the user is left with the knowledge that template and target can be made to match (together with an example of a match).

The similarity algorithm incorporated in ***Spartan*** attempts to examine template-target pairs that are most likely to lead to matches before examining pairs that are less likely. It also allows early termination upon finding a single good match, or a user-specified number of good matches*. Taken together, these can lead to one or two orders of magnitude speedup in the analysis.

* Criteria for early termination are specified using program options.
See **Similarity Analysis Options** in **Appendix D**.

State

Specification of **State** (electronic state) is by way of a pull-down menu. The default setting **Ground** (state) may be changed to **First Excited** (state) by *clicking* on ⌄ to the right of the box and back again to **Ground** (state) by *clicking* on ⌄ . In general, all models are available for **Ground** (state) calculations. Switching to **First Excited** (state) results in access to only two classes of models that handle excited state calculations, time-dependent density functional models* and configuration interaction models.

* Excited state calculations with time-dependent density functional models that incorporate explicit account of dispersion (ωB97X-V, B97M-V and ωB97M-V) are not presently supported.

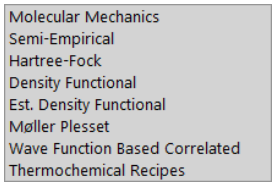
In Gas/Solvent

By default, all calculations are performed in the gas state. ***Spartan*** also supports energy, gradient and frequency calculations using both Hartree-Fock or density functional methods in the presence of solvent using the C-PCM continuum model. This is specified from the menu to the right of **in**, for **Energy**, **Equilibrium Geometry**, **Transition-State Geometry** and **Energy Profile** tasks. This depends only on the dielectric constant of the solvent.

Solvent choices are limited to non-polar (small dielectric, 7.4 for THF) polar (large dielectric, 37 for DMF) and water (dielectric constant of 78). Continuum models such as C-PCM do not account for explicit solvent-solute (or solvent-solvent) interactions such as hydrogen bonding. A non-hydrogen bonding solvent such as DMSO with a large dielectric constant will behave similarly to a hydrogen bonding solvent such as water. Additional solvent options are also provided.*

Type of Calculation

Specification of type of calculation is by way of a series of menus on two lines to the right of **Calculate**. Menus on the first line designate the task at hand, indicate whether it applies to ground or excited electronic state and specify solvent (if any). The class of methods along with associated "details" are specified in the menus on the second line. Available options for ground state energy calculations:

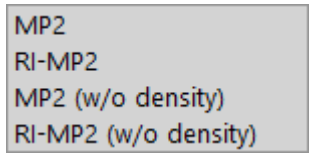


Computational options available for excited states and energy calculations include Configuration Interaction and (**T**ime **D**ependent) **D**ensity **F**unctional **T**heory.

Often choice of a model will prompt for additional specification:

Molecular Mechanics

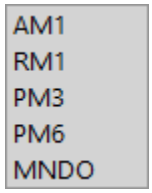
Møller Plesset



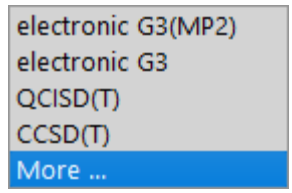
---

* Additional solvent choices and solvation models are also available via type-in from the **Options** field. A discussion is provided in the Properties FAQ, accessible from **Spartan'26 Help** in the **Help** menu.

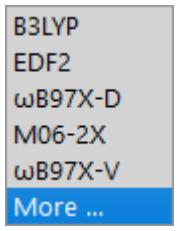
Semi-Empirical



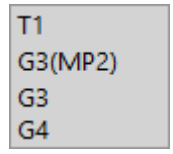
Wave Function Based Correlated



Density Functional



Thermochemical Recipes



Configuration Interaction

Selection of entry **More** which appears at the end of the list in several drop down menus leads to a dialog that provides a wider selection of methods:

Selection of **Equilibrium Geometry**, **Transition State Geometry** or **Energy Profile** removes the **Wave Function Based Correlated** and **Thermochemical Recipes**. Selection of **Equilibrium Conformer** or **Conformer Distribution** leaves only **Molecular Mechanics**, **Density Functional**, and **Est. Density Functional**, and selection of **NMR Spectrum** leaves only **Density Functional**.

Molecular mechanics models have proven suitable for determining equilibrium geometries of organic molecules. Calculations of compounds comprising hundreds or even thousands of atoms are possible, one or two orders of magnitude more than what is practical with even the simplest quantum mechanical models. Because molecular mechanics models require explicit bond-type and identification, they may not be used to obtain reaction energies, except where bond-types are identical. For example, they can be used to assess differences in energies among stereoisomers. Perhaps the most important use of molecular mechanics is as a first step in determining equilibrium conformation or distribution of low-energy conformers, where it serves to remove high energy conformers.*

* While present-generation molecular mechanics calculations are not reliable enough to consistently identify the lowest-energy conformer, let alone account quantitatively for conformer energy differences, they appear to be suitable for distinguishing between conformers that are good candidates from those that are not.

Semi-empirical models are the simplest methods based on quantum mechanics. They are applicable to molecules containing 100 - 200 atoms, but not to molecules containing thousands of atoms, for example, proteins, at least not routinely. The MNDO, AM1, RM1, PM3 and PM6 models generally provide geometries in good accord with experimental structures. The PM3 and PM6 models have also been parameterized for transition metals and generally provide a reasonable account of equilibrium geometries. While semi-empirical models are generally suitable for evaluation of such properties as polar surface area that depend solely on geometry, none of the present-generation of models is suitable for the calculation of relative energies as might be required to conclude whether a chemical reaction is weakly or strongly *exothermic*, thermoneutral or weakly or strongly *endothermic*, or whether one isomeric product of a reaction is likely to be more or less stable than another, or to identify the lowest-energy tautomer. Finally, semi-empirical models do not provide an acceptable account of conformer energy differences in flexible molecules. In fact, they do not appear to be superior to molecular mechanics models. Semi- empirical models are available for the calculation of IR spectra but do not provide a very good account. They are not available for the calculation of Raman, NMR or UV/visible spectra.

Hartree-Fock models follow from the **Schrödinger equation** by requiring that the electrons be independent particles, the **Hartree- Fock approximation**. Here, the motions of electrons in molecules (molecular orbitals) are approximated by a sum of the motions of electrons in atoms (atomic orbitals). A second approximation, termed the **LCAO** or **L**inear **C**ombinations of **A**tomic **O**rbital **approximation**, distinguishes different Hartree-Fock models. Different models use different basis sets, that is, different numbers and kinds of atomic orbitals. Two of the most popular small basis sets are 3-21G and 6-31G*. The 3-21G split-valence basis set uses one atomic orbital to describe each non-valence electron and two atomic orbitals to describe each valence (core) electron, that is, it comprises one 1s atomic orbital and two sets of 2s and 2p atomic orbitals.

$$1s$$
$$2s,\ 2p_x,\ 2p_y,\ 2p_z$$
$$2s',\ 2p_x',\ 2p_y',\ 2p_z'$$

The 6-31G* polarization basis set increases the flexibility by including d type atomic orbitals that, while not occupied in the atom, are used in molecules.

$$1s$$
$$2s,\ 2p_x,\ 2p_y,\ 2p_z$$
$$2s',\ 2p_x',\ 2p_y',\ 2p_z'$$
$$d_{xx},\ d_{xy},\ d_{xz},\ d_{yy},\ d_{yz},\ d_{zz}$$

Larger basis sets involving even higher splitting of valence shells, including higher-order (f and g type) functions as well as diffuse functions have been defined and are available in ***Spartan***. Among the most commonly used are the 6-311+G(2df,2p) basis sets proposed by Pople and the cc-pVQZ basis set proposed by Dunning. Description of their makeup is

provided in **Topics** available under the **Activities** menu and in ***A Guide to Molecular Mechanics and Quantum Chemical Calculations***[*] which is provided as a PDF under the **Help** menu.

3-21G, 6-31G* and larger basis set Hartree-Fock models generally provide good account of the geometries of organic molecules. This extends to unusual or unstable systems for which experimental data may be limited (and hence may be poorly represented in parameterization of the semi-empirical models). However, Hartree-Fock models are not successful in accounting for the equilibrium geometries of inorganic and organometallic compounds incorporating transition metals. The 3-21G and 6-31G* models provide at best a qualitative account of thermochemistry. For example, they are usually able to say whether a particular reaction is weakly or strongly *exothermic* or weakly or strongly *endothermic*, and to properly order the stabilities of closely-related isomeric products. Hartree-Fock models are, however, generally able to properly describe the energies of ***isodesmic*** reactions, that is, reactions where formal bond count is maintained.

The most conspicuous and most important flaw of Hartree-Fock models is their inability to properly account for the energetics of reactions where bonds are made or broken. This includes activation energies, that is, the difference in energy between reactants and transition state. The problem can be traced to the Hartree-Fock approximation, which in effect replaces interactions between electrons by interactions between one electron and the average field created by all other electrons. As a consequence, electrons get in each other's way to a greater extent than they should, leading to overestimation of the electron-electron repulsion energy, most significantly for electrons that are paired. Because bond dissociation reduces the number of electron-pairs (by one), the overestimation will be greater for reactants than for products, meaning that the calculated dissociation energy will be too small.[**]

Hartree-Fock models are available for calculation of IR, Raman, NMR and UV/visible spectra, the last in combination with CIS models for calculations involving excited states. IR (Raman) frequencies are typically overestimated by 10-15% and NMR chemical shifts show large variations from experimental values.

---

* This reference was written in 2003, and as such it does not include a full assessment or description of many of the newer computational features included in ***Spartan'26***. In particular many functionals and extended basis sets and no neural networks are not covered. An updated version is in preparation.

** Consider the bond dissociation energy of H_2. The (limiting) Hartree-Fock energy of the products (hydrogen atoms each with only one electron) is exact, while the energy of the reactant (H_2 with two electrons) is too large. Therefore, the calculated bond dissociation energy will be too small.

Correlated Models

Models that lessen the effects of the Hartree-Fock approximation are termed **correlated models** and divide into two broad categories, **density functional models** and **wave-function-based models**.

Wave Function Based Models

Wave function-based models, typified by **Møller-Plesset models** and **configuration interaction models** extend the flexibility of Hartree-Fock models by mixing ground and excited-state wave functions. They are significantly more costly than Hartree- Fock models and may or may not be more costly than density functional models (see discussion following).

However, they offer the advantage over the latter by providing a clear path to improvement. Both the so-called full configuration interaction model and the Møller-Plesset model taken to infinite order lead to the exact energy.

Density Functional Models (DFT)

Density functional models explicitly introduce an approximate ("empirical") correlation term (a functional). The advantage of density functional models over wave-function-based correlated models is their computation cost, which is of the same order as Hartree-Fock models. The disadvantage of density functional models is that it is not apparent how to improve on a particular choice of functional.

While it is common to view density functional models as an extension to Hartree-Fock models in that an extra term, the so- called exchange correlation functional is introduced, in practice, some functionals comprise distinct exchange and correlation components while others do not draw a boundary. What is the "correct" exchange/correlation functional? The quest has gone on for several decades and hundreds of functionals have been proposed. What follows is short classification of these methods.

Local Density Approximation (LDA)

The first functional proposed stems from a purely hypothetical problem in which a uniform gas of non-interacting electrons moves in a positively charged field. An analytical solution for the exchange energy is available and takes the form of the density to the 4/3's power. The form of the correlation energy may be arrived at through numerical simulation and is dependent on only the local density at each point.

Functionals that depend only on the electron density (ρ) are said to follow the local density approximation and are referred to as LDA functionals. LDA functionals are not generally suitable for accurate calculation of the geometries, energies and properties of molecules and have largely been supplanted by a variety of more-flexible forms. These have traditionally been divided into several classes, although the lines between them are blurred. In rough order of increasing complexity, these are:

Generalized Gradient Approximation (GGA)

These functionals depend on the gradient of the electron density, in addition to the density itself. GGA functionals have been around since the mid 1980's and were the first to provide a reasonable account of the energies of chemical reactions. In this sense, they were instrumental in drawing attention to density functional theory as a viable "low-cost" alternative to wave function based correlation techniques. The BLYP functional is perhaps the most well-known of the GGA functionals.

Global Hybrid Generalized Gradient Approximation (GH- GGA)

GH-GGA functionals (more commonly simply referred to as hybrid functionals) replace a fixed fraction of the exchange by the "exact" Hartree-Fock exchange, the fraction being a parameter. It is likely that it was the introduction of this class of functionals that caused the community to recognize (or at least admit) that density functional theory was "semi-empirical" in nature, in the sense that it is "parameterized". Adding the Hartree-Fock exchange is "costly", but led to significant improvements in the description of reaction energies. While GH-GGA functionals were introduced in the early 1990's, they remain a mainstay in the application of density functional theory to chemistry. The B3LYP functional, in particular, is still widely used even though there are now much better choices.

Range Separated Hybrid Generalized Gradient Approximation (RSH-GGA)

The defining idea behind range separated GGA hybrids is that the "optimum" amount of Hartree-Fock exchange varies with electron- electron distance, from a small percentage in the long range limit to a large percentage in the short range limit, in the extreme from 0% to100%. Both ωB97X-D and ωB97X-V functionals are range- separated GGA hybrids, and both incorporate additional features to account for dispersive interactions. We consider the ωB97X-D functional combined with the 6-31G* basis set to be a good model for geometry calculation, and the ωB97X-V functional with a large basis set such as 6-311+G(2df,2p) to be a good model for energy calculation.

***meta* Generalized Gradient Approximation (mGGA)**

A *meta* GGA functional not only depends on the electron density and its gradient (as does a GGA functional) but also on the Laplacian (second derivative) of the electron density. As such, it can be construed as the next logical step beyond GGA in constructing a Taylor series expansion of the electron density. More commonly, *meta* GGA functionals are viewed as adding the so-called kinetic energy density to GGA. In either case, the addition can be constructed as second-order term in a Taylor series expansion of the electron density. The B97M-V functional is an example of a "pure" meta functional.

Global Hybrid *meta* Generalized Gradient Approximation (GH-mGGA)

These are strictly akin to global hybrid GGA (GH-GGA) functionals in that a fixed percentage of the Hartree-Fock exchange is introduced. The "only" difference is that a *meta*

GGA functional including a second-order term replaces a GGA functional. The M06-2X functional that is widely accepted as an excellent choice for thermochemical comparisons is an example of a GH-meta GGA functional.

Range Separated Hybrid *meta* Generalized Gradient Approximation (RSH-mGGA)

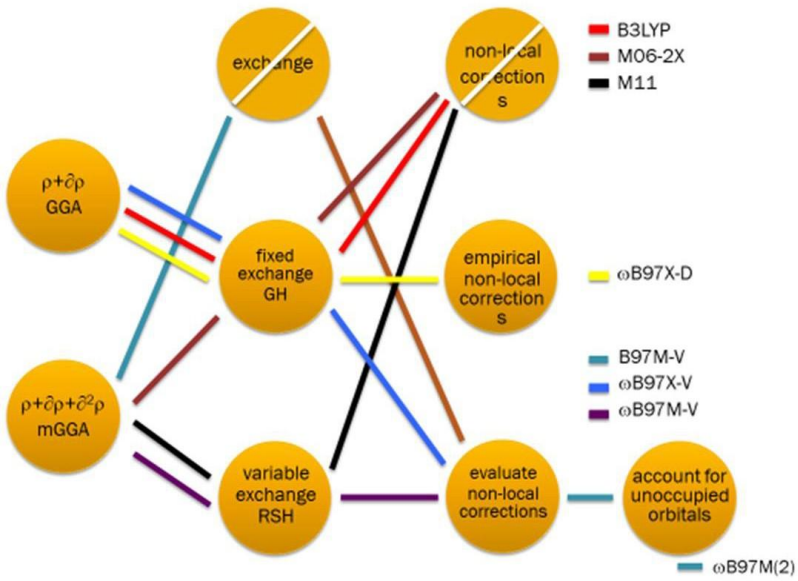
A range-separated *meta* GGA (RSH-mGGA) is identical to a range-separated GGA (RSH-GGA) functional except that a *meta* GGA functional has replaced the underlying GGA functional. M11 and ωB97M-V are examples of a range-separated meta GGA functionals.

Double Hybrid *meta* Generalized Gradient Approximation (DH-RSH-mGGA)

This class of functionals introduces a term that depends on excitations from occupied to unoccupied molecular orbitals. ωB97M(2) is an example of a double hybrid density functional model. At present, it is available for energy calculations only.

Non Local Corrections

Aside from the classifications outlined above, functionals may also be described as either strictly "local" or accounting for long- range dispersive van der Waals interactions. "Local" functionals are described in terms of a single integral over the three spatial coordinates, whereas non local functionals involve a double integral over two sets of coordinates. This may add significant cost to the calculations. For example, the range-separated GGA hybrid ωB97X-V functional uses the VV10 non-local correlation functional is 3-5 times more costly than the parent ωB97X functional. An alternative and less costly way to account for dispersive interactions is to add an empirical correction to the functional. So-called Grimme corrections are designated by appending "-D" as in ωB97X-D, or ""D3" as in B3LYP-D3, to the end of the functional.



Which classes of functionals are "best" and which functionals within each class are "best"?

1. LGA functionals are not suitable for chemical applications. While they are generally able to reproduce molecular geometry, better documented Hartree-Fock models with very small basis sets are equally successful. Of greater importance, LGA functionals present a poor account of the energetics of chemical reactions.
2. While GGA functionals provide marked improvement over LGA functionals they have largely been supplanted by GH- GGA functionals that are typically only slightly more costly. B3LYP, in particular, may be viewed as the first functional to be widely used (and continue to be used) among chemists, providing solid accounts of both geometry and reaction energies.
3. RSH-GGA functionals offer noticeable improvement over GH- GGA functionals for reaction energies, albeit with significant (factor of nearly 2) increase in computational cost. We consider the ωB97X-D functional to be a likely contender to replace B3LYP for routine chemical applications.
4. *meta* GGA, GH-meta GGA and RSH-meta GGA functionals, typified by B97M-V, M06-2X and M11 and ωB97M-V, respectively, appear to offer little improvement over RSH- GGA functionals.
5. Double hybrid RSH meta GGA functionals such as ωB97M(2) offer some improvement for reaction energies and conformational energy differences but at a significant computational cost.

"Cost" of Density Functional Models

The computation "cost" of density functional models depends on the class, the number of basis functions, η, and the number of points in the numerical integration grid, κ. GGA and pure meta- GGA functionals that do not require the Hartree-Fock exchange formally scale as $O(\eta^2\kappa)$. The other functionals combine this dependence with the cost of the Hartree-Fock exchange, which formally scales as $O(\eta^4)$ but in practice is $O(\eta^3)$ or lower. Finally, functionals such as ωB97X-V that directly account for dispersion have a step that scales as $O(\eta^2\kappa^2)$. This typically dominates the calculation. Times for calculation of the energy together with its gradient ("one step" in the optimization of molecular geometry) with the 6-31G* basis set for morphine ($C_{17}H_{19}NO_3$) follow. The ωB97M(2) functional is not included as it is presently not implemented for gradient calculation. The RI-MP2 model is provided for comparison.

B3LYP (GH-GGA)	7 mins.
ωB97X-D (RSH-GGA)	15 mins.
ωB97X-V (RSH-GGA + dispersion)	41 mins.
ωB97M-V (mGGA)	50 mins.
M06-2X (GH-mGGA)	19 mins.
M11 (RSH-mGGA)	22 mins.
RI-MP2	10 mins.

Density functional and wave function-based correlated models make use of the same basis sets as Hartree-Fock models, except that minimal and split-valence basis sets do not yield satisfactory results. Among the most popular models are the B3LYP/6-31G* and ωB97X-D/ 6-31G* density functional models and the MP2/ 6-31G* (second-order) Møller-Plesset model for equilibrium and transition-state geometry calculations, the ωB97X-V/ 6-311+G(2df,2p) density functional model for reaction energy calculations and the ωB97X-V/6-311+G(2df,2p)[6-311G*] and ωB97M-V/6-311+G(2df,2p)[6-311G*] dual basis set density functional models for conformer energy (Boltzmann weight) calculations* (Dual basis sets are described later in this chapter).

Density functional models and MP2 models provide an excellent account of the equilibrium geometries of organic molecules and are almost always to be preferred over Hartree-Fock models for this purpose. The one exception may be geometry calculation in very large molecules where practical consideration may require use of a very small basis set. Here the Hartree-Fock model with the 3-21G basis set may be the best choice. Note, however, that while density functional models provide a good account of molecules incorporating transition metals (and lanthanides), MP2 models do not. Both density functional models and MP2 models offer a good account of thermochemistry, including for reactions in which bonds are made or broken, activation energies being an important example.

Density functional models, with the exception of double-hybrid models, are available for the calculation of IR, Raman, NMR and UV/visible spectra, the last in conjunction with so-called time-dependent density functional models for calculations involving excited states. They are the methods of choice.

* The observation that geometry calculations require less costly models than either reaction or conformer energy calculations suggests that significant computer time saving can be achieved by using two different models to complete a task.

Thermochemical Recipes

A number of thermochemical recipes (combinations of different quantum chemical methods) have been formulated in an attempt to reproduce experimental heats of formation to within 4-8 kJ/mol. (Except for very simple molecules, errors in experimental heats are seldom less than this.) We review the details for G3(MP2), the simplest of the Gx recipes that have been formulated. (More costly procedures, G3, G4(MP2) and G4 are broadly similar.) G3(MP2) reproduces experimental heats of formation for a wide variety of small molecules to within 6 kJ/mol (mean absolute error), with only a small number of systems showing errors twice this amount. G3(MP2) may easily be applied to molecules comprising up to 10-15 heavy (non- hydrogen) atoms, but applications to molecules comprising more heavy atoms quickly become impractical.

Three different energies are required in the G3(MP2) recipe:

E_A MP2/6-31G* using the MP2/6-31G* equilibrium geometry

E_B MP2/6-311++G(2df, 2p) using the MP2/6-31G* equilibrium geometry

E_C QCISD(T)/6-31G* using the MP2/6-31G* equilibrium geometry

These are combined in order to take into account both the effects of increased basis set size and of electron correlation beyond the MP2 level. Note, that the two effects are treated independently in order to avoid having to perform a QCISD(T) calculation using a large basis set.

$$E = E_A + (E_B - E_A) + (E_C - E_A)$$

Two corrections are needed to bring the calculated energy in line with what is actually measured experimentally: a correction for zero-point energy; $E_{zero\text{-}point}$, and a correction to the enthalpy (energy) to account for finite temperature (298K), H(T).

$$E_{zero\text{-}point} = \frac{1}{2} h \sum_{i}^{vibrations} \nu_i$$

$$\Delta H(T) = H_{trans}(T) + H_{rot}(T) + \Delta H_{vib}(T) + RT$$

$$H_{trans}(T) = \frac{3}{2} RT$$

$$H_{rot}(T) = \frac{3}{2} RT \text{ (RT for a linear molecule)}$$

$$\Delta H_{vib}(T)^* = H_{vib}(T) - H_{vib}(0)$$

$$= Nh \sum_{i}^{vibrations} \frac{\nu_i}{(e^{h\nu_i/kT} - 1)}$$

h is Planck's constant, k is the Boltzmann constant, R is the gas constant and N is Avogadro's number. Both zero-point and temperature corrections require vibrational frequencies. These are obtained from Hartree-Fock/6-31G* calculations and require the corresponding equilibrium geometry.

To summarize, the G3(MP2) recipe requires two equilibrium geometry calculations (at HF/6-31G* and MP2/6-31G*), a HF/6-31G* frequency calculation, a MP2/6-311++G (2df, 2p) energy calculation and a QCISD(T)/6-31G* energy calculation. The QCISD(T) calculation dominates with increasing molecular size. Next most costly is the MP2/ 6-31G* geometry optimization, although both the large basis set MP2 energy and the HF/6-31G* frequency may be significant.

T1 is a simplified recipe intended to closely reproduce G3(MP2) heats of formation rather than experimental heats. It eliminates the QCISD(T)/6-31G* energy calculation (E_C), the zero-point energy calculation ($E_{zero\text{-}point}$) and the temperature correction to the enthalpy ($\Delta H(T)$), and replaces the large basis set MP2 energy calculation (E_B) with a dual-basis-set RI-MP2 calculation, where G3MP2large is the large basis set (the same basis set that is used in the G3(MP2) recipe) and 6-311G* is the small basis set. To compensate, an empirical correction based on the Hartree-Fock and RI-MP2 energies and Mulliken bond orders is introduced. This follows from a linear regression analysis involving G3(MP2) heats of formation for ≈1,050 molecules. T1 uses HF/6-31G* equilibrium geometries, instead of MP2/6-31G* geometries.

The T1 recipe reproduces G3(MP2) heats of formation with a mean absolute error of <1 kJ/mol. It reproduces the full set of ~2,000 experimental heats of formation in the NIST thermochemical database with a mean absolute error of 9 kJ/mol, comparable to the error of the G3(MP2) calculation. The T1 recipe also seems to provide a good account of conformational energy differences (although the experimental dataset that has been examined is very limited). This has been employed to identify the best (lowest heat of formation) conformer for the molecules in the Spartan Spectra and Properties Database.

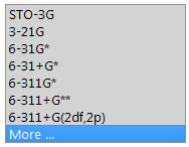
Except for semi-empirical methods and for thermochemical recipes, either **restricted** or **unrestricted scf methods** may be used. The former restricts paired electrons to the same orbital, whereas the latter allows the electrons to occupy different orbitals. The default (singlets with restricted methods and non-singlets with unrestricted methods) may be overridden using the **SCF** keyword from the **Options** field (**Appendix D**). Semi-empirical methods and thermochemical recipes are limited to unrestricted methods.

The semi-empirical models have essentially a built-in basis set. The Molecular Mechanics options do not incorporate basis sets. The Thermochemical Recipes use a series of models with basis sets pre-specified. The remaining quantum chemical models require the additional specification of basis set. **Spartan** incorporates a variety of common basis sets and includes the ability to import and use your own.*

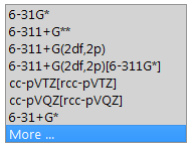
* See the Spartan **Help** content, **Help** menu > **Spartan'26 Help**.

Basis Sets

Except for molecular mechanics and semi-empirical methods and for thermochemical recipes, a basis set needs to be specified. Menus provide a number of choices.* For Hartree-Fock methods:



For all other methods:



The 3-21G(*) basis set for second-row and heavier main-group elements incorporates a set of d-type functions. While these are not occupied in the atomic ground state, they have been found necessary for proper description of bonding in molecules. 6-311++G (2df,2p) uses the 6-311++G (3d2f,2p) basis set for second-row elements.

The combination of a method, for example, ωB97X-D, and a basis set, for example, 6-31G*, constitutes a **theoretical model** or more simply a **model**. The nomenclature is to separate method and basis set by a "/", for example, ωB97X-D/6-31G*. By convention, specification of Hartree-Fock (HF) as a method is optional. That is, specification of basis set alone refers to Hartree-Fock models, for example, both HF/6-31G* and 6-31G* refer to combination of the Hartree-Fock method and the 6-31G* basis set. Methods where basis sets are not relevant for example, MMFF (molecular mechanics), or are not specified, for example, PM3 (semi-empirical) are themselves models.

* Other basis sets may be conveniently accessed by *clicking* on **More**. These include the Pople 6-31G or 6-311G basis sets followed by: (*i*d*j*f) or (*i*d*j*f, *k*p*l*d) where *i* is the number of sets of d functions on non-hydrogen atoms, j is the number of sets of ***f*** functions on non- hydrogen atoms, *k* is the number of p functions on hydrogen atoms and *l* is the number of sets of d functions on hydrogen atoms. If 0 eliminate p, d, f; if 1 eliminate the number; 6-31G or 6-311G with "+" or "++" inserted before G. The Dunning cc-pVxZ and aug-cc- pVxZ (x = 2,3,4,5) and Ahlrich, Weigend def 2 basis sets are also available.

Dual Basis Sets

If turned on (from the Basis Set Selector dialog), **Dual Basis** signifies that a Hartree-Fock, density functional, MP2 or RI- MP2 energy calculation (only) with a "large" basis set will be approximated by a calculation in which SCF convergence is first achieved using a "smaller" basis set and then corrected perturbatively for the effects of basis set extension. While the dual basis set approximation significantly affects total energies (and calculations with and without the dual basis set approximation cannot be mixed in energy comparisons), it has only a small effect on calculated relative energies and can reduce the cost of large-basis-set energy calculations by two to ten times (depending on the choice of the small and large basis sets)*. The conjunction of the 6-311G* basis set (small) with the 6-311G(2df,2p) basis set (large) is commonly employed for energy calculations with the ωB97X-V density functional model.

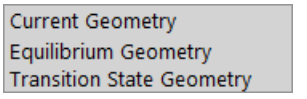
Pseudopotentials

Pseudopotentials are employed only where all-electron basis sets are unavailable and are used in conjunction with all-electron basis sets for elements where they are available. This is primarily for elements heavier than Kr, most importantly fourth-row main-group elements, second and third-row transition metals, and lanthanides. For example, an ωB97X-D/6-31G* calculation on ruthenium ethylene tetracarbonyl would employ a pseudopotential only for ruthenium and the all-electron 6-31G* basis set for hydrogen, carbon and oxygen. The overall method is referred to as ω**B97X-D/6-31G***. Pseudopotentials pertain to Hartree-Fock, density functional, MP2 and wave function based correlated calculations only, but not to RI-MP2 calculations or calculations using the STO-3G and 3-21G(*) basis sets.

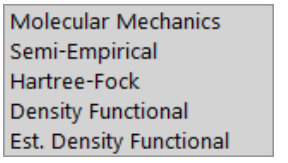
Start From

Energy calculations (only) from one model may be preceded by equilibrium or transition-state-geometry calculations at another (simpler) model.

The menu immediately to the right of **Start From** provides three choices:

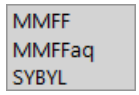


Selection of **Equilibrium Geometry** leads to a new set of menus, the leftmost being the class of methods.

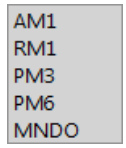


* It is important that the small basis set be a subset of the large basis set. The selections provided in the menus and in the "more" dialogs satisfy this condition.

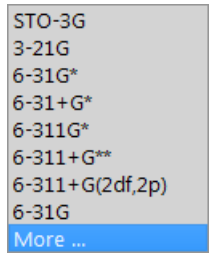
Selection of **Molecular Mechanics** leads to a menu.



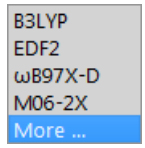
Selection of **Semi-Empirical** leads to a menu.



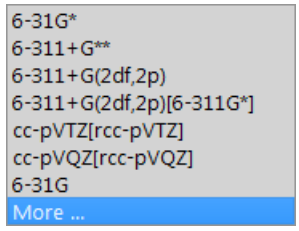
Selection of **Hartree-Fock** leads to a menu of available basis set (with **More** providing access to additional choices).



Selection of **Density Functional** leads to two menus, the first comprising available functionals (with **More** providing access to additional choices):



and the second comprising available basis sets (**More** provides additional options).



These selections are the same for **Transition State Geometry** except that **Molecular Mechanics** is not available.

Finally, ***Est. Density Functional*** leads to a single selection: **ωB97X-D/6-31G***.

Subject to

Spartan allows calculations to be carried out in the presence of geometrical constraints and/or with atoms which have been frozen in place.

If *checked*, **Constraints** signifies use of any previously defined constraints on distances, angles and dihedral angles into equilibrium and transition-state geometry optimization, conformation searching and generation of energy profiles. This does not apply to energy calculation, NMR spectrum calculation, similarity library generation or similarity analysis. See **Constrain Distance**, **Constrain Angle** and **Constrain Dihedral** under the **Geometry** menu (**Chapter 19**) for information about constraining geometrical parameters.

> Note that NOEs are not treated as constraints, but rather as post-calculation filters in establishing conformer distributions.

If *checked*, **Frozen Atoms** signifies that the coordinates of any atoms that have previously been frozen will not be moved during equilibrium and transition-state geometry optimization, conformation searching and generation of energy profiles. This does not apply to energy calculation, NMR spectrum calculation, similarity library generation or similarity analysis. See **Freeze Center** under the **Geometry** menu (**Chapter 19**) for information about freezing atoms. See also discussion in **Chapter 23** regarding the freezing of heavy atoms in returned hits from the ***CSD***.

Compute

Entries under this section request calculation of infrared, Raman, NMR and UV/visible spectra as well as a variety of descriptors used in QSAR analyses, and in the case of transition state optimizations, optionally compute an **I**ntrinsic **R**eaction **C**oordinate or IRC.

IR

If *checked*, calculates vibrational frequencies and infrared intensities together with the corresponding vibrational modes. These are available both in the **Spectra** dialog (**Spectra** under the **Display** menu; **Chapter 22**) and in the output (**Output** under the **Display** menu; **Chapter 22)**. Thermodynamic properties (enthalpy, entropy, heat capacity and Gibbs energy) are available from the **Thermodynamics** tab of the **Molecule Properties** dialog accessible from **Properties** under the **Display** menu (**Chapter 22**)*.

A handful of commonly used density functional models have been uniformly scaled to correct for systematic errors in **IR** frequency calculations. If a scaling factor has been employed, it will be listed in both the **Molecule Properties** dialog and the **Output Summary**.

* Entropy (and Gibbs energy) calculations are problematic in large part due to the harmonic approximation. Discussion is provided under **Properties and Spectra** in **Appendix A**.

Model:	**Scale Factor**:
EDF2/6-31G*	**.962**
B3LYP/6-31G*	**.952**
ωB97X-D/6-31G*	**.942**

Scale values may be changed from the **Thermodynamics** tab of the **Molecule Properties** dialog (**Properties** under the **Display** menu; **Chapter 22**).

Vibrational modes may be animated and an IR spectrum displayed from the **Spectra*** dialog (**Spectra** under the **Display** menu; **Chapter 22**).

Second derivatives required for frequency calculation are evaluated analytically for molecular mechanics, semi- empirical, Hartree-Fock and density functional models (except ωB97M(2)), but this involves numerical differentiation of analytical gradients for RI-MP2 and MP2 models. This is much more costly in terms of computation. Frequency calculations for double-hybrid density functional models and higher-order wave-function-based correlated models are not available.

Raman

If *checked*, calculates vibrational frequencies (the same as infrared frequencies) and Raman intensities. These are available both in the **Spectra** dialog (**Spectra** under the **Display** menu; **Chapter 22**) and in the output (**Output** under the **Display** menu; **Chapter 22**). The Raman spectrum may be displayed from either the simplified or more general form of the **Spectra** dialog (**Spectra** under the **Display** menu; **Chapter 22**). Note that request for a Raman spectrum also results in an IR spectrum.

NMR

If *checked*, calculates NMR chemical shifts and empirical estimation of HH and CH coupling constants. These are then available in the **Spectra** dialog (**Spectra** under the **Display** menu; **Chapter 22**), in the output (**Output** under the **Display** menu; **Chapter 22**), from the **Atom Properties** dialog (accessible from **Properties** under the **Display** menu; **Chapter 22**) and as atom labels (**Configure...** under the **Model** menu; **Chapter 18**).

* A simplified spectra dialog is presented by default. A more general spectra dialog can be utilized by turning off the spectra pane checkbox. (**Settings** tab, **Preferences** dialog, see **Options** menu, **Chapter 25**).

^{13}C (proton decoupled) and ^{1}H (either with or without the influence of three-bond HH coupling) spectra from Hartree-Fock or density functional calculations may be displayed from either the simplified or more general form of the **Spectra** dialog (**Spectra** under the **Display** menu; **Chapter 22**)*. DEPT spectra (^{13}C spectra in which carbons with one or three attached protons point up, carbons with two attached protons point down and quaternary carbons are absent) may only be displayed from the general form of the **Spectra** dialog. Finally, 2D COSY (proton vs. proton) and HSQC and HMBC (proton vs. ^{13}C) spectra using either empirically estimated or calculated HH and CH coupling constants may be displayed from both simplified and more general forms of the **Spectra** dialog.

Uncorrected chemical shifts (in particular, ^{13}C shifts) are not sufficiently accurate to be useful for NMR spectra prediction. ^{1}H, ^{13}C and ^{19}F shifts calculated from the EDF2/6-31G* model have been corrected using a scheme involving both topology and bond orders to neighboring atoms. ^{13}C shifts only from the B3LYP/6-31G*, ωB97X-D/6-31G* and ωB97X-D/ 6-311G* models have been corrected by a more recent scheme involving topology and bond lengths (as opposed to bond orders) to neighboring atoms. Corrections for other quantum chemical models are not available. Errors for corrected shifts are generally a half to a third of those for uncorrected shifts making them more appropriate for NMR spectra prediction.

> Note that ^{13}C chemical shifts from the B3LYP/6-31G* model in previous versions of ***Spartan*** (prior to ***Spartan'18***) were corrected using a less accurate scheme based on local topology.

NMR line intensities are assumed to be proportional to the number of equivalent carbons or hydrogens.

Coupling Constants (Empirical, Calculated (Fermi Contact) Calculated

Unless otherwise specified, coupling constants are not calculated, but 3-bond HH coupling constants as well as both 2 and 3-bond CH coupling constants are available from empirical relationships. Longer range HH and CH coupling constants as well as CC coupling constants may be calculated. Coupling constant calculation within the Fermi contact approximation is 3-5 times more costly than chemical shift and without this approximation, an order of magnitude more costly. It is advised that this be requested only when necessary. Splitting patterns are based on the number of neighboring hydrogens and on estimated or calculated 3-bond HH coupling constants.

---

* Chemical shifts for other nuclei are available in the **Output** dialog (**Output** under the **Display** menu; **Chapter 22**) and may also be attached as labels (**Configure...** under the **Model** menu; **Chapter 18**). Chemical shifts for proton, carbon, nitrogen, fluorine, and phosphorous are given relative to standard compounds, for example, TMS in the case of both proton and carbon. A table of standards is provided in **Appendix L**.

Coupling Constants - Empirical, Calculated (Fermi Contact), Calculated

Unless otherwise specified, coupling constants are not calculated, but 3-bond HH coupling constants as well as both 2 and 3-bond CH coupling constants are available from empirical relationships. Longer range HH and CH coupling constants as well as CC coupling constants may be calculated. Coupling constant calculation within the Fermi contact approximation is 3-5 times more costly than chemical shift and without this approximation, an order of magnitude more costly. It is advised that this be requested only when necessary. Splitting patterns are based on the number of neighboring hydrogens and on estimated or calculated 3-bond HH coupling constants.

UV/vis

If *checked*, specifies that a series of excited-state energy calculations will be performed (following a ground-state calculation) and state-to-state energy differences will be computed. Excited state energies and absorption intensities are available both in the **Spectra** dialog (**Spectra** under the **Display** menu; **Chapter 22**) and from the output (**Output** under the **Display** menu; **Chapter 22**). A UV/visible spectrum can be displayed from either the simplified or general form of the **Spectra** dialog (**Spectra** under the **Display** menu; **Chapter 22**). The default number of excited states is 6 but may be changed using the **UV States** keyword (UVSTATES=X) from the **Options** field (see discussion later in this chapter and in **Appendix D**). UV/visible spectra calculations are available only for Hartree-Fock models (CIS for the excited states) and density functional models (time-dependent density functional theory for the excited states).

QSAR

If *checked*, specifies calculation of a number of QSAR descriptors that would not otherwise be available. These include descriptors based on the electron density surface, the electrostatic potential map and the local ionization potential map, as well as polarizability and logP.*

Intrinsic Reaction Coordinate

An additional control, **IRC** (**I**ntrinsic **R**eaction **C**oordinate), appears when a transition-state geometry has been requested. If *checked*, **IRC** signifies that the resulting transition state will be used to generate a pathway (an **intrinsic reaction coordinate**) leading first to reactant and then to product. (It is advisable to compute the IR spectrum prior to calculation of an intrinsic reaction coordinate.)

---

* LogP is estimated according to the method of Ghose, Pritchett and Crippen (*J. Computational Chem.*, **9**, 80 (1988)). An alternative procedure due to Villar (*J. Computational Chem.*, **6**, 681 (1991); *Int. J. Quantum Chem.*, **44**, 203 (1992)) is available.

The sequence of steps: reactant ⟶ transition state ⟶ product, will be placed in a new file ***document.IRC. identifier*** where ***document*** is the name of the document submitted, ***IRC*** designates the origin as an IRC calculation and ***identifier*** is the molecule identifier within the document. The default number of steps (40) may be changed using the keyword **IRCSTEPS** (see **Appendix D**). **IRC** is not available for semi-empirical models.

Total Charge

Total charge on the molecule. The vast majority of molecules are uncharged (neutral).

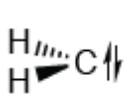
Molecules constructed from any of the 3D model kits (or from ChemDraw) are assumed to be neutral, that is **Total Charge** is **Neutral**. **Total Charge** may be changed either by *clicking* on the pull-down menu to the right of the box (**Cation**, **Dication**, or **Anion**, **Dianion**) or by typing a number in the box.

Molecules constructed from the 2D sketcher will, in almost all cases, "know" the total charge and the entry in the **Calculations** dialog will be correct.

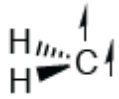
Total Charge is ignored for molecular mechanics calculations.

Unpaired Electrons

The vast majority of stable molecules have an even number of electrons arranged in pairs. These are referred to as singlets. Molecules with an even number of electrons but with two unpaired electrons are referred to as triplets. For example, methylene, CH_2, may either exist as a singlet or as a triplet.



singlet methylene



triplet methylene

The most common odd-electron molecules are referred to as doublets or more commonly, radicals. Here, all but one of the electrons is paired.

Molecules constructed from any of the 3D model kits (or from ChemDraw) assume that all electrons are paired, that is, **Unpaired Electrons** is **0**. **Unpaired Electrons** needs to be changed by *clicking* the menu to the right.

Molecules constructed from the 2D sketcher will, in almost all cases, "know" the number of unpaired electrons and the entry in the **Calculations** dialog will be correct.

Unpaired Electrons is ignored for molecular mechanics calculations.

Note that versions of ***Spartan*** prior to ***Spartan'14*** used the term **Multiplicity** to refer to the number of unpaired electrons: singlet = no unpaired electrons; doublet = 1 unpaired electron; triplet = 2 unpaired electrons.

Options

Program options may be specified using keywords entered into the **Options** field. Keywords may be either single words or expressions. Keyword=N indicates an integer argument, and keyword=F indicates a real argument. **Appendix D** contains a listing of commonly-used keywords. Additional keywords are referenced in the FAQ files accessed from ***Spartan's*** **Help** menu.

Global Calculations

If *checked*, signifies that settings in the **Calculations** dialog are to be applied to all molecules in the document.

To exit the **Calculations** dialog click on **Submit**, **Cancel** or **OK** at the bottom right of the dialog, or on . (**Submit** and **OK** are not available if the job is already executing.) Clicking on **OK** or on **Submit** overwrites any previous information. Additionally, **Submit** enters the job in the execution queue (see discussion later this chapter). Clicking on **Cancel** or on exits the **Calculations** dialog without saving any changes.

Surfaces

Spartan provides graphical display of the HOMO and LUMO among other molecular orbitals, the electron density, the spin density for molecules with unpaired electrons, the electrostatic potential, and the local ionization potential. Composite and custom maps are also supported. New in ***Spartan'26*** are additional gradient option for property maps. Now available in red-green-blue, red-white-blue, black-violet-sand, and gray-scale. User's can import their own customized gradient mapping (contact support@wavefun.com if this is of interest).

The concentration of electrons at any point in space may be determined using X-ray diffraction. While this is non-zero everywhere, the highest concentrations are in the immediate vicinity of the atoms. We refer to the ***electron density*** or more simply ***density*** as a surface that contains most (>99%) of a molecule's electrons and roughly corresponds to a space-filling model, that is, a van der Waals surface. It reveals overall molecular size and shape and demarks the steric barrier seen by encroaching molecules. A surface that encloses ~80% of molecules electrons reveals connections between atoms ("bonds") and will be referred to as the **bond density**.

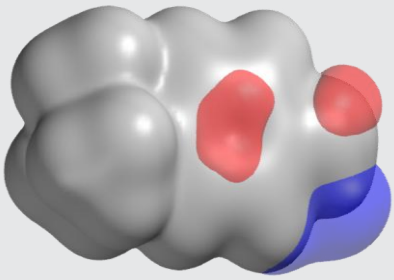
The ***spin density*** is the difference in the number of electrons of α and βspin at a point in space, locating unpaired electron(s) in a radicals or triplets.

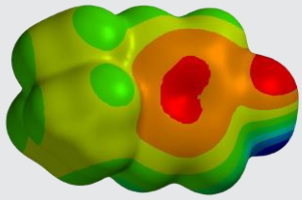
The ***electrostatic potential*** is the energy of interaction of a positive point charge with a molecule, assuming a fixed electron distribution. It represents a balance between repulsive interactions involving positively- charged nuclei and attractive interactions involving negatively-charged electrons. Regions where the balance tips toward attraction are said to be electron rich (basic) and subject to attack by electrophiles, while regions where the balance tips toward repulsion are said to be electron poor (acidic) and subject to attack by nucleophiles. Electron-rich regions such as lone pairs are frequently located outside the van der Waals surface, which suggests that they may be easily identified by constructing a surface of negative (attractive) electrostatic potential. Areas of positive electrostatic outside the van der Waals surface may be used to identify electron poor regions, such as those associated with acidic hydrogens. Note however, that the electrostatic potential is also positive (repulsive) throughout the region inside the van der Waals surface.

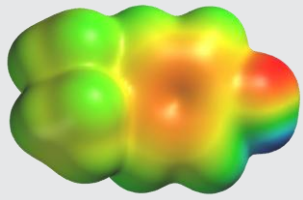
The ***local ionization potential*** indicates the ease or difficulty of electron removal (ionization) at a point in space. Like the negative regions of the electrostatic potential, regions of low local ionization potential are likely to be subject to attack by electrophiles. Regions of high local ionization potential have not proven to be useful indicators.

Note that neither the electrostatic potential nor the local ionization potential are experimental observables, although both relate to quantities that can be given clear chemical interpretation.

Aside from the electron density, any of the quantities listed above may be mapped onto any surface (except a molecular orbital surface). The only maps to have actually received widespread attention are those based on the electron density surface (depicting overall molecular size and shape). The most common maps are the electrostatic potential map, the local ionization potential map and the LUMO map. The ***electrostatic potential map*** "paints" the value of the electrostatic potential onto an electron density surface. By convention, colors toward red depict negative potential*, while colors toward blue depict positive potential, and colors in between (orange, yellow, green) depict intermediate values of the potential. Thus, an electrostatic potential map for *p-tert*-butylphenol will show oxygen to be red, its attached (acidic) hydrogen to be blue, the π faces of benzene to be orange or yellow and the *tert*-butyl group to be green. These are shown below with the banded (default) style as well as a continuous gradient style (with bands turned off).







An alternative representation using only shades of red and blue to designate negative and positive areas, respectively, and white to designate all other areas is also available (see **Surface Properties** from the **Properties** menu; **Chapter 22**).

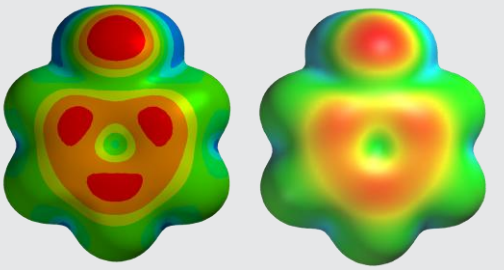
The main advantages that electrostatic potential maps have over separate electron density and electrostatic potential surfaces are clarity and compactness. The main disadvantage is that it provides information only about the contact surface and does not reveal how far electron-rich and electron-poor areas extend beyond the surface.

An alternative to an electrostatic potential map, referred to as an ***exposed electrostatic potential surface***, is a composite of three different surfaces: an electron density surface depicting overall molecular size and shape, a negative electrostatic potential surface identifying electron-rich regions and a positive electrostatic potential surface identifying electron- poor regions. These surfaces need to be generated and then displayed simultaneously. The electron density may either be displayed as an opaque solid or as a transparent solid (in order for the molecular skeleton to be seen inside). The two potential surfaces are best represented as transparent solids, the negative surface colored red and the positive surface colored blue, for example, for *p-tert*-butyl phenol.

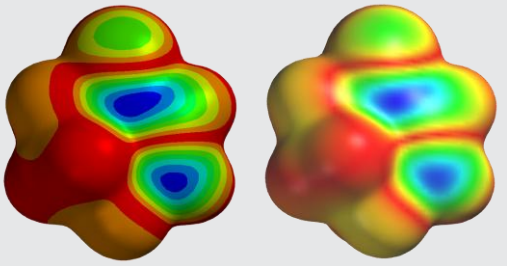
Some customization is required for best visualization of an exposed electrostatic potential map. Request a **density** surface and two **potential** surfaces from the **More Surfaces**. . . menu. All three should have "none" selected under properties options. First, *check* the box to the left of one of the **potential** surfaces. Select **Properties** from the **Display** menu and *click* on the surface to bring up the **Surface Properties** dialog. Remove the negative sign from the number to the right of **IsoVal:** and change the **Style** from **Solid** to **Transparent**. *Click* the downward arrow in the lower right corner of the **Surface Properties** dialog to toggle this to the **Surface Style** dialog. Use the slider bars to set the color of the surface to blue (Red 0%, Green 0%, and Blue 100%). Click the upward arrow in the lower right corner to return to the **Surface Properties** view. From the **Surfaces** dialog, *check* the box next to the other **potential** surface. A gray surface will appear above and below the aromatic ring and in the area around the lone pairs on the oxygen atom. *Click* on this surface. From the updated **Surface Properties** dialog, select **Transparent** for the surface **Style**. Again, *click* the downward arrow to toggle to the **Surface Style** view and set the color for this surface to red (Red 100%, Green 0%, Blue 0%). Finally, from the **Surfaces** dialog, *check* the box next to the **density** surface.

Three surfaces are now simultaneously displayed, electron density in a neutral (gray) color, the exposed negative electrostatic potential in red (indicating a potential hydrogen bond acceptor and/or positive ionizable group), and the exposed positive electrostatic potential in blue (indicating a potential hydrogen bond donor and/or negative ionizable group). This kind of multiple display may also be constructed with other combinations of surfaces.

Note that the exposed electrostatic potential surface provides the same information as the electrostatic potential map. Note also the broad similarity of this graphic to CFD representations discussed previously (see **CFD's** under the **Model** menu in **Chapter 18**). Hydrogen-bond acceptors and/or positive ionizable sites are associated with protruding negative potential surfaces, whereas hydrogen-bond donors and/or negative ionizable sites are associated with protruding positive potential surfaces.



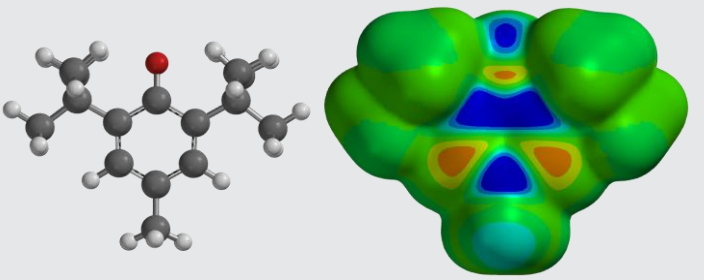
The ***local ionization potential map*** "paints" the value of the local ionization potential onto an electron density surface. By convention, colors toward red indicate low ionization potential, while colors toward blue indicate high ionization potential. Thus, the local ionization the local ionization potential map for aniline shows that the *ortho* and *para* ring positions have a lower ionization potential than the *meta* positions, consistent with the known directing ability of an amino group in electrophilic aromatic substitution.



The ***LUMO map*** "paints" the absolute value of the lowest- unoccupied molecular orbital (the LUMO) onto an electron density surface. By convention, colors near blue indicate high concentration of the LUMO, while colors near red indicate low concentration. Given that the LUMO designates space available for a pair of electrons, a LUMO map indicates where nucleophilic attack would likely occur.

while colors near red indicate low concentration. Given that the LUMO designates space available for a pair of electrons, a LUMO map indicates where nucleophilic attack would likely occur. For example, a LUMO map for cyclohexenone shows concentration in two regions, one over the carbonyl carbon and the other over the β carbon, consistent with both carbonyl addition and Michael (conjugate) addition. Note, however, that the LUMO may not be the "important" unoccupied molecular orbital in so far as designating the site of nucleophilic addition. This may instead be an orbital of higher energy (LUMO+1, LUMO+2, ...). These surfaces are also easily displayed, and some exploration may be needed before settling on an appropriate representation of nucleophilic reactivity.

The ***spin density map*** "paints" the value of the spin density onto an electron density surface. It is applicable only to molecules with unpaired electrons and will otherwise not appear in the menu. By convention, colors near blue indicate high concentration of spin density, while colors near red indicate low concentration. For example, a spin density map for the radical resulting from loss of hydrogen from 3,5-di-*tert*- butylhydroxytoluene (BHT) shows that the spin has delocalized from oxygen onto the *ortho* and *para* ring positions.



This radical would be expected to be particularly stable, which explains why BHT acts as an antioxidant (scavenging less favorable localized radicals).

All of these maps use an electron density surface to delineate overall molecular size and shape. However, not all regions on this surface are actually accessible and therefore available for interaction with their environment (or with an incoming reagent). ***Spartan*** allows these regions to be identified.*

Surfaces (including those underlying maps) connect points of equal value (they are isosurfaces), and may be displayed as an arrangement of dots, a mesh, or an opaque or translucent solid. Examples of graphical output in orthogonal projection are provided in **Figure 21-1**. Surfaces (and maps) may also be rendered in perspective (see **Chapter 18**) and in stereo (see **Chapter 2**).

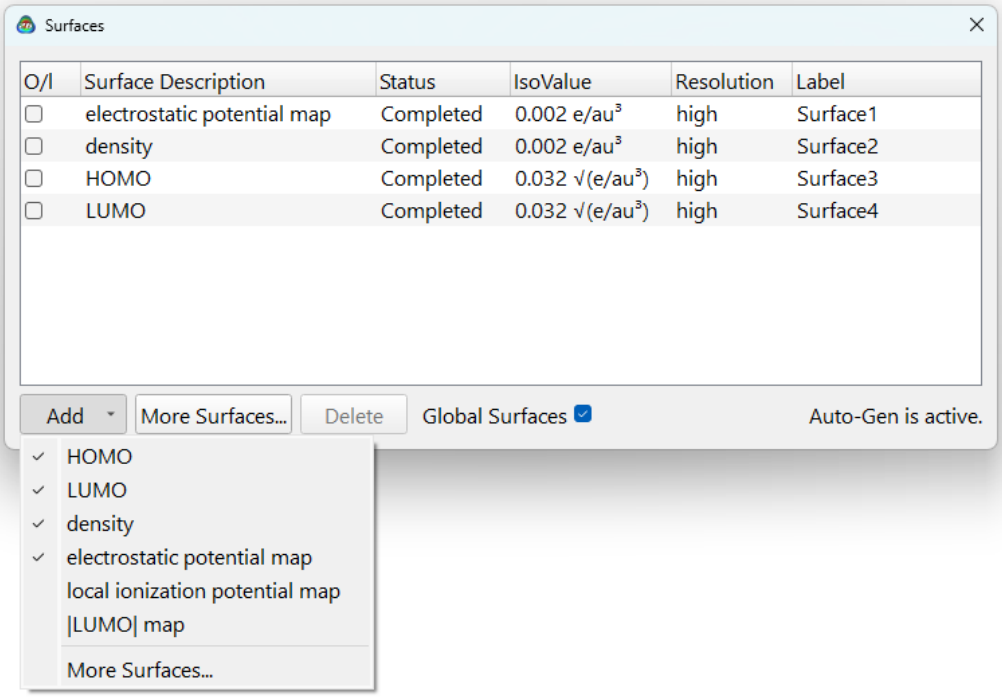
Calculated quantities may also be displayed as two dimensional contour plots (slices). Unlike surfaces and maps, these can be translated, rotated and zoomed independently of the molecular skeleton. An example of a slice display is provided in **Figure 21-1**.

Several different surfaces, maps and slices may be simultaneously displayed. In addition, any of the usual structure models may be displayed along with the graphic. The total display can become very complex, and selective use of meshes and/or translucent solids (as opposed to opaque solids) may facilitate visualization.

* By default, a region on a density surface is designated as inaccessible if a sphere of radius 1.0 Å centered on a line normal to the surface and touching a point in the middle of the region, impinges on any other regions of the density surface. The sphere radius may be changed in the **Settings Preferences** dialog (**Preferences** under the **Options** menu; **Chapter 25**).

Discussion of the utility of graphical models for describing molecular structure and chemical reactivity is provided in ***Topics*** available under the **Activities** menu.

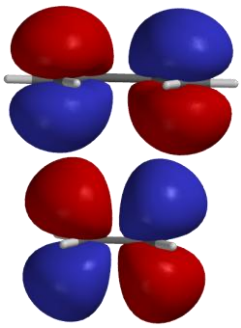
Selection of **Surfaces** from the **Setup** menu results in display of the **Surfaces** dialog.



This contains a box at the top for listing requested surfaces and property maps.

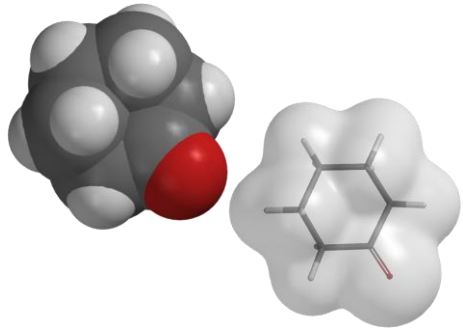
Figure 21-1: Examples of Graphical Displays Available in *Spartan*

Frontier orbitals for a symmetry-allowed Diels-Alder reaction,



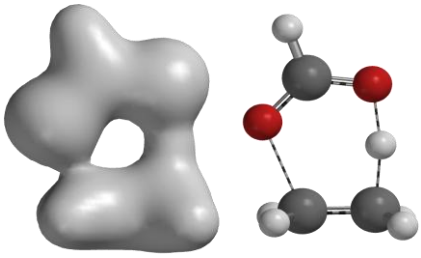
showing interaction of the HOMO of 1,3-butadiene and the LUMO of ethylene.

Space-filling model and electron density surface of cyclohexanone,



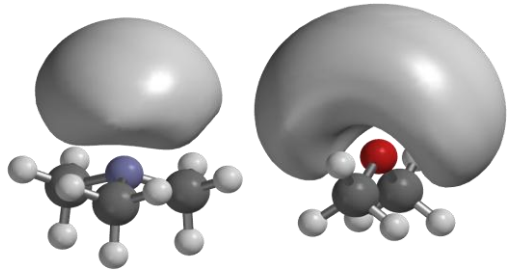
showing overall molecular size and shape.

Electron density surface (0.08 electrons/ au³) of transition structure for pyrolysis of ethyl formate,



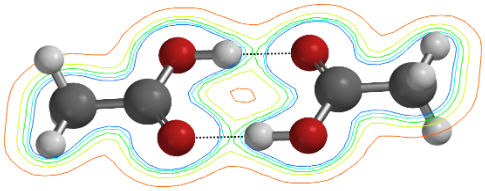
showing bonding in the transition state.

Electrostatic potential surfaces (-40 kJ/ mol) of trimethylamine (left) and dimethyl ether (right),



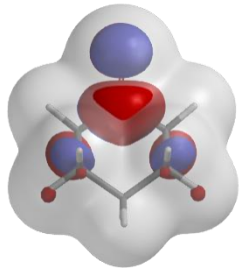
showing the lone pairs on nitrogen and oxygen, respectively.

Electron density slice for acetic acid dimer,



showing hydrogen bonding.

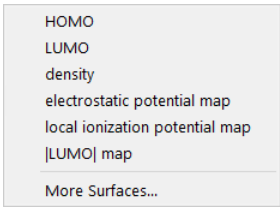
Simultaneous display of the LUMO and the electron density surfaces of cyclohexanone,

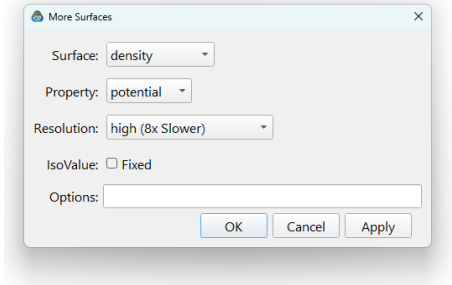


showing accessibility for nucleophilic attack.

Common Surfaces and Property Maps

Add at the bottom of the dialog is used to specify a number of commonly-used graphical surfaces and property maps*. *Clicking* on it leads to a menu.

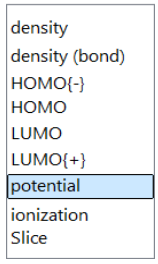




Selection of all but the last entry in the menu leads to a request for the analogous surface or map. This will be calculated at medium resolution unless a different resolution has been previously selected (see below). Additional surfaces and maps or the same surfaces or maps at different resolution may be requested by selecting **More Surfaces...** from the menu (or by *clicking* on **More Surfaces...** at the bottom of the **Surfaces** dialog). This leads to the **Add Surfaces** dialog that contains three menus and a check box:

Surface

Available surface types appear under the **Surface** menu.



Density portrays overall molecular size**, **density (bond)** locates bonds, **HOMO{-}**, **HOMO**, **LUMO**, **LUMO{+}**, **SOMO** are molecular orbitals, **potential** is the electrostatic potential, **ionization** is the local ionization potential and **spin density** is the spin density. **SOMO** (the singly-occupied molecular orbital) and **spin density** entries appear only for molecules with unpaired electrons.

> Unlike the molecular orbitals, the electron density and the electrostatic potential, the local ionization potential does not go to zero with increasing distance. This makes its use as a surface problematic.

* Additional selections are provided if the molecule has unpaired electrons.

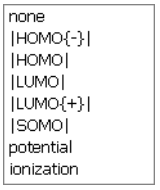
** An "electron density" surface, an electrostatic potential surface and an electrostatic potential map may also be constructed based on a molecular mechanics calculation. In this case, the density will be based on van der Waals radii and the potential will be based on charges.

Selection of **HOMO{-}** and **LUMO{+}** results in display of a box to decrement the HOMO and increment the LUMO. This allows any molecular orbital to be specified.

Slice designates that a plane will cut through the graphic defined by **Property** (see discussion following).

Property

Properties to be color coded onto surfaces to make property maps appear in the **Property** menu.

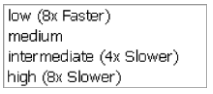


Available properties are the molecular orbitals (**HOMO{-}**, **HOMO**, **LUMO**, **LUMO{+}**, **SOMO**), the electrostatic potential (**potential**), the local ionization potential (**ionization**) and the spin density (**spin density**). **none** indicates that no property is to be mapped onto the surface). **SOMO** (the singly-occupied molecular orbital) and **spin density** entries appear only for molecules with unpaired electrons. **SOMO** (the singly-occupied molecular orbital) and **spin density** entries appear only for molecules with unpaired electrons. As with **Surface** above, selection of **HOMO{-}** and **LUMO{+}** leads to a decrement (increment) box.

A spin control designating **Alpha** or **Beta** will be displayed if **Unpaired Electron** (in the **Calculations** dialog) is set to a value other than **0**, or if **SCF=UNRESTRICTED** has been specified in the **Options** field (**Calculations** dialog; see **Appendix D**) and if **HOMO{-}**, **HOMO**, **LUMO** or **LUMO{+}** has been selected for **Surface** or for **Property**. *Clicking* on it toggles it between **Alpha** and **Beta**. **Alpha** designates that the molecular orbital either to be displayed as a surface or mapped as a property corresponds to α spin; **Beta** designates that the molecular orbital corresponds to β spin.

Resolution

Selection of surface resolution is from the **Resolution** menu.



Medium resolution generally is intended for routine work, low resolution is used to get rough images very quickly, while intermediate and high resolution may be employed to obtain publication quality graphics. Both calculation time and disk storage increase significantly with increasing resolution.

IsoValue

This allows a fixed isosurface to be generated in place of a three- dimensional grid of points (from which different isosurfaces, and/ or slices can be constructed).* Fixed isosurfaces are particularly attractive for property maps (that are generally based on a density surface representing molecular size and shape) as they require significantly less computer time and disk storage. Note that a fixed density surface needs to be based on a specific isovalue (see below) and not on percentage enclosure of the total number of electrons.**

Checking the box to the left of **Fixed** leads to display of the isovalue (0.002 electrons/au^3 in the case of density which corresponds to >99% enclosure of the total number of electrons). This value may be changed prior to surface generation by editing the text box that appears to the right. **Fixed** may not be specified in conjunction with **Slice** as a **Surface**.

Following **Surface**, **Property**, **Resolution** and (optionally) spin selection, *clicking* on **OK** adds the requested graphic to the list and removes the (**Add Surfaces**) dialog. *Clicking* on **Apply** adds the requested graphic to the list but leaves the dialog on screen. *Clicking* on **Cancel** does not add a graphic to the list but removes the (**Add Surfaces**) dialog.

The process (*clicking* on **Add...**, surface and property selection, *clicking* on **OK** or **Apply**) may be repeated as required.

An existing graphic or a request for a graphic may be deleted from the list by first highlighting (*clicking* on) it and then *clicking* on **Delete**.

Global Surfaces

If *checked*, **Global Surfaces** signifies that the requested surfaces will be be calculated for all members of the list.

Only one copy of the **Surfaces** dialog may appear on screen, and any actions relate to the currently selected molecule. The dialog may be removed by *clicking* on .

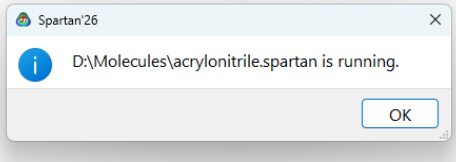
---

* The "electron density" surface resulting from a molecular mechanics calculation is fixed as it is based on use of van der Waals radii.

** When **Fixed** is not checked, isovalues may be computed on-the-fly to enclose a specified percentage of the total number of electrons. Discussion is provided in the **Surface Properties** dialog (**Properties** under the **Display** menu; **Chapter 22**).

Submit

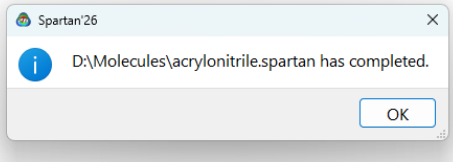
Following setup of a molecular mechanics or quantum chemical calculation, or a similarity analysis, as well as any requests for properties, spectra and/or graphical displays, the required calculations will begin when **Submit** is selected. If the job has not previously been named, selection of **Submit** triggers a request for a name. If the document being submitted comprises a single molecule and the molecule exists in the ***Spartan Spectra and Properties Database***, the name in the database will be presented as a default name. Otherwise, the default name presented will be ***spartan*** for the first job and ***spartanX*** (X an integer starting with 1 for all successive jobs). After a name has been provided (or if a name already exists or if the default name is accepted) a dialog appears indicating that the job has actually been submitted.*



Click on **OK** to remove it. After a job has started, and until it has completed, all associated files will be designated read only.

> In addition to the local (Windows, Macintosh or Linux) machine, ***Spartan***'s computational codes may run on a remote (licensed) server. The only exception is the similarity analysis code which at present must be run on the local machine. If remote submission has been set up (see **Available Servers** from the **Preferences** dialog under the **Options** menu; **Chapter 25**), selection of **Submit** will also request a name for the destination machine.

Another dialog appears following completion of a calculation.

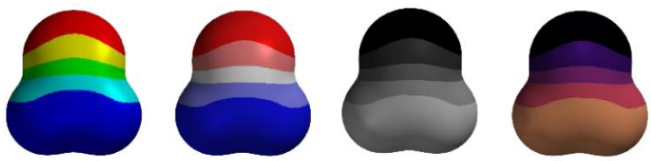


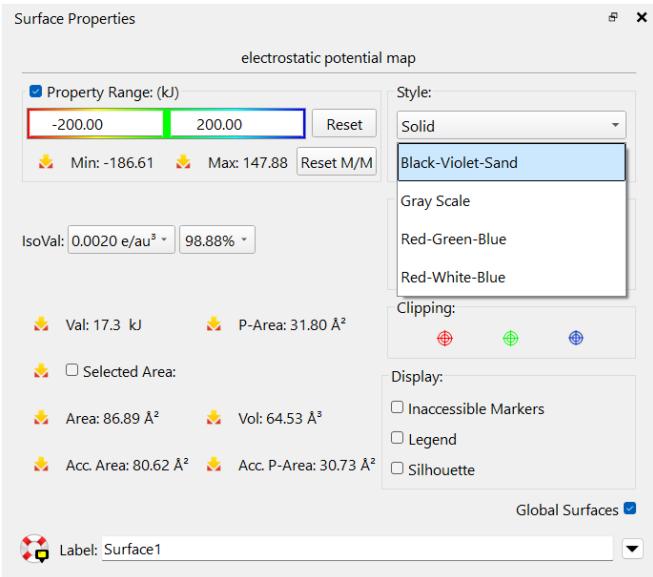
* The job is submitted to a job queue and will be submitted for execution only when released from this queue. See **Monitor** under the **Options** menu (**Chapter 25**) for discussion.

Click on **OK** to remove it.

Some types of jobs, including requests for a conformer distribution and an energy profile, lead to an additional document being created for each molecule in the original document. These new documents are named ***document.task.identifier*** where ***document*** is the name given to the original document, ***task*** is the type of job, ***Conf*** for a conformer distribution, ***NMR*** for an NMR spectrum calculation and ***Prof*** for energy profile, and ***identifier*** identifies the molecule inside the original document.

In the case where there is only one molecule in the original document, "***.identifier***" is not included in the name and a query dialog is provided asking whether the resulting ***Spartan*** document is to be opened. Otherwise, the resulting documents (one for each molecule in the original document) will need to be opened explicitly (**Open** under the **File** menu; **Chapter 16**).

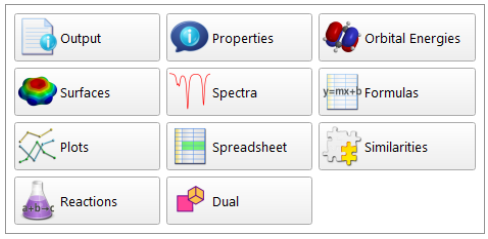




Shown above, new color options for property maps, accessible from the **Surface Properties** dialog. For easy access, *right-click* on a displayed surface and choose **Properties**.

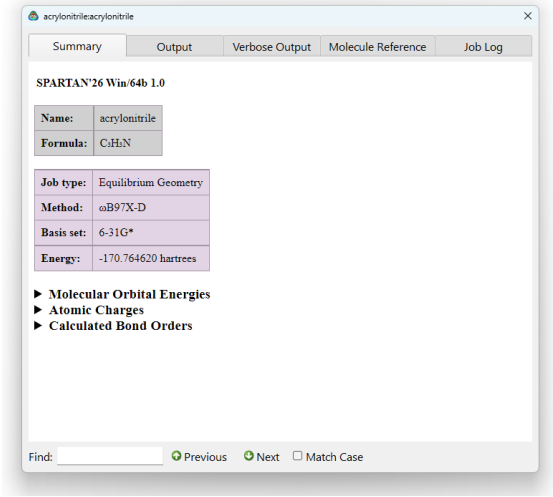
Chapter 22
The Display Menu

*Functions available under the **Display** menu provide for text, dialog, spreadsheet and graphical displays following molecular mechanics or quantum chemical calculations. Functions are also available to query a variety of on-screen objects, display and compare IR, Raman, 1 and 2D NMR and UV/visible spectra, animate vibrational motions, construct relationships among calculated properties or between reactants and products, perform linear regression and prepare plots from spreadsheet data, display similarities between molecular structures and between molecular structures and pharmacophores and calculate reaction energies.*



Output

Selection of **Output** opens a summary window.



The tabs at the top of the dialog select the type of output.

Summary output is initially displayed and provides a snapshot of calculated data, in particular, the energy and any spectral quantities. It is generated in real time in the graphical interface using information provided by ***Spartan***'s compute codes as well as any filters present in the interface.* ***Spartan*** provides savable/printable tables of IR/ Raman, NMR (chemical shifts and HH and CH coupling constants) and UV/visible spectra.** **Output** provides standard text output. **Verbose Output** contains a more granular (and much longer) output, which is mainly provided to aid in tracking down problems in "failed" calculations. This is "pruned" by default unless **Keep Verbose** is checked in the **Setting Preferences** dialog (**Preferences** under the **Options** menu; **Chapter 25**).*** **Molecule Reference** provides the literature reference for data retrieved from CSD and/or PDB. **Job Log** contains diagnostic information. The contents of the output window may be printed by *right clicking* inside it and selecting **Print** from the menu or copied to the clipboard by *right clicking* inside it and selecting **Copy** from the menu. **Find...**, **Find Next**, and **Save As PDF** functions are also available from the contextual menu. An output window is associated with each document and pertains to the selected molecule. Several output windows associated with different ***Spartan*** documents may be opened at once. To close *click* on at the top right.

Text output is not accessible from the **Output** menu for jobs that are actively running or are in the execution queue. However, standard **Output** and **Verbose Output** for jobs that are actively running (both locally or remotely) may be examined in "real time" by accessing the **Monitor** under the **Options** menu (**Chapter 25**). In addition, the current structure (in a geometry or transition-state optimization) can be examined and

* For example, calculated infrared/Raman frequencies and NMR chemical shifts are "corrected on the fly" as they are opened in the interface. This means that, to take advantage of future improvements in the correction scheme, you do not need to recalculate the uncorrected spectra. Coupling constants are not corrected.

** Contents of the tables in the summary output may be copied onto the clipboard by clicking on and then pasted into Excel (or another spreadsheet) for further analysis.

*** Irrespective of the setting, the last 100 lines of the verbose output are kept if the jobs fails.

queried. Note that in most cases, the **Monitor** presents text output in a cascading mode, you will need to "toggle" the triangle to the left of the document name to get to specific molecule output, and in some cases, an additional "toggle" of the molecule may be necessary in the case of multi-step calculations (conformational jobs, for example).

Properties

Spartan provides specialized dialogs for reporting (and in cases where the underlying calculation is not affected), changing the properties of molecules, atoms, bonds, surfaces and constraints, and for editing CFD's and substituent libraries. For plots created in ***Spartan***'s main window (or brought into this window from the spectra and plot panes), **Properties** may be used to change default plot styles, plot limits and fitting functions. **Properties** dialogs also quantify the difference between calculated and experimental NMR chemical shifts (assuming that the latter are available), display 2D drawings, alter model style and colors, atom labels and conformer settings as well as "fix" problems with structure, for example, bond typing and bond lengths to hydrogen that may be in error. Note that the last of these will affect the underlying calculation. Only one **Properties** dialog may be open, and this refers either to the selected molecule (**Molecule Properties**), or to the selected component (atom, bond, etc.) or attribute (graphical surface, substituent, etc.) of the selected molecule (**Atom Properties**, **Bond Properties**, **Surface Properties**, **Substituent Properties**, etc.), or to a plot, including spectra plot (**Plot Properties**, **Curve Properties**, etc.), or fitting function (**Regression Properties**). Selection of a different molecule leads to the dialog of whatever tab was previously selected in the **Molecule Properties** dialog for that molecule. Dialogs that refer to components/attributes of the (newly selected) molecule follow by *clicking* on the component/attribute.

With the **Properties** dialog on screen, *clicking* on a component/ attribute brings up the appropriate **Properties** dialog. For example, *clicking* on an atom brings up the **Atom Properties** dialog. *Clicking* on a different component/attribute brings up the appropriate **Properties** dialog. *Clicking* a second time on the same component reverts to the dialog of whatever tab was previously selected in the **Molecule Properties** dialog (with the exception of the clicking on a surface – which will update surface properties based on the location the user *clicks* upon.

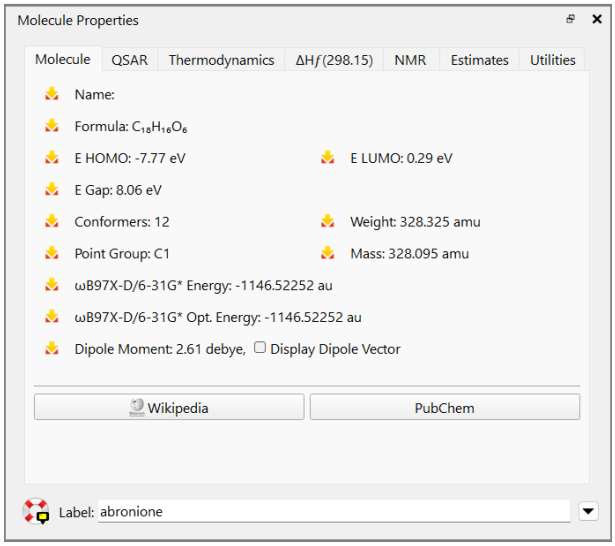
Properties dialogs have an associated **Utilities** or **Style** dialog. These provide additional style and color controls useful for highlighting (or de-emphasizing) a particular molecule, component or attribute, and are reached by *clicking* on ▼ at the bottom right of the appropriate **Properties** dialog. Return to the **Properties** dialog follows from *clicking* on ▲ at the bottom right of the associated **Utilities**/**Style** dialog.

The **Properties** (or **Utilities**/**Style**) dialog may be removed by *clicking* on ✕ .

Molecule Property Tabs:

The **Molecular Properties** dialog comprises six tabs: **Molecule**, **QSAR**, **Thermodynamics**, ***ΔHf* (298.15)**, **NMR**, **Estimates**, and **Utilities**. Entries under the **Molecule** tab relate to common molecular properties, only some of which depend on the selected level of calculation.

Molecule



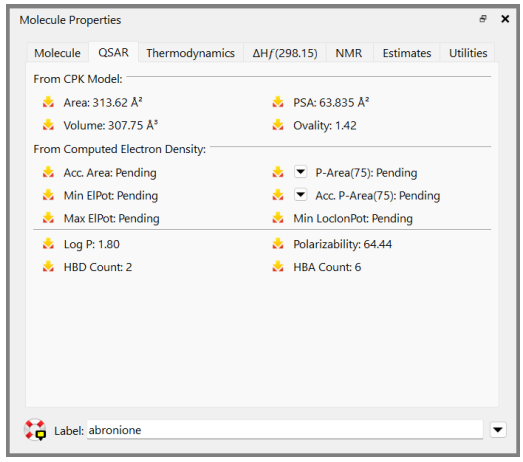
Molecule properties include the name and molecular formula, the energy (in au), the heat of formation (in kJ/mol), or the sum of the strain energy and the intramolecular interaction energy (in kJ/ mol) depending on the theoretical model, the HOMO and LUMO energies (in eV), the dipole moment (in Debye), the molecular weight* and mass* (both in amu), the point group, the predicted number of conformers, an Energy and an Opt. Energy.

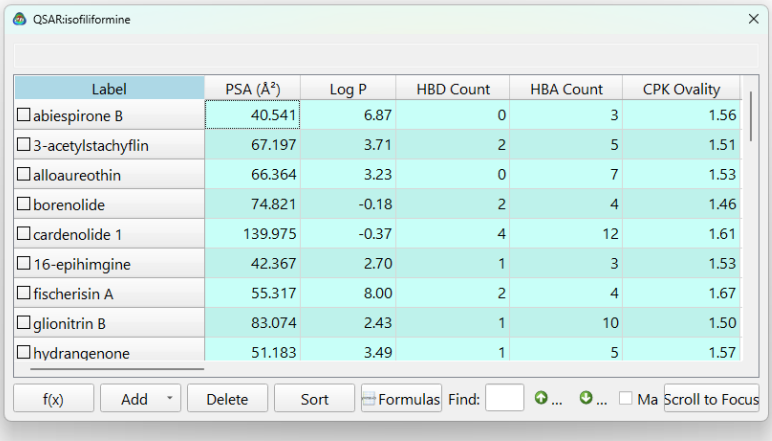
In most cases the Energy and Opt. Energy fill be the same, however it is not uncommon to optimize geometry with a less rigorous model and perform energy calculation with a more rigorous model (this can be set from as a single "calculation". These may be posted to the spreadsheet using the buttons, or *dragged* into the spreadsheet, the formula editor (see **Formulas** in this chapter), or onto any of the dialogs associated with the Spartan Spectra and Properties Database (see **Database** under the **Search** menu; **Chapter 23**).

The dipole moment vector may be added to the model by *checking* the box to the left of **Display Dipole Vector**. Buttons at the bottom of the dialog access available Wikipedia (based on name) and PubChem pages (based on InChI string). **Label** identifies the molecule in a document and appears in the first column of the spreadsheet, it can be overwritten from either location (see **Spreadsheet**, later this chapter).

QSAR

Entries under the **QSAR** tab provide additional properties, some of which may be important in QSAR/regression analyses. The internal **Spreadsheet** includes the ability to perform regression analysis, making quantitative structure-activity relationship studies relatively easy to assess.





Label	PSA (Å²)	Log P	HBD Count	HBA Count	CPK Ovality
abiespirone B	40.541	6.87	0	3	1.56
3-acetylstachyflin	67.197	3.71	2	5	1.51
alloaureothin	66.364	3.23	0	7	1.53
borenolide	74.821	-0.18	2	4	1.46
cardenolide 1	139.975	-0.37	4	12	1.61
16-epihimgine	42.367	2.70	1	3	1.53
fischerisin A	55.317	8.00	2	4	1.67
glionitrin B	83.074	2.43	1	10	1.50
hydrangenone	51.183	3.49	1	5	1.57

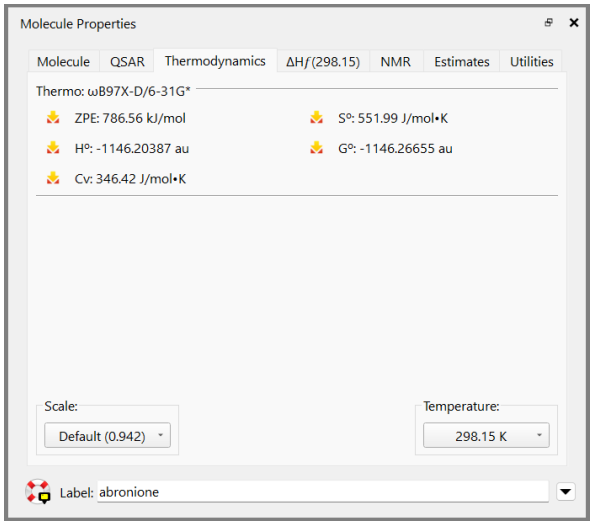
* Molecular weight is what would be obtained from looking at a macroscopic sample, that is, consisting of a mixture of different atomic isotopes, for example, ^{35}Cl and ^{37}Cl. Molecular mass is what would be observed from a mass spectrometry experiment of the dominant parent ion.

These include: the area, volume, polar surface area (PSA)* and ovality obtained from a space-filling model, as well as other structure-dependent indicators: LogP, polarizability and the number of hydrogen-bond donor (HBD) and acceptor sites (HBA).** All of these are independent of the level of calculation. Quantities that depend on the level of calculation and based on the electron density surfaces as well as on electrostatic potential maps are also available: the accessible area, the polar area and accessible polar area corresponding to absolute values of the electrostatic potential greater than 75, 100 and 125 kJ/mol (selection is made by repeated *clicking* on ▼), the minimum and maximum values of the electrostatic potential (as mapped onto an electron density surface) and the minimum value of the local ionization potential (as mapped onto an electron density surface). These quantities are not calculated unless explicitly requested by checking **QSAR** inside the **Calculations** dialog (**Calculations...** under the **Setup** menu; **Chapter 21**).

Thermodynamics

Entries under the **Thermodynamics** tab provide the zero-point energy, the enthalpy, the constant volume heat capacity, the entropy and the Gibbs energy. Except for the zero-point energy, all depend on temperature. The default setting (298.15 K) may be changed. All require vibrational frequencies, which can be requested from the **Calculations...** dialog (**Calculations...** under the **Setup** menu; **Chapter 21**).

Note that calculated frequencies are consistently larger than experimental values due in part to the harmonic approximation. They can be (and often are) scaled by a number <1. Several common models including EDF2/6-31G*, B3LYP/6-31G*, and ωB97X-D/6-31G* are automatically scaled. Other models can be scaled by typing into the keypad.



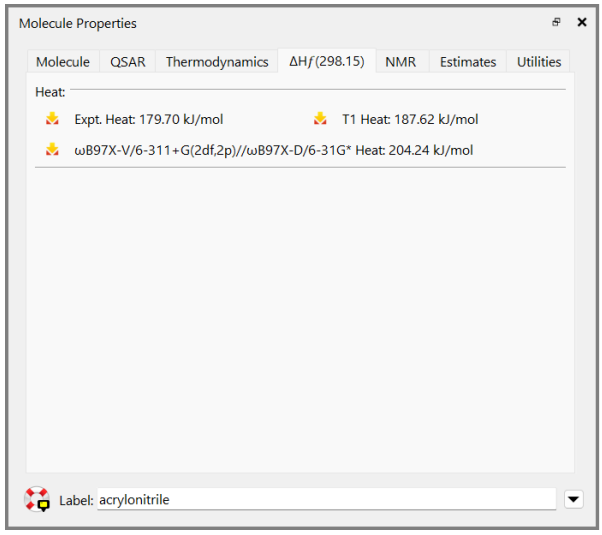
* Polar surface area is defined as the area due to nitrogen and oxygen and any attached hydrogens. Polar surface areas corresponding to arbitrary alternative definitions are available for posting into the spreadsheet using the **PAREA** function. See **Table 22-3**.

** Counts of hydrogen-bond donors and acceptors follow from atomic connectivity together with special case designations for common organic functional groups.

As discussed under **Properties and Spectra** in **Appendix A**, the entropy and Gibbs energy are subject to considerable uncertainty due to the underlying harmonic approximation. Specifically, low-frequency vibrational motions which dominate the entropy and significantly contribute to the temperature dependence of the enthalpy are poorly described by the harmonic approximations.*

ΔH°*f*(298.15)

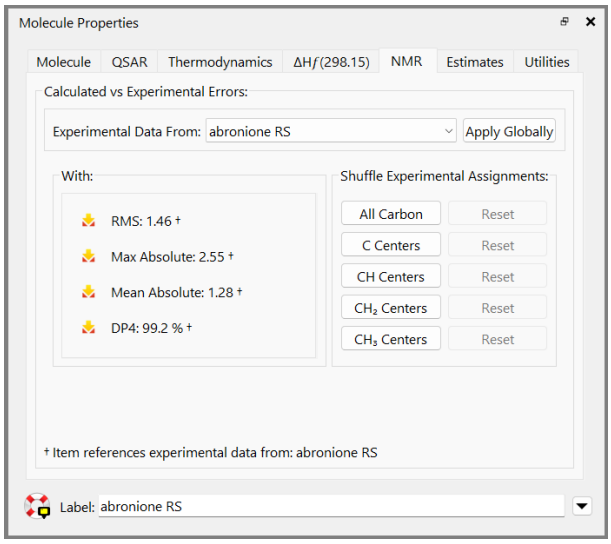
Previous versions of ***Spartan*** displayed heat of formation data in the main **Molecule** tab of the **Properties** dialog; this has become its own tab current versions. Available are **Expt. Heat** (if included in the NIST database), along with **T1** heat of formation as well as the heat of formation from the ωB97X-V/6-311+G(2df,2p)//ωB97X-D/6-31G* - if available in the **SSPD**.



NMR

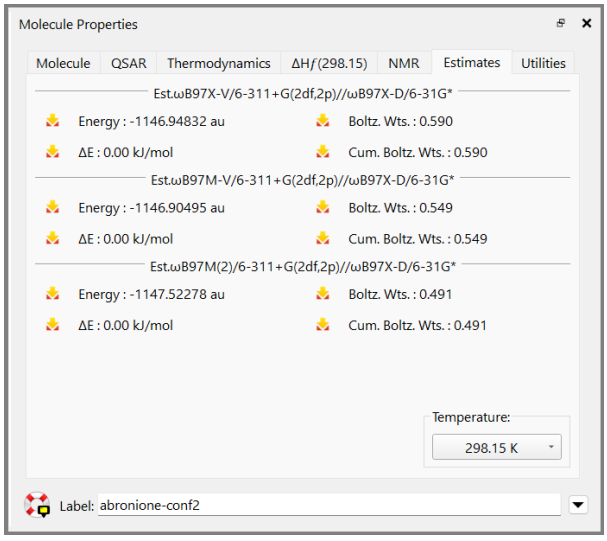
This reports the RMS, maximum absolute and mean absolute errors in either or both calculated proton and ^{13}C chemical shifts from experimental values (assuming that the experimental data have been obtained from NMRShiftDB or an external source and was entered manually). Chemical shift misassignments can and do occur (especially for quaternary carbons) and controls are available to shuffle ^{13}C experimental shifts assuming a fixed hydrogen count (C, CH, CH_2, CH_3) or ignoring hydrogen count. Reported errors will change accordingly. Shuffling experimental proton chemical shift assignments is presently not supported.

* In practice, enthalpy is clamped at 1/2 RT (1.24 kJ/mol at 298K) and entropy is clamped at 1/2R (4.15 J/K mol).



Estimates

This tab provides estimates from neural network models, provided the file has an ωB97X-D/6-31G* structure (geometry), energy estimates to three better QM models, and in the case of a conformer distribution file, Boltzmann Weights and Cumulative Boltzmann Weights are provided.



Utilities

This allows a variety of attributes of the selected molecule or all molecules in the selected document to be adjusted.

Set Standard Labels relabels the molecule such that carbons start with C1, nitrogens with N1,... Hydrogens start with H. Those associated with carbon take its number and are followed by a, b, c if there is more than one, for example, H5a. Hydrogens associated with elements other than carbon follow H by the appropriate atomic symbol and number and a, b, c if there is more than one hydrogen, for example, HN2a.

Add Hydrogens adds any hydrogens that are missing, for example, in older structure drawn from the Cambridge Structural Database. ***Requires recalculation***.

Reset Colors reverts to standard atom and bond colors.

Reset Model reverts to standard model type.

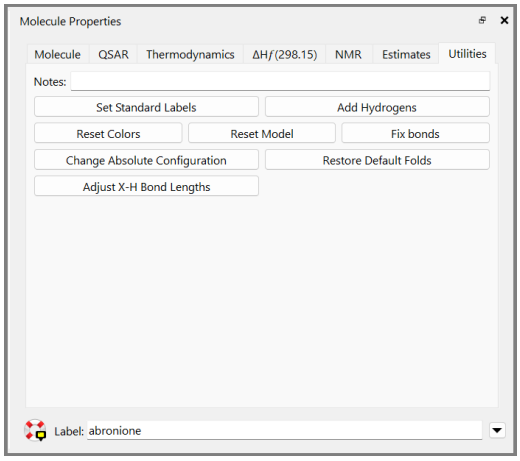
Fix Bonds repairs bonds that were either missing or were typed incorrectly. ***Requires recalculation***.

Change Absolute Configuration inverts all chiral (R/S) centers resulting in a change of absolute configuration. ***Requires recalculation***.

Restore Default Folds reverts single bond folds values and ring puckering designates to standard values.

Adjust X-H Bond Lengths sets bond lengths to hydrogen to "reasonable" values. Important for structures obtained from X-ray diffraction, for example, from the Cambridge Structural Database, where bonds to hydrogen are consistently too short. ***Requires recalculation***.

Molecule Utilities



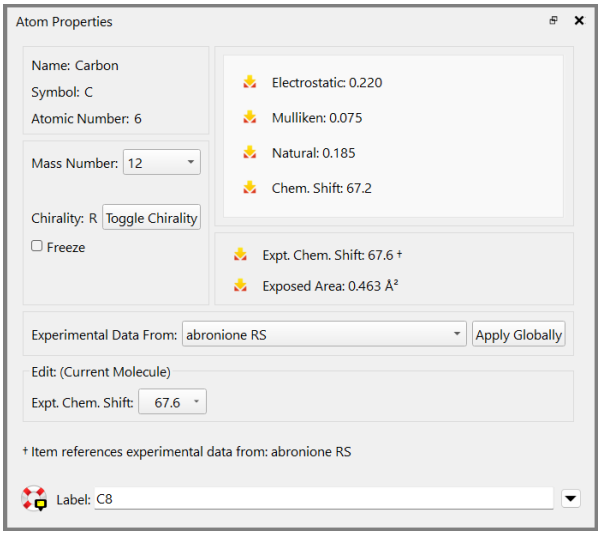
Clicking on ▼ at the bottom right of any of the **Molecule Properties** dialog tabs brings up a panel with lesser accessed **Utilities**.

This contains controls which are less frequently needed than those under the **Utilities** tab, specifically controls that provide alternative displays of aromatic molecules (Delocalized or Localized), information about amino acids in polypeptides and about CFD's, and also allows for toggling to the alternate enantiomer.

Atom Properties

Selecting an atom with a **Properties** dialog on screen, or selecting **Properties** from the

Display menu following selection of an atom, leads to the **Atom Properties** dialog:



This displays the element name and symbol, atomic number, mass number (isotope), R/S chirality, electrostatic-fit, Mulliken and natural charges (in electrons), calculated and experimental NMR chemical shifts (in ppm relative to standard) and the exposed surface area of a space-filling model (in $Å^2$). Experimental chemical shifts appear if proton or ^{13}C spectra have been retrieved from freely available online databases or have been manually entered.

> Proton and ^{13}C chemical shifts from B3LYP/6-31G*, EDF2/6-31G*, ωB97X-D/6-31G* and ωB97X-D/6-311G* calculations have been empirically corrected (see **Appendix A**). Corrections use both topology and bond orders or bond lengths to neighboring atoms. The correction schemes for B3LYP and ωB97X-D are significantly improved over that for EDF2.

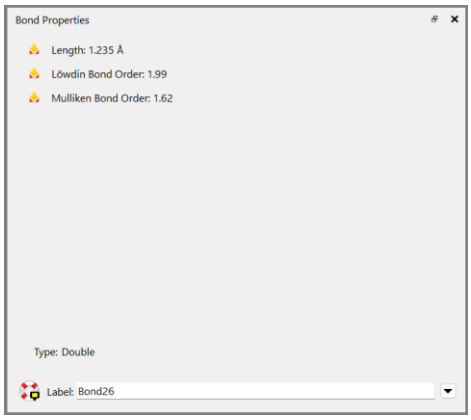
Database values may be altered. Charges, chemical shifts and exposed area may be posted into the spreadsheet. The dialog also allows changing the element and/or the isotope and the default label.

Checking **Freeze** freezes the atom (see **Freeze Center** in **Chapter 19**).

Clicking on ▼ at the bottom right of the **Atom Properties** dialog brings up the **Atom Style** dialog (not shown). This contains controls for selecting atom color and changing model style. *Clicking* on ▲ at the bottom right of this dialog returns to the **Atom Properties** dialog.

Bond Properties

Selecting a bond with a **Properties** dialog on screen, or selecting **Properties** following selection of a bond leads to the **Bond Properties** dialog:

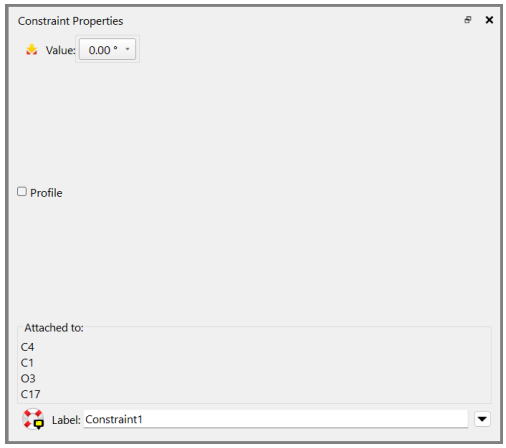


This displays the bond length, bond type and Mulliken and Löwdin bond orders.

Clicking on ▼ at the bottom right of the **Bond Properties** dialog brings up the **Bond Style** dialog (not shown). This contains controls for changing bond color. *Clicking* on ▲ at the bottom right of this dialog returns to the **Bond Properties** dialog.

Constraint Properties

Selecting a constraint marker with a **Properties** dialog on screen or selecting **Properties** from the **Display** menu following selection of a constraint marker, leads to the **Constraint Properties** dialog. This allows setting the value of a constraint, posting it to the spreadsheet and changing the default constraint label. *Checking* **Profile** leads to an expanded dialog:



This allows specifying a sequence of constraints for an energy profile (see **Calculations...** under the **Setup** menu; **Chapter 21**). The value of the starting constraint is given in the lefthand box to the right of **Value**, and the value of the ending constraint is given in the righthand box to the right of **Value**. The number of steps in the profile is given in the box to the right of **Steps**. Initially, the numbers in both boxes to the right of **Value** will be the same, and **Steps** will be set to **1**. These may be altered by *clicking* inside the appropriate box and entering the desired values from the number pad that appears. This functionality may also be accessed from **Constrain Distance** (**Angle**, **Dihedral**) under the **Geometry** menu (**Chapter 19**). An energy profile may involve constraints on more than one geometrical variable (changed in lock-step with one another). In this case, the constraint ranges for the individual variables need to be selected. If **Global Steps** is *checked*, the independent variables are moved in concert, meaning that the number of steps must and will be the same for each variable. If *unchecked*, the variables are moved independently, and the number of steps may differ from one to another. In this case, the total number of steps in the profile is the product of the number of steps for each variable. Be careful as this number will quickly become very large.

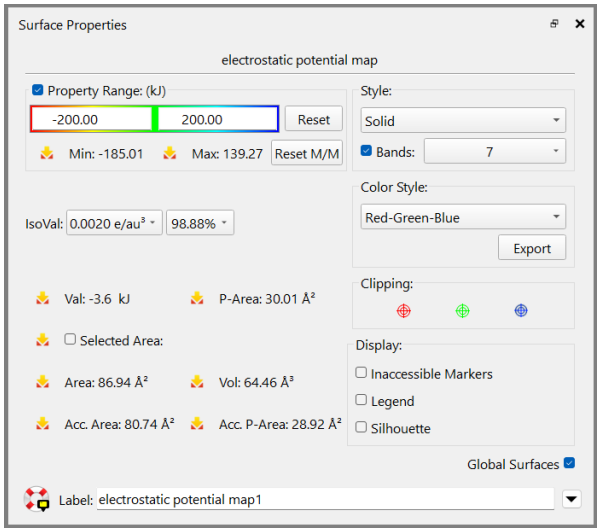
Clicking on ▼ at the bottom right of the **Constraint Properties** dialog brings up the **Constraint Style** dialog (not shown). This contains controls for selecting the color of the constraint marker. *Clicking* on ▲ returns to the **Constraint Properties** dialog.

Point, Ligand Point and Plane Properties

Selecting a user-defined point, ligand point or plane with a **Properties** dialog on screen or selecting **Properties** from the **Display** menu following selection of a point, ligand point or plane, leads to the **Point Properties**, **Ligand Point Properties** or **Plane Properties** dialog (not shown). These allow changing point or plane labels and new in ***Spartan'26***, specifying **Planes** as opaque or transparent. *Clicking* on ▼ at the bottom right of the **Point** (**Ligand Point**, **Plane**) **Properties** dialog leads to the **Point** (**Ligand Point**, **Plane**) **Style** dialog (not shown), with controls for adjusting point (plane) colors. *Clicking* on ▲ returns to the **Point** (**Ligand Point**, **Plane**) **Properties** dialog.

Surface Properties

Selecting a graphical display with a **Properties** dialog on screen, or selecting **Properties** from the **Display** menu following selection of a graphical display, leads to the **Surface Properties** dialog:



This allows changing display style (**Style**), either dots, or a mesh or transparent or opaque solid, selecting between continuous and banded displays (**Bands**) for property maps and, in the case of the latter, setting the number of bands, specifying isovalue, and as an alternative for electron density surfaces, percentage of the electrons contained inside the surface, turning on mapped properties, setting the property range of maps, and changing default surface labels. **Reset M/M** will set the property range to the minimum/maximum values. To revert to the default values, *click* the **Reset** button. The dialog reports (and optionally posts) the area (**Area**), volume (**Vol**) and accessible area (**Acc. Area**)* of the surface, the polar area (**P-Area**) and accessible polar area (**Acc. P-Area**) of an electrostatic potential map**, the value of the mapped property at the cursor position (Val)*** and the area enclosed in a selected band (**Selected Area**). The dialog also allows posting of the minimum and maximum value of the mapped property. If *checked*, **Legend** displays a color scale. If *checked*, **Selected Area** outlines the band containing the cursor. If *checked*, **Display Accessible Markers** reveals regions that are inaccessible. If *checked*, **Global Surfaces** designates that the settings in this dialog apply to all molecules in the document.

---

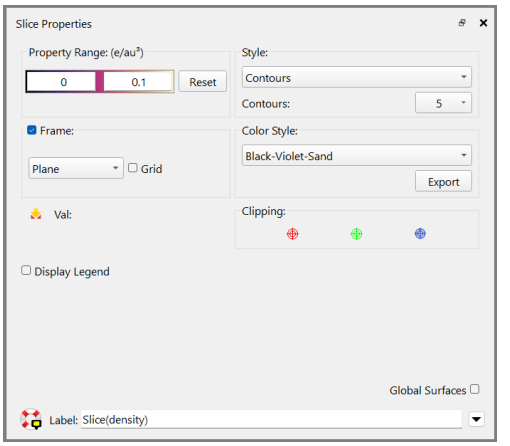
* A region on a density surface is designated as inaccessible if a sphere of radius 1.0 Å centered on a line normal to the surface and touching a point in the middle of the region, impinges on any other regions of the surface. The default radius (**Accessible Area Radius**) may be changed in the **Settings Preferences** dialog (**Preferences** under the **Options** menu; **Chapter 25**).

** This is defined as that part of the surface area for which the absolute value of the electrostatic potential is > 100 kJ/mol. The cutoff (**Polar Area Range**) may be changed in the **Settings Preferences** dialog (**Preferences** under the **Options** menu; **Chapter 25**).

*** To determine property value at another position, *click* on it. To bring up the **Molecule Properties** dialog, *click* on the background.

Up to three independent clipping planes operate in toggle mode. Following selection, a plane may be rotated, translated or zoomed with the usual mouse operations.

Where the surface is a “slice”, the **Slice Properties** dialog appears:



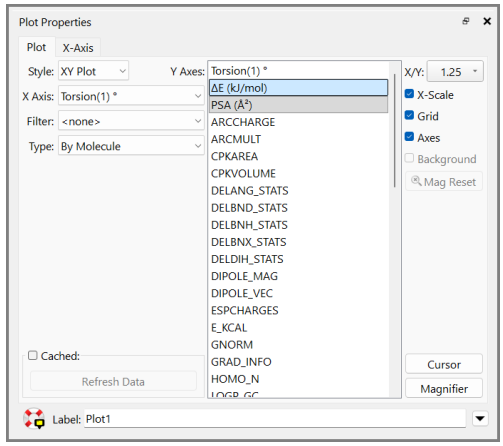
This contains controls to set the range of the property and to change the number of contours displayed and to the type of slice (plane, cylinder or sphere). The value of the property at the cursor position is displayed and, together with the minimum and maximum values of the property, may be posted to the spreadsheet. Check boxes allow for a legend, a frame around the slice and for a grid.

Clicking on ▼ at the bottom right of the **Surface Properties** dialog brings up the **Surface Style** dialog (not shown). This contains controls for selecting surface color. *Clicking* on ▲ at the bottom right of this dialog returns to the **Surface Properties** dialog.

Curve and Plot Properties

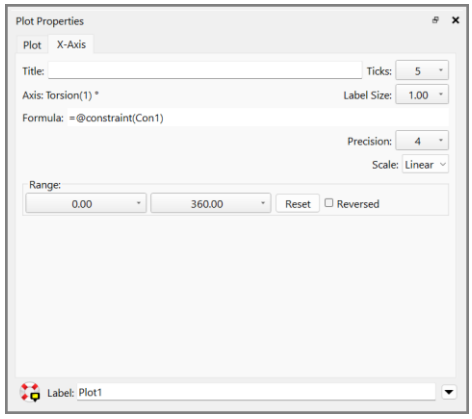
Plots created either using data from the spreadsheet or from spectra may be presented in separate windows or in ***Spartan***’s main window (see discussion later in this chapter). The following discussion relates only to the latter and concerns the way in which plots may be edited using a set of closely-related dialogs. These refer to curves (a relationship between variables), and plots (a set of relationships that share variables).

Selecting a plot frame with a **Properties** from the **Display** menu dialog on screen, or selecting **Properties** following selection of a plot frame, leads to the **Plot Properties** dialog (**XY Plot** dialog shown):



The **X Axis** menu and the **Y Axes** list duplicate functionality under the **Plots** dialog (see **Plots...** later in this chapter) and allow specification of a molecular property to serve as the X axis of the plot and one or more properties to serve as the Y axes. The content of the **Y-Axes** list depends on selection from the **Type** menu: **By Molecule** and **Free Form** reference explicit entries in the spreadsheet (see discussion later in this chapter) while **By Atom** references ***Spartan's*** file of atom-based properties (the ***Property Archive***). Initial selections may be altered as desired. In addition, the aspect ratio of the plot may be changed, the numerical scale of the X (horizontal) axis turned on or off, a grid added to the plot, and the default plot label changed. **Magnifier** brings up a frame with which to magnify selected areas of a plot. The magnifier may be positioned by *clicking* on the frame and then moving the mouse while holding down the right button. Magnifier selection results in the **Plot Magnifier Properties** dialog (not shown) with a single control to alter the width. The magnifier may be removed from the plot using **Delete** from the **Build** menu or the **Delete** key.

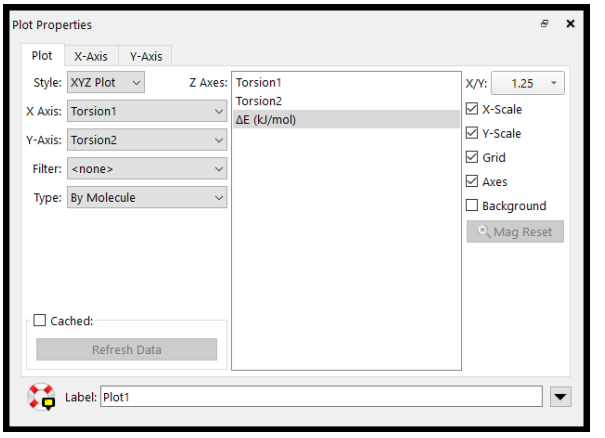
The range along the X (horizontal) axis may be altered from default settings by *clicking* on the **X-Axis** tab.



To alter the range, *click* inside the **Range** boxes, and enter the desired values into the number pad that appears. Controls are also available to alter the number of tic marks on the X axis, the label size and precision and change the scale from linear to logarithmic.

Clicking on ▼ at the bottom right of the **Plot Properties** dialog brings up the **Plot Style** dialog (not shown). This contains controls for selecting the color of the plot frame, as well as the label associated with the X axis. *Clicking* on ▲ at the bottom right of this dialog returns to the **Plot Properties** dialog.

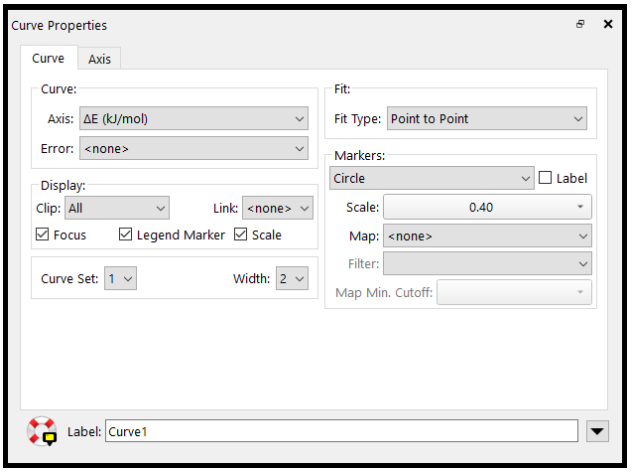
When the **Plots Plane** options is turned off (see **Options** menu; **Chapter 25**), **XYZ Plots** are enabled. Selection of **XYZ Plot** leads to an extended **XYZ Plot Properties** dialog:



This is identical to the **XY Plot Properties** dialog, except that it includes an additional menu (**Y Axis**) from which a property needs to be selected. In addition, the **Y Axes** list in the **XY Plot Properties** dialog has been renamed to **Z Axes**.

Ranges along X and Y axes may be adjusted by *clicking* on **X-Axis** and **Y-Axis** tabs to the left of the dialog. The resulting dialog (not shown) contains controls that are analogous to those previously discussed for XY plots.

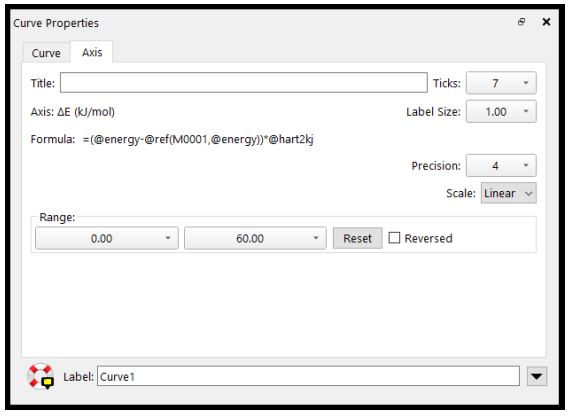
Selecting an individual curve in a plot with a **Properties** dialog on screen, or selecting **Properties** from the **Display** menu following selection of an individual curve, leads to the **Curve Properties** dialog:



With the **Curve** tab selected, this allows changing the quantity that is plotted (**Axis**),

provides a scale and error bars (obtained from the column in the spreadsheet designated in the **Error** menu) and turns clipping on. Controls are available to designate the kind of markers for the data points, request marker labels, and designate the kind of fit to the data points and the line width of the fitting curve*. If *checked*, **Focus** marks the selected molecule on the curve.

Clicking on the **Axis** tab changes the contents of the dialog and allows the scale along the Y axis of an XY plot or Z axis of an XYZ plot to be altered from the default (minimum to maximum value of the Y or Z):



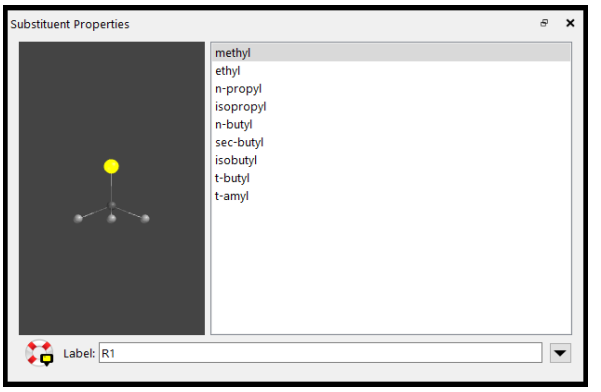
To alter the range, *click* inside the boxes underneath **Range** and enter the desired values using the number pad that appears. Controls are also available to alter characteristics of the Y (XY plot) or Z axis (XYZ plot) labels.

Clicking on ▼ at the bottom right of the **Curve Properties** dialog (with either tab selected) brings up the **Curve Style** dialog (not shown). This contains controls for selecting the color of the curve (and the label associated with the Y (or Z) axis). *Clicking* on ▲ at the bottom right returns to the original dialog.

* Fitting functions include linear, quadratic and cubic least square, a Fourier series, a Lorentzian or Gaussian. The data points may also be connected by straight lines (point-to-point) or by a smooth curve (cubic spline) or be presented as a skyline. Finally, the fitting curve can be removed.

Substituent Properties

Selecting a substituent icon with a **Properties** dialog on screen, or selecting **Properties** from the **Display** menu following selection of a substituent icon, leads to the **Substituent Properties** dialog:



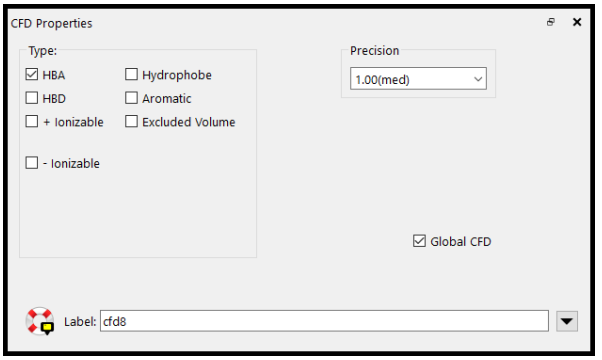
This allows substituent libraries already attached to a molecule to be edited, and the label on the substituent icon to be changed. It does not allow for attachment points to be reassigned. The dialog comprises a scroll box (listing the substituent names), a viewing window (showing a model of the substituent) and a label field.

To delete a substituent from the library, *click* on its name in the scroll box, then *right click* and select **Delete** from the contextual menu that appears. To add a new substituent that has previously been copied to the clipboard, *right click* inside the scroll box and select **Paste** from the contextual menu. Note that if the clipboard contains several molecules, all will be added to the substituent library. To add a substituent (or list of substituents) in a ***Spartan*** document to the library, *right click* inside the scroll box, select **Append** from the contextual menu and choose a document from the browser. Alternatively, *drag* the document that contains the substituent(s) from a window into the scroll box.

Clicking on at the bottom right of the **Substituent Properties** dialog brings up the **Substituent Style** dialog (not shown). This contains controls for selecting substituent (icon) color. *Clicking* on returns to the **Substituent Properties** dialog.

CFD Properties

Selecting a chemical function descriptor (CFD) or of an element of a pharmacophore with a **Properties** dialog on screen, or selecting **Properties** from the **Display** menu following selection of a CFD or pharmacophore element, leads to the **CFD Properties** dialog:

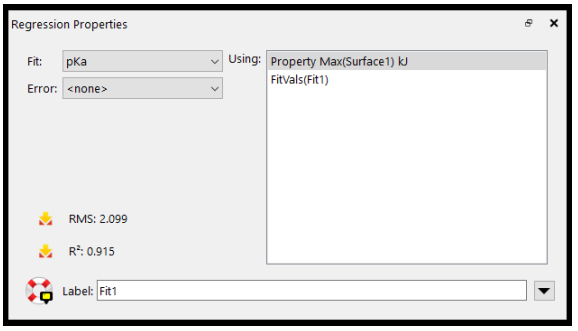


This allows changing the definition of the CFD, its label and the precision of the fit (the bigger the radius of the CFD, the less precise the fit). If *checked*, **Global CFD** specifies that CFD definitions will apply to all molecules in the document.*

Clicking on ▼ in the dialog brings up the **CFD Style** dialog (not shown). This contains controls for selecting CFD color. *Clicking* on ▲ in this dialog returns to the **CFD Properties** dialog.

Regression Properties

Following a linear regression analysis, a new row, labeled **Fit1****, appears near the bottom of the spreadsheet. This contains information about the fit. *Clicking* on this line with a **Properties** dialog on screen or selecting **Properties** from the **Display** menu after *clicking* on the line, leads to the **Regression Properties** dialog.



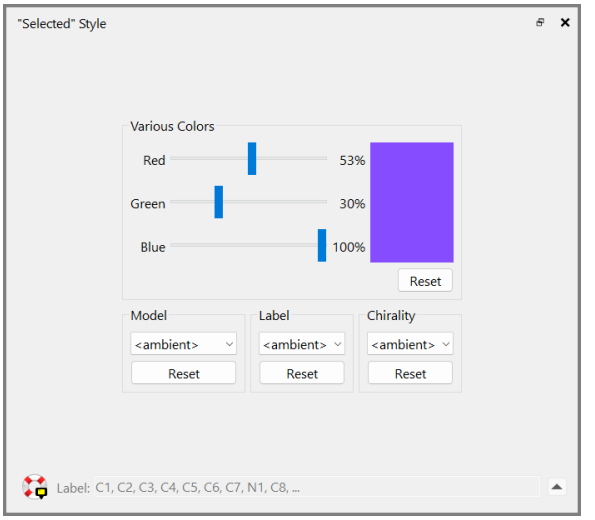
This reports RMSD and R^2, as well as allows for changing what is to be fit (**Fit**) and what it is to be fit to (**Using**). The error statistics will immediately update.

* This is generally applicable only where the molecules are conformers, or the CFD labels have been adjusted to associate positions on different molecules. Label reassignment is accomplished using the **Atom Properties** dialog (see discussion earlier in this chapter).

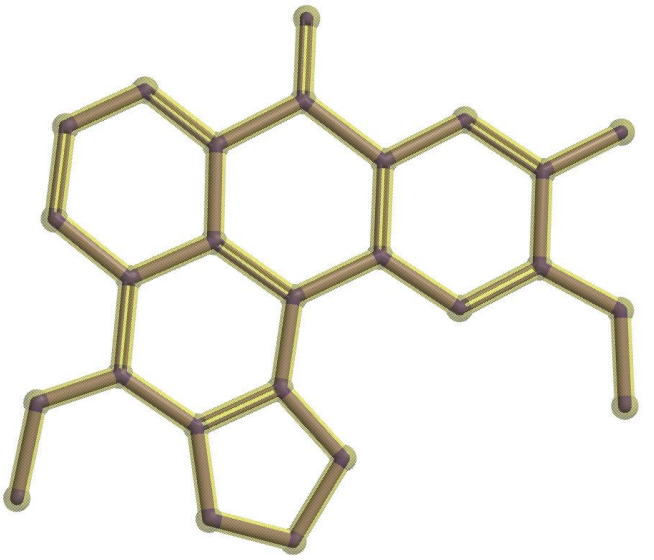
** More precisely, a row will be written for each fit, and labeled **Fit1**, **Fit2**, . . .

“Selected” Style

The **“Selected” Style** dialog is accessed either by defining a selection box (see **Mouse/Keyboard Operations** in **Chapter 1**) with a **Properties** dialog on screen, or by selecting **Properties** from the **Display** menu after defining a selection box.

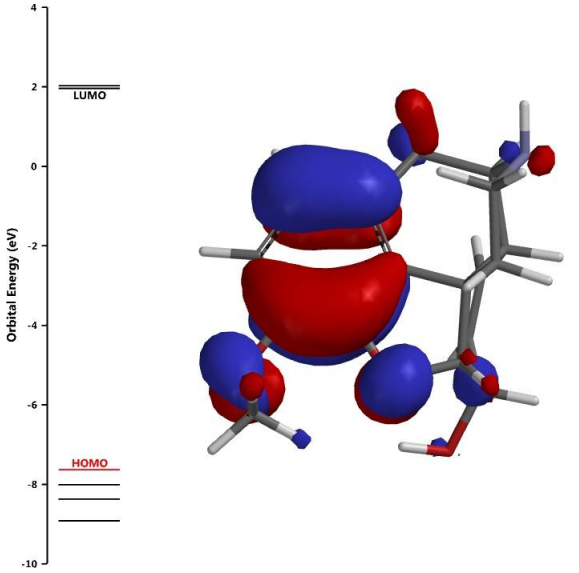


This contains controls for changing the color and model type of whatever is included in the selection box.



Orbital Energies

Selecting **Orbital Energies** leads to the display of an orbital energy diagram (this assumes that a wave function is available). This comprises up to twelve occupied molecular orbitals and two unoccupied molecular orbitals, the highest-occupied (HOMO) and lowest-unoccupied (LUMO) being explicitly designated.*



Clicking on an energy level in the diagram leads to display of the corresponding molecular orbital. This may be manipulated in the usual way and the display style altered from the menu at the bottom right of the screen. The energy (in eV) of the orbital is provided at the left of the menu. After one energy level has been selected and the associated orbital displayed, moving the mouse up or down over the diagram while holding down the left button ("swiping") then releasing the button selects the next higher or lower energy level.

For touch screen devices, moving a finger up or down over the diagram then lifting selects the next higher or lower energy level.

* Additional occupied and unoccupied molecular orbitals may be displayed by *clicking* on the orbital energy plot to bring up the **Orbital Energies Plot Properties** dialog and altering the contents of the **Maximum Occupied Levels** and **Maximum Unoccupied Levels** menus, respectively. The latter is restricted to 10 levels.

Surfaces ( )

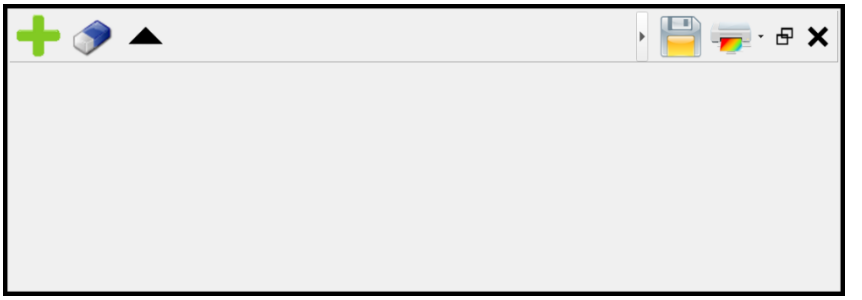
This accesses the same dialog previously described in **Chapter 21**.

Spectra

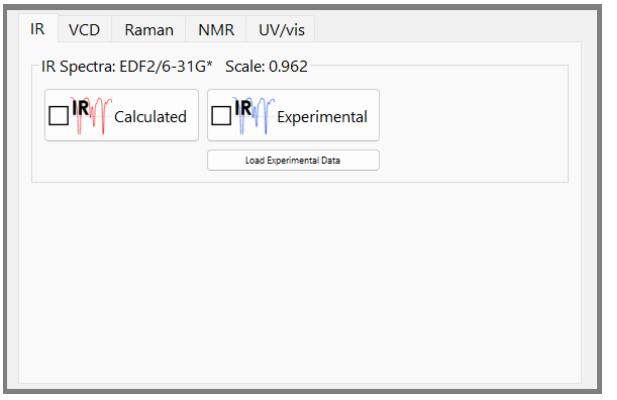
Spartan displays calculated IR, Raman, 1 and 2D NMR and UV/ visible spectra. Spectra need to have been previously requested from the **Calculations** dialog (**Calculations...** under the **Setup** menu; **Chapter 21**) or accessed from the Spartan Spectra and Properties Database. In addition, manual input as well as on-line access and display of experimental IR, NMR and UV/visible spectra from publicly available collections is provided, facilitating comparison with calculated spectra. ***Spartan*** allows two different modes for spectra display; a general presentation scheme and a simpler and perhaps more familiar mode. The choice is up to the user in the form of a preference. The simpler presentation mode is described first.

Simplified Presentation of Spectra

With **Spectra Pane** in the **Settings Preferences** dialog checked (**Preferences** under the **Options** menu; **Chapter 25**), selecting **Spectra** from the **Display** menu leads to an empty display pane at the bottom of the screen:

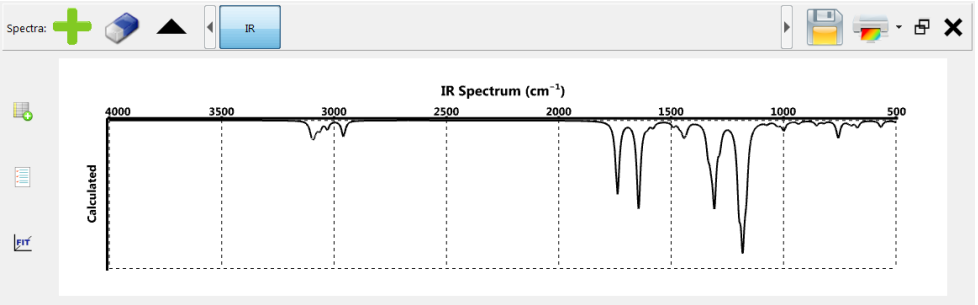


The only accessible control (in a bar at the top of the pane) is (add a spectrum). *Clicking* on this leads to a selection palette:

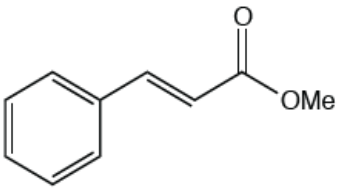


Once a tab is selected, check boxes are available for the types of spectra for which calculations (and in most cases experimental data) are available: IR, Raman, proton NMR with and without three-bond HH coupling ^{13}C NMR, COSY, HSQC and HMBC 2D NMR and UV/vis. COSY and HMBC spectra require coupling constants are available in addition to chemical shifts. COSY requires 3-bond HH coupling constants and HMBC requires both 2 and 3-bond CH coupling constants. The entry is "red" if a calculation has actually been performed and the corresponding spectrum is available. The checkboxes (at right) list the types of spectra for which experimental spectra ***may be available*** (manually input or from on-line public databases): IR, proton NMR, ^{13}C NMR and UV/vis. In addition, there is an entry for an experimental Raman spectrum, selection of which leads to a file browser from which a JCamp (.dx) file containing the spectrum may be chosen. These are shown in blue.

The procedure for displaying either a calculated or experimental spectrum (or both) is independent of the type of spectrum. For the purpose of illustration, we use the IR spectrum of methyl *trans*- cinnamate. A calculated spectrum is displayed by *clicking* on the appropriated (red highlighted) entry, following which the palette is dismissed. *Clicking* on IR results in an IR spectrum.



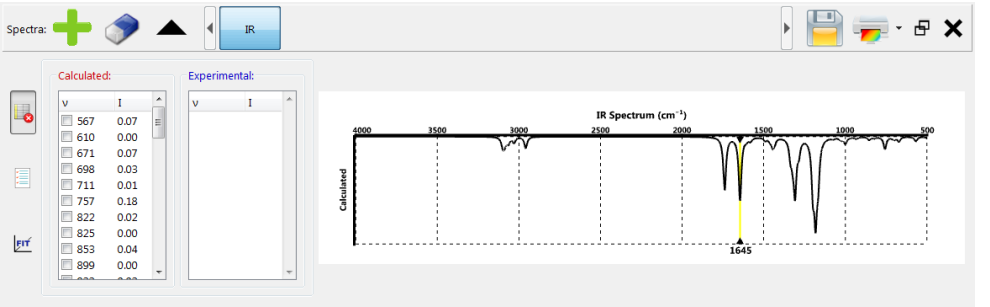
Clicking on the probe indicator selects it. Once selected, moving the mouse while holding down the left button moves the probe (triangle markers at the top and bottom of the spectrum) over the spectrum. When positioned directly over a special line, the triangles are darkened and connected by a vertical yellow line, and a numerical value for the line is provided at the bottom of the spectrum. In the case of an IR spectrum, this is a frequency in cm^{-1} and corresponds to a particular vibration of the atoms in the molecule. The molecular model (above the spectrum) vibrates to show this motion. For methyl *trans*-cinnamate, the line near 1645 cm^{-1} closely corresponds to the C=C stretch while the line near 1739 cm^{-1} almost entirely corresponds to the C=O stretch.



Moving the mouse while holding down the right button slides the viewable scale from 4000 cm^{-1} to 500 cm^{-1} but does not change the overall range (of 3500 cm^{-1}). The range is changed by using the scroll wheel. The original settings may be restored by *clicking* on in the bar at the top of the spectra pane.

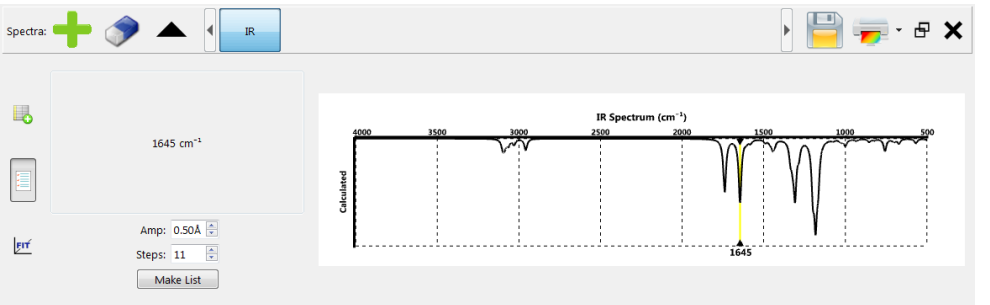
On touch screen devices, move one finger over the spectrum to position the cursor, move two fingers to slide the viewable scale and pinch two fingers to change the range.

In the case of IR and Raman (only) three buttons appear at the left of the spectrum , and . *Clicking* on leads to a scrollable panel at the left of the spectrum.



This contains a listing of calculated infrared frequencies and intensities. *Checking* the box to the left of an individual frequency moves the cursor on the spectrum over this line and animates the vibrational motion.

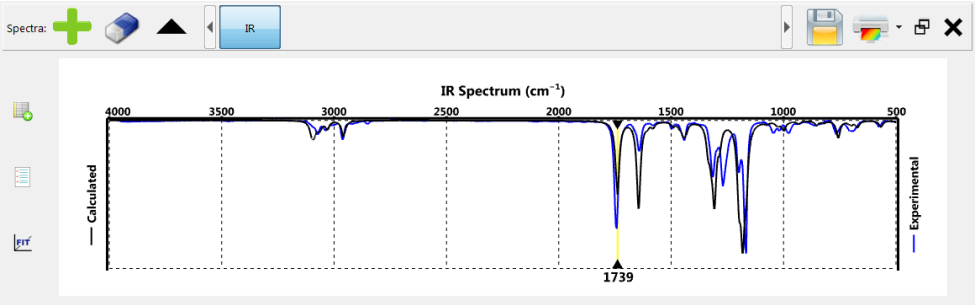
With a frequency selected, *clicking* on the **Make List** icon leads to a new panel.



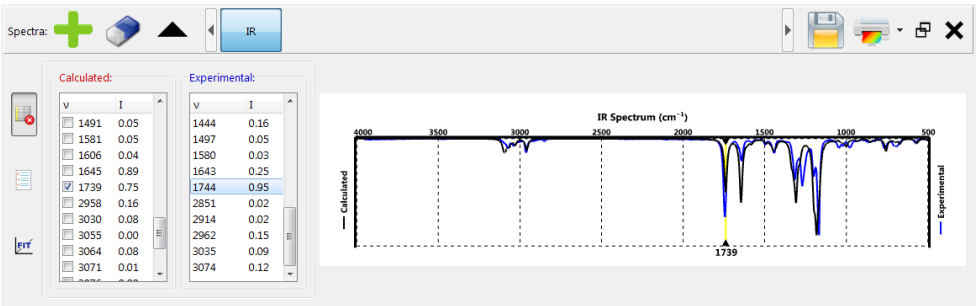
This is used to make a list of structures centering around the minimum (or maximum in the case of a transition state) with control over the amplitude of vibration (maximum displacement in Å) and number of steps. *Clicking* on the **Make List** button leads to a separate document.

If it exists, an experimental spectrum from the freely available NIST database may be superimposed on top of the calculated spectrum. The IR spectrum of methyl *trans*-

cinnamate is available. *Click* on and select from the palette.



Clicking on the **Tables** icon brings up two lists of frequencies and intensities, the list on the left corresponding to the calculated spectrum (as previously described) and the list on the right corresponding to the calculated spectrum.



The "experimental" frequencies and intensities have been obtained by fitting the measured infrared or Raman plot to a set of Lorentzian functions.

In the absence of an experimental infrared or Raman spectrum, the righthand list will be empty. You can drag a JCamp (.dx) file on top of the list. It will take a few seconds for the spectrum to be displayed as a fit needs to be performed.

Calculated infrared (and Raman) spectra that are initially displayed have been broadened to account for finite temperature. Unless they come from a B3LYP/6-31G*, EDF2/6-31G* or ωB97X-D/6-31G* calculation, they are otherwise unaltered. Infrared (and Raman) frequencies from B3LYP/6-31G*, EDF2/6-31G* and ωB97X-D/ 6-31G* models have been uniformly scaled to account for known systematic errors. In any event, a calculated infrared or Raman spectrum can be fit to the corresponding experimental spectrum (with line width and scale factors as adjustable parameters), by *clicking* on the button at the left of the spectra pane.

Additional spectra may be requested by *clicking* on in the bar above the spectra pane and then *clicking* on the appropriate entry in the resulting palette. Each new spectrum adds a tab to the bar, although calculated and experimental spectra share the tab. Switching between tabs changes the display. A spectrum can be deleted by *clicking* on in the bar above the pane. If calculated and experimental spectra were displayed, both will be deleted.

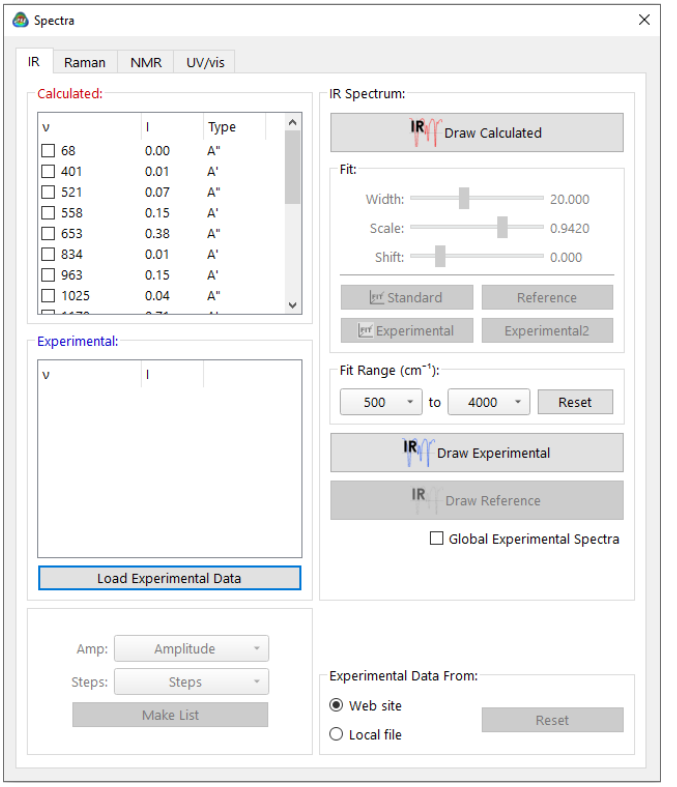
Controls at the top right of the spectra pane allow saving the spectrum as a PNG, JPEG or Bitmap image file , printing the file , detaching the spectrum pane from the main window and closing the pane .

Generalized Presentation of Spectra

With **Spectra Pane** (in the **Settings** tab of the **Preferences** dialog) unchecked (**Preferences** under the **Options** menu; **Chapter 25**), selecting **Spectra** from the **Display** menu leads to a tabbed dialog each of which pertains to a different type of spectra, IR, Raman, NMR and UV/vis. These allow different spectra for the same or for different molecules to be displayed on screen at the same time (as opposed to the one-spectra-at-a-time restriction in the simplified mode) and allow additional controls for adjusting the displays as well as for fitting calculated to experimental spectra.

IR, Raman (these two dialogs are nearly identical and will be discussed together)

Clicking on the **IR** (or **Raman**) tab leads to the **IR Spectra** (or **Raman Spectra**) dialog. This displays the calculated and (if available) experimental infrared spectra and allows for animated display of the vibrational modes as well as for generation of a sequence of structures along a vibrational coordinate.



At the left-hand side of the dialog is an ascending list of frequencies (in cm^{-1})* together with

(infrared) intensities and symmetry labels. A frequency is selected for display (animation) by *clicking* on it. *Clicking* again deselects it (stops the animation).

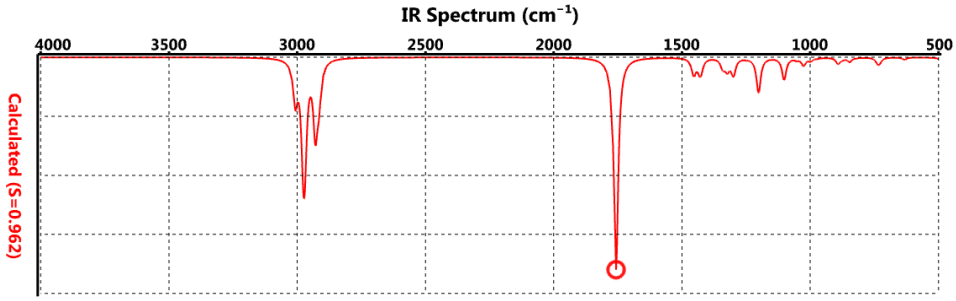
The amplitude of vibrational motion (the maximum displacement away from equilibrium of any pair of atoms) may be changed from the default amplitude (0.5Å which is much larger than appropriate for vibrational motion at room temperature) by altering the contents of the box to the right of **Amp** at the lower left of the dialog. The default number of steps that make up the display (11) may be changed by altering the contents of the box to the right of **Steps**** at the lower left of the dialog.

Clicking on **Make List** creates new document containing a sequence of structures corresponding to motion along the selected vibrational mode.

The original document is not closed, and additional sequences (corresponding to different vibrations) can be created.

Note that the information required to produce vibrational motions (animations) and to construct lists of structures along different vibrational coordinates is available as part of the entry in the Spartan Spectra and Properties Database for the ωB97X-D/6-31G* (see discussion in **Chapter 23**).

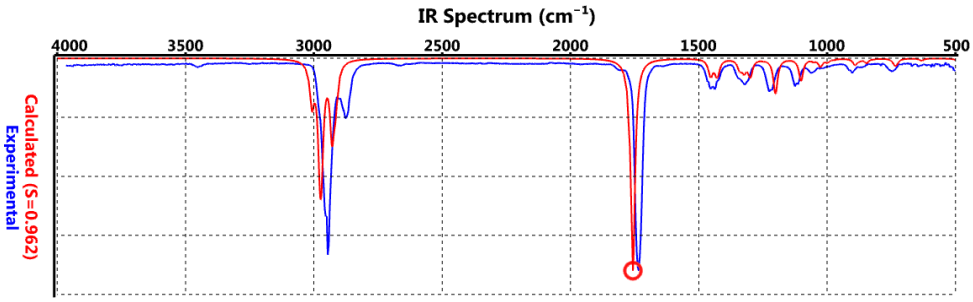
The calculated spectrum may be drawn by *clicking* on **Draw Calculated** at the top right of the dialog.



The experimental spectrum (if available) may be drawn on top of the calculated spectrum (or by itself) by *clicking* on **Draw Experimental** at the center right of the dialog (IR spectrum shown).

* Imaginary frequencies, for example, corresponding to the reaction coordinate for a transition state, will appear first in the list, and will be designated by the letter ***i*** in front of the frequency value.

** It is recommended that the number of steps be odd ensuring that the center point corresponds to the actual equilibrium or transition-state structure.



Finally, a reference spectrum (a spectrum shared by all molecules in the document; see discussion of the Spartan Infrared Database in **Chapter 23**) if available, may be displayed along with the calculated and/or experimental spectra by *clicking* on **Draw Reference**.

Draw operations for both calculated and experimental IR and Raman spectra, apply to all molecules in the document (not just the selected molecule). Note, however, that the reference spectrum is common to all molecules in the document.

> Experimental IR spectra are accessed from the NIST (National Institute of Standards and Technology) website (https://webbook.nist.gov/chemistry). This comprises approximately 7,000 IR spectra, primarily for organic molecules. Alternatively, the user may supply a spectra file (see **Appendix J**). Selection (website vs. file) is made at the bottom right of the dialog. **Global Spectra** controls whether spectra access pertains to all molecules in the document (checked) or just the selected molecule (unchecked). Note that experimental spectra, once retrieved either on-line or from a file, are kept in the document (and need not be retrieved again if they are deleted). They can be removed from the document by *clicking* on the **Reset** button.

The default range for calculated, experimental and (for infrared only) reference IR and Raman spectra is from 4000 cm^{-1} to 500 cm^{-1}. This corresponds to the range commonly measured and reported. This range may be changed by selecting **Properties** from the **Display** menu, *clicking* on the horizontal axis to bring up the **Plot Properties** dialog, *clicking* on the **X Axis** tab at the left of the dialog and editing the values under **Range** (see discussion earlier in this chapter). The spectrum is treated like any other graphical object and may be translated and scaled using the appropriate mouse operations. Touch screen operations also apply.

> On touch screen devices, move two fingers across the plot to translate it and pinch two fingers to scale it. It may not be rotated.

The calculated spectrum may be removed by *clicking* on **Delete Calculated** in the dialog. Similarly, the experimental spectrum may be removed by *clicking* on **Delete Experimental**

and the reference spectrum may be removed by *clicking* on **Delete Reference**. The last spectrum (calculated, experimental or reference) removed will also cause removal of the plot axes. These operations apply to all molecules in the document (not just the selected molecule). An individual (calculated or experimental) spectrum or entire plot may also be removed by selecting **Delete** from the **Build** menu and then *clicking* on a spectrum or on the plot axes, respectively, or holding down the **Delete** key while *clicking* on the spectrum or plot axes.

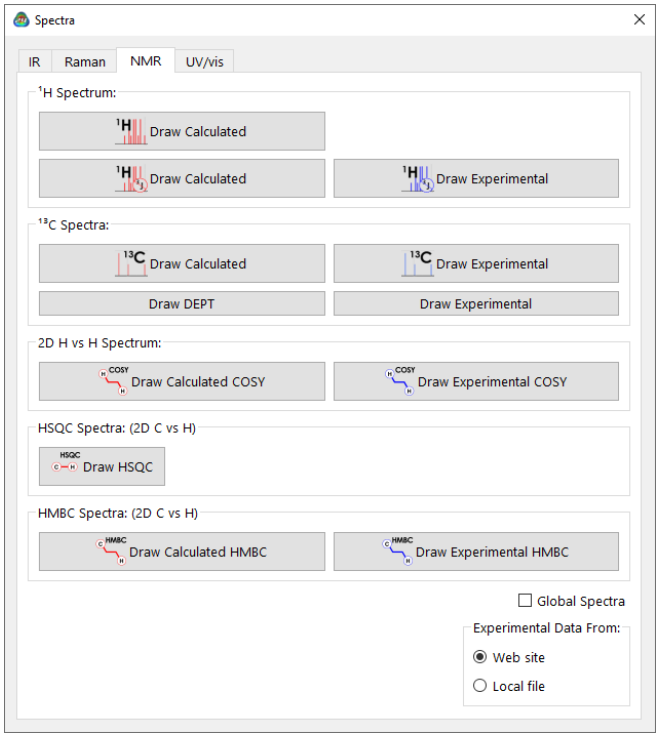
Calculated IR (Raman) frequencies exhibit systematic errors. Frequencies obtained from (limiting) Hartree-Fock models are typically 10-12% larger than measured frequencies, while frequencies obtained from density functional and MP2 models are typically 3-6% larger than measured frequencies. Systematic errors may be revealed (and quantified) by first simultaneously displaying calculated and experimental spectra and then scaling one of the spectra to provide a best visual fit. To scale the calculated spectrum, use the slider bar to the right of **Scale** near the top right of the dialog.

> A scaling factor may be applied to frequencies, and to all properties that make use of frequencies, prior to reporting using the **FreqScale** keyword typed into the **Options** box in the **Calculations** dialog (**Calculations...** under the **Setup** menu, **Chapter 21**).

A second slider bar marked **Temp** controls peak width. This is loosely connected with the temperature at which the experimental measurement is carried out. Low "temperature" (slider to the left) will produce sharp spikes, whereas high "temperatures" (slider to the right) will produce broad bands. The default setting is intended to give calculated spectra that broadly resemble experimental infrared spectra. Both **Scale** and **Temp** slider bars actually control the fit of the calculated frequencies and intensities to a Lorentzian function. The default settings can be restored by *clicking* on the **Standard** button. A best fit to an experimental spectrum superimposed onto the calculated spectrum can be obtained by *clicking* on the **Experimental** button. A best fit to an unknown reference spectrum (associated with all the molecules in the document rather than a single molecule) can be obtained by *clicking* on the **Reference** button.

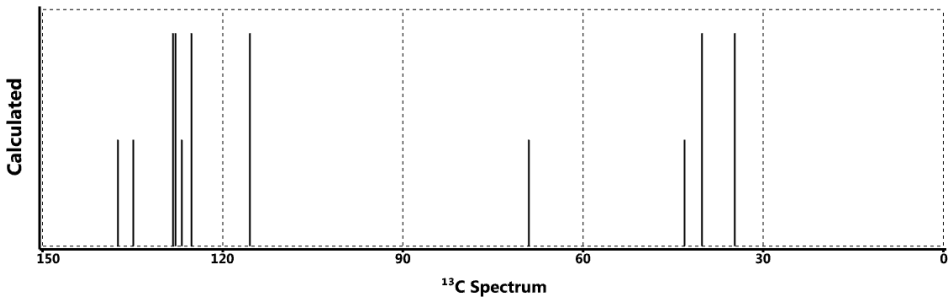
NMR

NMR spectra may be subdivided into four categories*: ^{1}H spectra with or without 3-bond HH coupling (the former being what is observed experimentally), ^{13}C and DEPT spectra, COSY spectra, and HSQC and HMBC spectra. Access to Experimental 1D spectra proton and ^{13}C is provided either from an on-line database, by manually attaching chemical shifts to either 2D sketches or 3D models or from a local file. While there is no provision for entering numerical values for experimental 3-bond HH and 2 and 3-bond CH coupling constants, couplings that are experimentally observed (or at least assigned) can be marked to either 2D sketches or 3D models. This allows construction and display of experimental COSY and HMBC spectra. *Clicking* on the **NMR** tab leads to the **NMR Spectra** dialog.

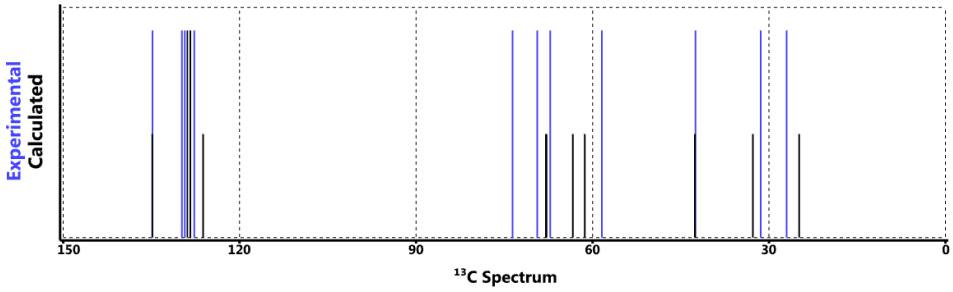


Calculated proton, ^{13}C and DEPT spectra may be drawn by *clicking* on their **Calculated** and **DEPT** buttons, respectively.* Proton and ^{13}C from the B3LYP/6-31G*, EDF2/6-31G*, ωB97X-D/6-31G*, ωB97X-V/6-31G* and ωB97X-D/6-311G* models are corrected based on topology and bond orders. ^{19}F shifts from the EDF2/6-31G* model are also corrected. Shifts from other models are not corrected.

* Chemical shifts for other nuclei are reported in the summary output (see **Output** earlier in this chapter) and may be displayed as labels (see **Configure...** under the **Model** menu; **Chapter 18**).



Experimental ^{13}C and DEPT spectra (if available) may be drawn together with or independently of the calculated spectra by *clicking* on **Draw Experimental**. Draw operations apply to all molecules in the document.



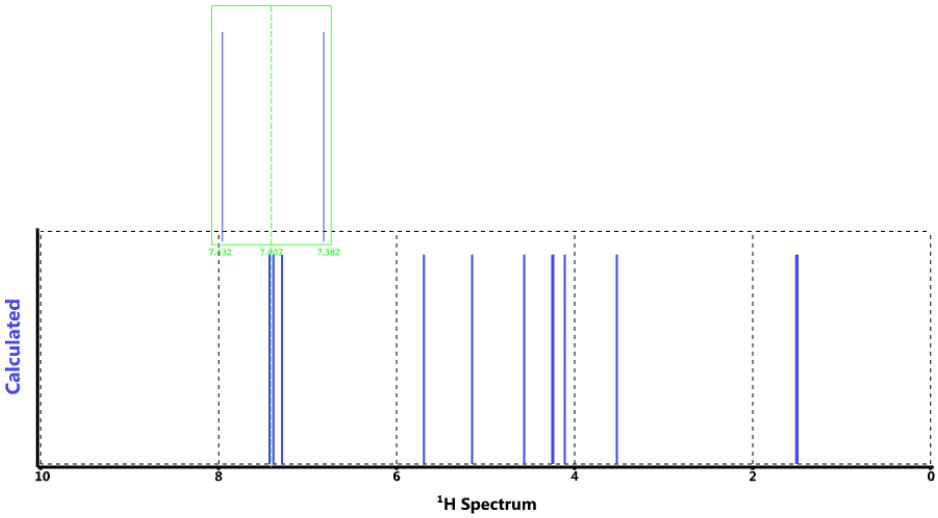
Experimental proton and ^{13}C spectra are accessed from the freely available NMR shift database (https://nmrshiftdb.uni-koeln.de). This database contains over 50,000 experimental spectra, all for organic molecules. Alternatively, may attach experimental shifts to either 2D sketches or 3D models or the user may supply a spectra file (**Appendix J**). Selection (website vs. file) is made at the bottom right of the dialog. **Global Spectra** controls whether spectra access pertains to all molecules in the document (checked) or just to the selected molecule (unchecked).

Proton spectra may be drawn to reflect spin-spin coupling due to neighboring nuclei. Coupling constants are not calculated but are estimated empirically based on three-dimensional geometry. A magnifier is available under the **Plot Properties** dialog (see discussion earlier in this chapter).

---

* DEPT spectra (both calculated and experimental) are simply derived from the corresponding ^{13}C spectra by drawing lines corresponding to quaternary and methylene carbons “down” rather than “up”.

Note that experimental spectra, once retrieved either on-line or from a file, are kept in the document (and need not be retrieved again if they are deleted). They can be removed from the document by *clicking* on the **Reset** button.

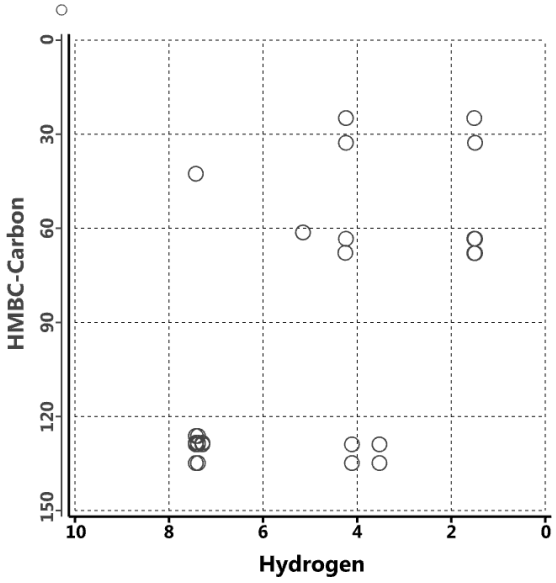


Default ranges for proton spectra (0 to 10 ppm) and ^{13}C spectra (0 to 150 ppm) may be changed by selecting **Properties** from the **Display** menu, *clicking* on the horizontal axis to bring up the **Plot Properties** dialog, *clicking* on the **X Axis** tab at the left of the dialog and editing the values under **Range**. Spectral plots may be translated and scaled either with the mouse or via two finger touch-screen operations.

A calculated spectrum may be removed by *clicking* on **Delete Calculated** (that has replaced **Draw Calculated**) and an experimental spectrum removed by *clicking* on **Delete Experimental**. The last spectrum removed will also cause removal of the plot axes. These operations apply to all molecules in the document. An individual (calculated or experimental)

spectrum or entire plot may also be removed by selecting **Delete** from the **Build** menu (or holding down the **Delete** key) and then *clicking* on the spectrum or on the plot axes, respectively.

2D spectra are available from **Draw Calculated** buttons. COSY spectra identify protons three bonds removed, HSQC spectra identify protons directly attached to a carbon and HMBC spectra identify protons that are three bonds removed from a carbon. Note that proton spectra "broadened" because of the inclusion of spin-spin coupling are not used in construction of 2D spectra. The corresponding proton and carbon spectra axes may be displayed along with 2D spectra by selecting them in addition to the 2D spectra. 2D spectra (and the 1D axes spectra) may be removed using the **Delete Calculated** buttons (that have replaced the **Draw Calculated** buttons). Experimental 2D spectra displays are not yet available. An HMBC spectrum is shown below.

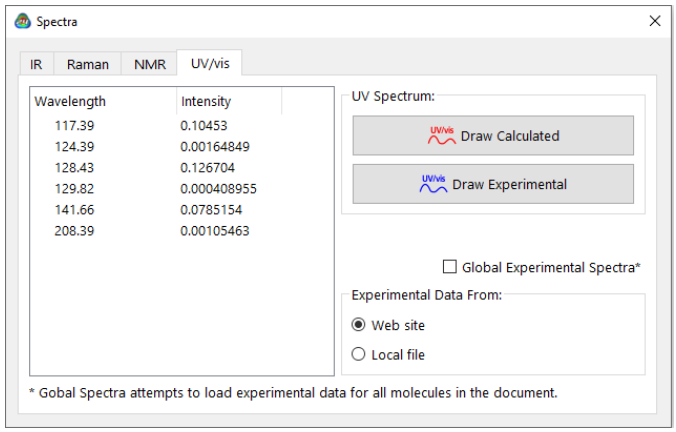


Uncorrected proton, ^{13}C, DEPT, that is, from models other than B3LYP/6-31G*, EDF2/6-31G*, ⎤ωB97X-D/6-31G* and ωB97X-D/6-311G**, are likely to show significant systematic errors in chemical shifts. To scale either the calculated or experimental spectra in order to provide a best visual fit, *click* on the spectra (not on an axis) to select and move the mouse up and down while simultaneously depressing the **Shift** and **Alt** (**option** on Mac) keys

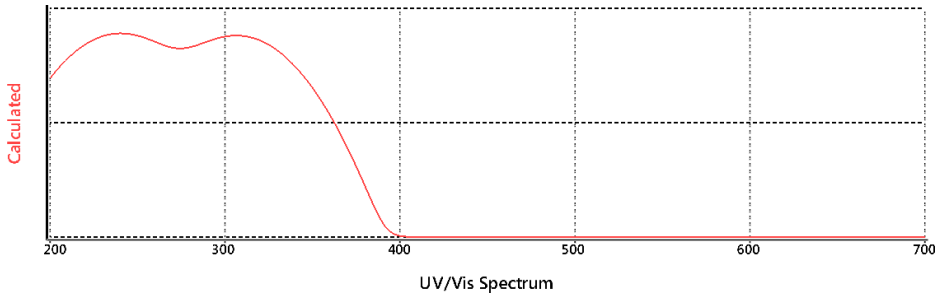
in addition to the right button. The scale factor (multiplying the chemical shifts) will be displayed on the vertical axis. Scaling applies uniformly to all molecules in the document.

UV/vis

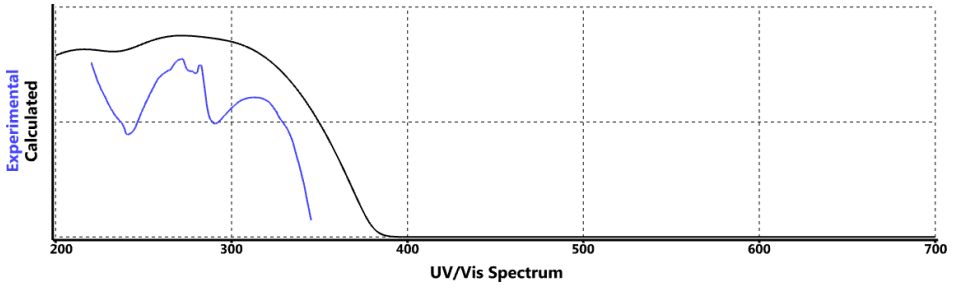
Clicking on the **UV/vis** tab leads to the **UV/vis Spectra** dialog that provides for display of UV/visible spectra.



This lists wavelengths and intensities. *Click* on **Draw Calculated** to draw the calculated spectrum.



Click on **Draw Experimental** to draw the corresponding experimental spectrum. Draw operations for both calculated and experimental UV/visible spectra and apply to all molecules in the document.



Experimental UV/visible spectra are accessed from the NIST (National Institute of Standards and Technology) website (https://webbook.nist.gov/chemistry). This comprises approximately 1,500 UV/visible spectra, primarily for organic molecules. Alternatively, the user may supply a spectra file (see **Appendix J**). Selection (website vs. file) is made at the bottom right of the dialog. **Global Spectra** controls whether spectra access pertains to all molecules in the document (checked) or just to the selected molecule (unchecked). Note that experimental spectra, once retrieved either on-line or from a file, are kept in the document (and need not be retrieved again if they are deleted). They can be removed from the document by *clicking* on the **Reset** button.

The default range for both calculated and experimental UV/ visible spectra is from 200 nm to 700 nm. This corresponds to the range commonly measured and reported. This range may be changed by selecting **Properties** from the **Display** menu, *clicking* on the horizontal axis to bring up the **Plot Properties** dialog, *clicking* on the **X Axis** tab at the left of the dialog and editing the values under **Range** (see discussion earlier in this chapter). The spectrum is treated like any other graphical object and may be translated and scaled using the appropriate mouse or touch- screen operations.

The calculated spectrum may be removed by *clicking* on **Delete Calculated** (that has replaced **Draw Calculated**) in the dialog. Similarly, the experimental spectrum may be removed by *clicking* on **Delete Experimental** (that has replaced **Draw Experimental**). The last spectrum removed will also cause removal of the plot axes. These operations apply to all molecules in the document (not just the selected molecule). An individual (calculated or experimental) spectrum or entire plot may also be removed by selecting **Delete**

from the **Build** menu (or holding down the **Delete** key) and then *clicking* on the spectrum or on the plot axes, respectively.

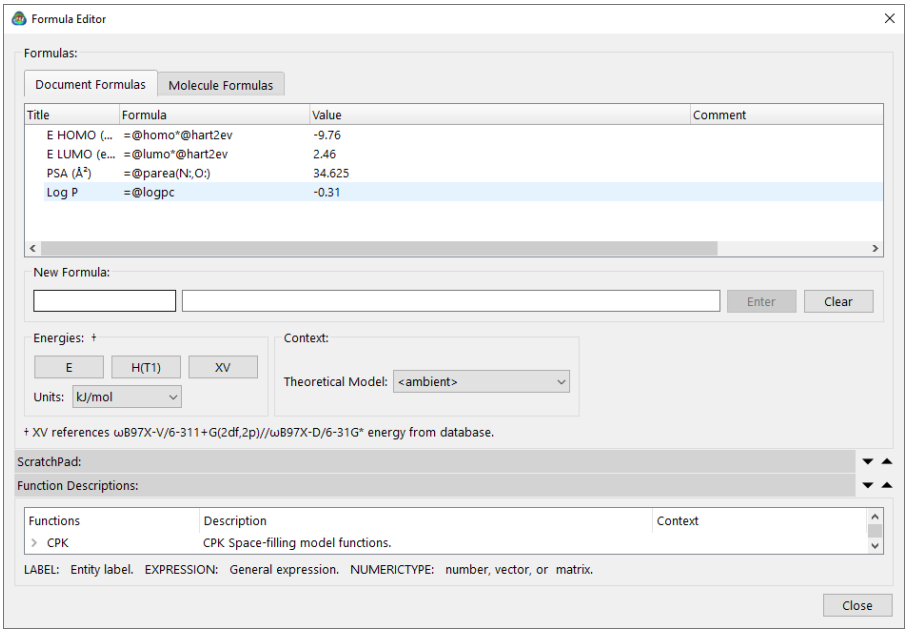
To help reveal similarities and differences between calculated and experimental UV/visible spectra, one or the other may be scaled. To scale, *click* on the spectra (not on an axis) to select and move the mouse up and down while simultaneously depressing the **Shift** and **Alt** (**option** on Mac) keys in addition to the right button. The scale factor will be displayed on the vertical axis. Scaling applies to all molecules in the document.

Only a single copy of one of the dialogs under **Spectra** may appear on screen; scaling will relate to the currently selected molecule. The dialog may be removed by *clicking* on .

Formulas

The molecular mechanics and quantum chemical methods available in ***Spartan*** are able to provide a wide variety of molecular and atomic properties and QSAR descriptors. Quantities related to structure are available through functions under the **Geometry** menu (**Chapter 19**), while most other properties may be accessed from **Properties** dialogs (previously discussed in this chapter) or from the spreadsheet (to be discussed later in this chapter). In some cases, however, it is not the properties or descriptors themselves that are of interest, but rather *relationships* among them. For example, while the energies of the HOMO and LUMO of a molecule may not be of interest, the difference in their energies (the HOMO-LUMO gap) may be. Another important kind of relationship involves a change in property between reactants and products of a chemical reaction, in particular, the change in energy.

Spartan provides a formula editor to define relationships among properties of a single molecule or of a single property between the reactants and products of a chemical reaction. The former is discussed here. and the latter will be discussed in the next chapter in the context of ***data mining***. The **Formula Editor** dialog is accessed by selecting **Formulas** from the **Display** menu.



The bottom half of the dialog contains controls to build formulas, while the top half stores the formulas. Aside from the energy and T1 heat of formation*, properties need to be dragged onto one of three "buttons" under **Properties** at the lower right of the **Formula Editor** dialog from its on-screen display (for example, a bond distance from its display at the bottom of the screen following selection of **Measure Distance** from the **Geometry** menu) or from one of the **Properties** dialogs (for example, the dipole moment in the **Molecule Properties** dialog). Following this, the button takes the name of the property. The property obtained from ***Spartan*** is now available for use in a formula. Energy and T1 heat of formation are given their own "permanent" buttons in the dialog.

Theoretical Model (under **Context** in the lower center of the **Formula Editor** dialog) needs to be specified from among the models stored in databases available to ***Spartan*** (see discussion in **Chapter 23**). **Ambient** refers to the theoretical model actually used for the calculation of the selected molecule and should not be selected if a calculation has not actually been run. The **Applied to** menu under **Context** will be grayed out if the molecule that ***Spartan*** is pointing to does not contain reaction information. Use of the formula editor to relate information for the reactants and products of chemical reactions is discussed in **Chapter 23**.

* Heats of formation from the T1 recipe are available for most molecules in the SSPD.

Pressing either the **E** or **H(T1)** buttons or a **Property** button onto which a property has been *dragged*, builds a formula and places it inside the second box at the center of the dialog. The default name (**Fx**, x=0, 1, 2...) appears in the first box. Either the formula or its name may be edited. Pressing **Enter** (**return** on Mac) stores the formula with the default or edited name. Pressing **Clear** removes the formula. *Double clicking* on a formula in storage brings it back to the boxes at the center of the formula editor (but does not remove it from storage). The formula can then be edited and returned to storage (by *clicking* on **Enter**). If the name of the formula has not been changed, it will replace the original in storage. If it has been changed, it will be treated as a new formula. Once in storage, a formula can be *dragged* to wherever it is needed, in particular, into the spreadsheet, the **Plots** dialog or any of the dialogs associated with the SSPD. To remove a stored formula, *right click* on it and select **Delete** from the menu.

Formulas use the usual set of arithmetic operations and parentheses to connect different properties. In addition, Boolean operators and standard mathematical functions designated in **Table 22-1** may be used. These need to be typed into the box at the center of the dialog. For example, to construct a formula for the HOMO-LUMO gap, first *drag* the LUMO energy from the **Molecule Properties** dialog onto one of the properties buttons in the **Formula Editor** dialog. *Pressing* the button leads to the text string ***@LUMO*@hart2eV*** at the center of the dialog. Insert (type) "**-**" at the end of the string. Next, drag the HOMO energy from the **Molecule Properties** dialog onto a properties button in the editor and *press* the button. The text string is now ***@ LUMO*@hart2eV-@HOMO*@hart2eV***. Note that while orbital energies are evaluated in atomic units (hartrees), the formula includes a conversion factor ***@hart2eV*** to bring them into electron volts (eV). You can change units by manually editing the text string using the expressions provided in **Table 22-2** (preceded by **@**). Change the title to the left of the string from **Fx** to ***HOMO/LUMO gap*** and *press* the **Enter** (**return**) key to the right of the string. The full formula is moved to the storage area at the top of the dialog. Note that the formula will be evaluated and numerical results presented if the appropriate data for the selected molecule with the selected theoretical model are available in the database (in this case, HOMO and LUMO energies).

A number of the quantities that are calculated by ***Spartan*** are not available from the graphical interface but must be extracted from the archive of properties maintained by ***Spartan***. Examples include orbital energies other than those corresponding to the HOMO and LUMO, and the polar surface area defined to include sulfur and any attached atoms in addition to nitrogen, oxygen and any attached hydrogens (the default). However, these and other quantities may be used in formulas by manually editing the text string. The expressions provided in **Table 22-3** need to be preceded by a **@**. For example, ***@ HOMO(-1)*** accesses the energy of the molecular orbital immediately below the HOMO, while ***@PAREA(7,8,16,7:,8:,16:)*** accesses the polar area extended to sulfur and any attached hydrogens as well as the polar area of nitrogen, oxygen and any attached hydrogens. The integer arguments correspond to atomic numbers (sulfur is 16) while the integer arguments followed by ":" specify attached hydrogens (16: means include hydrogens attached to sulfur).

Table 22-1: Arithmetic & Boolean Operations and Mathematical Functions

arithmetic operations	**Boolean operations**
+ addition	> greater than
- subtraction	>= greater than or equal to
* multiplication	< less than
/ division	<= less than or equal to
^ raise to a power	== equal to
	!= not equal to
	| or
	& and

mathematical functions

ABS(x)	absolute value	LN(x)	natural logarithm
ACOS(x)	inverse cosine	LOG(x)	log (base 10)
ASIN(x)	inverse sine	SIN(x)	sine
ATAN(x)	inverse tangent	SQRT(x)	square root
COS(x)	cosine	TAN(x)	tangent
EXP(x)	exponential		

Table 22-2: Conversion Factors and Constants

ANGS2AU	Ångstroms to atomic units
AU2ANGS	atomic units to Ångstroms
EV2HART	eV to atomic units (hartrees)
EV2KCAL	eV to kcal/mol
EV2KJ	eV to kJ/mol
HART2KCAL	atomic units (hartrees) to kcal/mol
HART2EV	atomic units (hartrees) to eV
HART2KJ	atomic units (hartrees) to kJ/mol
KCAL2EV	kcal/mol to eV
KCAL2HART	kcal/mol to atomic units (hartrees)
KCAL2KJ	kcal/mol to kJ/mol
KJ2EV	kJ/mol to eV
KJ2HART	kJ/mol to atomic units (hartrees)
KJ2KCAL	kJ/mol to kcal/mol
PI	π

Formulas are used in conjunction with spreadsheet manipulations and plotting (both discussed later in this chapter) and with data mining (discussed in **Chapter 23**).

Spreadsheet

Associated with each ***Spartan*** document (including documents comprising a single molecule) is a spreadsheet. This may be displayed by selecting **Spreadsheet**.

The spreadsheet comprises a series of rows (corresponding to different molecules in the document and columns (corresponding to different molecular properties). This gives rise to cells, the number of which is the product of the number of rows (molecules) and the number of columns (molecular properties). The spreadsheet may be expanded or contracted by positioning the cursor at one of the corners, *pressing* the left mouse button and *dragging* the mouse.

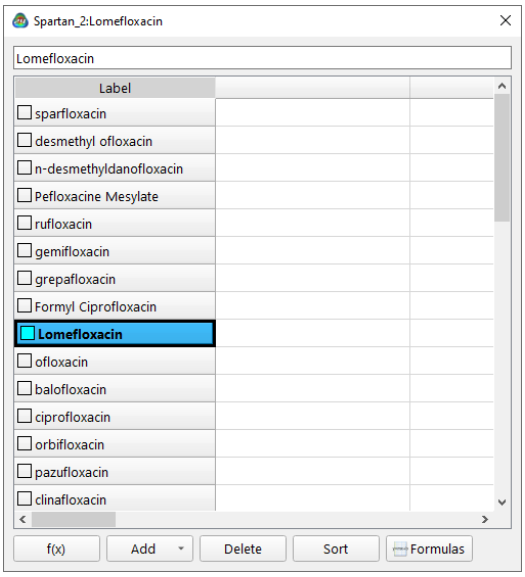
Only one molecule from one document may be selected (although several molecules may be simultaneously displayed). Molecule selection follows either by *clicking* on the spreadsheet cell containing the molecule label or identifier (leftmost column), or by using the and buttons or the scroll bar at the bottom left of the screen. Molecules may be ***animated*** (stepped through in succession) using the button at the bottom left of the screen. Animation speed may be adjusted from the **Settings Preferences** dialog (**Preferences** under the **Options** menu; **Chapter 25**). Selection of a new molecule in the document results in deselection of the previously selected molecule. A molecule may be designated for "permanent" display by *checking* the box to the left of its identifier in the spreadsheet. The molecules in a document may either be translated and rotated in concert or manipulated independently. This is controlled by **Coupled** (which operates in toggle mode) under the **Model** menu (**Chapter 18**). By default (**Coupled** checked) molecules move in concert. *Uncheck* **Coupled** to move them independently.

Table 22-3: Calculated Quantities

AREA	area of a user-defined plane ($Å^2$)
ELECTRONEGATIVITYEV	QSAR descriptor -(HOMO+LUMO)/2 (eV)
HARDNESSEV	QSAR descriptor -(HOMO-LUMO)/2 (eV)
HOMO(-n)EV	energy of nth orbital below the HOMO (eV)
HOMOBETA(-n)EV	energy of the nth orbital below the β HOMO (eV)
HYPERPOLARIZABILITY	beta hyperpolarizability parameter
INERTIA(i)	principal movements of inertia from largest (i=1) to smallest (i=3)
ISOTOPE(i)	mass number of atom i
LENGTH (i)	length of bond i (Å)
LOGPC	LogP from Crippen model
LOGPV	LogP from Villar model
LUMO(+n)EV	energy of nth orbital above the LUMO (eV)
LUMOBETA(+n)EV	energy of nth orbital above the β LUMO (eV)
PAREA (i, j, . . .)	surface area of space-filling model due to atoms of type i, j, ..., (atomic numbers or elemental symbols). i:, j:, ... signifies that attached hydrogens are also included. The surface area of nitrogen, oxygen and bonded hydrogens is the polar surface area. ($Å^2$)
POLARIZABILITY	alpha polarizability parameter

Table 22-4: Specialty Functions

AVG (column name)	average of values in column
FITVAL (fit name)	column of fit values from regression analysis
MIN (column name)	minimum of values in column
MAX (column name)	maximum of values in column
NUM (column name)	number of defined entries in column
ROW	the number of the row in the spreadsheet
ROW(molecule name)	the number of the row of molecule
REF(i, x)	the value of the x referenced to row i
STDEV (column name)	standard deviation of values in column
SUM (column name)	sum of values in column

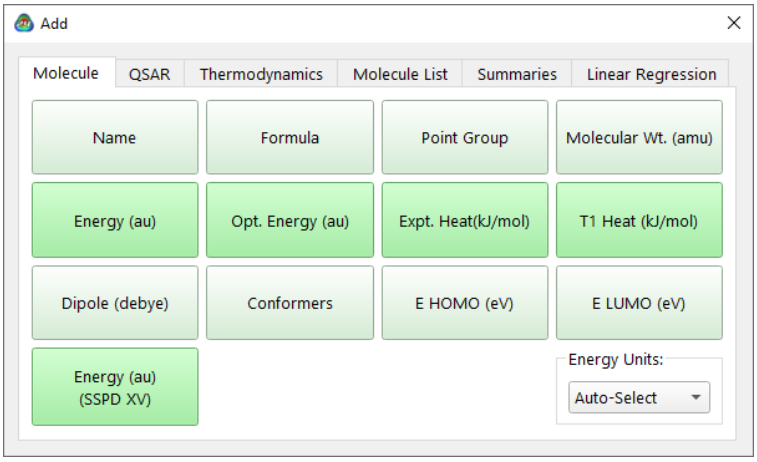


Upon initial entry, all columns of the spreadsheet except the leftmost column, are blank. The leftmost column contains a label that may be changed either by directly typing a new label into the spreadsheet or into the **Label** box in the **Molecule Properties** dialog (see discussion earlier in this chapter). Additionally, default identifiers (***M0001, ...***) may be replaced by chemical names if the molecule exists in the Spartan Spectra and Properties Database (SSPD) *by clicking* on **Label** (the leftmost header cell), then *right click* inside the header cell of the leftmost column and again select **Rename Selected using SSPD**.

Information may be added to the spreadsheet in several different ways:

From the Add Dialog

A selection of molecular properties may be entered into the spreadsheet by first *clicking* on the header cell of an empty column and then *clicking* on **Add...** at the bottom of the spreadsheet. Alternatively, *right click* inside the header cell and then to select **Add...** from the menu that results. This leads to a multi-tab dialog with **Molecule** selected.

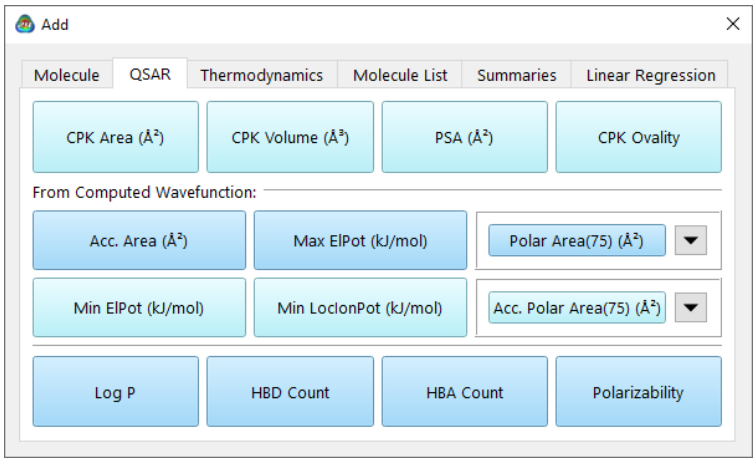


Name	molecule name as it appears in SSPD
Formula	molecular formula
Molecular Weight	molecular weight (in amu/Daltons)
E	energy (heat of formation, strain energy)
Pt Group	symmetry point group
Expt. Heat (kJ/mol)	experimental **ΔH°*f*** in kJ/mol (if available)
T1 Heat (kJ/mol)	T1 heat of formation in kJ/mol
Conformers	number of conformers
E HOMO	energy of highest-occupied molecular orbital
E LUMO	energy of lowest-occupied molecular orbital
Dipole	dipole moment (in debyes)
E ωB97X-V/ 6-311+G(2df,2p)	energy from SSPD (with ωB97X-D/6-31G* geometry)

One or more properties may be added to the spreadsheet by *clicking* on their entries.

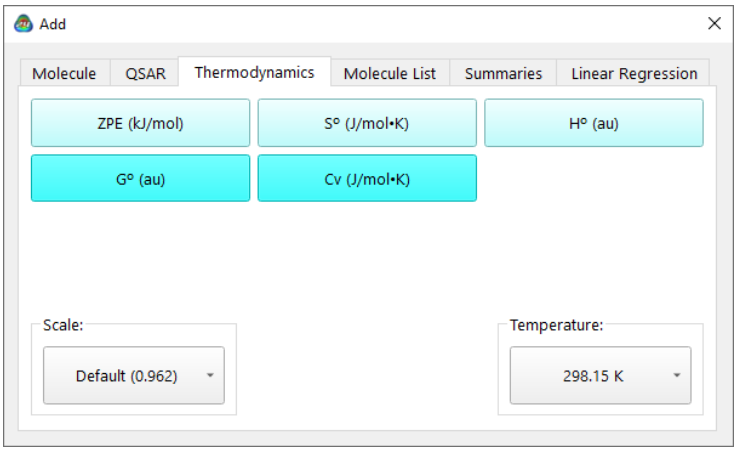
An **Energy Units** menu is available to change from the defaults (auto select) to au, kJ/mol or kcal/mol.

Clicking on the **QSAR** tab leads to another dialog.



CPK Area (Å^2)	surface area of a space-filling (CPK) model (in Å^2)
CPK Volume (Å^3)	volume of a space-filling (CPK) model (in Å^3)
PSA (Å^2)	polar surface area of a space-filling (CPK) model (in Å^2). Defined as the area due to electronegative atoms (N, O) and hydrogens attached to the atoms
CPK Ovality	measure of deviation from a spherical shape, where 1.0 = a sphere and values > 1.0 indicate deviation
Acc. Area (Å^2)	accessible surface area of an electron density surface (in Å^2). Surface is defined by electron density of 0.002 electrons/au^3 and accessible corresponds to a probe with a 1.0 Å radius. Probe radius may be changed in the **Settings** tab of the **Preferences** dialog (**Options** menu; **Chapter 25**)
Max ElPot (kJ/mol)	maximum value of the electrostatic potential on an electron density surface in kJ/mol)
Min ElPot (kJ/mol)	minimum value of the electrostatic potential on an electron density surface in kJ/mol)
Min LocIonPot (kJ/mol)	minimum value of the local ionization potential on an electron density surface in kJ/mol)
Polar Area(75) (Å^2)	accessible area of the region on an electrostatic potential map where the absolute value of the potential is >75 kJ/mol in (Å^2). The value of the potential may be changed to 100 or 125 kJ/mol. Probe radius may be changed.
LogP	octanol water partition coefficient
HBD Count	number of hydrogen-bond donor sites
HBA Count	number of hydrogen-bond acceptor sites
Polarizability	polarizability

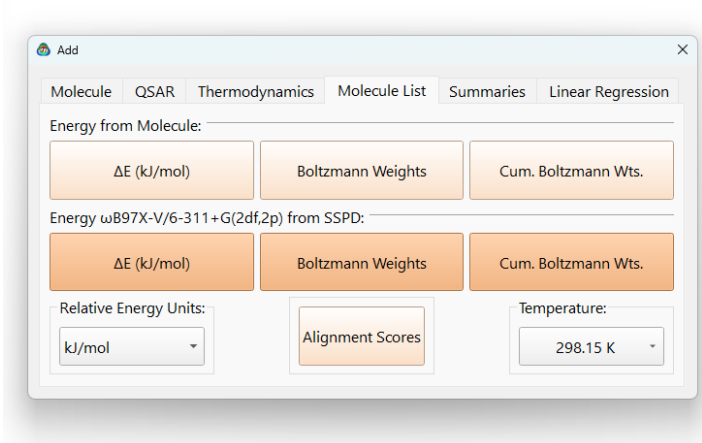
Clicking on the **Thermodynamics** tab leads to another dialog.



ZPE (kJ/mol)	zero-point energy (in kJ/mol)
S° (J/mol•K)	entropy (in J/mol•K)
H° (au)	enthalpy (in au). Sum of electronic energy and zero-point energy adjusted for finite temperature
G° (au)	Gibbs energy (in au). Sum of enthalpy and entropy.
Cv (J/mol•K)	heat capacity at constant volume (in J/mol•K)

Scale Vibrational Frequencies by is used to scale calculated frequencies, where default applies to the ambient model and is available for B3LYP/6-31G*, EDF2/6-31G* and ωB97X-D/ 6-31G* models. **Temperature** is used to set temperature. Note that vibrational frequencies need to be available for all quantities under the **Thermodynamics** tab.

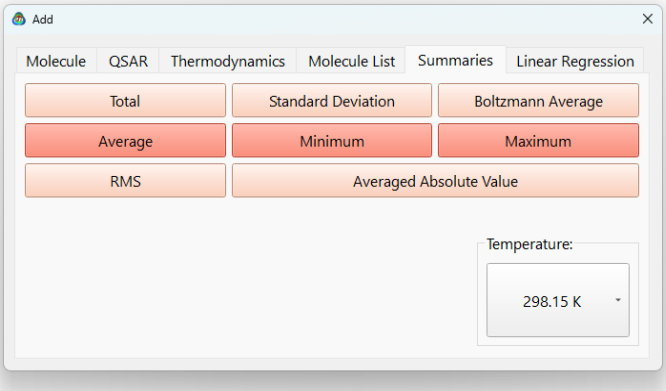
Clicking on the **Molecule List** tab leads to another dialog. This allows quantities for different molecules (or different conformers of the same molecule) in a list to be related.



rel. E (kJ/mol)	energy (heat of formation, strain energy) relative to selected molecule
Boltzmann Weights	Boltzmann weight based on relative energies
Cumulative Boltzmann Weights	Sum of the Boltzmann weights for the selected molecule and all molecules with lower energy than the selected molecule
Alignment Scores	$1-R^2/N$, where R^2 is the root mean square distance and N is the number of alignment centers. 1 is a perfect score
Relative ωB97X-V/ 6-311+G(2df,2p)//	energy from ωB97X-V/6-311+G(2df,2p)// ωB97X-D/6-31G* model relative to the selected molecule. Only applicable if data are available in SSPD

Units for relative energies and temperature for Boltzmann weights and cumulative Boltzmann weights may be selected from menus at the bottom.

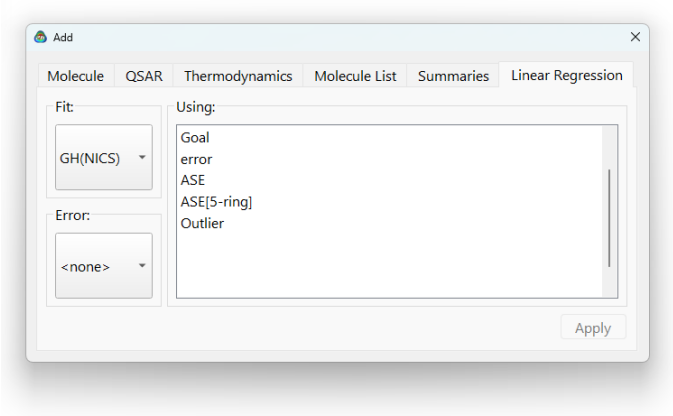
Clicking on the **Summaries** tab leads to another dialog.



Total	sum of column values
Minimum	minimum of column values
Standard Deviation	standard deviation of column values
Boltzmann Average	Boltzmann weighted average of column values
Average	average of column values
Maximum	maximum of column values
RMS	rms of column values

The temperature (for Boltzmann average) may be changed from the default value of 298.15K from the menu at the bottom of the dialog.

Linear regression analysis may be performed on data in the spreadsheet. *Clicking* on the **Linear Regression** tab brings up a dialog.



Select one entry from the **Fit** menu and one or more entries from the list under **Using**. *Clicking* on **Apply** performs the linear regression analysis and places the results in a row at the bottom of the spreadsheet identified by **Fit**. As many regression analyses as desired may be performed on the data in the spreadsheet. The individual results will be entered as separate rows in the spreadsheet, with names **Fit1**, **Fit2**, etc. Additional information about the regression analyses is available from the **Regression Properties** dialog (see discussion earlier in this chapter).

From Post Buttons

Post buttons found in a number of properties dialogs provide an alternative method to the **Add** dialog for entering calculated properties into the spreadsheet. Note that some properties may require user specification. These include individual bond distances, angles and dihedral angles (available from **Measure Distance**, **Measure Angle** and **Measure Dihedral** under the **Geometry** menu; **Chapter 19**), bond distance, angle and dihedral angle constraints (available from **Constrain Distance**, **Constrain Angle** and **Constrain Dihedral** under the **Geometry** menu; **Chapter 19**), atomic charges, chemical shifts (available from the **Atom Properties** dialog; this chapter), the accessible area of an electron density surface, the polar area and accessible polar area of an electrostatic potential map, the area of a selected region (band) of a banded property map, minimum and maximum property values on a map and the value of the property at a specific location on a property map (available from the **Surfaces Properties** dialog; this chapter). With the exception of the property value on a map and the area of a selected band, post generates an entire column. Where atom labels are involved, for example, in defining a specific distance, post can be expected to yield consistent results for all molecules in a document only where the molecules are closely related, for example, molecules resulting from a conformational search, or where labels have been explicitly reassigned*. The property value and the area of a selected band on a map is posted only for the selected molecule. Post buttons are also available for experimental heats of formation contained in SSPD, as well as for ωB97X-V/6-311+G(2df,2p) energies and T1 heats of formation.

From the Clipboard

Properties of one or more molecules in a document may be copied onto the clipboard and then pasted into individual spreadsheet cells. These include (but are not restricted to) bond distances, angles and dihedral angles (**Measure Distance**, **Measure Angle** and **Measure Dihedral** under the **Geometry** menu; **Chapter 19**), bond distance, angle and dihedral angle constraints (**Constrain Distance**, **Constrain Angle** and **Constrain Dihedral** under the **Geometry** menu; **Chapter 19**), atomic charges and chemical shifts (**Atom Properties** dialog; this chapter), and the value of a property on a property map and the area of a selected band (**Surface Properties** dialog; this chapter). To copy to the spreadsheet, first highlight the numerical value of the property in the appropriate screen location (distances, etc.) or dialog (charges, etc.), then select **Copy** from the **Edit** menu, then *click* on the appropriate (destination) cell in the spreadsheet, and finally select **Paste** from the **Edit** menu.

* Label reassignment is accomplished using the **Atom Properties** dialog (see discussion earlier in this chapter).

User-Defined Expressions

An expression may be entered either into a header cell (in which case it refers to all molecules) or into an individual cell (in which case it refers only to a single molecule). Expressions in the column header take the form ***name=formula***, where ***formula*** is made up of arithmetic operations (**Table 22-1**), specialty functions (**Table 22-4**), calculated quantities (**Table 22-3**), conversion factors and constants (**Table 22-2**) in addition to numerical values. References to specialty functions, calculated quantities and conversion factors and constants must be preceded by **@**. For example, ***mu=@DIPOLE*** typed into a header cell gives the dipole moment. Some functions have arguments, for example, ***c1*** and ***c2*** in the expression ***c12= @DISTANCE (c1,c2)*** refer to atoms c1 and c2, while ***3*** in the expression: ***orbitalE=@HOMO (-3)*** designates the energy of the molecular orbital three orbitals below the HOMO. It is necessary to *press* the **Enter** key (**return** key on Mac) following entry of the expression into a cell. The leading ***name=*** is optional for individual entries.

Formula construction inside the spreadsheet uses the same rules and references the same set of functions described in **Formulas** earlier in this chapter (in addition to the specialty functions in **Table 22-4**). It may be more convenient to construct expressions in the formula editor and *drag* them into the spreadsheet than to construct them in the spreadsheet. Note that spreadsheet expressions refer to the molecules in the spreadsheet and not to their reaction products.

Numerical Data

Numerical data may be entered by typing directly into the spreadsheet. A column header first needs to be specified. *Double click* on an empty column header cell, type in a name and *press* the **Enter** key (**return** key on Mac). Then, type the data into individual cells of the new column (*press* the **Enter** (**return**) key following each data entry). Or, use the ↓ key.

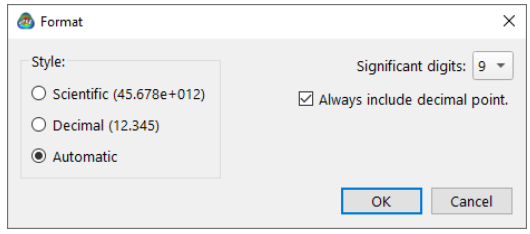
Each row in a spreadsheet corresponds to a molecule in a document, and new rows are automatically added in response to adding new molecules to the document. New molecules may be added by building (**Build New Molecule** under the **File** menu; **Chapter 16**) or sketching (**Sketch New Molecule** under the **File** menu; **Chapter 16**), by appending one or more existing documents each containing one or more molecules using either **Append Molecule(s)...** under the **File** menu (**Chapter 16**), or by *right clicking* inside the header cell of the first available row and selecting **Append** from the contextual menu that appears, by pasting from the clipboard, or by *dragging* from the file system. To copy a molecule onto the clipboard, first select (*click* on) it, and then select **Copy** from the **Edit** menu, or *click* on its identifier (left most column) in its spreadsheet, and then select **Copy** from the **Edit** menu. Alternatively, *right click* either on the molecule or on its identifier in the spreadsheet and select **Copy** from the menu that appears. Several molecules may be selected (and copied) at once using the **Shift** and **Ctrl** keys in the usual manner. To copy the contents of the clipboard to its destination, *click* on an empty row header in the spreadsheet (for the destination document), and then select **Paste** from the **Edit** menu. An alternative to the two-step **Copy-Paste** procedure is to *drag* the molecule or set of molecules from one spreadsheet to another.

A row (molecule) may be deleted from a spreadsheet, either by first selecting the molecule and then selecting **Delete Molecule** from the **File** menu (**Chapter 16**), or by first *clicking* on its identifier in the spreadsheet (leftmost column) and then either *clicking* on the **Delete** button at the bottom of the spreadsheet, or by *right clicking* on its identifier in the spreadsheet and then selecting **Delete Selected** from the contextual menu that appears. A warning is provided prior to deletion. An entire column in the spreadsheet may be deleted by first *clicking* inside its header cell and then *clicking* on the **Delete** button (or **Delete Selected** from the contextual menu).

Rows in the spreadsheet may be sorted according to the numerical values in any column either by first *clicking* inside the header cell and then *clicking* on the **Sort** button at the bottom of the spreadsheet or by *right clicking* inside the header cell and selecting **Sort** from the contextual menu that appears. The rows are placed in ascending order, the smallest (least positive) value of the selected property at the top, largest (most positive) value at the bottom. To sort in descending order, hold down the **Shift** key before *clicking* on the **Sort** button or selecting **Sort** from the contextual menu.

Information in one or more columns of the spreadsheet may be formatted by *right clicking* inside the header cell(s) and selecting **Format Selected** from the contextual menu.

Format as desired and *click* on **OK** to remove the dialog. The full contents of the spreadsheet may be formatted by *right clicking* inside the header cell for the left most column and then selecting **Format Selected** from the contextual menu.



A button at the bottom right of the spreadsheet toggles between numerical representation of data, **f(x)**, and formula presentation, **=?**.

The spreadsheet may be printed by *right clicking* inside the spreadsheet and selecting **Print** from the menu that results.

Spreadsheets are associated with individual documents and, where more than one document is open on screen, multiple spreadsheets may be displayed. A spreadsheet is removed when the associated document is closed and may also be removed by *clicking* on .

The contents of the spreadsheet may be brought into Excel™ using the clipboard. Select whatever cells are to be copied, select **Copy** from the **Edit** menu. Alternatively, *right click* with the proper cells selected and select **Copy** from the menu that appears. **Paste** into Excel.

The contents of an Excel spreadsheet may be brought into ***Spartan***. Copy whatever information is to be transferred to the clipboard, move into ***Spartan***, then *click* on the appropriate cell and select **Paste** from the **Edit** menu (or *right click* on the appropriate cell and select **Paste** from the menu that appears). Note, that information on the clipboard that goes beyond the number of rows in ***Spartan***'s spreadsheet will be ignored.

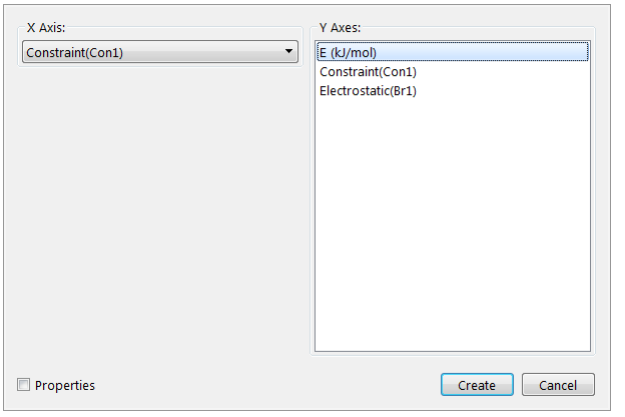
Plots...

Plots may be constructed from data in a spreadsheet and a variety of simple curves fit to these data. As with presentation of spectral data (see discussion earlier this chapter), ***Spartan*** allows both simplified and more general modes for presentation, the choice of which being left to the user.

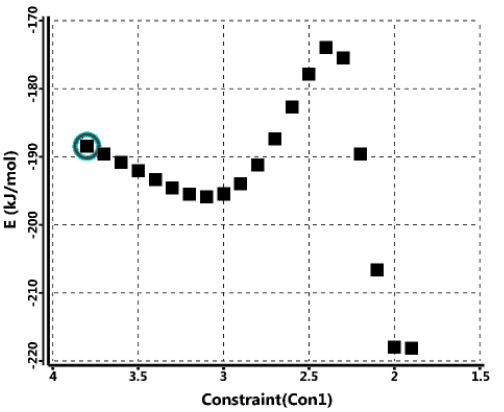
Simplified Presentation of Plots

With **Plots Pane** in the **Settings Preferences** dialog checked (**Preferences** under the **Options** menu; **Chapter 25**), selecting **Plots** from the **Display** menu leads to an empty display pane at the right of the screen.

Clicking on (add plot) in the bar at the top of the plots pane leads to a dialog.



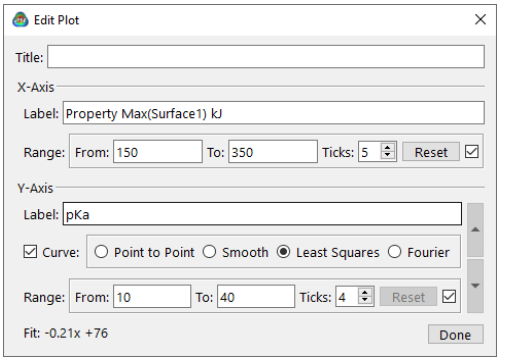
You need to select an item from the **X Axis** menu and one or more items from the **Y Axes** list, and then *click* on the **Add** button at the bottom of the dialog. A plot appears in the plot pane; the **Add Plot** dialog is dismissed.



By default, the scales for both horizontal and vertical axes are set to bound the data trying to provide limits and increments that are "rounded". Moving the mouse left and right while holding down the right button slides the horizontal scale but does not change the range. Similarly, moving the mouse up and down while holding down the right button slides the vertical scale. The horizontal range may be changed by moving the mouse left and right while holding down both the right button and shift key, and the vertical range changed by moving the mouse up and down while holding down both the right button and the shift key. The scroll wheel may be used to simultaneously change both horizontal and vertical ranges. The original settings may be restored by *clicking* on in the bar at the top of the plots pane.

On a touch screen device, move two fingers left and right and up and down to slide the viewable horizontal and vertical scales, respectively. Pinch two fingers left and right and up and down to change the horizontal and vertical scales, respectively.

The plot ranges may also be changed by *clicking* on in the bar at the top of the plots pane.



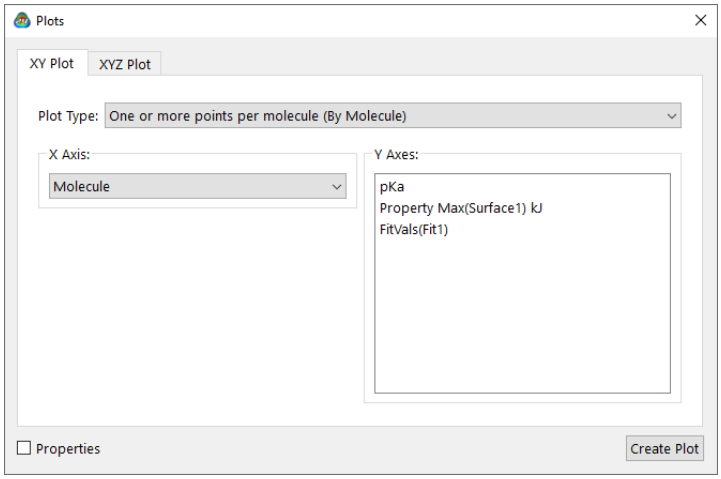
The resulting dialog also allows axis labels to be altered (from their initial values designated in the spreadsheet) the number of "ticks" of horizontal and vertical axes to be changed and

a plot title to be added. Finally, the "curve" can be changed to **Point to Point**, **Smooth**, **Least Squares** or **Fourier**.

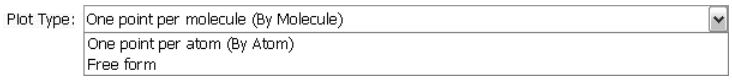
Additional plots may be added by *clicking* on in the bar at the top of the plots pane. Each plot is given a tab. Only one plot may be displayed at a time as controlled by which tab is selected. The selected (and displayed) plot may be deleted by *clicking* on .

Generalized Presentation of Plots

With **Plots Pane** in the **Settings Preferences** dialog unchecked (**Preferences** under the **Options** menu; **Chapter 25**), selecting **Plots** from the **Display** menu leads to the **Plots** dialog.



A tab at the top of the dialog selects between **XY Plot** (two dimensions) or **XYZ Plot** (three dimensions). Common to both is the **Plot Type** menu.



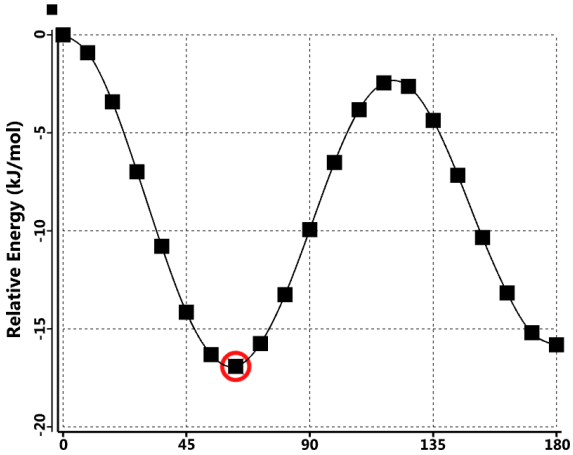
This distinguishes between plots based on the entire set of molecules in a document (**By Molecule**), to those based on the atoms in a single molecule (**By Atom**). The latter is the mode used to plot calculated IR, Raman and NMR spectra. The third choice (**Free Form**) is used to construct single molecule plots, for example, calculated, UV/ vis spectra as well as plots of experimental IR, Raman, NMR and UV/visible spectra. This is normally done automatically from the appropriate spectra dialog (see discussion earlier in this chapter).

By atom and free form plots make use of vector capabilities in ***Spartan's*** spreadsheet. See the FAQ ***Equations in Spartan*** available under **Help** from the **Help** menu (**Chapter 27**).

Also common to both is the **Properties** checkbox at the bottom left of the dialog. Unchecked, this provides access for the **X-Axis** (X-Axis and Y-Axis for an XYZ plot) menu and **Y-Axes** (**Z-Axes** for an XYZ plot) only to quantities that have been entered in the spreadsheet. Checked, it provides access to all quantities in the ***Spartan***'s Property Archive resulting from a calculation or from retrieval of information from SSPD.

XY Plots

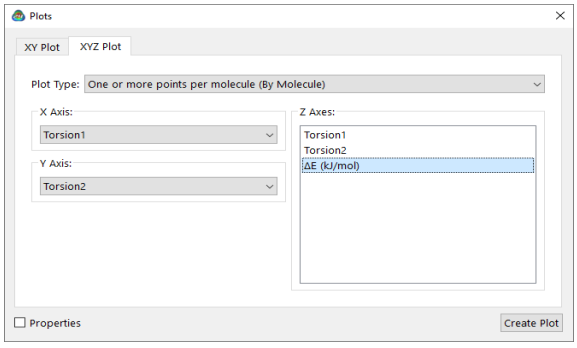
The **XY Plot** dialog incorporates an **X Axis** menu designating the molecular property to be displayed among the X (horizontal) axis, and a list of properties to be displayed along the Y (vertical) axis. These properties correspond one to one to the columns in the associated spreadsheet. To construct a 2D plot, select an item from the **X Axis** menu, then *click* on one or more items from the **Y Axes** list and finally *click* on **OK**. (Repeated *clicking* on a property in the **Y Axes** turns the list on and off.) The dialog is removed from the screen and a plot appears.



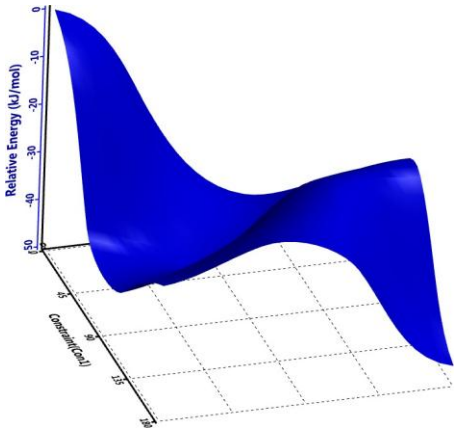
The plot may be moved about the screen. Select (*click* on) either the frame or on one of the curves (it will turn gold in response), then hold down the right mouse button and *drag* the mouse. The plot may also be scaled (expanded or shrunk) by first selecting either the frame or one of the curves and *dragging* the mouse while holding down both the right button and the **Shift** key. The plot may not be rotated. Touch-screen operations are also available. Move two fingers across the plot to translate it and pinch two fingers to scale it.

XYZ Plots

The XYZ Plot is reached by clicking on the **XYZ Plot** tab at the top of the dialog. This leads to a new dialog.



This is very similar to the XY Plot, except that it incorporates **X Axis** and **Y Axes** menus to designate quantities to be displayed along the X and Y axes, respectively, as well as a list of properties to be displayed along the Z axis. To construct a 3D plot, select one item from each of the **X Axis** and **Y Axis** menus, then *click* on one or more items from the **Z Axes** list and finally *click* on **OK**. A plot appears.



The plot can be moved about the screen. Select (*click* on) either the frame or on one of the curves (it will turn gold in response), then hold down the right mouse button and *drag* the mouse. The plot may also be scaled (expanded or reduced in size) by first selecting either the frame or one of the curves and *dragging* the mouse while holding down both the right button and the **Shift** key, or via the center mouse wheel.

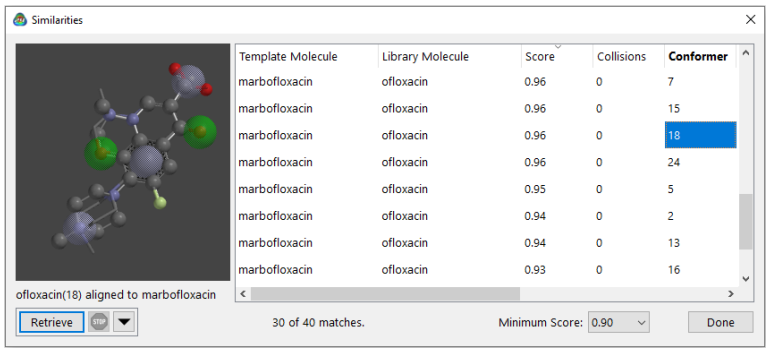
Finally, 3D plots (unlike 2D plots discussed earlier) may be rotated, by first selecting either the frame or one of the curves and *dragging* the mouse while holding down the left button. Touch-screen operations also apply. Move two fingers across the plot to translate it, move one finger to rotate it and pinch two fingers to scale it.

Plots initially presented connect the data points with cubic splines. A variety of curve fits (linear, least squares, Fourier, etc.) as well as a variety of different presentation formats are available under the **Plot Properties** and associated dialogs (see **Properties** earlier in this chapter).

A plot may be deleted by first selecting (*clicking* on) **Delete** from the **Build** menu , or holding down the **Delete** key, and then *clicking* on either the plot frame or on an individual curve. *Clicking* on the frame removes it and all associated curves, while *clicking* on a curve removes only this curve. However, the last curve deleted also deletes the frame. IR, Raman, NMR and UV/ vis spectral plots may also be removed from the corresponding **Spectra** dialog.

Similarities...

The results of a similarity analysis based either on selected atoms in a structure or on a set of CFD's as a template may be viewed using the **Similarities** dialog.

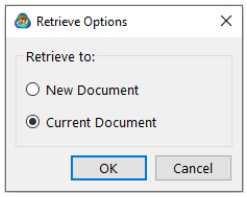


The box at the right of the dialog identifies each hit in terms of the identity of the **Template Molecule**, the **Library Molecule**, and in the case where a library molecule has been expanded in terms of a set of diverse conformers, the identity of the conformer. It also provides a score* and, in the case where excluded volume CFD's are present in the template, the number of collisions between these elements and the library molecule. Finally, it indicates whether the enantiomer for the molecule included in the library has been inverted. Entries in the dialog may be sorted by score or number of collisions (as well as by the names of the template and library molecules) by *clicking* on the appropriate column header at the top of the box.

A minimum score can be set from the **Minimum Score** menu at the bottom right of the dialog. The number of matches that satisfy this minimum is reported.

As hits are selected by *clicking* on them in the box, the library entry superimposed onto the template will be displayed in the window at the left of the dialog. This graphic may be manipulated (rotated, translated, scaled) with the usual mouse/keyboard commands (the cursor needs to be positioned inside the window). The display style cannot be altered, nor can any measurements be made.

The library entry associated with a particular hit may be retrieved (brought into ***Spartan***'s main window) by first selecting it and then *clicking* on the **Retrieve** button at the bottom of the **Similarities** dialog. The template cannot be retrieved. Retrieval can either be to a **New Document** or to the **Current Document** depending on the setting in the **Retrieve Options** dialog accessed by *clicking* on ▼ to the right of the **Retrieve** button.



* The score is defined as $[(1-R^2)/N]$-penalty, where R^2 is the root mean square distance between template and library molecule centers and N is the number of similarity centers. Whereas molecule alignment as a prelude to similarity scoring is based on atomic, CFD and pharmacophore centers ***only***, penalties are given for alignments that lead to unfavorable steric interactions between template and library entries or (in the case of molecule alignment by CFD's) to incorrect orientation of hydrogen bond donor or acceptor CFD's. These are subtracted from the score based on R^2. 1 is a perfect score.

Retrieval to the **Current Document** (that with the template) allows both the template and one or more hits from the library to be displayed simultaneously, with a full range of model styles and display options.

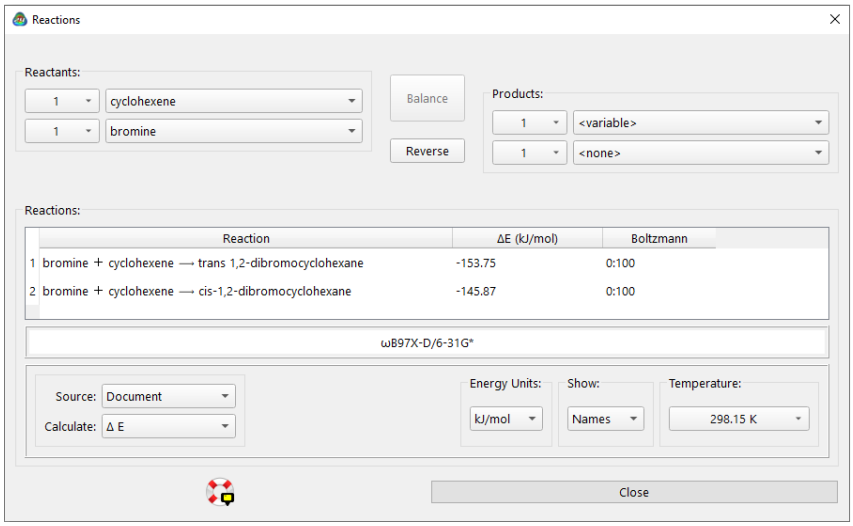
The **Similarities** dialog is closed by *clicking* on the **Done** button.

Reactions... ( )

Data entered in a ***Spartan*** document may be used to calculate reaction energies including activation energies.

$$\Delta E = E_{product1} + E_{product2} - E_{reactant1} + E_{reactant2}$$

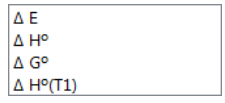
Selection of **Reactions...** from the **Display** menu leads to the **Reactions** dialog.



Two menus under **Reactants:** and two menus under **Products:** identify the reactants and products of reaction. They correspond to the labels (identifiers) of the molecules in the document, plus a null entry **<none>**. Note that the overall reaction needs to be mass balanced. An additional optional **<variable>** allows for all possible reactants and/or products leading to balanced reactions, to be displayed.

The **Source** menu identifies the source of the energies to be used in the reaction energy calculation: either the **SSPD** or the **Current Document** (or **Both**).

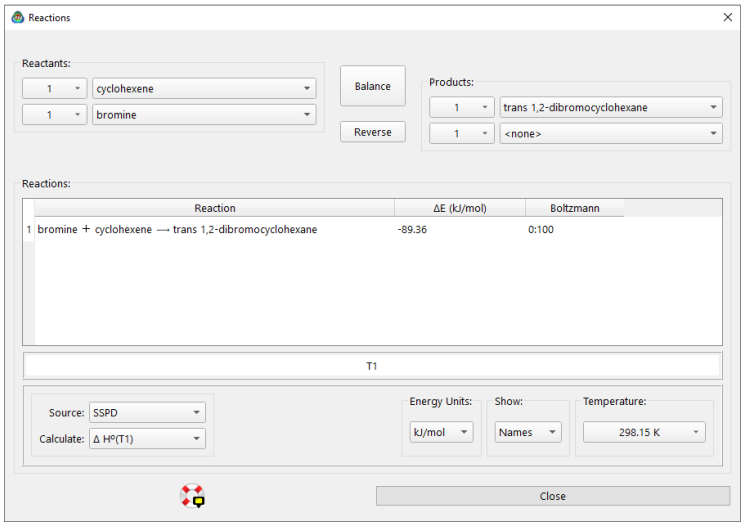
Selection of **SSPD** (from the **Source** menu) requires specification from **Calculate:** menu:



ΔE uses the ωB97X-V/6-311G(2df,2p) energy at the ωB97X-D/ 6-31G* geometry. ΔH° and ΔG° use the same energies but apply thermodynamic corrections based on EDF2-6-31G* frequencies.
ΔH°(T1) uses data from the T1 thermochemical recipe.

The **Energy Units** menu allows for selection of units.

A reaction energy is computed by selecting an energy from the drop- down menu next to **Calculate:** at the bottom left of the dialog.



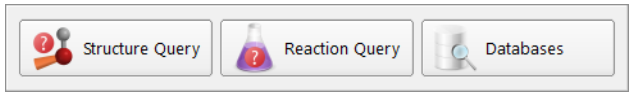
Substituted molecules can be used in place of real molecules as reactants and products. By accessing SSPD, this means that the energies for a series of related (by substitution) reactions can be computed at once.

The results of a reaction energy calculation may be printed by *right clicking* inside the display area of the **Reactions** dialog and selecting **Print** from the menu that results.

The **Reactions** dialog is closed by *clicking* on .

Chapter 23
The Search Menu

*Functions available under the **Search** menu provide for defining structure and reaction queries to be used in searches of databases licensed and attached to **Spartan**. These may include the Cambridge Structural Database (**CSD**) of experimental X-ray structures, the Spartan Spectra and Properties Database (**SSPD**) of calculated spectra, properties, QSAR descriptors and thermodynamic quantities, the Spartan Infrared Database (**SIRD**) of calculated infrared spectra, the NIST database of experimental infrared spectra (**XIRD**) and the Spartan Reaction Database (**SRD**) of calculated transition states. New in **Spartan'26** is the Natural Products Database (**NPD**), a collection of structures, associated experimental NMR shifts, and calculated shifts for some ≈3,000 or so published natural products. Also new in **Spartan'26** training sets from new neural network energy models which include energies from three high quality density functional models: **ωB97X-V**, **ωB97M-V**, and **ωB97M(2)** using the 6-311+G(2df,2p) basis set and starting from ωB97X-D/6-31G* geometries. The model headings may be optionally displayed as new search tabs alongside the above mentioned resources. The CSD is searchable by substructure, name, author and reference code, SRD is searchable by substructure, SSPD and the new energy model collections [**ωB97X-V**, **ωB97M-V**, and **ωB97M(2)**] may be searched by substructure, name, formula, molecular weight and isomer and the SIRD and XIRD are searched based on matching to an unknown infrared spectrum.*



Structure Query

This designates one or more free valences on a structure as attachment points. *Clicking* on the free valence marks it with an orange cone and designates it as an attachment point. This means that anything directly bonded to the valence

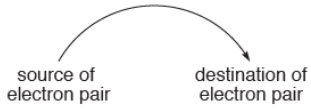
(including hydrogen) is an acceptable hit for a search on a database. For example, *clicking* on the *para* free valence in toluene designates that any *para*-substituted toluene (including toluene itself) is an acceptable hit. *Clicking* on both free valences of hydrogen peroxide means that all peroxides (including hydrogen peroxide and all hydroperoxides) are

acceptable hits. Free valences not designated as attachment points are assumed to be hydrogens. *Clicking* on a free valence previously designated as an attachment point removes the cone and its designation as an attachment point.

One or more structure queries may also be associated with a 2D sketch (see **Chapter 20**). These appear as ?'s on the sketch but as orange cones on the 3D structures.

Reaction Query

The main use of **Reaction Query** is to access a previously calculated transition state from the Reaction Database, or to identify similar transition states. As we shall see later in the chapter, it also serves to identify the product or products of a chemical reaction from the reactants to be used for mining the Spartan Spectra and Properties Database. The query comprises a set of curved arrows familiar to generations of organic chemists and formally corresponding to the movement of a pair of electrons. In general, the same *arrow-pushing* specifications described under the **Search Transition State** mode (**Chapter 20**, **Build** menu), are also applicable to the **Reaction Query** mode. The general strategy employed includes building either reactants or products, defining the reaction via definition of arrows that define the transition state, and then (typically) opening valences on reactants to broaden the search and potentially include related reactions among the results.



There are two possible sources of an electron pair and three possible destinations, leading to six combinations:

lone pair → lone pair	move lone pair
lone pair → bond	use lone pair to increase bond order
lone pair → space between atoms	use lone pair to create new (single) bond
bond → lone pair	decrease bond order to make lone pair
bond → bond	decrease order of one bond to increase order of another bond
bond → space between atoms	decrease order of one bond to make a new (single) bond

The first of these is a null operation, and its inclusion has no effect.

As presently implemented, the reaction query procedure does not distinguish between a lone pair and a radical center. Examples are provided at the end of this section.

Selecting **Reaction Query** results in a message at the bottom left of the screen.

Select atom or bond as tail.

The tail of the arrow corresponds to the source of the electron pair. If the source is a lone pair, then select the atom that holds the lone pair. If the source is a bond, then select the bond. *Clicking* on an atom or bond highlights (colors gold) the atom or bond and leads to a new message at the bottom left of the screen.

Select one atom, two atoms (a gap), or a bond as head.
If one atom, select it twice.

Clicking again on the same atom (or same bond) removes the highlighting and returns the first message. The head of the arrow corresponds to the destination of the electron pair. If the destination is an atom (leading to a lone pair), then *double click* on it. If the destination is an existing bond (leading to an increase in bond order from single $\rightarrow$ double or double $\rightarrow$ triple), then *click* on the bond. If no bond presently exists, *click* on the two atoms that will become bonded upon reaction. These operations result in a curved arrow being drawn on the reactant structure. This extends from an atom, or the center of a bond to an atom, or the center of a bond, or the center of a dotted line that has been drawn between atoms that are to be bonded. The original message returns to the bottom left of the screen.

Note that the head and tail do not need to reside on atoms or bonds on the same fragment. Also the tail may involve atoms on two detached fragments.

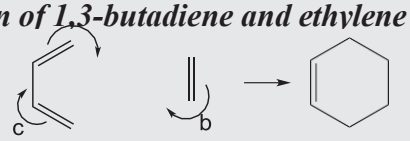
The process (tail selection followed by head selection) is repeated as necessary to fully define the reaction. Incorrect reaction arrows may be removed by selecting **Delete** from the **Build** menu followed by *clicking* on the arrow to be deleted. You need to again select **Reaction Query** in order to continue arrow selection. Alternatively, *click* on the arrow(s) to be deleted while holding down the **Delete** key.

Once defined, reaction queries can be used to search the Spartan Reaction Database (SRD) for transition-state structures associated with the defined reaction (discussed in **Databases**). Unless one or more structure queries have been specified, the search will only return an exact match (if it exists in SRD). More commonly, structure queries will be specified in which the objective is to provide a guess at a transition state based on its similarity to an entry in the SRD (see discussion in **Guess Transition State** later in this chapter).

One or more reaction queries may also be associated with a 2D sketch (see **Chapter 20**). These appear as curled arrows (↷) on the sketch and on the resulting 3D structures.

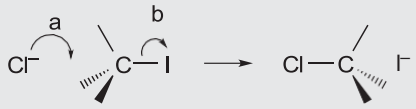
Examples of Reaction Queries

Diels-Alder reaction of 1,3-butadiene and ethylene



a, b. double bond to empty space leading to single bond and single bond

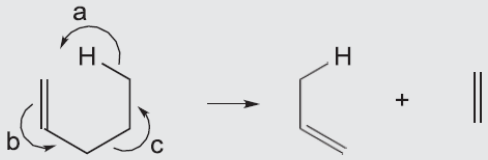
c. double bond to single bond leading to single bond and double bond



a. lone pair on Cl^- to empty space leading to ClC bond

c. CI bond leading to lone pair on I

Ene reaction of 1-pentene

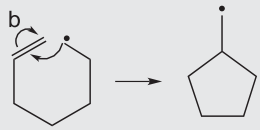


a. CH bond to empty space leading to CH bond

b. double bond to single bond leading to single bond and double bond

c. single bond to single bond leading to empty space and double bond

Ring closure of 1-hexenyl radical to methylcyclopentyl radical



a. radical center to empty space (between C_2 and C_6) leading to single bond
b. double bond to terminal carbon leading to single bond and radical center

Databases

This provides access to the Cambridge Structural Database (***CSD***) of experimental X-ray crystal structures, the Spartan Spectra and Properties Database (***SSPD***) of structures, IR and NMR spectra and a variety of molecular properties, the Spartan Reaction Database (***SRD***) of calculated transition states, the Spartan Infrared Database (***SIRD***) of calculated infrared spectra and the NIST database (***XIRD***) of experimental infrared spectra. New in ***Spartan'26*** are **ωB97X-V**, **ωB97M-V**, and **ωB97M(2)**, these include energy results from the 6-311+G(2df,2p) basis set starting from ωB97X-D/6-31G* geometries. The CSD may be searched by substructure, molecule, name, author name, and CSD reference code. SSPD, CSD and SRD may be searched based on substructure. SSPD may also be searched by name, formula, molecular weight and isomer. If a selected molecule on screen exists in SSPD, it may be replaced by a database entry and, even if it is not replaced, the molecule name in SSPD may be used. SRD may only be searched by substructure. SIRD and XIRD are searched by comparison with an input infrared spectrum. The ωB97X-V, ωB97M-V, and ωB97M(2) collections may be searched in exactly the same way as the SSPD.

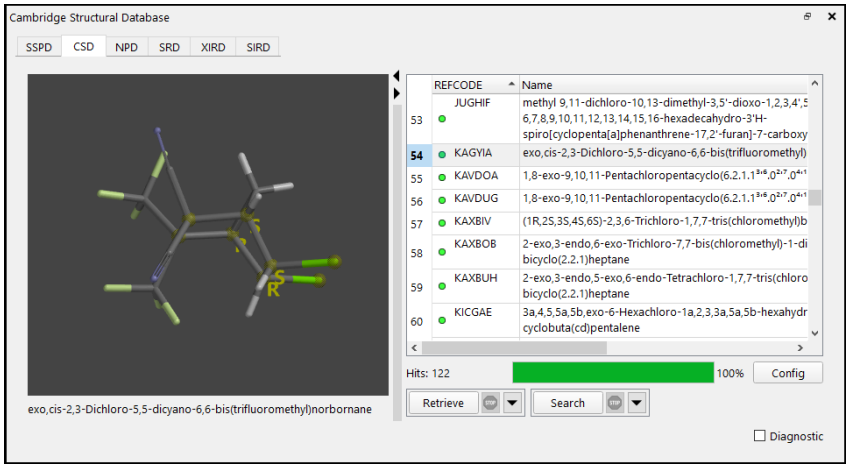
Cambridge Structural Database (CSD)

The **C**ambridge **S**tructural **D**atabase (**CSD**) is a collection of more than 1.25 million experimental X-ray crystal structures for organic and organometallic molecules.*

The collection is maintained by the Cambridge Crystallographic Data Centre (CCDC) and grows approximately 10% per year. CSD not only contains information about the molecular geometry, but also about the manner in which molecules pack in the solid state. In short, the CSD represents a gold mine of detailed experimental structural information and also serves to identify molecules that can be (and have been) synthesized and purified through crystallization.

* CSD is not included with ***Spartan*** but is available by subscription from CCDC (https://www.ccdc. cam.ac.uk/). For installation instructions, see **Appendix G**.

Tutorials illustrating the use of CSD from ***Spartan*** are provided in **Chapter 15**.

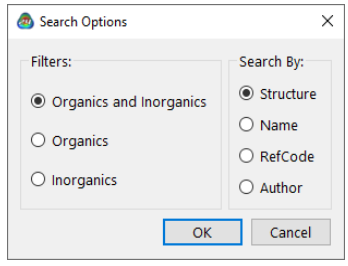


Substructure searching: A molecule is first constructed, and one or more attachment points are identified. *Clicking* on the **CSD** tab at the top of the **Database** dialog leads to the **CSD** dialog.

This contains a window at the left for previewing structures, a box at the right for listing hits on CSD and buttons at the bottom to control aspects of the search and subsequent transfer of CSD data into the file system. A substructure search is set up by designating one or more free valences as attachment points. Selecting **Structure Query** from the **Search** menu and then *clicking* on a free valence marks it with an orange cone. *Clicking* again removes the cone and designation as an attachment point. Anything may be grown off an attachment point (including hydrogen), although substituents are not permitted.

Free valences not designated as attachment points are assumed to be hydrogens.*

To limit a CSD search either to organic molecules (about half the total collection) or to molecules containing transition metals, lanthanides or actinides, bring up the **Search Options** dialog by *clicking* on ▼ to the right of the **Search** button at the bottom right of the dialog, and *check* **Organics** or **Inorganics**, respectively.

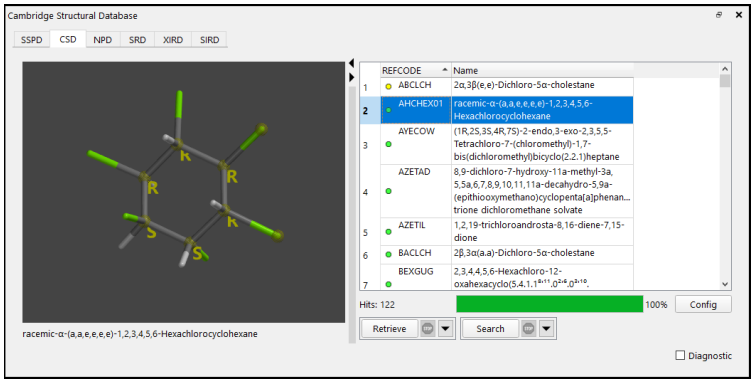


To search all molecules in CSD, *check* **Organics and Inorganics** from the **Search Options** dialog.

* ***Spartan***'s interface to CSD accesses only molecular information. Additional information related to packing in the crystal may be accessed using CCDC's program ConQuest (and other application), supplied as part of the CSD subscription.

After a substructure has been specified, a search of CSD is initiated by *clicking* on the **Search** button at the bottom right of the dialog. If there are hits to be found, these will appear in the scroll box at the right of the dialog. The search can be terminated at any time by *clicking* on STOP to the right of the **Search** button at the bottom of the dialog. Once completed (or stopped), hits designated by their CSD reference codes (**REFCODE**) and the compound names (**Name**) are displayed at the right of the **CSD** dialog.** The total number of hits is indicated immediately below. Structural data are available only for entries preceded by a filled yellow circle, although experimental references are available for all entries (**Molecule Reference** tab under **Output** in the **Display** menu; **Chapter 22**). A tube model for any entry for which a structure is available may be displayed in the window at the left of the dialog by *clicking* on its name in the box.

An open/white circle, it means there is a literature reference but no 3D structure (this is true of some small number of very old CSD entries). If a hit returns with a yellow circle, the substructure was found, but there is a difference in configuration (axial/equatorial, cis/trans, etc.). A green hit (the "norm") indicates the exact substructure was found (and is highlighted in a yellow halo in the 3D display of the hit.



The model can be manipulated (rotated, translated, scaled) inside the window with the usual mouse/keyboard touch commands.

The structure is ***exactly as it appears in CSD***. In particular, it may contain two or more different molecules, two or more different conformations of the same molecule and/or extraneous molecules (most commonly solvent molecules and counterions). Hydrogens will be present only where they have been assigned in the experimental structure. The structure may also contain errors in bonding due to uncertainties in the original data. Modifications to the CSD entry

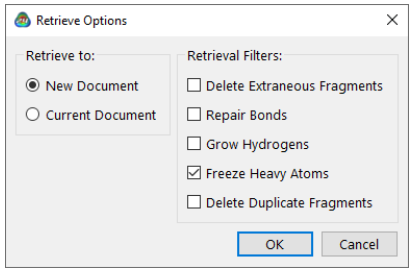
---

* If no attachment points are specified, a search for an exact match is carried out assuming that all atoms and bonds are exactly as given in the query and that all free valences are assumed to be hydrogens.

** Additional information is available from a contextual menu. *Right click* in the first blank header cell at the right of the **CSD** dialog.

The model can be manipulated (rotated, translated, scaled) inside the window with the usual mouse/keyboard touch commands.

The structure is ***exactly as it appears in CSD***. In particular, it may contain two or more different molecules, two or more different conformations of the same molecule and/or extraneous molecules (most commonly solvent molecules and counterions). Hydrogens will be present only where they have been assigned in the experimental structure. The structure may also contain errors in bonding due to uncertainties in the original data. Modifications to the CSD entry to address these issues may be made prior to transfer to the file system using the **Retrieve Options** dialog. This is accessed by *clicking* on ▼ to the right of the **Retrieve** button at the bottom of the **CSD** dialog.



(i) **Delete Extraneous Fragments**
If *checked*, deletes any detached molecules that do not contain the search query.

(ii) **Repair Bonds**
If *checked*, attempts to fix bonding errors based on the actual geometry in CSD and on normal valence rules.

(iii) **Add Hydrogens**
If *checked*, adds hydrogens wherever they appear to be missing, based on the actual geometry and on normal valence rules. This function is also available under **Add Hydrogens** in the Utilities tab of the **Molecule Properties** dialog (**Display** menu; **Chapter 22**).

(iv) **Freeze Heavy Atoms**
If *checked*, freezes all atoms except for hydrogens. This allows experimental X-ray hydrogen positions that are often poorly determined, to be refined using molecular mechanics or quantum chemical methods, while maintaining the heavy atom skeleton.*

* Hydrogen positions may be refined using molecular mechanics using **Minimize** under the **Build** menu; **Chapter 20**. Refinement using quantum chemical models is specified in the **Calculations** dialog (**Calculations...** under the **Setup** menu; **Chapter 21**) by way of **Frozen Atoms** (**Freeze Center** under the **Geometry** menu; **Chapter 19**).

(v) **Retrieve to:**

Selection of **New Document** means that the search results are to be stored in a new document and leads to a request for a document name, whereas selection of **Current Document** results in the retrieved molecules being appended to the end of the current document.

The **Retrieve Options** dialog is dismissed with all changes to existing settings maintained by *clicking* on **OK**. *Clicking* on **Cancel** or on also dismissed the dialog but any changes to settings are lost.

Retrieval is accomplished by first selecting (*clicking* on) one or more hits in the scroll box and then pressing the **Retrieve** button at the bottom of the **CSD** dialog. The **Shift** and **Ctrl** keys may be used in the usual way to specify retrieval of multiple hits. Retrieval can be stopped at any time by *clicking* on to the right of the **Retrieve** button at the bottom of the **CSD** dialog.

Any molecules retrieved from CSD will be identified by the internal (CCDC) reference code (REFCODE). The molecule name can be accessed by bringing up a spreadsheet and typing "***comp***" into an empty column header.

Spartan Spectra and Properties Database (SSPD)

The Spartan Spectra and Properties Database (SSPD) comprises over 317,000 molecules, which have been obtained from the density functional model, ωB97X-D/6-31G*. The ωB97X-D/6-31G* model has been shown to consistently yield reliable structures, spectra and properties. However, except where reactants and products are very similar, for example, stereoisomers, so as to benefit from error cancellation, it is *not* reliable for the calculation of reaction energies. With this in mind, heats of formation (in kJ/mol) from the T1 thermochemical recipe have been provided as a "property" for most molecules, additionally thermodynamic properties (based on IR from the both ωB97X-D/6-31G*) are available, as well as energies (in atomic units) from the ωB97X-V/6-311+G(2df,2p)//ωB97X-D/6-31G* density functional model for most molecules.

Additionally, experimental heats of formation for a small number of molecules from public NIST database are also available.

SSPD entries for most molecules use the best conformation assigned from the T1 thermochemical recipe rather than from MMFF molecular mechanics. T1 has been shown to reliably account for conformational energy differences in small molecules (where experimental data are available), which suggests that it should be able to reliably identify the lowest-energy conformer. Note, however, that T1 is presently defined only for uncharged, closed-shell molecules incorporating H, C, N, O, F, Si, P, S, Cl and Br. The conformations of molecules outside of these limits (perhaps 1.5 - 2% of the entries in SSPD) use the MMFF method for assigning the lowest-energy conformer.

The following recipe has been used to create the majority of individual entries in SSPD:

1. Perform a systematic conformer search using MMFF molecular mechanics for systems with up to 1,000 conformers. For systems exceeding 1,000 conformers, perform a systematic search in which conformers are removed at random (leaving approximately 1,000 conformers).
2. Keep the lowest-energy conformer and up to five additional (higher-energy) conformers. Choice is based on energy (lower is better) and on diversity (more diverse is better).
3. Perform up to six HF/6-31G* geometry optimizations.
4. Perform T1 calculations on the two lowest energy (HF/6-31G*) conformers.
5. Perform ωB97X-D/6-31G* and EDF2/6-31G* geometry optimizations on the lower energy (T1) conformer.
6. Perform ωB97X-V/6-311+G(2df,2p)//ωB97X-D/6-31G* energy calculations

This procedure for providing T1 heats of formation for conformationally flexible molecules is completely automated and has previously been described in **Calculations...** under the **Setup** menu (**Chapter 20**).

Additionally, the vast majority of SSPD entries include both the infrared* and NMR spectra. The vibrational motions associated with selected lines in the calculated infrared spectrum can be animated. Proton and ^{13}C NMR chemical shifts have been empirically corrected and 3-bond HH and both 2 and 3-bond CH coupling constants have been obtained from parameterized expressions. COSY, HSQC** and HMBC NMR spectra (2D spectra) to be generated on-the-fly from the information in SSPD.

Each SSPD entry includes information useful for QSAR type analyses, specifically, information based on electrostatic potential and local ionization potential maps, a variety of thermochemical quantities and (in specific cases) gas-phase acidities and basicities. Each SSPD entry also includes ωB97X-V/6-311+G (2df,2p)//ωB97X-D/6-31G* energy as well as both T1 and (where available) experimental heats of formation.

All SSPD entries include the wave function, allowing on-the-fly generation of any graphical surface and property maps available to ***Spartan***.

Because of inclusion of the wave function (and to a lesser extent inclusion of the vibrational motions for IR), SSPD has a significant disk footprint (≈13 GB). The in-memory footprint of SSPD is ≈300 or so MB.

Accessing and Processing Data From SSPD

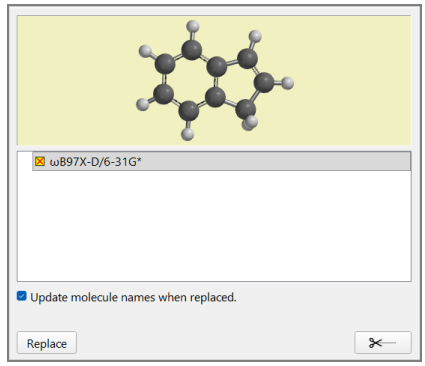
There are two ways to access data from SSPD. The first way is to replace a structure that has been built (or read in) with a matching database entry. All structures in a document for which there are matches in SSPD may be replaced in a single operation. The second way is to search SSPD for entries that share particular structural features or have part of a name in common or are isomers.

---

* Infrared spectra in SSPD are only available from the EDF2/6-31G* model.

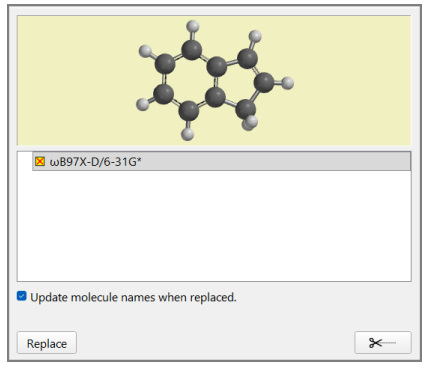
** HSQC spectra do not require coupling constants and all 1-bond CH interactions are assured to be observable.

Molecule Replacement



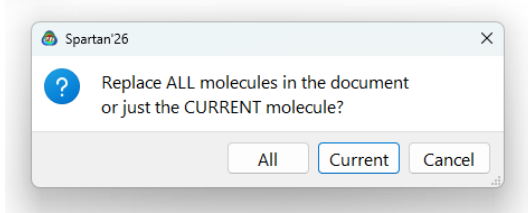
Molecule replacement is initiated by *clicking* on the molecule name displayed at the 1H-indene bottom of ***Spartan***'s screen. (The fact that a name appears means that the molecule is in SSPD.) *Clicking* on the name leads to a dialog at the bottom of the screen.

The model in the viewing window may be manipulated using the usual mouse/keyboard commands (you need to position the cursor inside the viewing area). Touch commands are also supported. Ball- and-spoke model style may not be changed.



The selected model may be replaced by the selected database entry by *clicking* on **Replace** at the bottom right of the dialog. (Replacement can be undone by selecting **Undo** from the **Edit** menu; **Chapter 17**.) If the box to the left of "***Update molecule names when replaced***" is *checked*, replacement destroys all information. If it is not checked, replacement destroys all information except the molecular identifier ("Label") in the leftmost column of the spreadsheet (see **Spreadsheet** under the **Display** menu; **Chapter 22**). The **Tear-Off** button moves the database preview into the ***Spartan*** interface where it behaves as a standard dialog. To close it, *click* on in the upper right corner.

Entries are selected as before, after which they may be examined in the viewing screen and used to replace the on-screen molecule.



In the event that the selected (on- screen) molecule is part of a multi- molecule document, a second dialog will result following *clicking* on **Replace**.

Clicking on **All** replaces all molecules for which database entries are available, while *clicking* on **Current** replaces only the selected entry. Molecules in the document for which there are no database entries are not affected. As with single molecule replacement, all information (with the possible exception of the molecule identifier) is replaced by that in the database.

Substructure and Other Modes of Searching SSPD

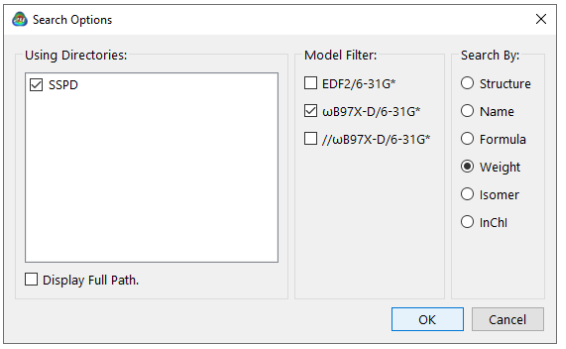
Access to data from SSPD follows from selecting **Databases** and then *clicking* on the **SSPD** tab (at the top left of the resulting window). This comprises three different panels, **Search**, **Plot** and **Analysis**, selection of which is controlled by tabs at the bottom left of the window. A search of SSPD is carried out from the **Search** panel, plotting of the search results from the **Plot** panel and regression analysis of search results from the **Analysis** panel.

Substructure searching of SSPD is similar to that already described for substructure searching of the Cambridge Structural Database (CSD). The only required input is a three-dimensional structure. While one or more growth points (see **Structure Query** earlier in this chapter) will normally be specified, they are optional and in their absence an exact search will be carried out. One or more substituents may be attached to the structure (see **Substituent Model Kit** under the **Build** menu; **Chapter 20**), in effect allowing multiple starting structures. (This capability is not available for CSD searches.) Name searching requires a name (or partial name), formula searching requires an exact formula, weight searching requires a molecular weight (returned are all hits within a few tenths of an AMU, and isomer searching requires a three-dimensional structure.

Following a search on SSPD, hits will be listed on the right and the structure of the selected hit will be shown in a window at the left. In addition to the molecule **Name**, the molecular weight and the theoretical model are available from a contextual menu. *Right click* in an empty header cell.

Controls along the bottom right access the **Search Options** dialog which specifies which directories to be searched, the kind of search, and the theoretical model(s) to be searched. They are also used to initiate and terminate a search and to designate the location for any results that are retrieved from the search.

Clicking on ▼ to the right of the **Search** button leads to the **Search Options** dialog, and the **Using Directories:** lists the SSPD source, check the **Display Full Path** to see the installation path. **Model Filters** specifies the theoretical models, EDF2/6-31G* EDF2/6-31G* (historical), ωB97X-D/6-31G* or //ωB97X-D/6-31G* (only those entries with ωB97X-V/6-311+G(2df,2p)//ωB97X-D/6-31G*) included in the **Search**.



Search By designates the type of search: **Structure**, **Name**, **Formula**, **Weight, Isomer, or InChI**.

(i) **Name**

Selection followed by *clicking* on **OK** in the **Search Options** dialog leads to a box **By Name** at the bottom right of the dialog. The search will return all entries that contain the string that is entered into the box. For example, *typing* toluene will not only result in toluene, but also in molecules such as *para*-toluenesulfonic acid and 4-chloro-2-fluorotoluene.

(ii) **Formula**

Selection followed by *clicking* on **OK** in the **Search Options** dialog leads to a box **By Formula** at the bottom right of the dialog. A formula of the form $C_cH_hHet_t$, where "Het" corresponds to elements other than carbon and hydrogen (each element needs to be specified explicitly). c, h, t are integers and are required. For example, the formula for nitrobenzene needs to be written as ***C6H5N1O2***. The order of atoms is irrelevant. The search will return all entries with this formula. For example, the formula ***C6H5N1O2*** will return not only nitrobenzene but also molecules such as phenyl nitrite and pyridine-4-carboxylic acid.

(iii) **Weight**

Selection followed by *clicking* on **OK** in the **Search Options** dialog leads to a box **By Weight** at the bottom right of the dialog. The search will contain all entries with the specified molecular weight ± 0.05 amu.

(iv) **Isomer**

Selection followed by *clicking* on **OK** in the **Search Options** dialog specifies that all isomers of the selected molecule will be returned. The result is the same as specification of a molecular formula.

(v) **InChI**

Selection followed by *clicking* on **OK** in the **Search Options** dialog specifies an exact InChI search be performed on the user provided InChI key.

Clicking on **OK** removes the **Search Options** dialog with all settings maintained. *Clicking* on **Cancel** or removes the dialog.

A search is carried out by *clicking* on the **Search** button at the bottom right of the database dialog, following which hits will appear at the right. The search may be terminated at any time by *clicking* on to the right of the **Search** button. Any hits that have already been identified will be kept. A tube model may be manipulated with the usual mouse/keyboard operations (the model style is fixed).

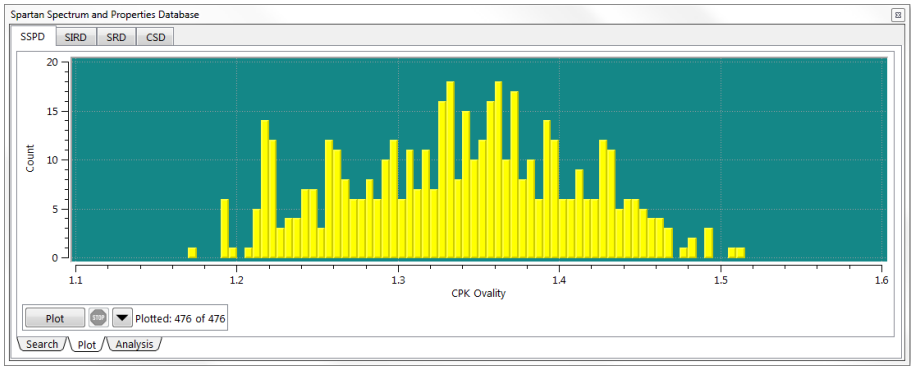
Properties may be added to the hit list by *dragging* them (names not numerical values) from one of the properties dialogs (see **Properties** under the **Display** menu; **Chapter 22**), the formula editor (**Formulas** under the **Display** menu; **Chapter 22**) or from the main ***Spartan*** screen, primarily entries associated with the **Geometry** menu; **Chapter 19**. The hit list can be sorted using the property, by *right clicking* on the column header cell and selecting **Sort** from the menu.

The results of a search on SSPD as well as any subsequent plotting and analyses are saved in the ***Spartan*** document corresponding to the original structure query or to the blank document that was used for a name or spectra query.

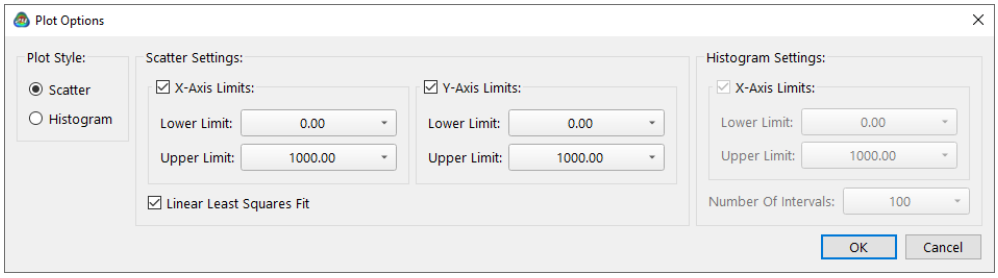
> A much larger number of molecules can be handled in the SSPD window than in ***Spartan's*** main window (several tens of thousands compared to a few thousand). This is because prior to retrieval each entry comprises very little information, whereas following retrieval each entry is much larger, the primary difference being inclusion of the wave function. Thus, plots and analysis can be carried out on large datasets.

Plotting Data Resulting from an SSPD Search

Data resulting from a search on SSPD may be displayed either as a histogram (property value vs. number of occurrences) or as a 2D scatter plot (property value vs. property value). Access to plotting is via the **Plot** tab at the bottom left of the dialog.



Selection of plot type and variable ranges is from the **Plot Options** dialog accessible from ▼ to the right of the **Plot** button.

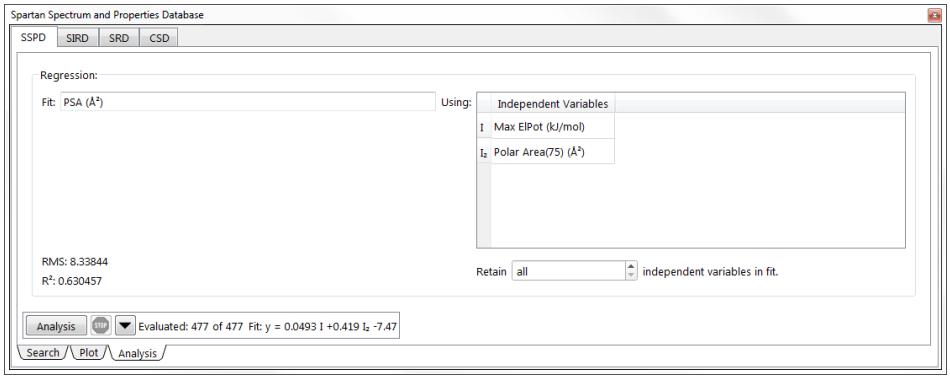


Default plot ranges are from minimum to maximum property values. *Checking* **Linear Least Squares Fit** (**Scatter Settings**) fits a straight line to the data. (A more general multi-variable regression analysis can be carried out from **Analysis**; see discussion following.) **Number of Intervals** (**Histogram Settings**) sets the number of bars in the histogram. *Clicking* on **OK** removes the dialog with all settings maintained. *Clicking* on **Cancel** or removes the dialog without maintaining the settings.

A plot is drawn by *clicking* on the **Plot** button at the bottom left of the dialog. The points will not appear instantaneously as the data need to be extracted from the file system. Plotting may be terminated at any time by *clicking* on to the right of the **Plot** button. If a least-squares line has been requested, it will be based on the accumulated data. At any time, the name of the molecule associated with a point on a scatter plot can be displayed by positioning the cursor over that point. The point can be deleted by *right clicking* and selecting **Delete** from the resulting menu. The plot will rescale as appropriate.

Regression Analysis of Data Resulting from an SSPD Search

A regression analysis of data resulting from a search may be carried out by selecting the **Analysis** tab at the bottom left of the dialog.



The name (not the numerical value) of the property to be fit is *dragged* from one of the properties dialogs, the formula editor or from the main ***Spartan*** screen, into the box to the right of **Fit**. The names of one or more properties to be used in the fit are *dragged* into the box to the right of **Using**. The fitting procedure (singular value decomposition) will use only variables that have been selected (*clicked* on) from among the list, up to a maximum number

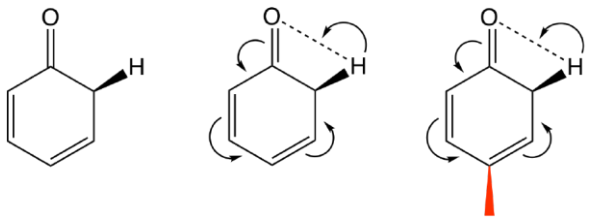
of independent variables (the most important in terms of their contribution to the fit from among this list) specified in the menu at the bottom right of the dialog. Note that the number of variables actually used in the fit may be smaller than the maximum number selected (from the menu at the lower right of the dialog), as variables that are linearly dependent and/or do not contribute significantly to the fit will be discarded. The analysis is started by *clicking* on the **Analysis** button at the bottom left of the dialog. As with plotting, this does not occur instantaneously as the underlying data needs to be extracted from the file system. Analysis may be terminated at any time by *clicking* on to the right of the **Analysis** button, following which results from a fit to a portion of the data will be displayed.

Upon completion, or upon user termination, the fit function will be displayed at the lower left of the dialog. This can be copied to the clipboard (*right click* and select **Copy** from the menu) and then pasted (*right click* and select **Paste** from the menu) in the **Plot** dialog.

Relationships Involving Chemical Reactions

We have previously discussed how the formula editor may be used to construct relationships among properties for a single molecule (see **Formulas** under the **Display** menu; **Chapter 22**). The formula editor can also be employed to construct relationships between the reactants and the products of a chemical reaction. This is possible because the products of a reaction are completely and unambiguously defined given the structure of the reactant (or reactants) and a set of curled arrows (see **Reaction Query** earlier in this chapter). The same information can also be used to guess a transition state for the reaction (see **Guess Transition State** later in this chapter).

For example, consider the difference in energies between 2,4-cyclohexadieneone and phenol. Whereas the normal preference is for keto forms to be favored over enols, here the reverse is certainly true as phenol benefits from aromatic stabilization. Three steps are needed to use SSPD to query the extent to which substituents alter the preference in the parent compound. First, a reactant must be built, reaction queries placed to designate the product of reaction and one or more structure queries (indicated below by red wedges) added to define the limits of the search. For example limit substituents in 2,4-cyclohexadieneone to the 4 position.



Next, a formula for the energy of reaction needs to be constructed. Inclusion of arrows extends the formula from the query itself (a 4-substituted 2,4-cyclohexadieneone), to the product of reaction (a *para*-substituted phenol) and to the difference between product and reactant. Selection of which is to be employed in the formula follows from the **Applied to**

menu under **Context** in the formula editor (highlighted only if arrows are present). In this case, the choices are (R) 2,4-cyclohexadieneone, (P1) phenol or (P-R) reaction. The formula for the reaction energy, ***@energy*@hart2kj***, appears in the box at the center of the dialog following selection of **(P-R) reaction** from the **Applied to** menu and *clicking* on **E** under **Energies**. The default name may be changed by editing the text inside the box to the left of the formula. *Clicking* on **Enter** enters the (renamed) formula ***@ rea(@energy*@hart2kj)*** into the list at the top of the formula editor.

Finally, a search may be carried out by *clicking* on the **Search** button inside the SSPD **Search** dialog. After it has completed, dragging the formula for the reaction energy onto an empty header cell of the hit list inside the SSPD **Search** dialog will result in reaction energies for all entries to appear.

As a second example, consider the energy for addition of hydrogen fluoride to propene, leading to 1-fluoropropane. Here there are two reactants (hydrogen fluoride and propene), but only the latter will be provided growth points*. Thus, the query is for information about the energy of HF addition as a function of substitution on the alkene.

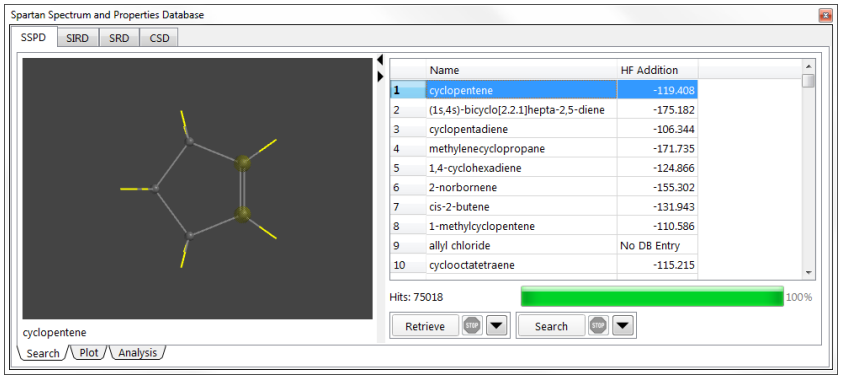
The query starts from an intramolecular complex of propene and hydrogen fluoride, oriented such that the C=C and H–F bonds parallel with fluorine on the unsubstituted side of the alkene. Two arrows (reaction queries), one from the CC bond to the space between the unsubstituted carbon and fluorine and the other from the HF bond to the space between the substituted carbon and hydrogen, define the product (1-fluoropropane). Structure queries attached to all three open valences extend the search to all available substituted propenes (for which analogous substituted 1-fluoropropanes are also available).

A formula for the energy of reaction may be constructed using the same sequence as in the previous example, selecting **(P-R) reaction** from the **Applied to** menu under **Context** in the formula editor, *clicking* on **E**, changing the default name and *clicking* on **Enter**.

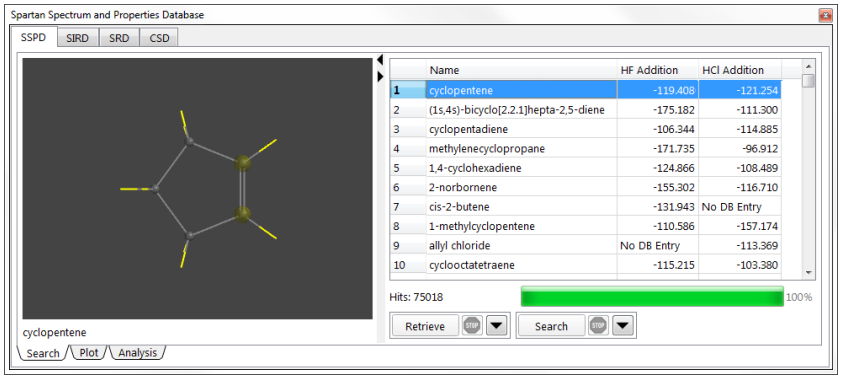
Following a search using the query, dragging the formula for the reaction energy onto an empty header cell of the hit list inside the SSPD **Search** dialog adds a column of reaction energies to the list.

---

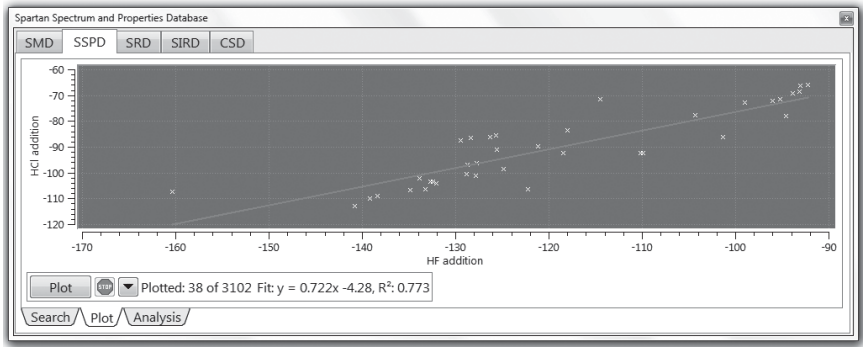
* While there can be two or more reactants, only one can be provided structure queries, that is, all other reactants need to be single molecules.



It is possible to apply different queries to the same set of search results, for example, to ask about the energy of addition of HCl (instead of HF) to a substituted propene. Whereas the query changes (the reaction of HCl and propene rather than of HF and propene), the underlying search (on propene) does not. The result is two sets of energies. (Results designated ***No DB Entry*** indicate that the required information is not available.)

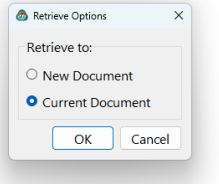


A plot of HCl vs. HF reaction energies from the SSPD **Plot** dialog shows the extent to which they are actually correlated.



The analysis and plotting capabilities applied to the data in SSPD are certainly more difficult to master than most of the other aspects of ***Spartan***. However, the size and diversity of the collection of organic molecules in SSPD means that the capabilities may be of considerable value in exploring molecular properties and reaction energies in advance of calculations on closely-related systems of interest.

Retrieving Data from an SSPD Search



Prior to retrieval, it is necessary to specify a destination from the **Retrieve Options** dialog. *Click* on ▼ to the right of the **Retrieve** button.

Selection of **New Document** results in a new document; selection of **Current Document** results in the retrieved molecules being appended to the end of the current document. The **Retrieve Options** dialog is dismissed by *clicking* on **OK**. *Clicking* on **Cancel** or on ✕ also dismisses the dialog but any changes to settings are lost.

To retrieve, select (*click* on) one or more hits in the box and *press* the **Retrieve** button. The **Shift** and **Ctrl** keys may be used to specify retrieval of multiple hits. Retrieval may be stopped at any time by *clicking* on STOP to the right of the **Retrieve** button.

Spartan Reaction Databases (SRD)

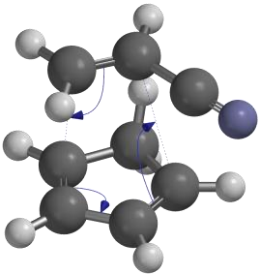
The **S**partan **R**eaction **D**atabase (**SRD**) is a collection of transition states for approximately 1,500 organic and organometallic reactions obtained from one or more of the following theoretical models: AM1, PM3, HF/3-21G and HF/6-31G*.* Each entry contains the infrared spectrum, which should be characterized by a single imaginary frequency corresponding to the reaction coordinate.

An entry in SRD is in terms of a three-dimensional structure (corresponding to the reactants or the product), together with an appropriate set of curved arrows (connecting the reactants to the product or vice versa) and designation of theoretical model.

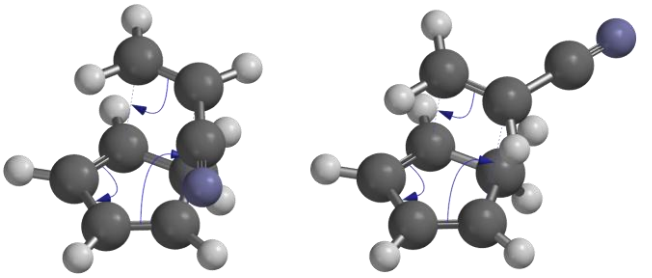
Instructions for drawing reaction arrows has been provided earlier in this chapter (**Reaction Query**).

For example, an entry for the Diels-Alder reaction of cyclopentadiene and acrylonitrile is given below.

* Efforts are underway to update SRD to include ωB97X-D/6-31G* entries.



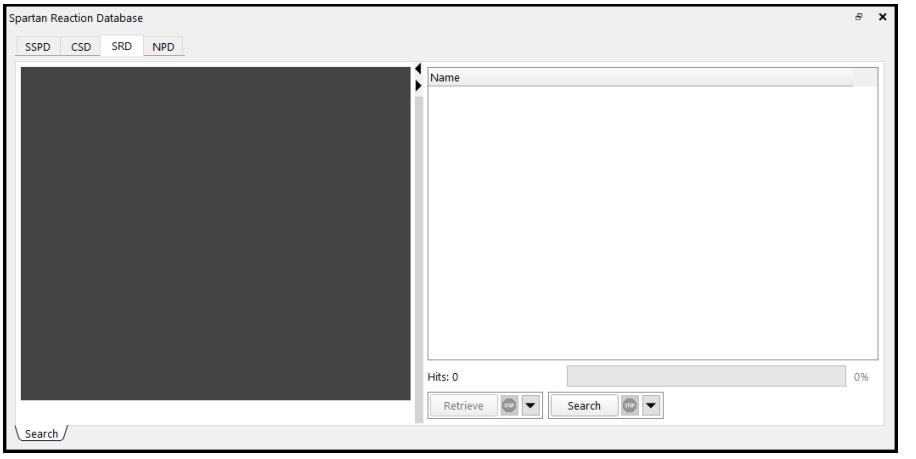
Note, that stereoisomers are distinguished by the way the reactants are positioned. For example, reactions leading to *endo* and *exo* Diels- Alder adducts are distinct.



Like SSPD, SRD is open-ended and users may construct their own data collections with additional reactions and/or results from additional theoretical models on existing reactions (**Appendix H**).

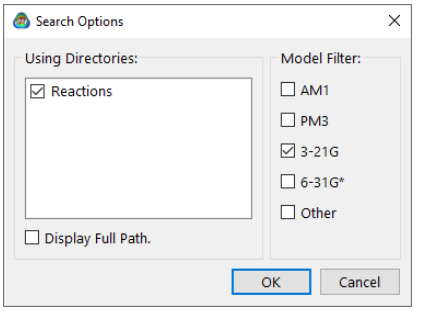
Substructure Searching

Substructure searching from SRD is very similar to that from SSPD. Access to SRD is provided by *clicking* on the **SRD** tab at the top of the **Database** dialog. This gives rise to the **SRD** dialog.



The **SRD** dialog contains a window on the left for previewing structures and a box on the right for listing hits on SRD, as well as buttons at the bottom to control various aspects of the search and subsequent transfer of data into ***Spartan***'s file system.

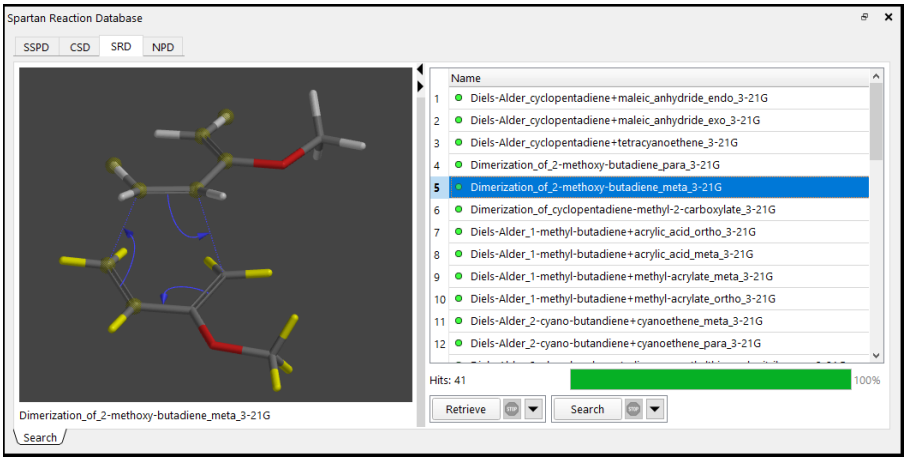
A substructure search is set up by specifying one or more attachment points on the reactant structure (see **Structure Query** earlier in this chapter) that has been augmented with reaction arrows (see **Reaction Query** earlier in this chapter). Structure queries and reaction arrows can also be added to 2D sketches (see discussion under **2D Sketch Palette** in **Chapter 20**).



Prior to starting the actual search, it is necessary to specify which theoretical model or set of models are to be included. This is done from the **Search Options** dialog and is reached by *clicking* on ▼ to the right of the **Search** button at the bottom right of the dialog. **Using Directories** indicates the libraries to be searched.

One is supplied with ***Spartan*** and users may replace or supplement this with their own collections (see **Appendix H**). *Checking* one or more boxes under the **Method Filters** specifies the theoretical model(s) to be searched. *Clicking* on **OK** removes the dialog. (*Clicking* on **Cancel** or ✕ also removes the dialog, but changes are not saved.)

Following specification of structure/reaction arrows, attachment points and designation of theory level(s), a search on SRD is carried out by *clicking* on the **Search** button at the bottom right of the dialog. If there are hits, they will begin to appear in a few seconds in the box at the right of the dialog. The search may be terminated at any time by *clicking* on STOP to the right of the **Search** button at the bottom right of the dialog.

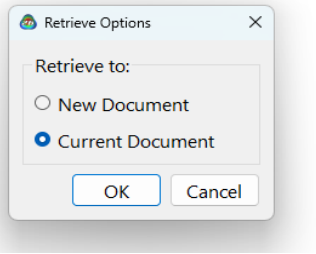


Database hits are displayed in the box at the right of the dialog. They may be sorted by name, molecular weight and theoretical model by *clicking* on the appropriate tab. A 3D structure for a particular hit may be previewed by *clicking* on its name in the box. A tube model of

the transition state with bonding consistent with the reactant and associated reaction arrows is displayed in the window at the left of the dialog, and its name appears immediately below the window. The model may be manipulated inside the window in the usual way, but the model style cannot be changed.

Retrieving Data from an SRD Search

Prior to retrieval of one or more database hits, it is necessary to specify a destination. This is accomplished with the Retrieve Options dialog, brought up by *clicking* on ▼ to the right of the Retrieve button.



Selection of **New Document** results in a new document; selection of **Current Document** results in the retrieved molecules being appended to the end of the current document. The **Retrieve Options** dialog is dismissed by *clicking* on **OK**. *Clicking* on **Cancel** or on ✕ also dismisses the dialog but any changes to settings are lost.

Retrieval is accomplished by first selecting (*clicking* on) one or more hits in the box and then pressing the **Retrieve** button at the bottom of the **SRD** dialog. The **Shift** and **Ctrl** keys may be used to specify retrieval of multiple hits. Retrieval may be stopped at any time by *clicking* on STOP to the right of the **Retrieve** button.

Spartan Infrared Database (SIRD)

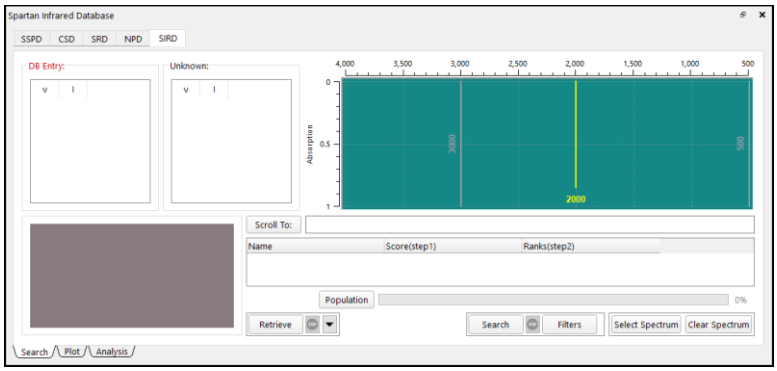
The **S**partan **I**nfrared **D**atabase (**SIRD**) is actually a second access point into the Spartan Spectra and Properties Database (SSPD) discussed earlier in this chapter. It comprises infrared spectra for >317,000 molecules calculated from the EDF2/6-31G* model. Unlike the other databases of calculated information attached to ***Spartan***, SIRD is searched not on substructure but rather on a measured (or calculated) infrared spectrum. This is input in JCAMP format (.dx).* The search is carried out in two steps: a fast search on the full database is performed first to identify a small set of possible matches which are then examined more thoroughly.

Additionally, all descriptions applicable to the **SIRD** are also true of the **XIRD** (the only difference is that the source for **XIRD** searches is the NIST database of some ≈7,000 experimental IR spectra.

Spectra Searching

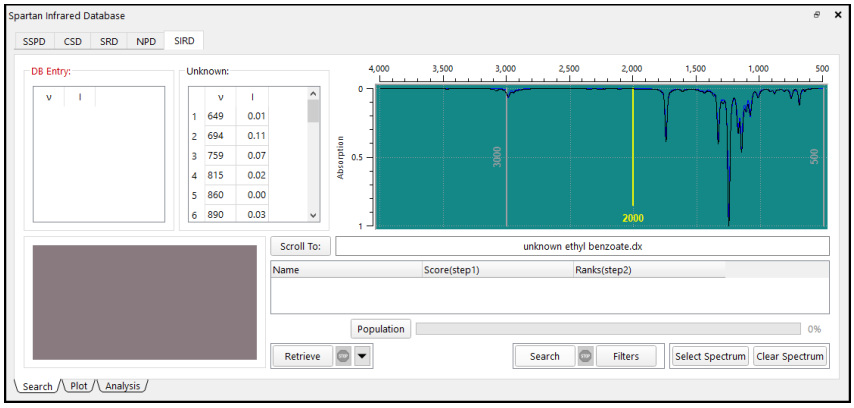
Access to SIRD is provided by *clicking* on the **SIRD** tab at the top of the **Database** dialog. A ***Spartan*** document must be open even if it is an empty document (**New** from the **File** menu). This gives rise to the **SIRD** dialog.

* A sample .dx file is provided in **Appendix J**.



This dialog is quite unlike those associated with other databases accessible from ***Spartan***, although it shares two aspects: a window at the lower left for examining the 3D structure of the selected hit and a box at the lower right for listing all hits. The window at the top right is for examining the match between the unknown (measured) infrared spectrum and that of a hit in the database, and boxes at the top left list the frequencies and intensities for both unknown and database spectra. Finally, buttons at the bottom of the dialog access the unknown spectrum, provide (optional) filters for its search against spectra in the database and control transfer of any search results into ***Spartan***'s file system.

A search is initiated by *clicking* on **Select Spectrum** at the bottom right of the dialog. This leads to a file browser that only recognizes .dx files. Selection of a file in the usual manner followed by *clicking* on **Open** (or *double clicking* on the name) displays the unknown spectrum (in blue) in the window at the top right of the dialog with the name underneath.

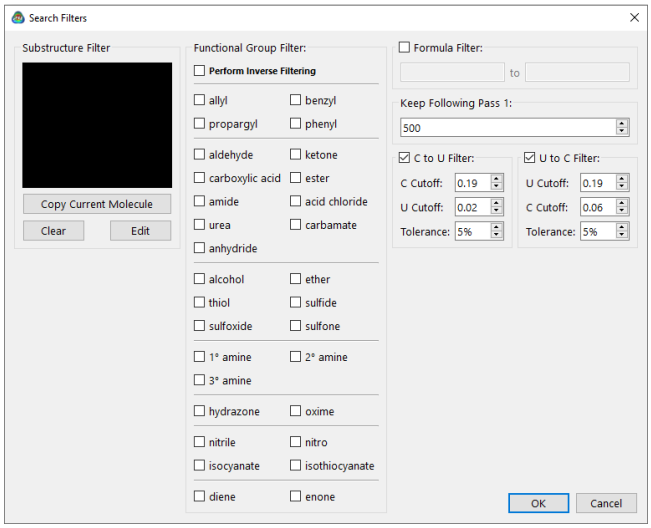


the spectrum has been fit to a Lorentzian function, allowing a set of vibrational frequencies and intensities to be extracted are displayed in the box marked **Unknown:** immediately to the left of the spectrum.* Features in the unknown spectrum with very low intensity cannot be fit. The fit is shown in black.

* The number of frequencies obtained by the fit will almost certainly be smaller than the actual number of vibrational frequencies (3N-6 for a non-linear molecule with N atoms).

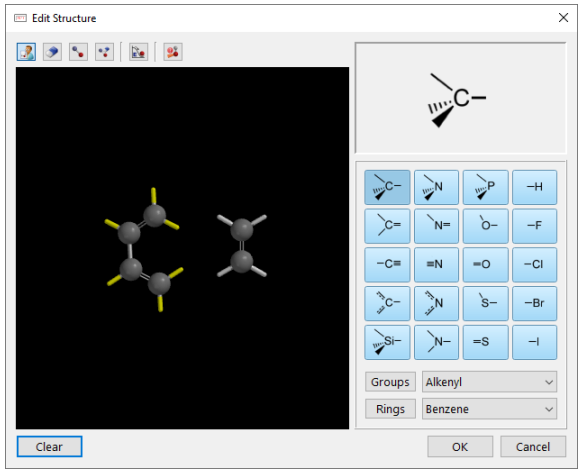
Filters

Conditions under which the search is carried out may be specified by *clicking* on the **Filters** button at the bottom of the dialog. (This step is optional.)



The **Filters** dialog contains a window for constructing a substructure filter on the left, a listing of common functional groups in the middle, and controls that limit the molecular formula on the top right. Controls at the bottom right govern the number of preliminary hits that are moved from the first pass of the search to the (more costly) second pass and the number of hits that are finally kept.

To build a substructure, *click* on **Edit** below the (empty) **Substructure Filter** window at the top left of the screen. Alternatively, you can start with the currently selected molecule by *clicking* on **Copy Current Molecule** prior to *clicking* on **Edit**.



The resulting dialog contains most of the functionality in the organic model kit (see **Chapter 20**) as well as the ability to define one or more attachment points (see **Structure Query** earlier in this chapter). The substructure needs to be in the hit, and if no attachment points are designated, it is the only acceptable hit. *Click* on **OK** to exit the build panel.

Functional Group Filters allow the search to be restricted to molecules that incorporate one or more of the functional groups. *Check* to select and *uncheck* to deselect.

Formula Filter allows selection of the minimum and maximum molecular formula (or both) to be considered in the search. The format of the input is $C_cN_nO_o$ ..., where C, N, O, etc. are atomic symbols and c, n, o are the number of each of the designated atoms. Neither hydrogens nor any elements that are not explicitly designated are considered.

Keep Following Pass 1 controls the number of entries that are kept following the initial pass of ***Spartan***'s spectra matching algorithm. The larger the number, the more computer time will be needed, but the more likely it will be to produce a good hit (assuming that the unknown or a closely-related molecule is actually in the database).

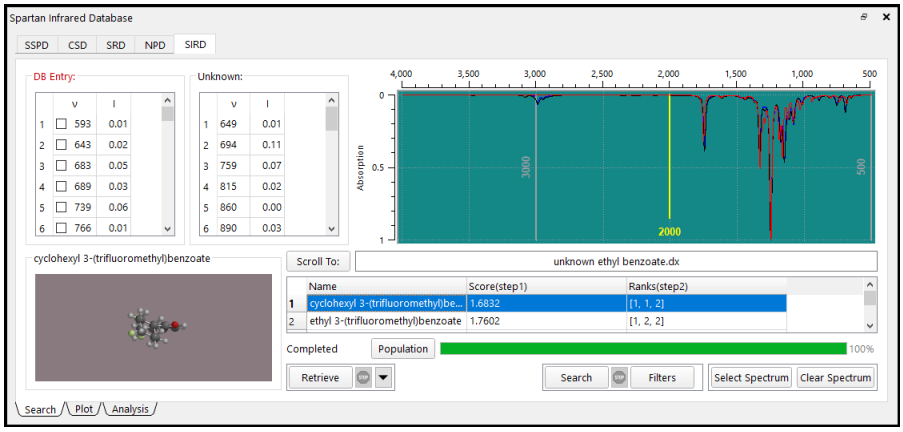
Keep Following Pass 2 controls the number of hits that are returned to the user. Does not affect the time required for the search.

Clicking on **OK** exits the dialog. *Clicking* on **Cancel** or also dismisses the dialog but any changes to the settings are lost.

In addition, the search may be restricted to a specific frequency range. This is controlled by the two grey slider bars displayed inside the dialog containing the spectrum of the unknown. Note that the experimental spectrum often concludes at 500-600 cm^{-1}.Frequencies below this range (the "far infrared") typically correspond to torsional motions and likely depend on conformation. Also, while the experimental spectrum usually extends to 4500 cm^{-1}, the region beyond ~2800 cm^{-1} is dominated by CH stretching vibrations and may be too crowded to be of value in spectral searching. Our experience suggests that a reasonable range for fitting calculated to experimental infrared spectra is from 700- 2200 cm^{-1}.

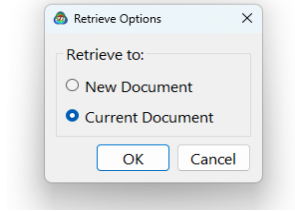
A search on the reference spectrum restricted by whatever conditions are specified in the **Search Options** dialog is carried out by *clicking* on the **Search** button at the bottom of the **SIRD** dialog. The search is carried out in two passes. The first pass may require several tens of seconds depending on the size of the database and the filters that have been set. The second pass may require several minutes (depending on the number of hits kept following the initial pass as specified in the **Search Options** dialog). Searching may be stopped at any time by *clicking* on to the right of the **Search** button at the bottom of the **SIRD** dialog. However no results will be returned if the search is stopped prematurely.

When the search has completed a listing of hits (the number of which is specified by **Keep Following Pass 2** in the **Search Options** dialog) ordered by their fit errors (lowest is best) appears in the box at the right of the **SIRD** dialog. As each hit is selected (*clicked* on in the list), its spectrum (displayed in red) will be superimposed onto the spectrum of the unknown.



Calculated vibrational frequencies and intensities for the hit are displayed in a box at the top left of the dialog (to the left of the corresponding values for the unknown). *Clicking* inside the box to the left of a (calculated) frequency animates the vibrational motion.

Retrieving Data from SIRD Searches



Prior to retrieval on one or more database hits, *click* on to the right of the **Retrieve** button to bring up the **Retrieve Options** dialog. Selecting **New Document** creates a new document, while selecting **Current Document** appends the retrieved molecules to the end of the current document. *Click* on **OK** to dismiss the dialog. *Clicking* on **Cancel** or also dismisses the dialog but any changes are lost.

Retrieval is accomplished by first selecting (*clicking* on) one or more hits in the box and then *pressing* the **Retrieve** button at the bottom of the **SIRD** dialog. The **Shift** and **Ctrl** keys may be used to specify retrieval of multiple hits. Retrieval may be stopped at any time by *clicking* on ( ) to the right of the **Retrieve** button.

Whichever database dialog is open may be closed by *clicking* on , or alternatively by again selecting **Databases** from the **Search** menu . The control operates in toggle mode.

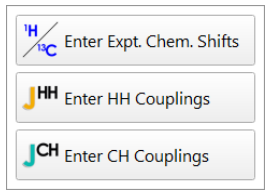
External Infrared Database (XIRD)

This uses identical search logic as the **Spartan Infrared Database (SIRD)** to search the public NIST collection of experimental infrared spectra. Note that retrieval of a hit transfers only a 3D structure.

Chapter 24

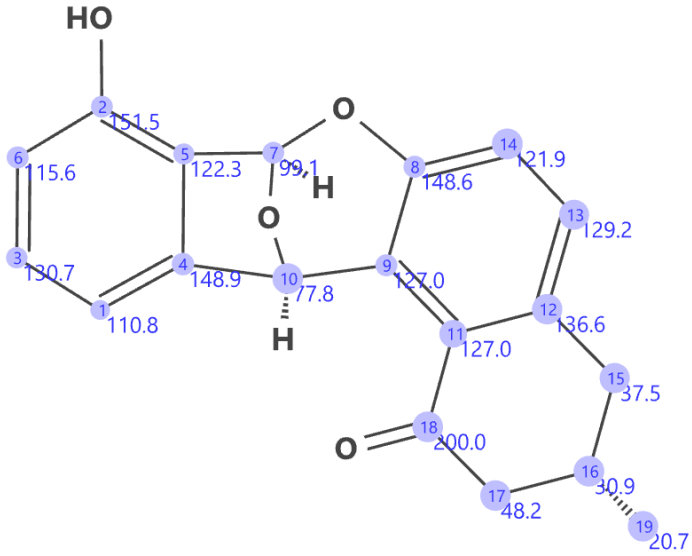
The NMR Menu

*Functions available under the **Data** menu allow for attaching experimental proton and ^{13}C chemical shifts to a 3D structure and for designating observed 3-bond HH couplings (for constructing COSY spectra) and 2 and 3-bond CH couplings (for constructing HMBC spectra).*

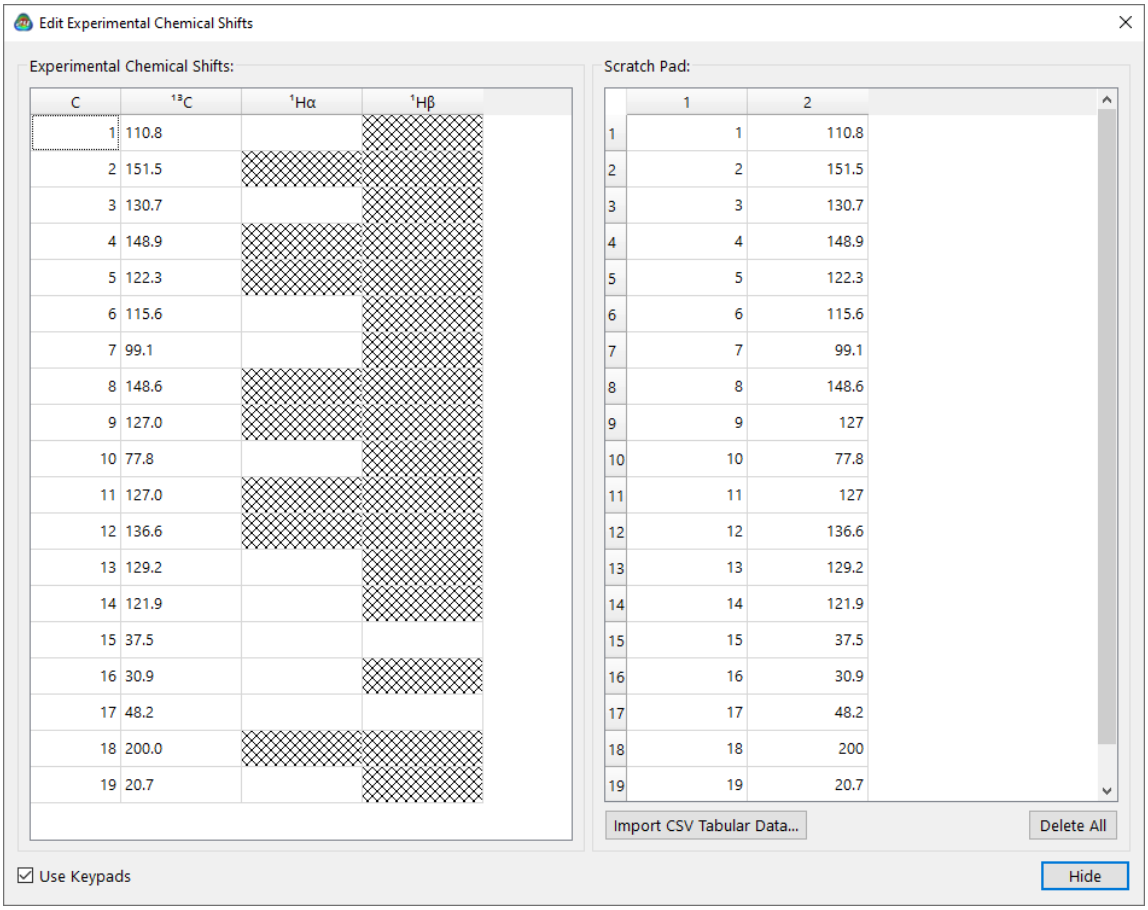


$^{1}H/^{13}C$

Input/import of experimental proton and ^{13}C chemical shifts is most easily accommodated using **Edit Experimental Chemical Shifts** tool. This is available in either 2D (Sketch) OR 3D (View) modes, shown below in 2D:



This is a cell-based panel providing for both manual entry, or import (or copy/paste) into a Scratch Pad area that can be copied from to transfer individual, multiple, or all shift values as a column:

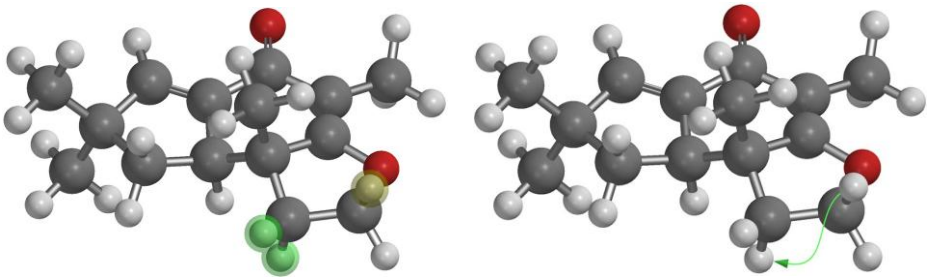


Experimental proton and/or ^{13}C chemical shifts may *also* be added to the 3D molecular model by *clicking* on the appropriate proton or carbon, *clicking* on **Edit** to the right of **Expt.Chem.Shift** at the bottom right of the screen, entering a shift value (2 decimal places for hydrogen and one for ^{13}C) into the number pad and finally *clicking* on **Enter**. As a result, the shift (in red label) will be attached to the atom.

Note that chemical shifts can also be added to a 3D structure from the **Atom Properties** dialog (**Display** menu; **Chapter 22**), and to a 2D drawing from **New Sketch** or in the **File** menu (**Chapter 16**) or **Edit Sketch** in the **Build** menu (**Chapter 20**). 2D representatives are certainly more familiar, an extra step is required for proton shifts as hydrogens are generally not shown.

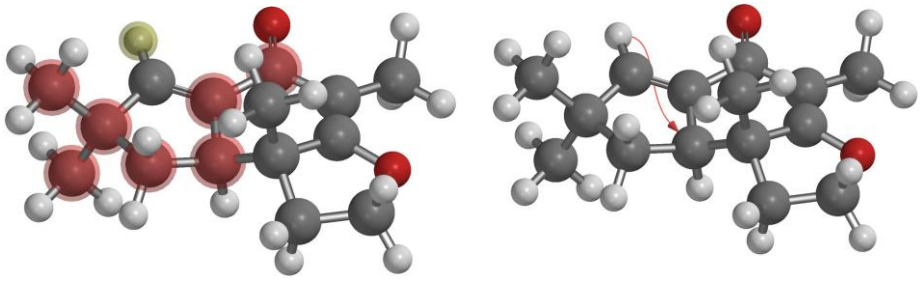
J^{HH}

Observed 3-bond HH couplings (not numerical values of HH coupling constants) may be placed on a 3D model by *clicking* on one of the hydrogens following which all possible hydrogens which might couple through 3 bonds will be highlighted in green. *Clicking* on one of these will connect the two with a double headed green arrow.



J^{CH} J^{CH}

In a similar way, observed 2 and 3-bond CH couplings are added to a 3D model by *clicking* on the carbon following which all possible hydrogens 2 or 3 bonds removed will be highlighted in red. *Clicking* on one of these will connect the two with a red arrow with the head being on the carbon.



Both HH and CH couplings may be added to a 2D drawing from **New Sketch** in the **File** menu (**Chapter 16**) or **Edit Sketch** in the **Build** menu (**Chapter 20**). This requires added steps as hydrogens are not generally shown in 2D drawings.

Chapter 25

The Options Menu

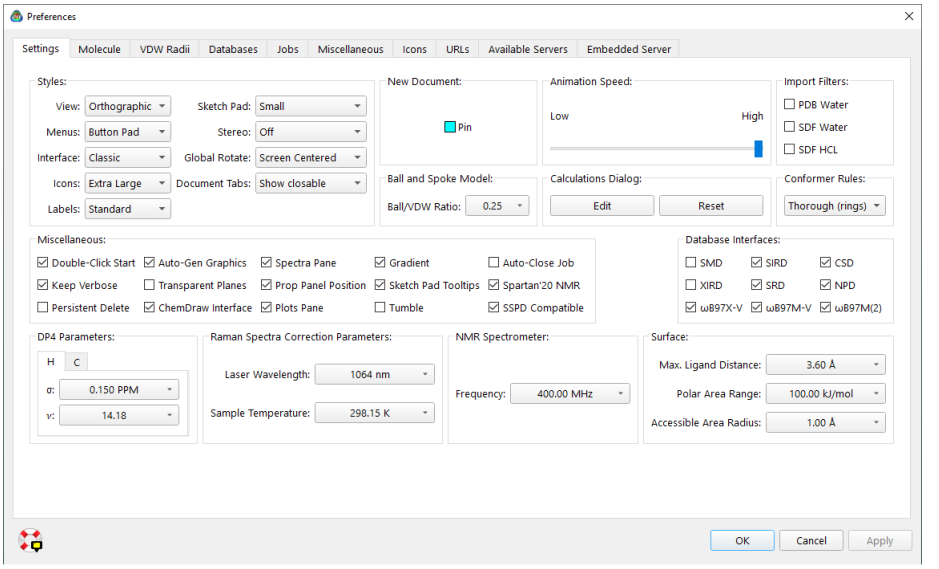
*Functions under the **Options** menu* set default colors, fonts, user preferences and van der Waals radii, locate databases, identify remote servers, set job queue's, set icon displays and identify/change URL's for on-line accesses and configure **Spartan'26**. They also allow for changing default colors and fonts and for monitoring executing jobs.*

Preferences... *

This sets up preferences relating to the graphical user interface (**Settings**), and to molecule displays (**Molecule**). It permits changes to default van der Waals radii used for space-filling models as well as for calculating molecular surface areas and volumes (**VDW Radii**). It also specifies the locations of databases (**Databases**), sets job queues (**Jobs**), specifies miscellaneous features (**Miscellaneous**), specifies which icons are to be displayed (**Icons**) and specifies URLs (**URLs**) for on-line connections, identifies available (remote) compute servers (**Available Servers**) and configures ***Spartan'26*** as a server for remote devices (**Embedded Server**). Selection results in one of ten **Preference** panels, depending on which tab is. *Clicking* on a tab brings up the associated preferences. To exit the **Preferences** dialog *click* on **OK**. *Clicking* on **Cancel** or exits the dialog without instituting any changes.

* Consistent with MacOS **Preferences** are also located under the ***Spartan'26*** menu in the Macintosh version.

Settings Tab



Style

View: Orthogonal/Perspective
Controls the view of structural models and graphics.

Menus: Classic List/Button Pad
Controls presentation of menus either as lists (Classic List) or as icon palettes (Button Pad). The latter is likely to be better suited to touch screen computers and tablets.

Interface: Classic/Touch
Under the **Touch** setting, some menu/dialog items (including up/ down arrows) are displayed in a larger size to better support touch screen computers and tablets.

Icons: Small/Medium/Large/Extra Large/Jumbo
Controls size of program icons in the tool bar.

Sketch Pad: Small/Medium/Large
Controls the size of the sketch pad (palette of sketch tools) for ***Spartan***'s 2D builder.

Stereo: Off/Red-Cyan
Turns stereographic display on and off.

Global Rotate: Screen Centered/Molecule Centered

Screen Centered rotates all molecules about a common center.
Molecule Centered rotates each molecule about its own center.

Document Tabs
Hide will revert display style behavior to that of versions of ***Spartan*** prior to ***Spartan'14***, **2014**. That is, all open documents will be visible when in **View** mode. If **Show** is *checked*,

this displays a tab at the bottom of the screen for each open document. This allows for displaying molecules from documents other than the currently selected document by *checking* the box to the left of the tab. **Show closable** is the same as **Show** but includes the ability to close the document by *clicking* on the button at the right of the tab.

New Document: Pin
If *checked*, defaults to display of any new documents (from building or brought in from the **File** menu) irrespective of whether or no they have been explicitly selected. Does not affect the status of existing open documents. **Pin** setting is only applicable if either **Show** or **Show closable** is selected (see **Documents Tabs** above).

Animation Speed
Slider bar controls the speed for animations.

Ball and Spoke Model
Sets the ratio of the radius of the balls in ball-and-spoke model to the VDW radius. The default value is 0.25.

Conformer Rules: Normal/Skeletal/Thorough
Chooses between rule sets for conformational searching. **Normal** is the default and should be used for **Equilibrium Conformer** and **Conformer Distribution** calculations where the Monte Carlo approach is involved. **Skeletal** (in versions prior to ***Spartan'18*** this was called **Trimmed**) eliminates degrees of freedom and should be used for **Similarity Library** calculations where a systematic approach is carried out. **Thorough** considers twist-boat conformers of six-membered rings (in addition to chair conformers).

Calculations Dialog
Controls the default task that is displayed upon entering **Calculations...** from the **Setup** menu. The default is Equilibrium Geometry with the ωB97X-D/6-31G* model, this can be modified by clicking the **Edit** button, which leads to a sample **Calculations** dialog. *Clicking* on **Reset** restores the default task.

Miscellaneous:

Double-Click Start
If *checked*, *double-click* as opposed to *single-click* is required to place the initial atom, group, ring, ligand, etc. on screen when using the 3D builder (consistent with the 2D Sketch builder behavior).

Auto-Gen Graphics
If *checked*, graphics calculations will be performed by the graphical interface (without having to submit a calculation) as long as the information necessary to generate the surface is available (a previous calculation has been run, or the molecule has been retrieved from the SSPD). Note, however, that graphics calculations will not be auto-generated on documents containing more than 25 molecules. These will need to be submitted as a calculation.

Spectra Pane
If *checked*, utilizes a simple dedicated Spectra dialog for visualization and exploration of calculated and experimental spectra. If *unchecked*, the presentation reverts to the spectra display style available in versions prior to release of ***Spartan'14***. The latter allows finer control over details in the display (axis ranges, line width, etc.) which can be transferred back to the simpler display.

Gradient
If *checked*, the single color background by a color gradient background.

Keep Verbose
If *checked*, keeps extended Q-Chem output. Normally discarded upon successful completion of submitted calculations, this may be useful for identifying the source of problems for calculations that have not successfully completed or have led to suspicious results. (The last 100 lines of the verbose output is automatically kept for a job that has abnormally terminated). Verbose output significantly increases the size of the ***Spartan*** document.

Transparent Planes
If *checked*, planes (**Define Plane** under the **Geometry** menu; **Chapter 18**) will be transparent rather than opaque.

Prop Panel Position
If *checked*, retains the on screen position of the Properties dialog between ***Spartan*** sessions.

Sketch Pad Tooltips
If *checked*, provides brief description of the tools available in the sketch pad by hovering over the icon.

Persistent Delete
If *checked*, delete function is persistent. If not *checked*, delete will revert to the previously selected function after a single delete operation.

ChemDraw Interface (Windows only)
If *checked*, adds **ChemDraw** as a tab at the top of the model kit. This allows for use of the ChemDraw program (version 19 or newer) as an alternative for sketching molecules. (All ***Spartan*** versions open **ChemDraw** files).

Plots Plane
If *checked*, utilizes a simple dedicated Plots dialog for visualization and exploration of user generated plots. If *unchecked*, the presentation reverts to the plots style available in versions prior to release of ***Spartan'14***. The latter allows finer control over plot details (axis ranges, markers, etc.) which can be transferred back to the simpler plots presentation. Note: support for 3D (XYZ) plots is not yet implemented in the Plots Pane and is only available if this preference not *checked*.

Tumble
If *checked*, allows automatic tumbling of molecules. To tumble a molecule, select it, *press* the left mouse button, move the mouse and release the button. To stop tumbling, *left click*.

To start tumbling on touch screen devices, swipe one finger over the screen. To stop, tap anywhere on screen.

Auto-Close Job
If *checked*, upon job completion, the job completed dialog remains for five seconds, then closes.

Spartan'20 NMR
If *checked*, the **NMR Spectrum (Spartan'20)** task is enabled from the **Calculations** dialog (closely reproducing the 2019 literature reference).

SSPD Compatible
If *checked*, the **SSPD Compatible** task is enabled from the **Calculations** dialog (allowing the user to generate identical calculations to those included in the current SSPD, as well as earlier versions). The minimum requirement is an equilibrium geometry from ωB97X-D/6-31G*.

Databases
Controls what database options are available from the **Databases** entry (**Search** menu, see **Chapter 23**). Turns on or off access to Spartan Molecular Database (SMD) – a legacy collection, Spartan Infrared Database (SIRD), External Infrared Database (XIRD), Spartan Reactions Database (SRD), the Natural Products Database (NPD), and the three new energy databases: ωB97X-V, ωB97M-V, and ωB97M(2), as well as the Cambridge Structural Database (CSD). The CSD must be licensed through the Cambridge Crystallographic Data Centre, in order to access this resource.

Import Filters PDB Water Filter
If *checked*, removes water molecules from imported PDB files.

SDF Water Filter
If *checked*, removes water molecules from SD files.

SDF HCL Filter
If *checked*, removes hydrochloric acid molecules from SD files.

DP4 Parameters H|C
Select proton (H) or ^{13}C (C) to alter respective DP4 parameters.

α
Alter α parameter.

γ
Alter γ parameter.

IR/Raman Spectra Correction Parameters Laser Wavelength
Raman intensities are dependent on the wavelength of the laser in the spectrometer. This setting allows for adjustment of calculated intensities based on **Laser Wavelength**. *Click* inside the box and use the number pad that appears.

Sample Temperature
Raman intensities are also dependent on temperature. This setting allows for adjustment of **Sample Temperature** for Raman spectra calculations. *Click* inside the box and use the number pad that appears.

NMR Spectrometer Frequency
The frequency of the spectrometer, used in coupling constant calculations.

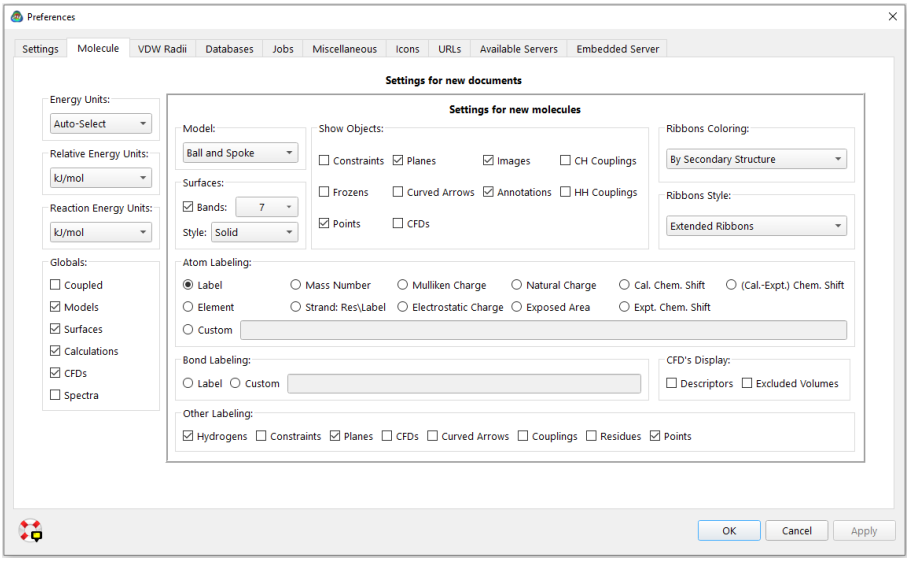
Surface

Max. Ligand Distance (Å)
Sets the maximum distance (in Å) for an amino acid residue (in a protein) to be recognized as bonded to a ligand. *Click* inside the box and use the number pad that appears.

Polar Area Range (kJ/mol)
Sets the potential (in kJ/mol) for calculating polar area from the electrostatic potential map. The range is given as a single number but represents the range between –value to +value, for example, the default range of 100 kJ/mol means a range from -100 to +100 kJ/mol. Values above and below the range are considered when determining polar area. *Click* inside the box and use the number pad that appears.

Accessible Area Radius (Å)
Sets sphere radius (in Å) for determining accessible area, the default is 1.0Å. *Click* inside the box and use the number pad that appears.

Molecule Tab



This specifies default settings for model appearance. These settings may be overridden for individual molecules in a document using entries under the **Model** menu and for specific portions of a molecule using **Utilities/Style** dialogs associated with **Properties** dialogs (**Properties** under the **Display** menu; **Chapter 22**).

Energy Units

Auto Select, where units depend on the model and on whether they refer to absolute or relative quantities, ***au*** (atomic units), ***kJ/ mol*** and ***kcal/mol***

Relative Energy Units

au (atomic units), ***kJ/mol*** and ***kcal/mol***

Reaction Energy Units

au (atomic units), ***kJ/mol*** and ***kcal/mol***

Model: Wire/Ball and Wire/Tube/Ball/Spoke/Space Filling/Line

Controls default model style.

Surface Style

If **Bands** is *checked*, this specifies that graphical surfaces, for example electrostatic potential maps, are to be displayed in terms of a series of color bands, rather than as a continuous spectrum. The number of bands is selected from the menu to the right. This setting can be changed at the

document level and individual surface level as well (**Surface Properties** under the **Display** menu; **Chapter 22**).

Show Objects: Constraints / Frozens / Points / Planes / Reactions / CFD's / Images / Annotations / CH Couplings / HH Couplings
If *checked*, constraints and frozen markers, points and planes, reaction arrows, CFD's and attached images, user-supplied annotations and experimental CH and HH couplings will always be shown as part of the model. Otherwise, they will be shown only in the appropriate mode.

Atom Labeling: Label / Element / Mass Number / Mulliken Charge / Electrostatic Charge / Natural Charge / Strand: Residue / Label / R/S / Exposed Area / Cal Chem Shift/ Exp Chem Shift / (Cal-Exp) Chem Shift / Custom
Controls default label type. **Custom** allows the user to enter an expression designating a label. These are of the same form as used in the spreadsheet. See the section on **User-Defined Expressions** under **Spreadsheet** in the **Display** menu (**Chapter 21**).

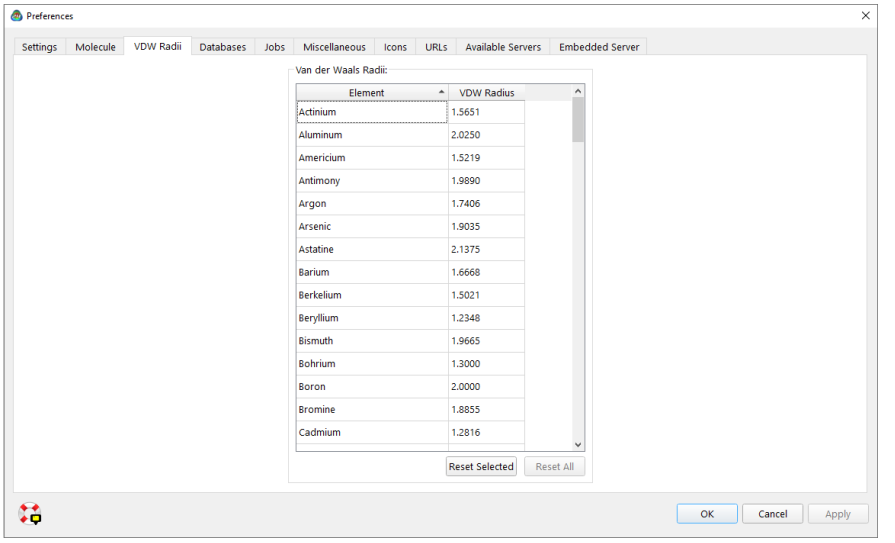
Bond Labeling: Label/Custom
Controls default label type. **Custom** allows the user to enter an expression designating a label. These are of the same form as used in the spreadsheet. See the section on **User-Defined Expressions** under **Spreadsheet** in the **Display** menu (**Chapter 21**).

Other Labeling: Hydrogen/Constraint/Plane/CFD/Reaction/ Coupling/Residue/Point
If *checked*, corresponding labels will be shown.

CFD Display: Descriptors/Excluded Volume
If *checked*, corresponding labels will be shown.

vdW Radii Tab

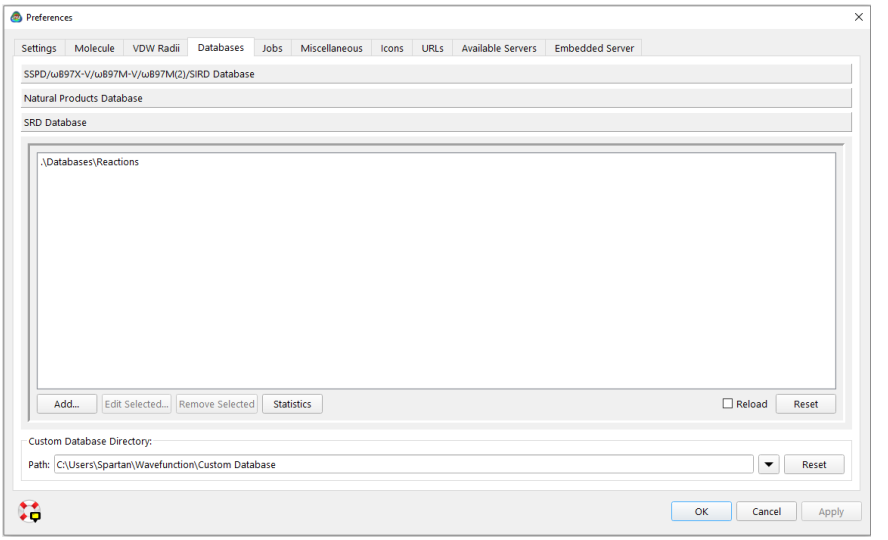
This provides a list of van der Waals radii



To order the list by element name *click* on **Element**, and by atomic radius *click* on **vdW Radius**. Individual entries may be changed from default values by first *clicking* on the entry and then entering a new value. The currently selected entry may be returned to its default radius by *clicking* on **Reset Selected** at the bottom of the dialog, and the full set of radii may be returned to their default values by *clicking* on **Reset All** at the bottom of the dialog.

Databases Tab

This allows setting up of paths for Wavefunction database directories.



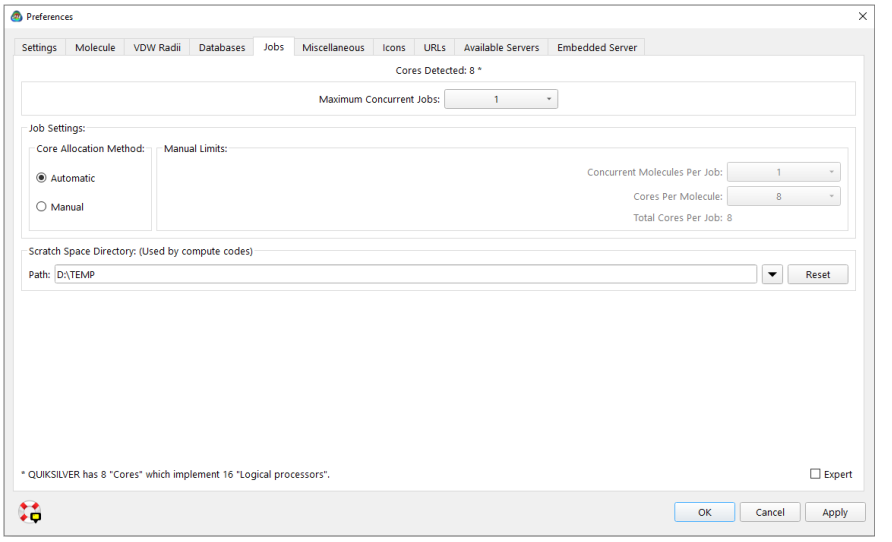
Several directories may appear for each database. Specification of which directory (or directories) will be searched in the **Search Options** dialogs associated with SSPD, ωB97X-V, ωB97M-V, ωB97M(2), SIRD, SRD and XIRD (**Databases** under the **Search** menu; **Chapter 23**). Note that SIRD is the same database as SSPD*, and only the method of access (spectral matching vs. substructure matching) are different.

User-defined collections may be added (see **Appendix H**).

* More precisely, it is the subset of entries in SSPD that contain infrared spectra.

Jobs Tab* (Platform dependent, see comments in **Appendix K**)

This allows setting of job limits.



Maximum Concurrent Jobs (Windows)
Designates the maximum number of jobs that can be run at one time.

> You need to use caution that the maximum number of jobs does not significantly exceed the number of physical processors (cores) on your computer and also that you have sufficient memory to support this maximum number.

Job Settings
Toggles between core/thread allocation method, either: **Automatic** or **Manual**. **Automatic** will assign all available cores to a single document (single molecule inside a ***Spartan*** document if there is only one molecule in the document) and will attempt to optimally divide total cores/threads among the molecules in the document to maximize performance when there are multiple molecules in a document. **Manual** will follow settings under **Manual Limits**.

Manual Limits
Concurrent Molecules Per Job (Windows) sets the number of molecules inside a ***Spartan*** document that that will run simultaneously and **Cores Per Molecule (Windows)** sets the number of cores used on a single molecule.

If you are simultaneously working with two or more documents, you need to use caution. It is very easy (and, depending on available memory, potentially very inefficient) to demand more threads than are actually available. This is not an issue if you are working only with a single document even if it contains more than one molecule.

Scratch Space Directory: (used by compute codes)
Allows users to customize the temporary location ***Spartan*** uses during calculations. The default is the same location your operating system uses. Configuring to utilize a solid state drive has been shown to increase performance.

Checking **Expert** at the bottom right brings up an extended dialog, containing controls for individual job memory and scratch disk limits. Unless you've been advised to modify these from the default values, it is likely these will only be needed in exceptional circumstances.

Memory Limit
Memory (RAM) in units of MB, in general Spartan will request the amount RAM needed to complete a calculation (upon submittal), in cases where this value is modified, use must take into account the setting under **Maximum Concurrent Jobs** – for example, if memory limit is set to 15000 MB (15 GB), but the maximum concurrent jobs setting is 4, Spartan will expect the machine to have as much as 60 GB of RAM available if four jobs are submitted.

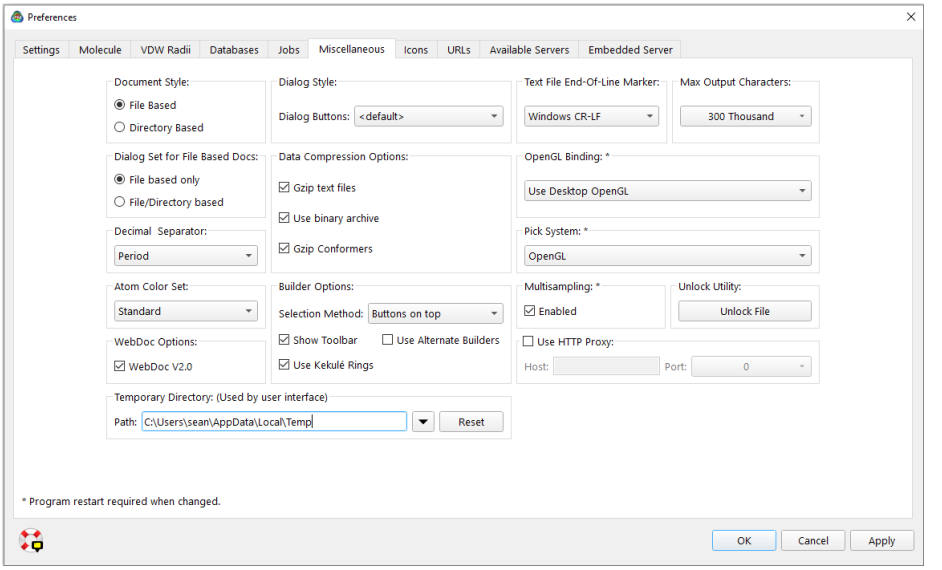
Scratch Disk Space
Disk space in units of GB. In general, the default setting should be sufficient. Min requirements are 128 GB of available disk (not all calculations require significant disk resources).

---

* **Note:** Cores are physical processors (typically between 4 and 64 for current generation Intel and AMD chips) whereas threads are virtual processors, the number of which is an integer multiple of the number of cores (typically 2). Optimum performance in numerical applications such as ***Spartan*** benefits most (in parallel performance) with access to additional physical cores. It is not typically advisable to increase the total core count beyond the number of physical cores (performance cores) in the machine.

Miscellaneous Tab

Items here refer to primarily cross-platform applications.



Document Style

Toggles between **File Based** and **Directory Based**. The latter is for compatibility between Windows and previous Macintosh and Linux versions prior to ***Spartan'16***.

Dialog Set for File Based Docs

Where **Document Style** is set to **File Based**, restricts the file browser to read **File Based** documents only or both **File Based** and **Directory Based** documents.

Atom Color Set

Toggles between **Standard** atom colors shared by ***Spartan*** and ***Odyssey*** software and a color scheme used by older versions of ***Spartan*** (**Classic**).

Dialog Style

Menu selects between several dialog formats: **Windows**, **Macintosh**, and **KDE** and **Gnome** Linux. **Default** uses the style associated with the machine on which ***Spartan*** is installed.

Data Compression Options

Allows compression of text files (**Gzip text files**) and conformer file (**Gzip Conformers)** and switches to binary for full archive storage: (**Use binary archive**). By default all are selected to minimize file sizes of ***Spartan*** documents. Deselection may be required for file compatibility with much earlier versions of ***Spartan***.

Builder Options Selection Method

Toggles among **Buttons on top**, **Tabs on side** and **Menu** to control selection of the model kits in the 3D builder. The second and third choices use less vertical space and may be required on small-screen laptops and tablets.

Show Toolbar
If *checked*, places the toolbar (**Delete**, **Make Bond**, **Break Bond** and **Minimize**) at the bottom of the 3D builder. Can be removed to save space on small screens.

Use Kekulé Rings
If *checked*, aromatic rings from 3D builders are shown in Kekulé (alternating single and double bond) form. Otherwise, they are shown with partial double bonds. Aromatic rings from 2D sketcher are always shown in Kekulé form.

Use Alternative Builder
If *checked*, provides a different presentation of the model kits in
the 3D builder. Try both and see which one you like.

Temporary Directory (used by interface)
Allows users to specify the temp location used for graphics calculations when using the **Auto-Gen Graphics** preference (on by default). Configuring to utilize a solid state drive has been shown to significantly increase performance.

Use HTTP(S) Proxy
Allows setting up of an alternative path for access to external websites, for example, experimental spectra databases.

Text File End-Of-Line Marker
Controls line terminator for text files. May be needed to support file transfer to earlier versions of ***Spartan*** and to external programs specific to Windows, Mac or Linux environments.

WebDoc V2.0
If *checked* (the default), this optimizes file transfers to and from remote servers. It may need to be *unchecked* where the servers are running earlier versions of ***Spartan***.

Max Output Characters
Sets maximum number of characters for output files.

OpenGL Binding
Toggles between **Use Desktop OpenGL** or **Use Software OpenGL**. In cases where the local display adapter does not full support OpenGL (not uncommon among Integrated Intel Graphics), or if the program accessed via remote desktop, it may be necessary to switch to the Software OpenGL option, which utilizes provided OpenGL libraries in the Spartan installation, rather than the machine's display adapter driver set.

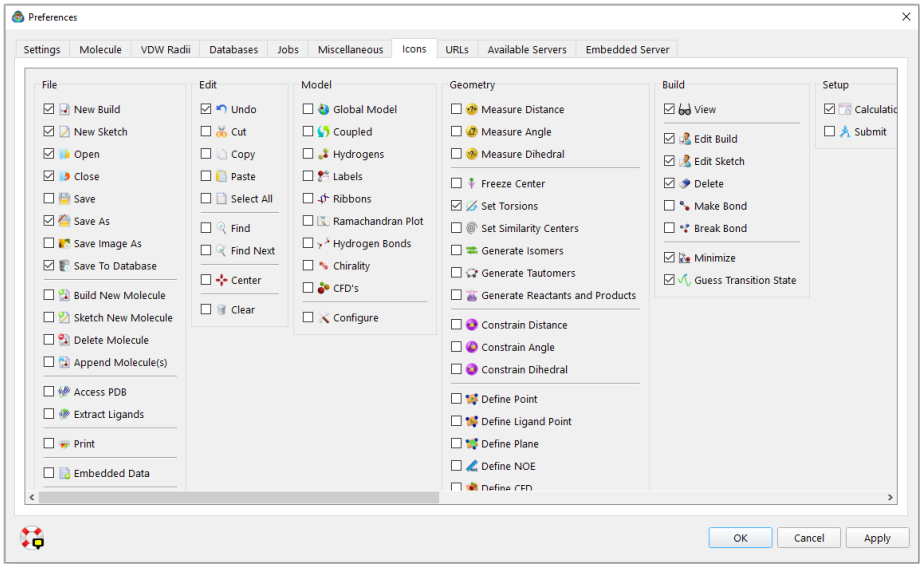
Pick System
Toggles between **OpenGL**, **Color** and **Color+Geometric** selection models. **OpenGL** is the standard but causes problems for some Intel on chip graphics, where either **Color** or **Color+Geometric** should be used. Graphics chip is automatically detected upon installation and this control should be properly set.

Multisampling
If *checked* invokes multisampling which results in improved graphical images (removes jagged lines). Likely to be unnecessary for very high-resolution displays. Will slow down graphics performance.

Unlock Utility
In some cases, .spartan (or spartan directories) can be locked. This may occur if a user attaches a file to email while it is still open (for example), or in some cases if the calculation fails due to an unexpected event (power failure or an OS system signal), the file may close with a "locked" flag in place. To open a file that presents a *locked*, Click **Unlock File** which leads to the File System, point to and open the file that reports as locked.

Icons Tab

Icons for all menu entries as well as for the **Spectra** panel are listed (you will need to use the horizontal slider bar to see them all). If *checked*, the icon will appear in the tool bar at the top of the ***Spartan*** screen and in the **Spectra** panel.

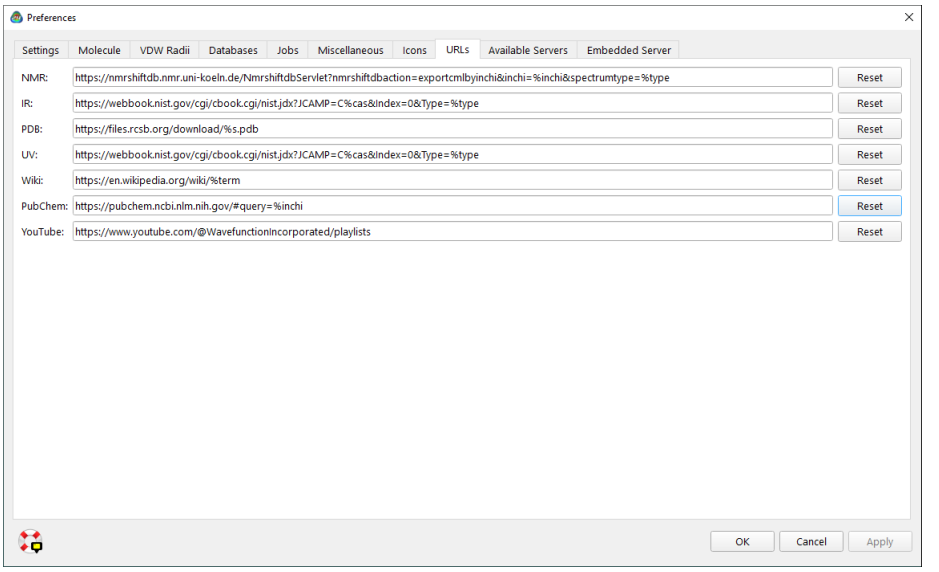


Icon display is limited to one "permanent" row and approximately 20 medium size icons will fit on screen. Exceeding screen space will result in the inclusion of arrows [>>] at the end of the toolbar. *Clicking* the [>>] button will expose any hidden icons.

Users have full control over which icons are permanently displayed in the icon bar, but all options are available from Spartan's main pull-down menus: **File**, **Edit**, **Model**, **Geometry**, **Build**, **Setup**, **Display**, **Search**, **NMR**, **Options**, **Activities**, and **Help**.

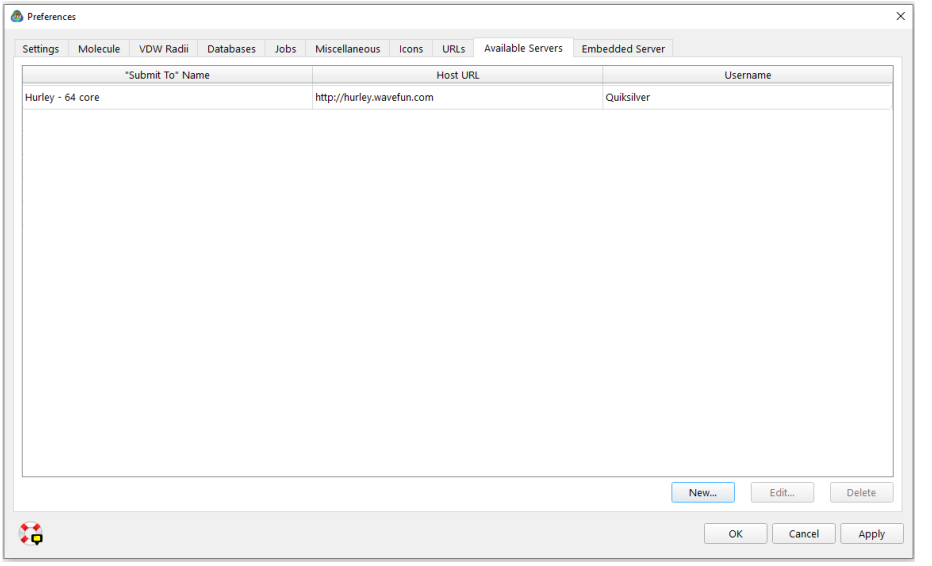
URLs Tab

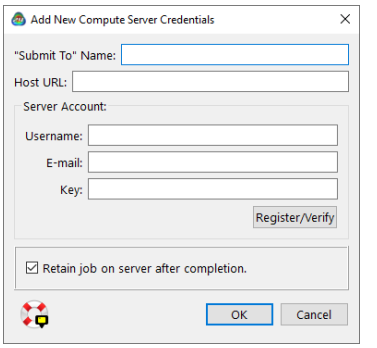
Lists URLs for access to experimental structural and spectral databases and to Wikipedia and PubChem sites along with Wavefunction's YouTube channel.



Available Servers Tab

In addition to processing molecular mechanics and quantum chemical calculations "locally", ***Spartan'26*** is able to submit calculations to a licensed remote server. Remote accounts need to be set up using the **Embedded Server** (Windows) tab in the **Preferences** dialog or **Service Control** (Macintosh or Linux) in the **Options** menu (see later in this chapter). **Available Servers Tab** is used to request accounts on one or more remote servers.





A server is added by first *clicking* on **New** at the bottom of the **Available Servers** dialog.

"Submit To" Name
The name of the server. This will appear in the **Submit to** menu inside the **Save As - Submit** dialog.

Host URL
The URL of the server. This can either be provided as a text string (FQDN, or Fully Qualified Domain Name suffix) or as an IP address. At present **http** and (new in ***Spartan'26***) **ssh** options are available for submission type.

Username
A text string (with **no spaces**) used by the server to identify you.

E-mail
Your e-mail address. This is presently a required field but is not actually utilized or stored anywhere beyond the user's machine.

Key
This is required only in the event that the server operates in high-security (Remote Registration: **Disabled**) mode. This is controlled by the administrator of the server from the **Embedded Server** tab in the **Preferences** dialog (Windows) or **Server Control** dialog (**Options** menu Macintosh and Linux), see later in this chapter.

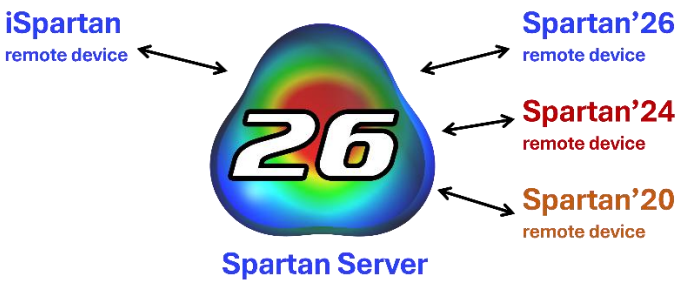
Retain Job On Server After Completion
If *checked*, retains the job on the server after it has completed. It can be retrieved at any time. Unless specifically desired, it is advised to leave this option off.

Clicking on **Register/Verify** will create/confirm an account (or will respond with any error conditions).

Does your organization you push out a custom application images? Customers with dedicated computational devices – a means for pre-populating registration fields and automatically registering upon initial job submission is available (via XML file). Contact ***support@wavefun.com*** for additional details.

Embedded Server Tab/Service Control

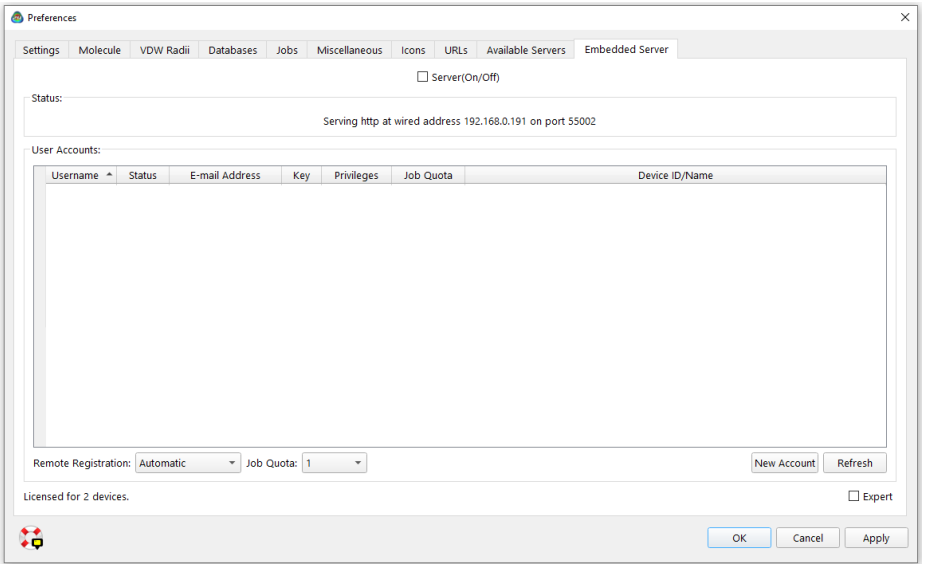
In addition to its local capabilities, ***Spartan'26*** be used as a server for licensed copies of ***Spartan'26*** running on other machines as well as (older) ***iSpartan*** running on iPads and iPhones (***remote devices***).



Operation requires communication between the server and the remote device. At the lowest level of security (the default), nothing is required at the server end (except turning on the server) and each of the remote devices simply needs to be registered on the host. Higher security levels require explicit action on the part of the administrator of the server, and the highest-level security requires explicit exchange of "keys" between the administrator of the server and the users of the remote devices. Content below describes general operations, see **Appendix K** for a complete description of functionality, including a description of platform dependencies.

Turning the Server On and Off

Checking the box to the left of **Server (On/Off)** in the **Embedded Server/Service Control** dialog turns the server on. *Unchecking* the box turns the server off. When switched ***on***, the IP of the server will be provided under **Status**. You will need to *click* the **Apply** button after making any changes. If the device has been setup as a computational server, additional details will apply, see Appendix K for more information on remote server options.



Remote Registration

From simplest and least secure to more complicated and most secure, there are three ways to "register" remote devices with the server: **Automatic**, **Sign-Off Required** and **Disabled**. This is designated in the **Remote Registration** menu at the bottom of the dialog.

Automatic. In this mode, anyone with ***Spartan'26*** or ***iSpartan*** can set up an account without the administrator needing to take any action. The only exception might be to limit the number of jobs that can be executed at one time by an individual account on the server. This is controlled by the **Job Quota** menu at the bottom of the dialog.

Sign-Off Required. In this mode, anyone can set up an account (as in **Automatic**), but the administrator must activate the account. Prior to activation, **Status** is **pending**. This may be changed to **activated** by *double clicking* inside the cell corresponding to the status of the pending account followed by *clicking* on **Accept** in the dialog that results and then *clicking* on **Apply** at the bottom of the **Server** dialog.

Disabled. In this mode, the administrator is required to set-up a user's account. This requires that the user supplies the administrator with the Machine ID of the computer. For ***Spartan'26***, this is available from the **About Spartan'26** dialog (**Help** menu). For iOS devices, leave the Machine ID blank (all zeroes). The Machine ID is used by the administrator to generate a **key** which is then sent back to the user. This needs to be entered along with the other information (associated with lower-security accounts) during the registration process (see discussion in **Available Servers** earlier in this chapter).

User Accounts

Associated with each user account is a series of attributes. These identify the account, provide contact information of the user and allow the server administrator to set limits. *Double-click* a field to edit.

Username
This is the name of the account provided by the user at the time the account was set up (spaces are ***not*** allowed).

Status
This is marked either **Activated** or **Pending**. **Activated** means that an account is available for use. **Pending** means that the administrator needs to take further action in order for the account to be available.

E-mail Address
This is the e-mail address of the user provided at the time the account was set-up.

Key
This is the key supplied to the user by the administrator and then used to set up an account. This only applies for the highest-security setting for **Remote Registration** (**Disabled**).

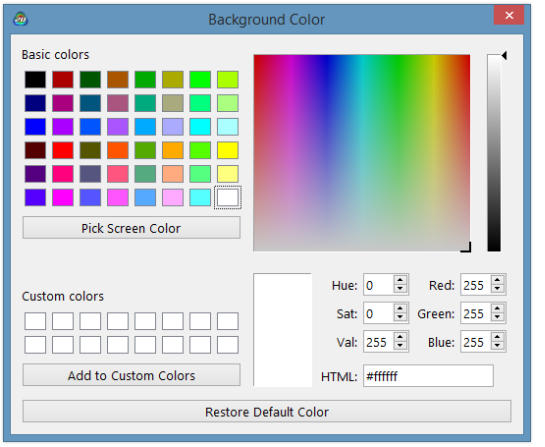
Privileges
This allows the administrator to give an account the permission to start jobs that are in the job queue.

Job Quota
This is the maximum number of jobs that a particular user account is permitted to run at one time. It overrides (for this account) the global job quota discussed earlier. The administrator selects a number (or "unlimited") from the menu inside the **Edit Job Quota** dialog.

Device ID/Name
This identifies the device (computer, tablet, or iOS device) associated with a user account.

Colors

This alters default colors. Selection leads to the **Colors** dialog.

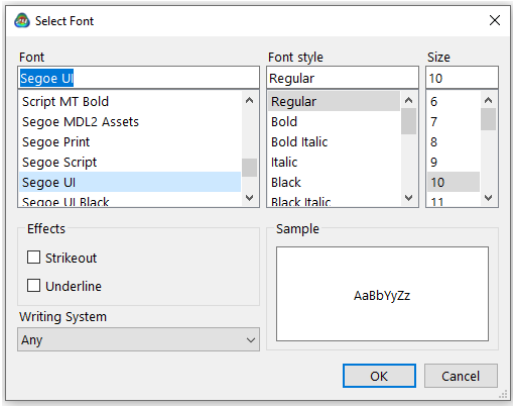


After selecting an object, its color may be set by choosing from the palette, moving the cursor inside the window of colors, or by selecting either a set of hue, saturation and values,

or red, green and blue settings or with the HTML designator. The default color may be reset by *clicking* on **Restore Default Color**. Color selection applies to all objects of the same type, for example, all carbon atoms, and not just to the selected carbon. Further control of colors is available from **Utilities/Style** dialogs associated with **Properties** dialogs (**Properties** under the **Display** menu; **Chapter 22**). *Clicking* on removes the dialog.

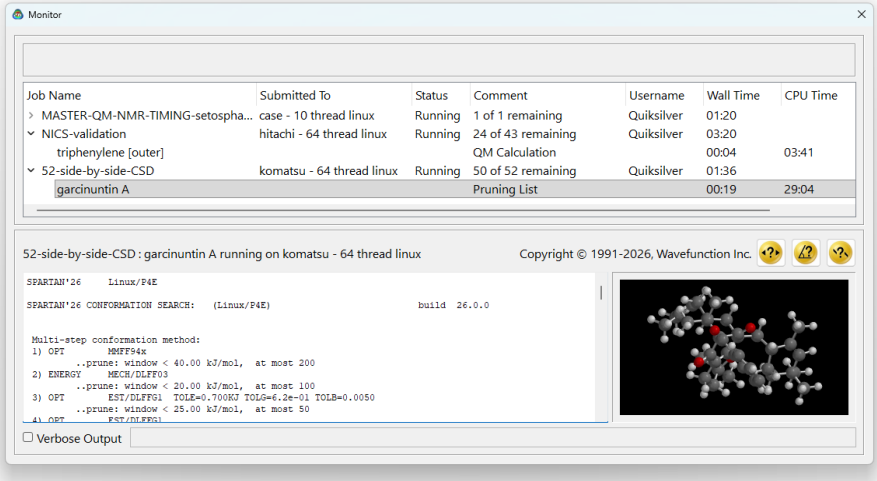
Fonts/Graphics Fonts. . . /

This selects fonts, style and size of labels attached to molecules (**Labels** and **Configure...** under the **Model** menu; **Chapter 18**), and plots (**Plots...** under the **Display** menu; **Chapter 22**). Selection leads to the **Fonts** dialog.



Selections are made from the **Font**, **Font Style** and **Size** menus. *Clicking* on **OK** dismisses the dialog with selections kept. *Clicking* on **Cancel** or on dismisses the dialog but selections are lost.

Monitor



This provides a listing of all executing/queued jobs and their status; both locally submitted and remotely submitted jobs are listed. To see accumulated output for an executing job, *click* on the triangle to the left of the document name and then select the name of the molecule. A ball-and-spoke model of the selected (executing) job will be displayed in a window to the right of the dialog. It can be manipulated using the usual mouse commands (you need to position the cursor inside the window). Touch-screen commands are presently limited to rotation (move one finger). Model style cannot be changed. Note that (except for molecular mechanics and semi-empirical calculations) the structure is updated throughout an equilibrium geometry of transition state optimization, and bond lengths, angles and dihedral angles can be queried.

To stop a job, *right-click* on its name, and then select **Terminate** from the **Actions** menu at the top of the dialog (or *right click* on its name and select **Terminate** from the menu that appears). To start a queued job (irrespective of the imposed queue limits; see previous discussions under **Jobs**), *right-click* on its name and select **Start** from the **Control Actions** menu.

The **Monitor** may be removed either by selecting **Exit** from the **File** menu or by *clicking* on at the top of the dialog.

Calculator

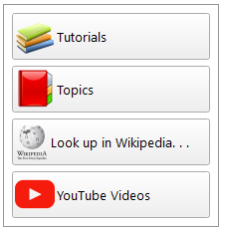
Selection brings up a **Calculator**. This functions the same way as a normal pocket calculator. The **Calculator** is removed by *clicking* on .

Icons

Toggles the set of icons that appear in the icon bar above the menus either on or off. For customization of specific icons, see **Preferences**, **Icons** tab (earlier this chapter).

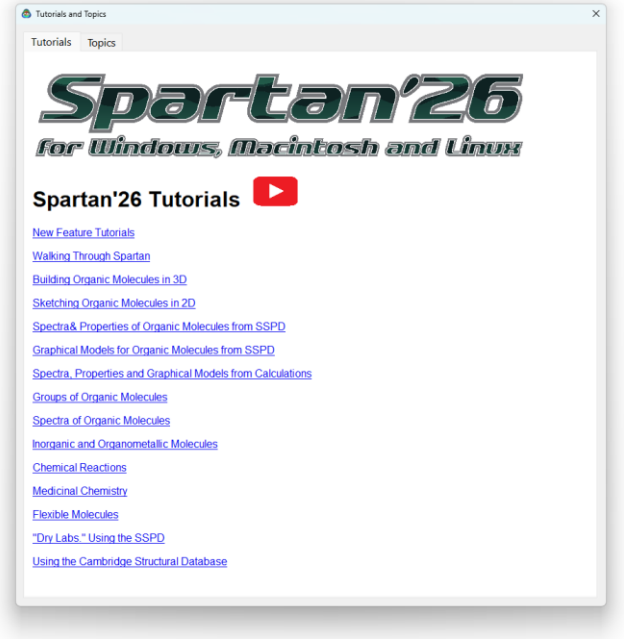
Chapter 26
The Activities Menu

*The **Activities** menu permits on-screen display of the full set of **Spartan** tutorials and a series of topics of practical relevance to molecular modeling. It also allows a Wikipedia page to be brought up (external to **Spartan**).*



Tutorials/Topics /

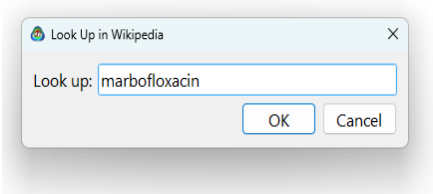
Selection of **Tutorials** or **Topics** brings up a dialog with either the **Tutorials** or **Topics** tab selected (**Tutorials** tab shown).



Selection of an entry results in opening of a chapter from the **Tutorials** as a pdf file which can be displayed alongside ***Spartan*** while working through the chapter's content.

Look Up in Wikipedia...

Selection results in a dialog.



By default, the query is the name of the selected molecule. *Clicking* on **OK** or entering a different query followed by *clicking* on **OK** leads to a Wikipedia page. This occupies a window that is external to ***Spartan***.

YouTube Videos

Selection results in a new browser window opening and automatic connection to Wavefunction's YouTube channel, including playlists with Tutorial videos.

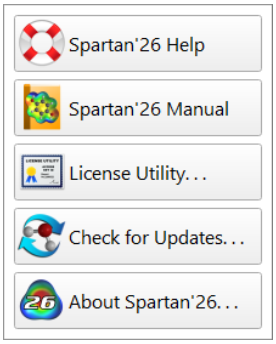
Chapter 27

The Help Menu

This chapter describes available help content.

Spartan'26 Help

This provides information relating to application of computational methods available in ***Spartan***, as well as technical details regarding the program's operation.



Several items are accessed under **Spartan'26 Help**; including ***General Operating Features***, ***Selecting a Theoretical Model***, ***Calculation***, and a number of FAQ documents (***License Questions, Properties, Quantum Mechanics Energy Questions***, ***Basis Sets***, ***Molecular Mechanics Questions***, ***Convergence Questions***, ***Conformation and Energy Profile Questions***, ***Memory Questions, Neural Networks, IR Searching, and Option Line Keywords***) related to computational questions. **Help** also provides a link to Wavefunction's website. Help files are HTML documents.

The ***Spartan Tutorial and User's Guide*** (this document) and ***A Guide to Molecular Mechanics and Quantum Chemical Calculations*** are also available (as PDF files) under **Help**.

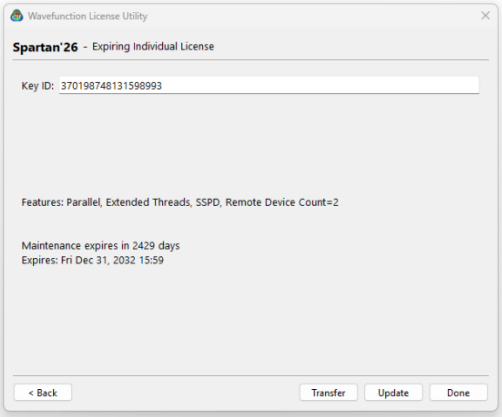
Finally, note that several dialogs, in particular, the **Calculations** dialog incorporate imbedded help messages. *Clicking* on at the upper right, followed by *clicking* on a menu, button, etc. in the dialog gives rise to a brief informative message about the object queried.

Spartan'26 Manual

Provides access to the ***Spartan'26 Tutorial and User's Guide*** (this document) as a PDF file.

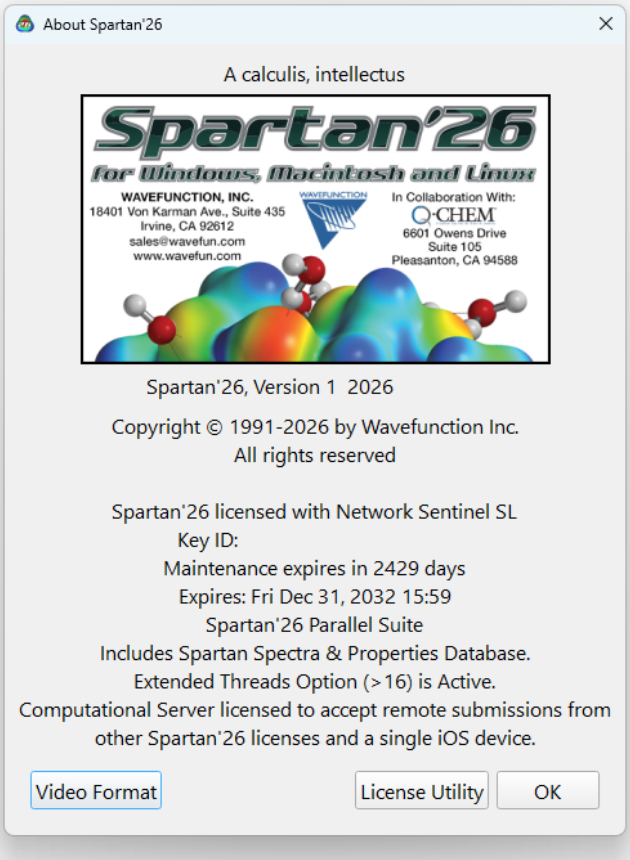
License Utility

Provides access to the **License Utility** dialog. From here you can access the **Key ID** for your license and reference available **Features** and **Maintenance** date, change license option from the originally specified choice (Individual or Network) by *clicking* the **Back** button, **Update** the license, or request a license **Transfer**.



About Spartan'26

Provides information about the user's release of ***Spartan***.



Appendix A

Topics

Essays address topics that are important to molecular modeling in general and to modeling applications of ***Spartan*** in particular. Most of these essays are at an elementary level and provide only fundamentals. More detailed information is provided in the essays dealing with flexible molecules and with NMR spectroscopy to address the protocols in ***Spartan***.

The firs essay deals exclusively with the machine learning developments, new in this release, **Neural Network Models**.

The next two essays are foundational. **Potential Energy Surfaces** relates two of the quantities that directly result from a quantum chemical calculation, geometry and energy. **Theoretical Models** outlines the steps taken in moving from the Schrödinger equation to practical techniques and broadly outlines the scope of these techniques.

Finding and Verifying Equilibrium and Transition-State Geometries and **Total Energies and Thermodynamic and Kinetic Data** provide specifics about the relationship between geometry and energy in the context of practical quantum chemical models. **Calculating Accurate Heats of Formation** details procedures that are available in ***Spartan*** to reliably reproduce the experimental heats of formation.

Dealing with Conformationally-Flexible Molecules describes requirements needed to identify the lowest-energy conformer of a flexible molecule. Multistep protocols first introduced with ***Spartan'18*** expanded the coverage to high-level quantum chemical models able to provide accurate Boltzmann-weighted averages.

Interpreting Conformational Preferences shows how information resulting from a calculated energy profile for single-bond rotation may be related to familiar chemical notions.

Calculating Infrared Spectra shows how the raw data resulting from a quantum chemical calculation may be combined with two empirical parameters to produce accurate infrared spectra. **Calculating NMR Spectra** describes a multi-parameter scheme to obtain ^{13}C chemical shifts into close agreement with experiment.

Calculating Chemical Shifts in Conformationally Flexible Molecules describes protocols for efficiently calculating NMR chemical shifts for molecules with multiple degrees of conformational freedom. One set of protocols introduced with ***Spartan'18*** obtains accurate Boltzmann weights to form proper averages while the set of protocols introduced with ***Spartan'20*** and expanded in ***Spartan'26*** enable much simpler and lower-cost methods to select the conformer that best agrees with measured shifts. **Assigning Stereo and Regioisomers from Calculated Chemical Shifts** ranks chemical shifts of isomers according to DP4 scores. **Calculating HH and CH Coupling Constants** assesses practical

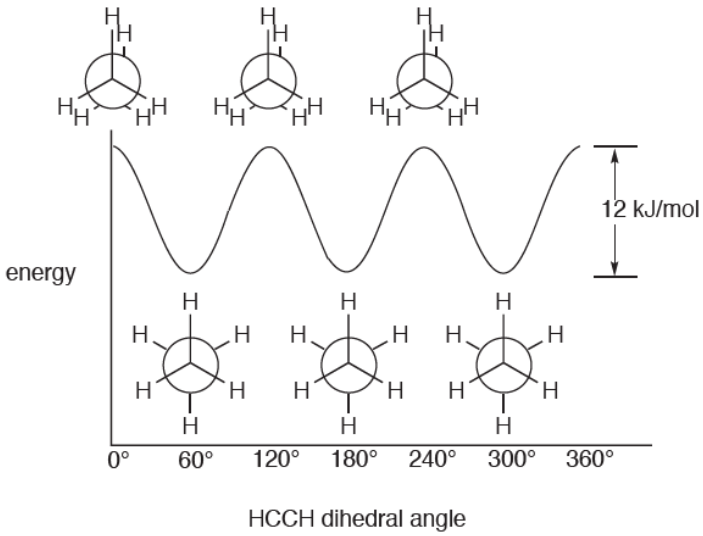
methods for calculation of 3-bond HH and 2 and 3-bond CH coupling constants and in addition describes an empirical scheme for obtaining CH coupling constants. **Calculating COSY and HMBC Spectra** describes presentation modes for 2D NMR spectra and comparison with experimentally assigned couplings.

Atomic and Molecular Orbitals describes how molecular orbitals arise from atomic orbitals in the context of the practical quantum chemical models introduced previously and illustrates the information that can be drawn from them. **Electron Densities: Sizes and Shapes of Molecules** describes molecular size and shape from the perspective of quantum mechanics and in so doing provides a "platform" from which to evaluate how a molecule interacts with its environment. **Electrostatic Potential Maps: Charge Distributions** describes a model that uses color to "map" the electrostatic potential (the energy of a positive point charge with the nuclei and electrons of a molecule) in order to distinguish neutral, positive and negative regions on an accessible surface. **Local Ionization Potential Maps and LUMO Maps: Electrophilic and Nucleophilic Reactivities** show how maps colored using the energy of ionization and the absolute value of the lowest-unoccupied molecular orbital, respectively, may be used to account for electrophilic and nucleophilic reactivity.

POTENTIAL ENERGY SURFACES

One Dimensional Energy Surfaces

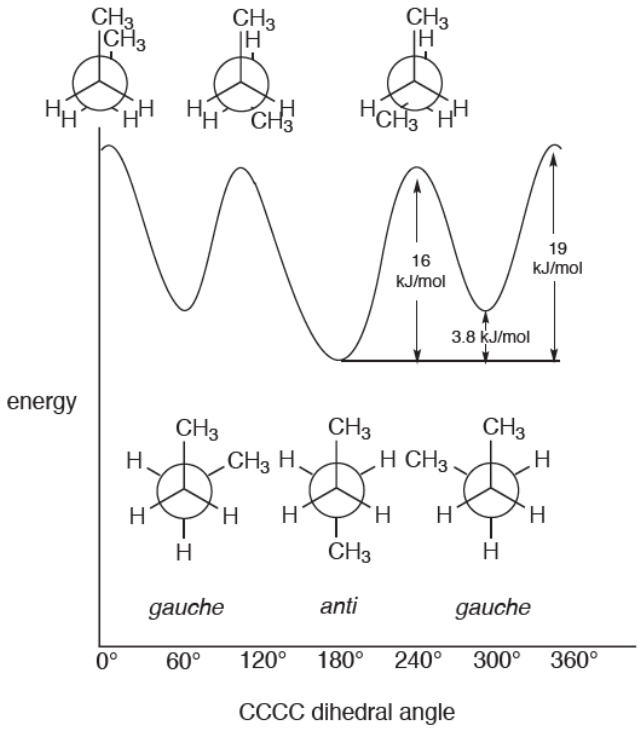
Every chemist has encountered a plot depicting the change in energy of ethane as a function of the angle of torsion (dihedral angle) around the carbon-carbon bond.



Full 360° rotation leads to three identical energy minima in which the hydrogens are staggered, and three identical energy maxima in which the hydrogens are eclipsed. The difference in energy between eclipsed and staggered structures of ethane, termed the barrier to rotation, is known experimentally to be approximately 12 kJ/mol. Note that any physical measurements on ethane pertain only to its staggered structure, or more precisely the set of three identical staggered structures. Eclipsed ethane ***does not exist*** in the sense that it cannot be isolated and characterized. Rather, it can only be imagined as a structure in between equivalent staggered forms.

> Open ***ethane rotation*** in the ***topics*** directory*. The image which appears is one frame in a sequence depicting rotation about the carbon-carbon bond in ethane. *Click* on the ◀ and ▶ keys at the bottom left of the screen to look at other frames. Verify that the staggered structures correspond to minima on the energy plot and that the eclipsed structures correspond to maxima. *Click* on the ▶ key to animate the sequence. Close ***ethane rotation*** when you are finished.

Somewhat more complicated but also familiar is a plot of energy vs. the dihedral angle involving the central carbon-carbon bond in *n*-butane.



* For Windows, the ***Topics*** directory is found in ***Program Files/Wavefunction/Spartan26***. It needs to be copied to another location available to the user prior to opening it in ***Spartan***. For Macintosh, this is located at the top of the ***Spartan26*** disc image. For Linux, the ***Topics*** directory is found in the install directory. Copy the ***Topics*** directory to a location that allows write permission, typically the user's home directory.

This plot also reveals three energy minima, corresponding to staggered structures, and three energy maxima, corresponding to eclipsed structures. In the case of *n*-butane, however, the three structures in each set are not identical. Rather, one of the minima, corresponding to a dihedral angle of 180° (the *anti*-structure), is lower in energy and distinct from the other two *gauche* minima (dihedral angles around 60° and 300°), which are identical. Similarly, one of the energy maxima corresponding to a dihedral angle of 0°, is distinct from the other two maxima (with dihedral angles around 120° and 240°), which are identical. As with eclipsed ethane, eclipsed forms of *n*-butane do not exist, and correspond only to hypothetical structures in between *anti* and *gauche* minima. Unlike ethane, which is a single compound, any sample of *n*-butane is made up of two distinct compounds, *anti n*-butane and *gauche n*-butane. The relative abundance of the two compounds as a function of temperature is given by the Boltzmann equation (see the topic ***Total Energies and Thermodynamic and Kinetic Data***).

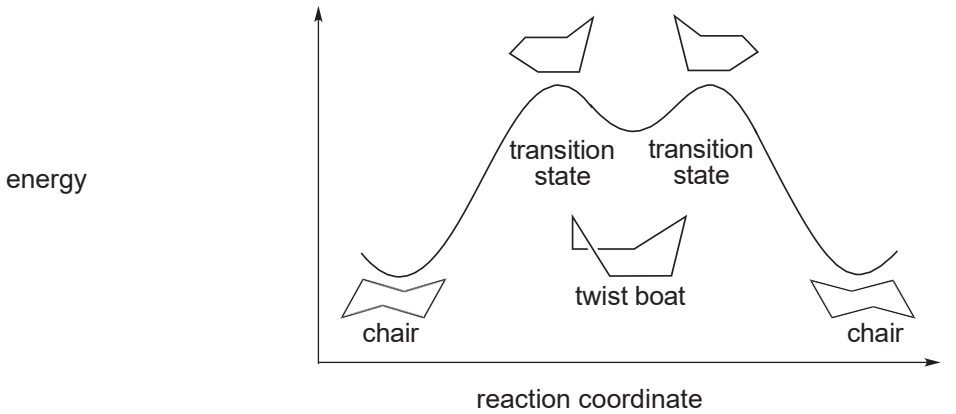
Open ***n-butane rotation*** in the ***topics*** directory. The image which appears is one frame of a sequence depicting rotation about the central carbon-carbon bond in *n*-butane. *Click* on the and keys at the bottom left of the screen to look at other frames. Verify that the staggered structures correspond to minima on the energy plot and that the eclipsed structures correspond to maxima. Also, verify that the *anti*-structure is lower in energy than the *gauche* structure. *Click* on to animate the sequence. Close ***n-butane rotation*** when you are finished.

The important geometrical coordinate in *n*-butane may clearly be identified as a torsion about the central carbon-carbon bond. This is an oversimplification, as bond lengths and angles no doubt change during rotation around the carbon-carbon bond.

Quantum chemical models available in ***Spartan*** are able to account for the subtle changes in bond lengths and angles which result from changes in conformation. Open ***n-butane geometry changes*** in the ***topics*** directory. The two plots depict the variation in central CC bond distance and in CCC bond angle as a function of the CCCC torsional angle. The variation in energy is superimposed on each plot. Note how closely the bond distance and energy changes parallel each other. Note also that the bond angle is insensitive to conformation except in the region of the *syn* (0° torsional angle) structure where it has opened up by several degrees. Close ***n-butane geometry changes*** when you are finished.

Many Dimensional Energy Surfaces

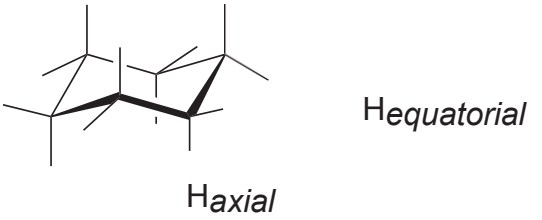
It will usually not be possible to identify a simple geometrical coordinate to designate a chemical transformation. A good example of this is provided by the potential energy surface for ring inversion in cyclohexane.



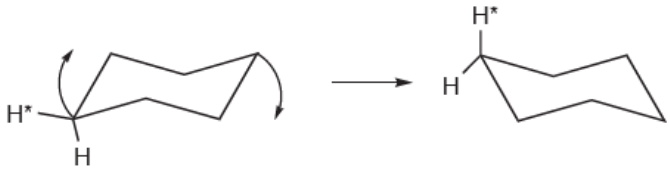
In this case, the geometrical coordinate connecting stable forms is not specified in detail (as it was in the previous two examples) but is referred to simply as the ***reaction coordinate***. The two energy maxima on the ***reaction coordinate diagram*** have been designated as ***transition states*** to indicate that their structures may not be simply described (as are the energy maxima for rotation in ethane and *n*-butane).

The energy surface for ring inversion in cyclohexane, like that for *n*-butane, contains three distinct energy minima, two of lower energy referred to as chair forms, and one of higher energy referred to as a twist boat form. In fact, the energy difference between the chair and twist boat structures is around 23 kJ/mol and, only the former can be observed at normal temperatures. For a discussion, see the topic ***Total Energies and Thermodynamic and Kinetic Data***.

All six carbons are equivalent in the chair form of cyclohexane, but the hydrogens divide into two sets of six equivalent *equatorial* hydrogens and six equivalent *axial* hydrogens.



However, only one kind of hydrogen can normally be observed, meaning that *equatorial* and *axial* positions interconvert via some low-energy process. This is the ring inversion process just described, in which one side of the ring is bent upward while the other side is bent downward.



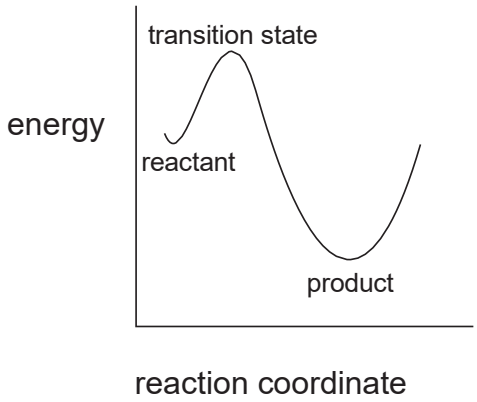
As shown in the potential energy diagram on the previous page, the overall ring inversion process appears to occur in two steps, with a twist boat structure as a midway point (an intermediate). The two (equivalent) transition states leading to this intermediate adopt structures in which five of the ring carbons lie (approximately) in one plane.

The energy profile for ring inversion in cyclohexane may be rationalized given what we have already said about single-bond rotation in *n*-butane. Basically, the interconversion of the reactant into the twist-boat intermediate via the transition state can be viewed as a restricted rotation about one of the ring bonds.

Correspondingly, the interconversion of the twist boat intermediate into the product can be viewed as rotation about the opposite ring bond. Overall, two independent bond rotations, pausing at the high- energy (but stable) twist-boat intermediate effect conversion of one chair structure into another equivalent chair, and at the same time switch *axial* and *equatorial* hydrogens.

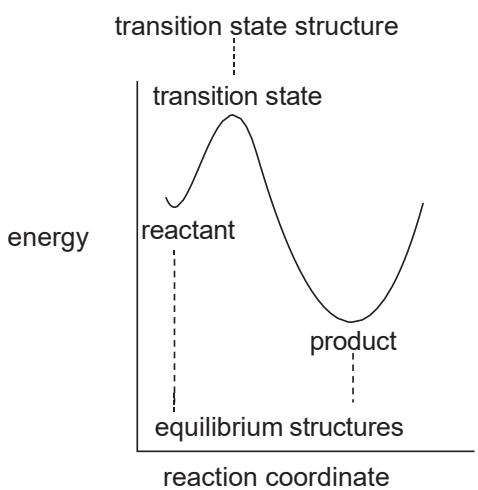
Open ***cyclohexane ring inversion*** in the ***topics*** directory. The image which appears is one frame in a sequence depicting ring inversion in cyclohexane. *Click* on the ◀| and |▶ keys at the bottom left of the screen to look at other frames. Verify that the three minima on the energy plot correspond to staggered structures and that the two maxima correspond to eclipsed structures. Also, verify that the twist boat structure is higher in energy than the chair structures. *Click* on ▶ to animate the sequence. Note that the overall ring inversion appears to occur in two steps, one step leading up to the twist boat and the other step leading away from it. Close ***cyclohexane ring inversion*** when you are finished.

Ethane, *n*-butane and cyclohexane are all examples of the types of motions which molecules may undergo. Their potential energy surfaces are special cases of a general type of plot in which the variation in energy is given as a function of reaction coordinate.

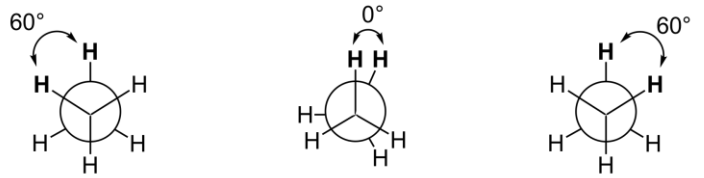


Diagrams like this provide essential connections between important chemical observables: structure, stability, reactivity, selectivity, and energy.

The positions of the energy minima along the reaction coordinate give the equilibrium structures of the reactant and product. Similarly, the position of the energy maximum gives the structure of the transition state. Both energy minima (which correspond to stable molecules) and the energy maximum (which may correspond to a transition state) are well defined. However, the path connecting them (reaction coordinate) is not well defined, in the sense that there are many possible paths. Liken this to climbing a mountain. The starting and ending points are well defined as is the summit, but there can be many possible routes.

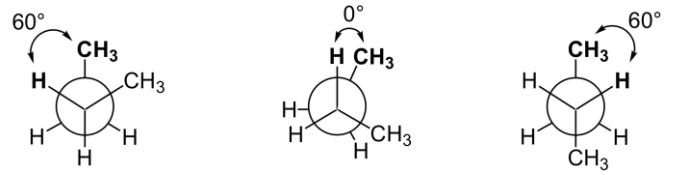


The reaction coordinate for some processes may be quite simple. For example, where the "reaction" is rotation about the carbon-carbon bond in ethane, the reaction coordinate may be thought of as the HCCH torsion angle, and the structure may be thought of in terms of this angle alone. Thus, staggered ethane (both the reactant and the product) is a molecule for which this angle is 60° and eclipsed ethane is a molecule for which this angle is 0°.



staggered ethane "reactant" | eclipsed ethane "transition state" | staggered ethane "product"

A similar description applies to "reaction" of *gauche n*-butane leading to the more stable *anti* conformer. Again, the reaction coordinate may be thought of as a torsion about the central carbon-carbon bond, and the individual reactant, transition state and product structures in terms of this coordinate.



gauche n-butane "reactant" | "transition state" | *anti n*-butane "product"

Equilibrium structure (geometry) may be determined from experiment, given that the molecule can be prepared and is sufficiently long-lived to be subject to measurement. On the other hand, the geometry of a transition state may not be experimentally established. This is simply because a transition state is not an energy well which can trap molecules. Therefore, it is impossible to establish a population of molecules on which measurements may be performed.

Both equilibrium and transition-state structures may be determined from quantum chemical calculations. The fact that a molecule may not be stable enough to be detected and characterized (or even exist) is not important. It would seem from our discussion that equilibrium and transition-state structures can be distinguished from one another simply by inspecting the shape of the potential energy surface in the vicinity of the structure. Of course, such a surface cannot actually be visualized for a system with more than one or at most two degrees of freedom. However, the set of frequencies associated with the vibrational motions around the structure, the same quantities measured by infrared spectroscopy, will all be real numbers for stable molecules (energy minima), whereas there will be one (and only one) vibrational frequency which is an imaginary number for a transition state. This does not guarantee that this is the transition state *of interest*, but if it is, the coordinate (vibrational motion) associated with it is the reaction coordinate. Further discussion is provided in the topic ***Calculating Infrared Spectra***.

THEORETICAL MODELS

A variety of different procedures based on quantum mechanics (so- called quantum chemical models) have been developed to calculate molecular structure and properties as well as infrared, NMR and UV/visible spectra. All follow from a deceptively simple looking ***Schrödinger equation*** first written down in 1927.

$$\hat{H}\,\Psi = \varepsilon\Psi$$

$\hat{H}$ (the ***Hamiltonian*** or more precisely ***Hamiltonian operator***) is the only known. It describes the kinetic energies of the particles that make up a molecule and the Coulombic interactions between the individual particles. Positively-charged nuclei repel other nuclei, and negatively-charged electrons repel other electrons, but nuclei
attract electrons. Ψ (the ***wave function***) is a function of the Cartesian coordinates, and ε (the ***energy***) is a number. The goal in solving the Schrödinger equation is to find a function that when operated on by
the Hamiltonian yields the same function multiplied by a number. Note that there are many (actually an infinite number of) solutions to the Schrödinger equation. These correspond to the ground and numerous excited states of an atomic or molecular system.

The energy of a molecule can be measured. On the other hand, the wave function has no physical meaning, and is not subject to experimental measurement, although the square of the wave function times a small volume element gives the probability of finding an electron inside that volume. This exactly corresponds to what is actually measured in an X-ray diffraction experiment.

The Schrödinger equation has been solved exactly for the hydrogen atom (a one-electron system), where the wave functions are familiar to chemists as the s, p, d, ... atomic orbitals. The lowest-energy s orbital corresponds to the ground state of the hydrogen atom, whereas the higher-energy solutions correspond to excited states.

s orbital *p orbital* *d orbital*

The Schrödinger equation may easily be written down for many- electron atoms as well as for molecules, although it cannot be solved. Even something as seemingly simple as the helium atom with only two-electrons presents an insurmountable problem. Approximations must be made.

Hartree-Fock Molecular Orbital Models

Hartree-Fock molecular orbital models or ***molecular orbital models***, as they are commonly referred to, were the first practical quantum chemical models to be formulated. They result from making three approximations to the Schrödinger equation:

1. Separate nuclear and electron motions. The ***Born-Oppenheimer approximation*** says "from the point of view of the electrons" the nuclei are stationary. This eliminates nuclear motion and leads to an ***electronic Schrödinger equation*** which can be solved for the H_2^+ molecule but cannot be solved for molecules with more than one-electron. The Born-Oppenheimer approximation is of little consequence for the description of molecular properties, for example, equilibrium geometries and reaction energies, and may be used without concern.

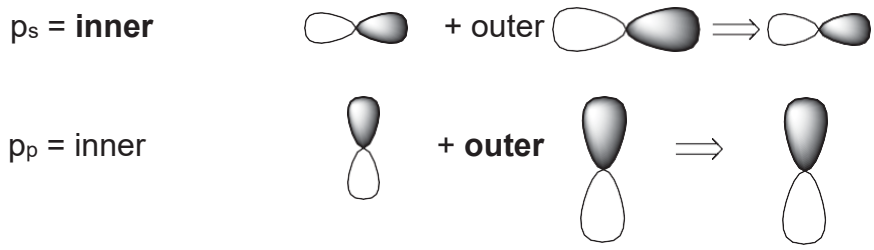
2. Separate electron motions. The ***Hartree-Fock approximation*** eliminates the need of having to simultaneously account for the motions of several electrons. It leads to a much simpler set of equations in which the motion of each electron in an environment made up of the nuclei and all the other electrons is sought.

3. Represent each one-electron solution or ***molecular orbital*** by a linear combination of atom-centered functions or ***atomic orbitals***. The ***LCAO*** (***L***inear ***C***ombinations of ***A***tomic ***O***rbitals) ***approximation*** reduces the problem of finding the best functional form for the molecular orbitals to the much simpler problem of finding the best set of linear coefficients. As the number and complexity of the atomic orbitals increases, the energy and other properties approach limiting values. However, computational cost also increases. The goal is to provide as few functions as possible to yield a value for the property of interest that adequately reflects its limit. Note, that the limiting values of properties are not expected to be the same as experimental values, but rather reflect the behavior of the Hartree-Fock model.

Basis Sets

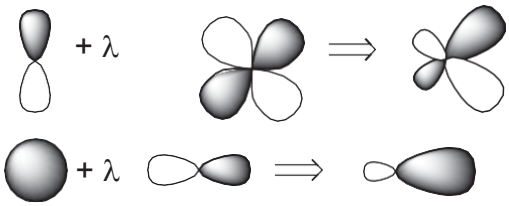
Gaussian functions are polynomials in the Cartesian coordinates times an exponential in the ***square*** of the distance from the origin. They are distinct although very closely related to the exact solutions of the hydrogen atom (exponential functions in the distance), and are labeled 1s, 2s, 2p, ... , the same nomenclature used to describe hydrogen atom solutions.

A ***minimal basis set*** includes only sufficient functions to hold all the electrons on an atom and to maintain its spherical shape. This involves a single 1s orbital for each hydrogen atom, and a set of five orbitals (1s, 2s, $2p_X$, $2p_Y$ and $2p_Z$) for each carbon atom. Because a minimal basis set incorporates only one set valence p functions, the components of which are the same size, atoms in nearly spherical environments will be better described than atoms in aspherical environments. A ***split-valence basis set*** addresses this problem by providing two different sets of valence p functions, one compact set and one loose set. This allows different linear combinations for different directions. For example, the compact p orbital can be emphasized to construct a σ bond while the loose p orbital can be emphasized to construct a π bond.



Because the functions in a minimal or split valence basis set are centered on the atoms, they may have difficulty describing electron distributions that fall in between atoms (that is, bonds). ***Polarization basis sets*** address this problem by providing a set of d-type functions (***polarization functions***) on main-group elements, and (optionally) a set of p-type functions on hydrogen. The resulting combinations can be thought of as hybrid orbitals, for example, the pd and sp hybrids shown below.

set of p-type functions on hydrogen. The resulting combinations can be thought of as hybrid orbitals, for example, the pd and sp hybrids shown below.



The so-called 6-31G* basis set will be used for the infrared and NMR calculations described in future topics. The number "6" to the left of the "-" in the name indicates that 6 functions are used to describe each inner-shell (core) atomic orbital. The numbers, "31" to the right of the "-" indicate that groups of 3 and 1 functions are used to describe each valence-shell atomic orbital. "*" designates that polarization functions are supplied for non-hydrogen atoms. Were two stars to be present (as in 6-31G**) this would indicate that p-type polarization functions would also be placed on hydrogen atoms.

The valence basis functions can be further split and additional polarization functions can be added including f-type functions. A commonly used basis set is designated 6-311+G(2df,2p). "311" indicates a triply-split valence, "2df" indicates that two sets of d-type functions and a set of f-type functions are added to the valence of heavy atoms, and "2p" indicates that two sets of p-type functions are added to the valence of hydrogen atoms.

Taken together, these three approximations lead to a set of equations known as the ***Roothaan-Hall equations***. They increase in computational cost as the cube of the size (number of basis functions) and can easily be applied to molecules incorporating up to 100 heavy (non- hydrogen) atoms.

Beyond Hartree-Fock Models

Of the three approximations we have made to reach Hartree-Fock models, the second is to be taken most seriously. According to the Hartree-Fock approximation, electrons "move independently", which means that both the electron-electron repulsion energy and the total energy will be too large. The limiting Hartree-Fock energy is, therefore, necessarily higher (less negative) than the experimental energy. ***Electron correlation*** is the term given to describe the coupling or correlation of electron motions. The ***correlation energy*** is defined as the difference between the Hartree-Fock energy and the experimental energy and is necessarily a negative quantity.

There are two conceptually different approaches for calculating the correlation energy, and numerous specific models arising from each of these approaches. Wave function-based models start from the Hartree-Fock wave function combining it with wave functions resulting from electron excitations from filled to empty molecular orbitals. Density functional models supplement the Hartree-Fock Hamiltonian.

Configuration Interaction Models

Configuration interaction models are archetypal of wave function based correlated models. In the unachievable limit, so-called ***full configuration interaction***, all possible single and multiple electron promotions from occupied to unoccupied Hartree-Fock molecular orbitals and assuming a complete basis set, the energy is same as would be achieved by solution of the electronic Schrödinger equation. More practical ***limited configuration interaction*** schemes have been formulated by limiting the number of electron simultaneously promoted (1-electron, 2-electron, ...) and the number of filled and unfilled molecular orbitals involved in the promotions.

Møller Plesset Models

An alternative and more commonly used wave function-based model is ***Møller-Plesset theory***. This assumes that the Hartree-Fock energy E_0 and wave function Ψ_0 are solutions to an equation involving a Hamiltonian, $\hat{H}_0$, that is very close to the exact Schrödinger Hamiltonian, $\hat{H}$. This being the case, $\hat{H}$ can be written as a sum of $\hat{H}_0$ and a small correction, V. λ is a dimensionless parameter.

$$\hat{H} = \hat{H}_0 + \lambda\hat{V}$$

Expanding the exact energy in terms of a power series of the Hartree-Fock energy yields:

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \lambda^3 E^{(3)} + \ldots$$

Substituting this expansion into the Schrödinger equation and collecting terms in powers of λ leads to an explicit expression for the energy correction. The sum of $E^{(0)}$ and $E^{(1)}$ is the Hartree-Fock energy. Including the next term gives rise to the so-called ***MP2*** (second-order Møller-Plesset) ***model***.

Both the simplest configuration interaction (limited to single and double electron excitations only) and the MP2 models increase in computational cost as the power of the total number of basis 5th functions. In practice, this limits them to molecules incorporating 25-50 heavy (non-hydrogen) atoms at most.

Density functional theory or simply **DFT**, is based on two theorems elaborated by Hohenberg and Kohn, which taken together, prove that the energy and other properties of a many-electron system in its ground state may be correctly and uniquely described in terms of a function of the electron density. The term "functional" or a function of a function arises because the electron density is itself a function of the three spatial coordinates. What the two Hohenberg-Kohn theorems imply is that the Schrödinger equation can actually be solved; that is, the completely "intractable" problem involving the coupled motions of n electrons in a static field due to the nuclei (a "molecule") may be replaced by an eminently "solvable" problem that treats the electrons as independent (that is, non-interacting) particles. Because the electron density is a function of only three coordinates, in effect a 3 dimensional problem is substituted for a 3n dimensional problem. Unfortunately, the Hohenberg-Kohn theorems do not tell us anything about the functional itself.

In the density functional formalism, the electronic energy, E^{el}, is written as a sum of the kinetic energy, E_T, the electron-nuclear interaction energy, E_v, the Coulomb energy, E_J, and a term combining the exchange and correlation energies, E_{XC}.

$$E^{el} = E_T + E_V + E_J + E_{XC}$$

What is "the" exchange/correlation functional? The quest has gone on for several decades and hundreds of functionals have actually been proposed and divided into several distinct classes.

Local Density Approximation (LDA)

The functional first proposed stems from a purely hypothetical problem in which a uniform gas of non-interacting electrons moves in a positively charged field. An analytical solution for the exchange energy is available and takes the form of the density raised to the 4/3's power. The correlation energy may be arrived at through numerical simulation and is also only dependent on only the local density at each point.

$$E_{XC} = E(\rho)$$

Functionals that depend only on the electron density (ρ) are referred to as LDA functionals. They are not very successful in describing the properties of molecules and have largely been replaced.

Generalized Gradient Approximation (GGA)

GGA functionals depend on the gradient of the electron density ($\nabla\rho$) in addition to the density itself.

$$E_{XC} = E(\rho, \nabla\rho)$$

Originating in the mid 1980's, they were the first approaches to provide a reasonable account of molecular properties and in particular the energies of chemical reactions. In this sense, they were instrumental in drawing attention to density functional theory as a viable "low-cost" alternative to wave function based correlation techniques, something that LDA functionals had failed to do. The BLYP functional is representative.

Global Hybrid Generalized Gradient Approximation (GH-GGA)

GH-GGA functionals often referred to as hybrid functionals) replace a fixed fraction of the exchange by the "exact" Hartree-Fock exchange, the fraction being an adjustable parameter. It is likely that it was the introduction of hybrid functionals that caused the community to recognize that density functional theory was "semi-empirical" in nature. Adding the Hartree-Fock exchange is "costly" but led to significant improvements in the description of reaction energies. While GH-GGA functionals were introduced in the early 1990's, they remain a mainstay in the application of density functional theory to chemistry. The B3LYP functional, in particular, is perhaps still more widely used than any other functional, even though there are now much better choices.

Range Separated Hybrid Generalized Gradient Approximation (RSH-GGA)

The idea behind range separated GGA hybrid functionals is that the "optimum" amount of Hartree-Fock exchange varies with electron- electron distance, from a small percentage in the long range limit to a large percentage in the short range limit, in the extreme from 0% to100%. Both ωB97X-D and ωB97X-V functionals are range- separated GGA hybrids. Both incorporate so-called local corrections to account for dispersive interactions (see discussion following).

***meta* Generalized Gradient Approximation (mGGA)**

A *meta* GGA functional not only depends on the electron density and its gradient (as does a GGA functional) but also on the Laplacian of the electron density. As such, it can be construed as the next logical step beyond GGA in constructing a Taylor series expansion of the electron density.

$$E = E(\rho, \nabla\rho, \nabla^2\rho)$$

More commonly, *meta* GGA functionals are viewed as adding the so-called kinetic energy density to GGA.

$$E = E(\rho, \nabla\rho, \nabla^2\psi)$$

In either case, the addition can be construed as second-order term in a Taylor series expansion of the electron density. The B97M-V functional is an example of a "pure" *meta* functional.

Global Hybrid *meta* Generalized Gradient Approximation (GH-mGGA)

These are strictly akin to global hybrid GGA (GH-GGA) functionals in that a fixed percentage of the Hartree-Fock exchange is introduced. The "only" difference is that a *meta* GGA functional including a second-order term replaces a GGA functional. The M06-2X functional, widely accepted as an excellent choice for thermochemical comparisons, is an example of a GH-*meta* GGA functional.

Range Separated Hybrid meta Generalized Gradient Approximation (RSH-mGGA)

A range-separated *meta* GGA (RSH-mGGA) functional is identical to a range-separated GGA (RSH-GGA) functional except that a *meta* GGA functional has replaced the underlying GGA functional. M11 and ωB97M-V are examples of range-separated *meta* GGA functionals.

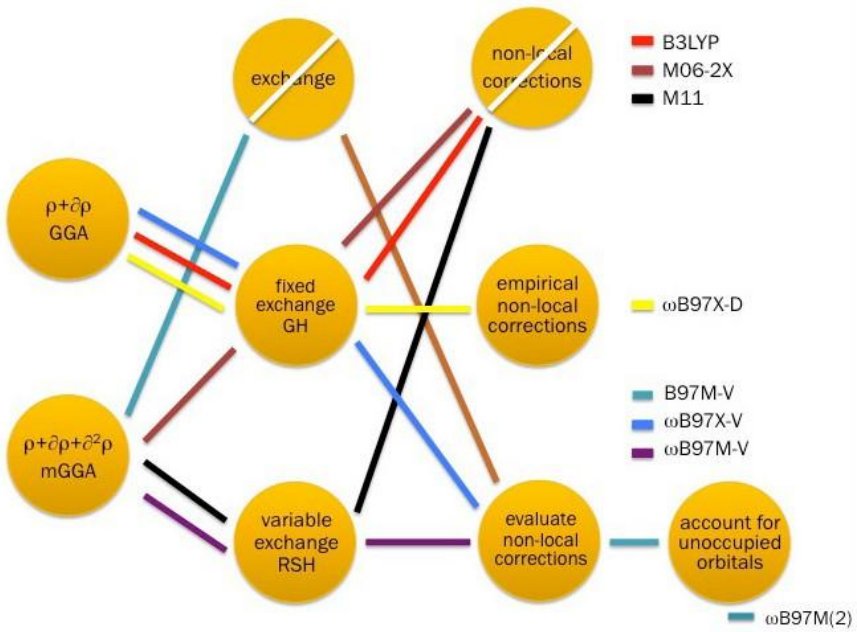
Double Hybrid *meta* Generalized Gradient Approximation (DH-RSH-mGGA)

A double hybrid *meta* GGA (DH-RSH-mGGA) functional accounts for contributions of unoccupied molecular orbitals as well as occupied orbitals, by way of a MP2 like approach. The ωB97M(2) functional is now supported for energy calculations only.

Non Local Corrections

The previous functional classes are all considered "local" in that they are described in terms of a single integral over the three spatial coordinates. In order to capture dispersive van der Waals interactions, so-called "nonlocal" correlation functionals are needed. These involve a double integral over two sets of coordinates; these may add significant cost to the calculations. For example, the range-separated GGA hybrid ωB97X-V functional used throughout this text, incorporates theVV10 non-local correlation functional and is 3-5 times more costly than the parent ωB97X functional. An alternative and less costly way to account for dispersive interactions is to add an empirical correction to the functional. So-called Grimme corrections are designated by appending "-D" as in ωB97X-D, or "D3" as in B3LYP-D3, to the
end of the functional.

A graphical summary of a number of commonly-used functionals stemming from different classes is provided in the figure below.



Which classes of functionals are "best" and which functionals within each class are "best"? A broad base of experience allows some general remarks to be made.

Numerical Integration

Unlike both Hartree-Fock and wave function based correlated models, density functional models cannot be evaluated wholly analytically. Some components require numerical integration, which introduces another variable into the mix, specifically the form and size of the integration grid. An oversimplified description places the grid points along a set number of angular directions and at a set number of distances from the origin. This is referred to as a ***Lebedev grid***.

"Cost" of Density Functional Models

The computation "cost" of density functional models depends on the class, the number of basis functions, η, and the number of points in the numerical integration grid, κ, where κ >> η. GGA and pure meta-GGA functionals such as B97M-V that do not require the Hartree-Fock exchange formally scale as O(η2κ). The other functionals combine this dependence with the cost of the Hartree-Fock exchange, which formally scales as O(η4) but in practice is O(η3) or lower. Finally, functionals such as ωB97X-V that directly account for dispersion have a step that scales as O(η2κ2). This typically dominates the calculation. Times (in minutes) for calculation of the energy together with its gradient ("one step" in the optimization of molecular geometry) for morphine (C17H19NO3) with the B3LYP (GH-GGA), ωB97X-D (RSH-GGA), ωB97X-V (RSH-GGA + dispersion), B97M-V (mGGA), M06-2X (GH-mGGA) and M11 and ωB97M-V (RSH-mGGA) functionals, as well as with RI-MP2, which formally scales as O(η5), with the 6-31G* basis set are provided in **Table**

A-1. These have been obtained using a single core of a 3.3 GHz Intel i7 5820K processor. **Table A-2**provides times for 6-311+G(2df,2p) calculation, a dual basis set 6-311+G(2df,2p) calculation and a dual basis set cc-pVQZ-g (cc-pVQZ minus the set of g functions) energy calculation, relative to the corresponding 6-31G* energy/gradient calculations. For example, the value of 26 for the dual basis set cc-pVQZ–g ωB97X-D energy calculation given in **Table A-2**, means that this calculation is 26 times more costly than the ωB97X-D energy/gradient calculation. Put another way, a structure optimization requiring 26 cycles would cost the same as an energy obtained using the larger basis set.

The manner of reporting is deliberate. The 6-31G* basis set is usually deemed satisfactory for calculation of equilibrium geometry, whereas larger basis sets (for example, 6-311+G(2df,2p) and cc-pVQZ-g) are known to be required for accurate descriptions of the energies of most chemical reactions. Determination of geometry for molecules of the complexity of morphine typically requires upwards of 10-20 energy/ gradient cycles, that is, an order of magnitude (or more) greater than the time reported in the first column of the table. Both parts must be considered in order to correctly access cost and practicality.

Table A-1: Times for energy and gradient calculations on morphine with B3LYP, ωB97X-D, ωB97X-V, B97M-V, M06-2X, M11 and ωB97M-V density functional models and RI-MP2 model with the 6-31G* basis set (need ωB97M-V times)

	6-31G* + gradient
B3LYP	7
ωB97X-D	15
ωB97X-V	41
B97M-V	50
M06-2X	19
M11	22
RI-MP2	10

Table A-2: Times for B3LYP, ωB97X-D, ωB97X-V, B97M-V, M06-2X, M11 and ωB97M-V density functional and RI-MP2 energy calculations with 6-311+G(2df,2p), dual 6-311+G(2df,2p) and dual cc-pVQZ-g basis sets relative to the corresponding times for energy and gradient calculations with the 6-31G* basis set give in Table A-1.

	6-311G(2df,2p)	6-311+G(2df,2p) [6-311G*]	cc-pVQZ-g [rcc-pVQZ]
B3LYP	26	5	48
ωB97X-D	14	3	26
ωB97X-V	6	2	10
B97M-V	5	2	6
M06-2X	13	3	21
M11	11	3	20
RI-MP2	18	3	33

Several conclusions may be drawn:

1. Times for calculation of the energy together with its gradient vary by close to an order of magnitude among the seven models. The B3LYP model is the least costly and the ωB97X-V, B97M-V and ωB97M-V models most costly. The latter result suggests that calculation of the gradient of dispersion functional common to both models is to blame. In terms of cost, the ωB97X-D model is closest to B3LYP. The surprising result is perhaps the good performance of the RI-MP2 model, surpassed only by B3LYP.

2. 6-311+G(2df,2p) energy calculations are roughly an order of magnitude more costly than the corresponding energy/gradient calculations. At the high end and low ends is the factor of 26 for the B3LYP model and a factor of 5 for the B97M-V model. Structure optimization typically requires upwards of 20 steps (energy/gradient) and will likely dominate the overall task.

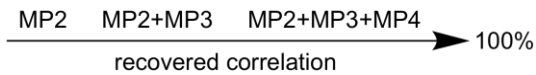
3. The dual basis set approximation applied to 6-311+G(2df,2p) basis set leads to significant cost savings. If this is the chosen energy method, the costly step will most certainly be determination of geometry.

4. Except for the ωB97X-V and B97M-V functionals, dual basis set cc-pVQZ-g calculations are roughly an order of magnitude more costly than the corresponding 6-311+G(2df,2p) calculations. This means that geometry and energy steps are likely to contribute equally to overall cost. The much smaller increase in cost for B97M-V is presumably due to its lack of the Hartree-Fock exchange.

5. The RI-MP2 model is clearly competitive, greater in overall cost only to the B3LYP models. Because of its inherent O(η5) scaling, RI-MP2 will eventually fall significantly behind density functional models in performance as molecular size is increased.

Disadvantages of Density Functional Models

Low computation cost relative to wave function based electron correlation models is the principal attraction of density functional models. Aside from the RI-MP2 model, density functional models are the only procedures that are both reliable and routinely applicable to molecules of moderate size. However, density functional models come with some disadvantages not shared by wave function based procedures. By far the most serious of these is that they do not offer a clear pathway to improvement. Whereas wave function based procedures return a higher and higher percentage of the correlation energy with increasing degree of complexity, for example, in successively adding orders in the Møller-Plesset models,

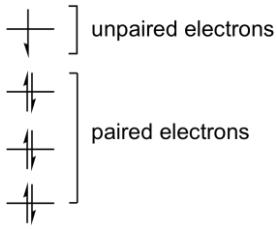


there is no guarantee that increasing the flexibility (and complexity) of a particular functional will actually lead to improvement. It is clear that present generation functionals are “better” than those of previous generations, for example, ωB97X-D and M06-2X are significantly better than B3LYP, in particular for calculation of reaction energies. What is less clear is whether the improvement is to large extent due to careful parameterization to high quality reference data (and to an increased number of parameters) rather than from insight into the “proper” functional form. Functional selection, even while guided by clear physical models, certainly has an uncertain component, and at least at present it is not possible to say from first principles which of several functionals is likely to provide the “best” results. This can only be done by way of thorough comparisons with “known” data.

In practice, density functional models increase in computational cost as the cube of the total number of basis functions (the same dependence seen for Hartree-Fock models). Because most functionals require calculation of the Hartree-Fock exchange energy, they are necessarily more costly than Hartree-Fock models but can easily be applied to molecules incorporating up to 100 heavy (non-hydrogen) atoms.

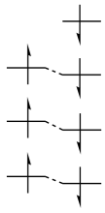
Models for Open-Shell Molecules

Thus far, discussion has been limited to molecules with closed- shell electron configurations, that is, with all electrons being paired. This covers the vast majority of organic molecules as well as most organometallic molecules. There are two ways to think about molecules with unpaired electrons. The obvious way is to insist that electrons are either paired or are unpaired.



This is referred to as ***restricted*** and the individual models as restricted models, for example, the restricted Hartree-Fock (or RHF) model.

The restricted procedure does not necessarily yield the lowest possible energy, simply because it forces the "paired" electron into the same spatial orbitals. Removing this constrains provides greater flexibility and generally lead to a lower energy. This is termed ***unrestricted*** and the resulting models are termed unrestricted models, for example, the unrestricted Hartree-Fock (or UHF) model.
Aside from yielding lower energy, unrestricted models are generally less costly than restricted models and are much more widely used and is the default procedure in ***Spartan***.



Semi-Empirical Molecular Orbital Models

The principal disadvantage of Hartree-Fock, density functional and MP2 models is their computational cost. It is possible to introduce further approximations together with empirical parameters in order to significantly reduce cost while still retaining the underlying quantum mechanical formalism. So-called ***semi-empirical molecular orbital models*** follow in a straightforward way from Hartree-Fock models:

1. Insist that basis functions on different atoms do not overlap ("see each other"). The so-called ***NDDO approximation*** is rather drastic but reduces the computation effort by more than an order of magnitude over Hartree-Fock models.
2. Restrict to a ***minimal valence basis set*** of atomic functions. This means that there are no inner-shell (core) functions in the basis set. As a consequence, the cost of doing a calculation involving a second-row element such as silicon, is no more than that incurred for the corresponding first-row element such as carbon.
3. Introduce adjustable parameters to reproduce specific experimental data. This is what distinguishes the various semi-empirical models currently available. Choice of parameters, more than anything else, appears to be the key to formulating successful semi-empirical models.

Molecular Mechanics Models

The alternative to quantum chemical models are molecular mechanics models. These do not start from the Schrödinger equation, but rather from a simple but chemically reasonable picture of molecular structure, a so-called ***force field***. In this picture, just as with a Lewis structure, molecules are made up of atoms (as opposed to nuclei and electrons), some of which are connected (bonded). Both crowding (van der Waals) and charge-charge (Coulombic) interactions between atoms are then considered, and atom positions are adjusted to best match known structural data (bond lengths and angles).

Molecular mechanics is much simpler than attempts at solving an "approximate" Schrödinger equation, but requires an explicit description of chemical bonding, as well as a large amount of information about the structures of molecules. This biases results and seriously limits the predictive value of molecular mechanics models. Nevertheless, molecular mechanics has found an important role in molecular modeling as a tool to establish equilibrium geometries of proteins and other large molecules. It has also been widely used to establish the preferred conformation of molecules with multiple degrees of conformational freedom and hundreds or even thousands of accessible conformers. With regard to the latter, comparisons with results of high-level quantum chemical calculations have, however, clearly shown that present-generation molecular mechanics models are often not satisfactory in identifying the "best" (lowest-energy) conformer and are not to be trusted for obtaining Boltzmann weighed averages. However, they have proven to be useful for identifying highly unfavorable conformers allowing some shortening of the initial list of possible structures.

Choosing a Theoretical Model

No single method of calculation is likely to be ideal for all applications. A great deal of effort has been expended to define the limits of different molecular mechanics and quantum chemical models, and to judge the degree of success of different models. The latter follows from the ability of a model to consistently reproduce known (experimental) data. Molecular

mechanics models are restricted to determination of geometries and conformations of stable molecules. Quantum chemical models also provide energy data, which may in turn be directly compared with experimental thermochemical data, as well as infrared, Raman, UV/visible and NMR spectra and properties such as dipole moments, which may be compared directly with the corresponding experimental quantities. Quantum chemical models may also be applied to transition states. While there are no experimental structures with which to compare (see the topic ***Potential Energy Surfaces***), experimental kinetic data may be interpreted to provide information about activation energies (see the topic ***Total Energies and Thermodynamic and Kinetic Data***).

Success is not an absolute. Different properties, and certainly different problems may require different levels of confidence to be placed in the calculation to actually be of value. Neither is success sufficient. A model also needs to be practical for the task at hand. Were this not the case, there would be no reason to look further than the Schrödinger equation itself. Models that may be practical for small to medium size organic molecules cannot be expected to be applied to proteins. Models that are successful and practical for organic molecules may not necessarily meet either criterion for inorganic molecules or transition-metal organometallics. Not only does the nature and size of the system needs to be considered, with due attention to the available computational resources and the experience (and patience) of the practitioner. Specifics aside, practical models usually share one common feature in that they are not likely to be the best possible treatments which have been formulated. Compromise is almost always an essential component of model selection. Continued advances in both digital computers and computer software will continue to raise the bar, but it will be some time before fully reliable models will be routinely applicable to all chemical systems of interest.

The MMFF molecular mechanics model generally provides a satisfactory description of equilibrium geometry where, at least for organic molecules . It has also proven to be suitable for removing high-energy conformers for molecules with multiple degrees of freedom in preface to quantum chemical calculations of conformer energy differences.

Semi-empirical model are appropriate for:

i) Equilibrium geometry determinations for large molecules.
ii) Transition-state geometry determinations.
iii) Equilibrium and transition-state geometry determinations involving transition metals.

Semi-empirical models are unsuitable for:

i) Calculation of reaction energies.
ii) Calculation of conformer energy differences.

Small basis set Hartree-Fock models such as HF/3-21G and HF/6-31G* are appropriate for:

i) Equilibrium and transition-state structure determinations of organic molecules.
ii) Calculation of energies of *isodesmic* reactions, including comparisons of regio and stereoisomers.

They are unsuitable for:

i) Calculation of reaction energies that involve bond making or breaking and calculation of absolute activation energies.

ii) Comparison of isomer energies of molecules with different bond types.

iii) Equilibrium and transition-state structure determinations for transition-metal organometallic molecules.

iv) Calculation of conformer energy differences.

Small basis set density functional models such as B3LYP/6-31G* and ωB97X-D/6-31G* are appropriate for:

i) Equilibrium and transition-state structure determinations for organic molecules as well as molecules incorporating transition metals.

ii) Calculation of all types of reaction energies, although some caution is needed.

iii) Calculation of conformer energy differences, although caution is needed.

The RI-MP2/6-31G* model is appropriate for:

i) Equilibrium and transition-state structure determinations except for molecules including transition metals.

ii) Calculation of all types of reaction energies, although some caution is needed.

It is unsuitable for:

i) Equilibrium and transition-state structure determinations for transition-metal organometallic molecules.

ii) Calculation of reaction and activation energies where transition-metals are involved.

The ωB97X-V/6-311+G(2df,2p) model is appropriate for calculating all types of reaction energies. Implemented together with a dual basis set, it provides a practical and reliable means to establish conformer energy differences.

FINDING AND VERIFYING EQUILIBRIUM AND TRANSITION-STATE GEOMETRIES

The energy of a molecule depends on its geometry. Even small changes in structure can lead to quite large changes in total energy. What is the "best" choice of geometry for use in a molecular modeling study? While experimental structures, where available, would at first glance seem to be ideal, there are multiple problems with this. First, while upwards of one million structures have been established experimentally, many, many more have not been. Second, the vast majority of experimental structures follow from X-ray crystallography on solid samples and may differ significantly from those of isolated molecules to which the calculations pertain. An additional problem with X-ray structures is that bonds to hydrogen are too short (by as much as 0.1 - 0.2Å). Finally, experimental data for reactive or otherwise short-lived molecules are scarce, and data for transition states are completely lacking. In the final analysis, there is no alternative to obtaining geometries from calculation. Fortunately, this is not difficult, although it may be demanding in terms of computer time.

Determination of geometry (geometry optimization) is an iterative process. The energy and energy gradient (first derivative of the energy with respect to all geometrical coordinates) are calculated for the initial geometry, and this information is then used to project a new geometry. This process continues until three criteria are satisfied. First, the gradient must closely approach zero. This ensures that the optimization is terminating in a flat region of the potential surface (either the bottom of an energy well in the case of equilibrium geometry or the top of an energy hill in the case of transition-state geometry). Second, successive iterations must not change any geometrical parameter by more than specified (small) value. Third, successive iterations must not change the total energy by more than a specified (small) value.

Equilibrium Geometries

In order for a geometry to correspond to an energy minimum, the curvature of the energy surface must be positive, that is, the structure must lie at the bottom of an energy well. The surface's curvature is defined by the ***Hessian*** (the matrix of second derivatives of the energy with respect to geometrical coordinates).

> What is actually done is to transform from the original coordinates to a new set of geometrical coordinates (***normal coordinates***) for which the Hessian will be diagonal, that is, all off-diagonal elements will be zero. In this representation, all (diagonal) elements must be positive for the geometry to correspond to an energy minimum. Normal coordinate analysis, as it is termed, is required for the calculation of vibrational frequencies, which relate directly to the square root of the elements of the (diagonal) Hessian. Positive Hessian elements yield real frequencies; negative Hessian elements yield imaginary frequencies.

Geometry optimization does not guarantee that the final structure has a lower energy than

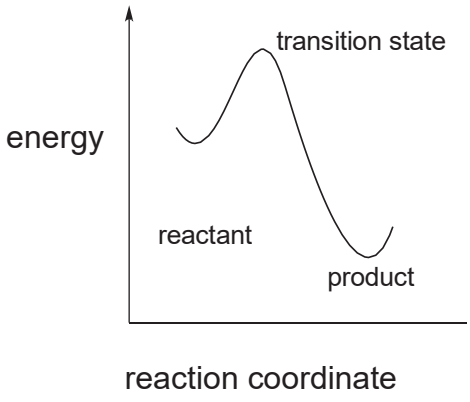
any other structure of the same molecular formula. All that it guarantees is a ***local minimum***, that is, a geometry with a lower energy than that of any similar geometry, although not necessarily the lowest energy geometry possible for the molecule. The number and types of chemical bonds are maintained in the optimization process as are single bond and flexible-ring conformers. Altering bond types would lead to isomers (stable molecules of the same molecular formula) which under normal conditions could not be reached, whereas altering conformation would lead to different "shapes" of the same molecule, which would be in equilibrium under the same conditions. Finding the best conformer or ***global minimum*** requires repeated optimization starting with different initial geometries corresponding to different initial conformers. Only when all local minima have been located is it possible to say with certainty that the lowest energy geometry has been identified. This process is termed ***conformational analysis***.

In principle, geometry optimization carried out in the absence of symmetry, that is, with C_1 symmetry, must result in a local minimum. On the other hand, imposition of symmetry may result in a geometry that is not a local minimum. For example, optimization of ammonia constrained to a planar trigonal geometry (D_{3h} symmetry) will result in a structure that corresponds to an energy maximum in the direction of motion toward a puckered trigonal geometry (C_{3v} symmetry). This is the transition state for inversion at nitrogen in ammonia.

The most conservative tactic is always to optimize geometry in the absence of symmetry. If this is not done, it is always possible to verify that the structure located indeed corresponds to a local minimum by calculating the vibrational frequencies on the final (optimized) structure. If one or more frequencies are imaginary, then the geometry does not correspond to an energy minimum.

Transition-State Geometries

Chemists recognize a transition state as the structure that lies at the top of a potential energy surface connecting reactant and product (see the topic ***Potential Energy Surfaces***).



More precisely, a transition state is a point on the potential energy surface for which the gradient is zero (just as it is for an equilibrium geometry), but for which the diagonal representation of the Hessian has one and only one negative element, corresponding to the reaction coordinate (see diagram above). All the other elements are positive. In other words, a transition state is a structure that is an energy minimum in all dimensions except one, for which it is an energy maximum. Mathematically, such a structure is referred to as a first-order saddle point.

The geometries of transition states on the pathway between reactants and products are not as easily anticipated as the equilibrium geometries of the reactants and products themselves. This is not to say that transition-state geometries do not exhibit the same systematic behavior as equilibrium geometries, but rather that there is not sufficient experience to identify what systematics do exist, and more importantly how to capitalize on structural similarities. It needs to be recognized that transition states cannot even be detected let alone characterized experimentally, at least not directly. While measured activation energies relate to the energies of transition states above reactants, and while activation entropies and activation volumes as well as kinetic isotope effects may be interpreted in terms of transition- state structure, no experiment can actually provide direct information about the detailed geometries and/or other physical properties of transition states. Quite simply, transition states do not exist in terms of a stable population of molecules on which experimental measurements may be made. Experimental activation parameters may act as a guide, although here too it needs to be pointed out that their interpretation is in terms of ***transition state theory***. This assumes that all molecules proceed over a single transition state (the high point along the reaction coordinate) on their way to products. Even then, experiments tell little about what actually transpires in going from reactants to products.

Lack of experience about "what transition states look like" is one reason why their detailed geometries are more difficult to obtain than equilibrium geometries. Other reasons include:

i) Algorithms for locating transition states are less well developed than procedures for finding equilibrium structures. After all, minimization is an important task in many diverse fields of science and technology, whereas saddle point location has few if any important applications outside of chemistry.

ii) It is likely that the potential energy surface in the vicinity of a transition state is "flatter" than the surface in the vicinity of a local minimum. After all, transition states represent a delicate balance of bond breaking and bond making, whereas overall bonding is maximized in equilibrium structures. As a consequence, the potential energy surface in the vicinity of a transition state is likely to be less well described in terms of a simple quadratic function (assumed in all common optimization procedures) than the surface in the vicinity of a local minimum.

iii) To the extent that transition states incorporate partially (or completely) broken bonds, it might be anticipated that very simple theoretical models lacking adequate treatment of electron correlation will not be able to provide entirely satisfactory descriptions.

In time, all of these problems will be overcome, and finding transition states will be as routine as finding equilibrium geometries is today. Chemists can look forward to the day when reliable tools become available for the elucidation of reaction mechanisms.

While the same iterative procedure previously described for optimization of equilibrium geometry applies as well to transition states, the number of steps required for satisfactory completion is likely to be somewhat larger. This is due to the factors discussed earlier. Note, however, that the task of transition state determination may be completely automated and needs no more human intervention than that involved in locating equilibrium geometries.

Having found a transition-state geometry, two tests need to be performed in order to verify that it actually corresponds to a proper transition state, and further that it actually corresponds to the transition state for the process of interest, that is, it smoothly connects energy minima corresponding to reactant and product:

i) Verify that the Hessian (matrix of second energy derivatives) yields one and only one imaginary frequency. This requires that a normal mode analysis be carried out on the proposed transition-state geometry. The imaginary frequency will typically be in the range of 400-2000 cm^{-1}, quite similar to real vibrational frequencies. In the case of flexible rotors, for example, methyl groups or floppy rings, the analysis may yield one or more additional imaginary frequencies with very small (<100 cm^{-1}) values. These can be difficult to "get rid of" simply because energy changes are likely to be very small and inside the precision settings of the optimization procedure. Small imaginary frequencies can almost always be ignored but make certain to verify what motions these small imaginary frequencies actually correspond to (see discussion following) before doing so. Most important, be wary of structures that yield only very small imaginary frequencies. This suggests a very low energy transition state, which quite likely will not correspond to the particular reaction of interest.

ii) Verify that the normal coordinate corresponding to the imaginary frequency smoothly connects reactants and products. A simple way to do this, which requires no additional calculations, is to animate the normal coordinate corresponding to the imaginary frequency, that is, to walk along this coordinate without any additional optimization. This does not necessarily require any additional calculations beyond the normal mode analysis already performed. "Incorrect" transition states located by calculation, that is, transition states that do not link the reactant to the expected product, may indicate new chemistry, so don't discard them too quickly! There are more costly procedures that actually involve "walking" the geometry down from the transition state to both reactants and products. In our view these are rarely worth the effort.

Reactions Without Transition States

Not all chemical reactions have transition states, and that the rates of some reactions depend only on the speed with which reactants diffuse into one another (so-called, ***diffusion controlled reactions***). In fact, reactions without energy barriers are quite common. Two radicals will typically (but not always*) combine without activation, for example, two methyl radicals to form ethane.

$$H_3C^{\bullet} + {}^{\bullet}CH_3 \longrightarrow H_3C{-}CH_3$$

Radicals will often add to paired-electron species with no (or very small) activation, for example, methyl radical and ethylene forming 1-propyl radical.

$$H_3C^{\bullet} + H_2C{=}CH_2 \longrightarrow H_3C{-}CH_2{-}CH_2^{\bullet}$$

Exothermic ion-molecule reactions that have activation energies in solution, do not necessarily have activation energies in the gas phase. Any complex of an ion and a neutral molecule is likely to be lower in energy than the separated species and the entire reaction coordinate for an ion-molecule reaction might lie below the energy of the separated reactants for example, nucleophilic attack by OH^- on CH_3Cl to give CH_3OH and Cl^-.

Failure to find a transition state, and location instead of what appears to be a stable intermediate or even the final product, does not necessarily mean failure of the theoretical model (nor does it rule this out). It may simply mean that there is no transition state!

Calculations Using Approximate Geometries

Given that small-basis set Hartree-Fock models, semi-empirical models and even molecular mechanics models generally provide geometries for organic molecules that are quite close to those obtained from Hartree-Fock, density functional and MP2 models, it is legitimate to ask whether or not structures from these techniques may be used as the basis for energy and property calculations.** It would be of great help were this the case as geometry optimization is a major cost in any modeling investigation.

"Exact" geometries must be used for frequency (infrared and Raman spectra) calculations. The reason for this is that frequencies are related to the first finite term in a Taylor series expansion of the energy (as a function of geometry). This is (assumed to be) the second-derivative term, which will not be true if the first-derivative term (the gradient) is not precisely zero. (Cubic and higher order terms are assumed to be small and are ignored.) Frequencies evaluated at non-equilibrium (or non-transition-state) geometries are meaningless.

---

* Exceptions may occur where both radicals are delocalized, for example, combing two benzyl radicals.

** Molecular mechanics models are not applicable to transition states as they have been parameterized to account for the structures of stable molecules. This is not to say that molecular mechanics parameterizations could not be developed for transition states, simply that they have not been.

TOTAL ENERGIES AND THERMODYNAMIC AND KINETIC DATA

In addition to molecular geometry, energy is certainly the most important quantity to come out of molecular modeling. Energy can be used to reveal which of several isomers is most stable, to determine whether a particular chemical reaction will have a thermodynamic driving force (an *exothermic* reaction) or be thermodynamically uphill (an *endothermic* reaction), and to ascertain how fast a reaction is likely to proceed. Other molecular properties, such as dipole moment, and infrared, Raman, UV/visible and most importantly NMR spectra are also of great interest, but energy plays a special role.

There is more than one way to express the energy of a molecule. Most common to chemists is the heat of formation, ΔH_f. This is the heat of a hypothetical chemical reaction that creates a molecule from well defined (but arbitrary) standard states of each of its constituent elements. Note that the heat of formation, which most commonly assumes a value between -1,000 and +500 kJ/mol, cannot be directly measured, but must be obtained indirectly.

An alternative, total energy, is the heat of a hypothetical reaction that creates a molecule from a collection of separated nuclei and electrons. Like the heat of formation, total energy cannot be measured directly and is used solely to provide a standard method for expressing and comparing energies. Total energies are always negative numbers and are much larger than the sum of the bond energies By convention, they are expressed in so-called atomic units* or au, but may be converted to other units as desired:

$$1 \text{ au} = 2625 \text{ kJ/mol}$$

It makes no difference which reference reaction (heat of formation or total energy) is used to calculate the thermochemistry of a balanced chemical reaction (reactant 1 + reactant 2 + . . . $\longrightarrow$ product 1 + product 2 + . . .):

$$\Delta E \text{ (reaction)} = E_{\text{product 1}} + E_{\text{product 2}} + \ldots - E_{\text{reactant 1}} - E_{\text{reactant 2}} - \ldots$$

Total energies will be used in the discussion that follows. A negative ΔE indicates an *exothermic* (thermodynamically favorable) reaction, while a positive ΔE an *endothermic* (thermodynamically unfavorable) reaction.

A special case involves energy differences among isomers. Each comparison may be viewed as a chemical reaction involving only two molecules:

$$\Delta E\text{(isomer)} = E_{\text{isomer 2}} - E_{\text{isomer 1}}$$

* The exact energy of hydrogen atom is -0.5 atomic units.

A negative ΔE means that isomer 2 is more stable than isomer 1. Total energies may also be used to calculate activation energies, $\Delta E^{\ddagger}$:

$$\Delta E^{\ddagger} = E_{\text{transition state}} - E_{\text{reactant 1}} - E_{\text{reactant 2}} - \cdots$$

Here, $E_{\text{transition state}}$ is the total energy of the transition state, and $E_{\text{reactant1}}$, $E_{\text{reactant2}}$, ... are the total energies of the reactants. Activation energies are expected to be positive numbers*, meaning that the transition state is less stable that reactants.

Reaction and activation energies are sufficient to know whether a reaction is *exothermic* or *endothermic* and whether it proceeds with small or large activation barrier. There are, however, situations where energies need to be replaced by Gibbs energies in order to take proper account contributions due to entropy.** For example, a proper account of the equilibrium concentrations of reactants and products requires calculation of the equilibrium constant, K_{eq}, which according to the Boltzmann equation, is related to the Gibbs energy of reaction, ΔG_{rxn}:

$$K_{eq} = \exp(-\Delta G_{rxn}/RT)$$

Here R is the gas constant, and T is the temperature (in K). At room temperature (298K) and for ΔG_{rxn} in au, this is given by:

$$K_{eq} = \exp(-1060\Delta G_{rxn})$$

ΔG_{rxn} has two components, the enthalpy of reaction, ΔH_{rxn}, and the entropy of reaction, ΔS_{rxn}. These are defined as follows:

$$\Delta G_{rxn} = \Delta H_{rxn} - T\Delta S_{rxn}$$

$$\Delta H_{rxn} \approx \Delta E_{rxn} = E_{\text{product 1}} + E_{\text{product 2}} + \cdots - E_{\text{reactant 1}} - E_{\text{reactant 2}} - \cdots$$

$$\Delta S_{rxn} = S_{\text{product 1}} + S_{\text{product 2}} + \cdots - S_{\text{reactant 1}} - S_{\text{reactant 2}} - \cdots$$

Although ΔG_{rxn} depends on both enthalpy and entropy, it is often assumed that the entropy contribution will be small and can be neglected. Further assuming that $\Delta H_{rxn} \approx \Delta E_{rxn}$, equilibrium constants can then be estimated according to the "Boltzmann" equation:

---

* Referred to separated reactants, "negative activation energies" are possible. This is due to the formation of a complex, the total energy of which is lower than the sum of the total energies of the reactants.

** Entropy contributions will be largest where the numbers of reactants and products differ. On the other hand, entropy might be expected to largely cancel where reactants and products are very similar.

$$K_{eq} \approx \exp(-\Delta E_{rxn}/RT) \approx \exp(-1060\ \Delta E_{rxn})$$

This equation may also be used to establish the equilibrium composition of a mixture of isomers:

$$\text{Isomer 1} \rightleftharpoons \text{Isomer 2} \rightleftharpoons \text{Isomer 3} \rightleftharpoons \ldots$$

$$\%\ \text{Isomer i} = \frac{100 \exp(-1060\ E_{\text{Isomer i}})}{\sum_k \exp(-1060\ E_{\text{Isomer k}})}$$

Isomer energies, E_{isomer}, are given in atomic units relative to the energy of the lowest-energy isomer. An important special case involves an equilibrium between two isomers:

$$\text{Isomer 1} \rightleftharpoons \text{Isomer 2}$$

$$\frac{[\text{ Isomer 1 }]}{[\text{ Isomer 2 }]} = \exp[-1060\ (E_{isomer1} - E_{isomer2})]$$

Reaction rate constants, k_{rxn}, are also related to Gibbs energies. As before, if entropy contributions can be neglected, the rate constant can be obtained directly from the activation energy, $\Delta E^{\ddagger}$, according to the Arrhenius equation:

$$k_{rxn} \approx (k_B T/h)[\exp(-\Delta E^{\ddagger}/RT)]$$

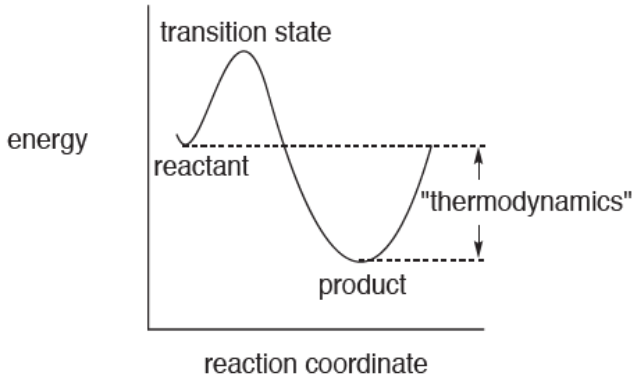
Here k_B and h are the Boltzmann and Planck constants, respectively. At room temperature and for $\Delta E^{\ddagger}$ in au, k_{rxn} is given by:

$$k_{rxn} = 6.2\text{x}10^{12} \exp(-1060 \Delta E^{\ddagger})$$

Another way to describe reaction rates is by half-life, $t_{1/2}$, the amount of time it takes for the reactant concentration to drop to one half of its original value. When the reaction follows a first-order rate law, rate = $-k_{rxn}$[reactant], $t_{1/2}$ is given by:

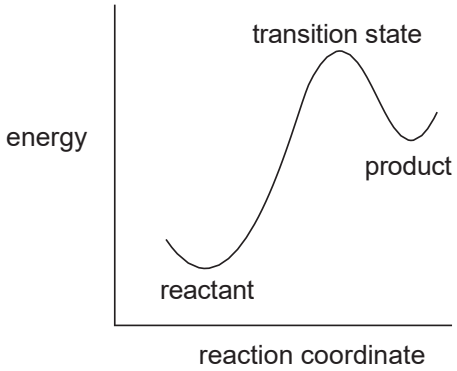
$$t_{1/2} = \ln 2/k_{rxn} = 0.69/k_{rxn}$$

It is useful to associate reaction energies and reaction rates with potential energy diagrams introduced earlier. The thermodynamics of reaction is given by the relative energies of the reactant and product on the potential surface.

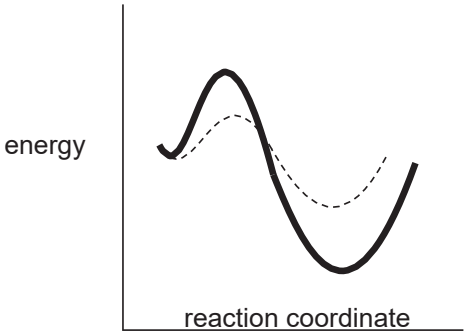


For situations such as bond rotation in ethane and ring inversion in cyclohexane, the reactant and product are the same, and the reaction is said to be ***thermoneutral***. The most common case, as depicted in the above diagram, is where the energy of the products is lower than that of the reactants. The reaction is said to be *exothermic*, and the difference in stabilities of reactant and product is simply the difference in their energies. For example, the "reaction" of *gauche n*-butane to *anti n*-butane is *exothermic*, and the difference in stabilities of the two conformers is simply the difference in the energies (~3.8 kJ/mol).

Chemical reactions can also be *endothermic*, which give rise to a reaction profile.

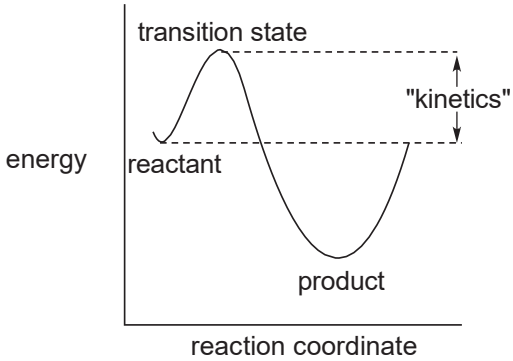


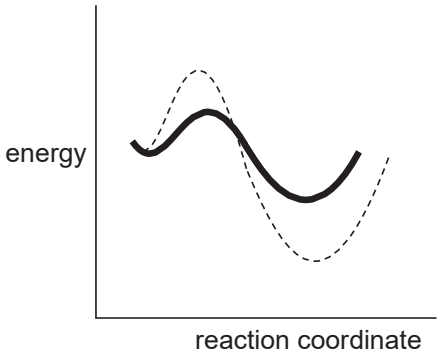
Where two or more different products may form in a reaction, thermodynamics tells us that if we wait long enough, the product formed in greatest abundance will be that with the lowest energy irrespective of pathway.



In this case, the product is referred to as the ***thermodynamic product*** and the reaction is said to be ***thermodynamically controlled***.

The energy of the transition state above the reactants (the activation energy) provides the connection with reaction rate (kinetics). Of course, such an interpretation is based on the notion that all reactions actually pass through a single transition state.





Aside from the obvious dependence on the concentration of reactants, absolute reaction rates also depend on the percentage that encounters between molecules will actually lead to reaction (the so-called pre- exponential or A factor). The hope is that pre-exponential factors for closely-related reactions will be very similar and therefore can be ignored. Thus, it is not at all evident that activation energies for disparate reactions will parallel measured reaction rates.

The product formed in greatest amount in a kinetically controlled reaction (the kinetic product) is that proceeding via the lowest energy transition state, irrespective of whether or not this is lowest energy product (the thermodynamic product).

Kinetic product ratios show dependence with activation energy differences which are identical to thermodynamic product ratios with differences in reactant and product energies.

THERMOCHEMICAL RECIPES AND CALCULATING ACCURATE HEATS OF FORMATION

The "energy" of a molecule is most commonly reported as a heat of formation. As noted in the previous topic, this is defined as the enthalpy at 298.15K of a hypothetical chemical reaction in which the molecule is transformed into a set of products that correspond to the most stable forms of its constituent pure elements at room temperature. For example, the heat of formation of ethylene corresponds to the enthalpy of a reaction to yield graphite and molecular hydrogen.

$$C_2H_4 \rightarrow 2C\ (\text{graphite}) + 2H_2$$

Differences in heats of formation between the products and reactants (reaction enthalpies) indicate the extent to which the reaction will be favorable (*exothermic*) or unfavorable (*endothermic*) and allow thermodynamic product distributions to be established.*

In almost all cases, the heat of formation is obtained from a heat of combustion. For example, the heat of formation of ethylene would likely have been established from its reaction with oxygen to produce carbon dioxide and water.

$$C_2H_4 + 3O_2 \rightarrow 2CO_2 + 2H_2O$$

Experimental heats of formation are available for ~2000 compounds. While much of the data is accurate to within 4-8 kJ/mol, a significant portion is subject to greater uncertainty. The most egregious source of error is that the reported heat actually does not correspond to the reported structure. More common sources of error include impure samples, incomplete combustion and most importantly, poorly characterized combustion products. Hydrocarbons and oxycarbons present fewest problems as combustion leads only to carbon dioxide and water. However, combustion of molecules with other elements may give rise to a complex mixture of products and greater uncertainty, with nitrogen compounds being particularly problematic.

Despite their fundamental importance, heats of formation are not routinely determined for new compounds. While the combustion experiment is straightforward and does not require particularly expensive instrumentation, accurate measurements may require (and will likely destroy) significant quantities of compound. Very few chemists are willing to part with hundreds of mg of a compound that they have just spent days, weeks or months preparing.

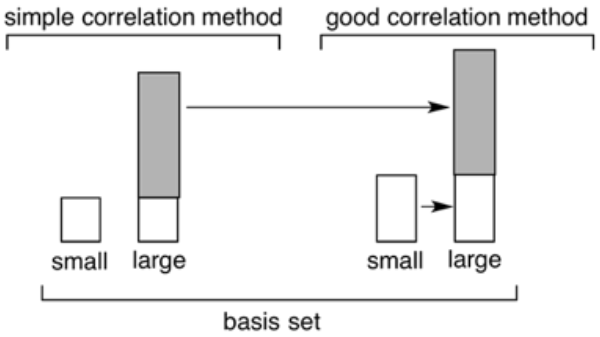
---

* Of course, other references are equally suitable for examining mass balanced reactions. For example, total energies from quantum chemical calculations are referenced to separated nuclei and electrons, in the case of ethylene.

$$C_2H_4 \rightarrow 2C^{6+} + 4H^{+} + 16e^{-}$$

Because of the lack of experimental data and serious concerns over the accuracy of part of the data that do exist, considerable attention has been directed at the use of quantum chemical calculations to obtain heats of formation. One avenue that has been pursued relies on the knowledge that CCSD (T) calculations (coupled cluster singles and doubles with triples corrections introduced perturbatively) reliably reproduce the energetics of a wide variety of organic reactions. The problem is that the singles and doubles contribution scale as $O(\rho\iota^6)$, where $\rho\iota$ is the number of basis functions, and the triples correction scales as $O(\rho\iota^7)$. This limits practical applications of the method to very small molecules (<15 non-hydrogen atoms).

A simplified account of the "solution" first proposed by John Pople is depicted in the next page. The essential idea is that the energy obtained using a "good" electron correlation method and a large basis set may be closely approximated by combining the energy calculated using the good correlation method with a small basis set with difference in calculated energies for a "simple" correlation method with small and large basis sets. In principle, only three calculations are required: small and large basis set calculations with the simple correlation method and a small basis set calculation with the good correlation method. In practice, some of the schemes that have been formulated do exactly this while others employ a combination of simple correlation methods.



The original Pople implementation and subsequent developments primarily by Larry Curtiss are now collectively known as Gx methods. G1 relied on MP4 as the "good" correlation method with a gradual shift to the more costly and more accurate CCSD(T) method. To some extent, the overall objective has also broadened, from a means to approximate correlation energy to a stand-alone source of thermochemical data (298° heats of formation). Our focus is on the former, as possible low-cost replacements of the CCSD(T) correlation method.

The two most widely-used methods are G3(MP2) and G3, and the two most recent additions to the series are G4(MP2) and G4. A summary of the procedures involved in calculating electronic energy using these four methods is provided on the next page. The full recipes lead to heat of formation in addition to electronic energy and require specifying an equilibrium geometry and a procedure for obtaining vibrational frequencies in addition to

“empirical” corrections to account for the standard states of each of the elements. While each the four procedures involves several steps, these can be carried out in proper sequence without user intervention.

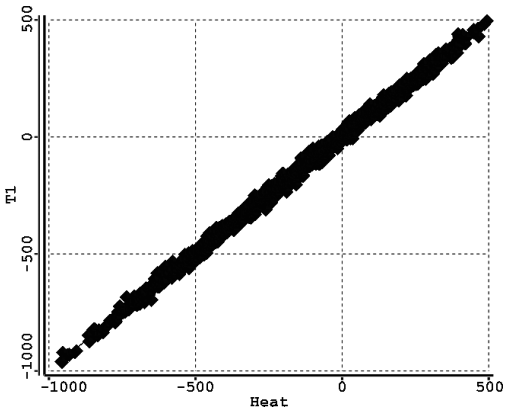
	simple correlation method	good correlation method	limiting steps
G3(MP2)	MP2	QCISD(T)/6-31G*	T in QCISD(T)
G3	MP2 & MP4	QCISD(T)/6-31G*	MP4 and T in QCISD(T)
G4(MP2)	MP2	CCSD(T)/6-31G*	T in CCSD(T)
G4	MP2 & MP4	CCSD(T)/6-31G*	MP4 and T in CCSD(T)

The T1 Thermochemical Recipe

The simplest of the GX recipes, termed G3(MP2), involves several costly steps, most significantly an MP2/6-31G* geometry calculation, a HF/6-31G* frequency calculation and a QCISD(T)/6-31G* energy calculation. In practice, G3(MP2) scales as the 7th power of size and is applicable only to molecules with molecular weights less than 150 amu. Clearly, even simpler procedures are required for routine application to larger molecules.

The goal behind the T1 recipe was to maintain the accuracy of G3(MP2) but at significantly reduced computation cost. It is limited to uncharged, closed-shell molecules comprising H, C, N, O, F, Si, P, S, Cl and Br. T1 substitutes the MP2/6-31G* geometry used in G3(MP2) by a HF/6-31G* geometry, eliminates both the HF/6-31G* frequency and the QCISD(T)/6-31G* energy calculations and approximates the MP2 energy calculation with the G3MP2 large basis set by an analogous calculation using a dual basis set RI-MP2 model. Taken together, these changes reduce computation time by 2-3 orders of magnitude, and T1 calculations on molecules in the molecular weight range of 400-500 amu are practical.

The T1 recipe, unlike G3(MP2), involves parameters, specifically atom counts, Mulliken bond orders and HF/6-31G* and RI-MP2 energies. These have been determined using linear regression as a best fit to G3(MP2) (not experimental) heats of formation for >1100 small molecules. It reproduces these values with mean absolute and RMS errors of 1.8 and 2.5 kJ/mol, respectively. More important, the T1 recipe reproduces experimental heats of formation for a set of >1800 diverse organic molecules from the NIST thermochemical database with mean absolute and RMS errors of 8.5 and 11.5 kJ/ mol, respectively. The plot provided below covers the data from -1000 to +500 kJ/mol.



Heats of formation from the T1 recipe are included as a property in the Spartan Spectra and Properties Database (SSPD), a database that presently comprises more than 300k entries.

DEALING with CONFORMATIONALLY FLEXIBLE MOLECULES

Very few molecules are rigid and adequately described in terms of a single conformer. More commonly, two or more distinct conformers resulting from rotation around single bonds or changes in flexible rings are likely to exist. For example and as previously discussed in the topic ***Potential Energy Surfaces***, three conformers result from rotation around the central carbon-carbon bond in *n*-butane, one *anti* conformer (CCCC dihedral angle = 180°) and two (equivalent) *gauche* conformers (CCCC dihedral angle = ± ~60°). Two chair conformers result from ring inversion in methylcyclohexane, one with the methyl group *equatorial* and the other with it *axial*. Because energy barriers to single-bond rotation and to ring inversion are generally very small (on the order of <10-20 kJ/mol), the different conformers will be in equilibrium and the influence of any particular conformer on the properties of the molecule will depend on its energy and on the temperature. As a consequence, experimental measurements will either reveal the presence of the individual conformers or result from a weighted average of conformers. For example, the IR spectrum of a flexible molecule will show features due to individual conformers, whereas the NMR spectrum of the same molecule will show an average. The different behavior may be traced to the different timescales of the two experiments; IR is fast whereas NMR is slow.

Identifying the lowest-energy conformer of a molecule with multiple degrees of conformational freedom presents a serious challenge for calculation. Even more so is providing accurate Boltzmann weights allowing proper description of average energies, properties and perhaps most important, NMR spectra. Until recently, molecular mechanics provided the only practical alternative for molecules with hundreds or even thousands of accessible conformers. Expectations were high for molecular mechanics force fields such as MMFF which were after all developed and parameterized to account for conformer energetics, albeit for very small molecules with a single degree of conformational freedom for which experimental data was available.*

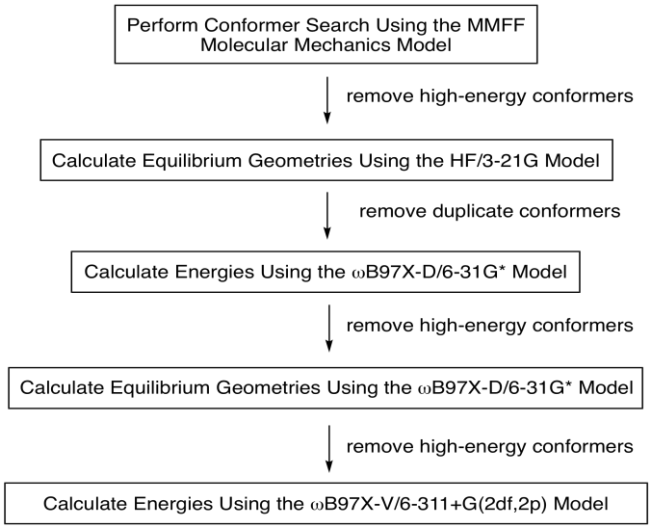
However, recent comparisons with conformer energies obtained from high-level quantum chemical calculations for (very few) molecules with multiple degrees of freedom show that MMFF actually provides a poor account. Furthermore, it has been shown that neither semi-empirical nor Hartree-Fock models provides a suitable account and that either density functional or MP2 models with moderate to large basis sets are required.

---

* While much is known about the preferred conformation of molecules with multiple degrees of conformational freedom in the solid (crystalline) state, almost exclusively from X-ray diffraction, there is very little known about preferred conformation of isolated molecules. Furthermore, there is very little overlap between gas and solid-phase structures because small molecules tend not to crystallize and larger molecules that do crystallize are not easily investigated by gas-phase techniques such as microwave spectroscopy. This means that it is difficult to separate inherent (gas-phase) conformational preferences from those associated with the requirements of crystal packing.

Direct application of correlated models is not presently practical or at least routine for molecules where hundreds or more conformers need to be examined. What is practical are multi-step procedures starting from an extensive molecular mechanics search and ending with an appropriate model for energy calculation. The idea is that intermediate steps with successively better models will gradually reduce the number of conformers that need to be considered (passed on to the next step). Several present recipes (available from menus) are supported in ***Spartan***. The simplest (least costly and likely least reliable) use either B3LYP/6-31G* or ωB97X-D in the final step and MMFF equilibrium geometries. Improvement (and increased computation cost) follows by using B3LYP/6-31G* or ωB97X-D geometries. The most reliable (and most costly) models use B97M-V/6-311+G(2df,2p)[6-311G*] or ωB97X-V/6-311+G(2df,2p)[6-311G*] for final energy (Boltzmann weight) calculations together with B3LYP/6-31G* or ωB97X-D/6-31G* geometries, respectively. Whatever the recipe employed, the resulting Boltzmann weights are available to determine average quantities.

One of the menu-selected multi-step protocols is outlined below.



Step 1: Conformer Search Using the MMFF Molecular Mechanics Model

The first and least costly step in the protocol involves a systematic conformational search using the MMFF molecular mechanics model, keeping all distinct conformers with energies less than 40 kJ/mol than that of the lowest-energy conformer found. The energy threshold is needed to control the number of conformers passed on to the next step, recognizing that a molecule with 8 degrees of conformational freedom may give rise to several thousand distinct conformers and one with 10 degrees of freedom to several tens of thousands of conformers. In addition, a limit on the total number of conformers needs to be imposed. If exceeded, a subset of the most "geometrically diverse" set is kept. This number is typically in the range of 200-1000 conformers.

Step 2: Calculate Conformer Geometries Using the HF/3-21G Model

Obtain the equilibrium geometry for each of the conformers using the Hartree-Fock method with the 3-21G basis set (HF/3-21G or simply 3-21G). Eliminate any conformers that are outside a 40 kJ/mol threshold as well as duplicate conformers. This step is necessary because MMFF commonly results in structures that are not energy minima, and further optimization using the HF/3-21G model allows these to be removed. Note, however, that this model is not able to reliably identify the "best" conformer let alone provide accurate Boltzmann weights, which is why the 40 kJ/mol energy threshold needs to be maintained.

Step 3: Calculate Conformer energies Using the ωB97X-D/6-31G* Model

Use the ωB97X-D/6-31G* model to calculate energies for each of the conformers, keeping only those that are within 15 kJ/mol of the lowest energy conformer. The much lower energy threshold reflects experience that, even with a small basis set, the ωB97X-D model provides conformer energy differences that are much more reliable than those from other MMFF or HF/3-21G models.

Step 4: Calculate Conformer Geometries Using the ωB97X-D/6-31G* Model

Recalculate equilibrium geometries using the ωB97X-D/6-31G* model, keeping only conformers that are within 10 kJ/mol of the lowest energy conformer. The slight lowering of the energy threshold anticipates that better equilibrium geometries will lead to more reliable conformer energy differences.

Step 5: Calculate Conformer energies Using the ωB97X-V/ 6-311+G(2df,2p)[6-311G*] Model

The final step in the protocol is to replace ωB97X-D/6-31G* conformer energies with those obtained from the ωB97X-V/6-311+G(2df,2p) model. In practice, the so-called dual basis set approximation is employed, reducing the cost of this step by as much as an order of magnitude.

In terms of computation, the last two steps in the protocol outlined above are the costliest: calculation of equilibrium geometries with the ωB97X-D/6-31G* model and calculation of energies with the ωB97X- V/6-311+G(2df,2p)[6-311G*] model.

A less costly (factor of two) protocol has also been examined and found to produce comparable albeit less reliable results. This replaces ωB97X-D/ 6-31G* by B3LYP/6-31G* and ωB97X-V/6-6311+G(2df,2p)[6-311G**] by B97M-V/6-311+G(2df,2p)[6-311G*].

Where only the identity of the lowest-energy conformer is required, ***Spartan*** employs the so-called ***Monte Carlo*** procedure. It moves randomly in a single (randomly chosen) dimension in conformational space deciding to abandon the move (returning to the starting point for another random move) or keep the move (using it for the starting point for the next random move) based on the energy relative to that of the lowest-energy conformer yet found. In practice, a Monte-Carlo search will nearly always find the lowest-energy conformer by considering only 5-10% of the possible number of conformers.

A Monte-Carlo search is probably not an appropriate starting point where the objective is to establish a Boltzmann distribution. The reason is that it rapidly finds the "best" region, at the expense of other regions, which, given the known deficiencies of molecular mechanics,

may exclude the actual important contributors (or even the actual most important contributor). The only alternative is a ***systematic search*** which "looks everywhere". However, this rapidly becomes impractical as the number of conformers increases very rapidly with the number of rotatable single bonds and flexible rings. The solution employed in ***Spartan*** is a so-called ***sparse systematic search*** where systematic steps are accepted (and energies evaluated) or rejected based solely on a random number criterion. For example, if at the outset only 10% of the possible conformers can be examined, then only those generating a random number between say 0 and 0.1 (out of a possible range of 0 to 1, will be accepted.

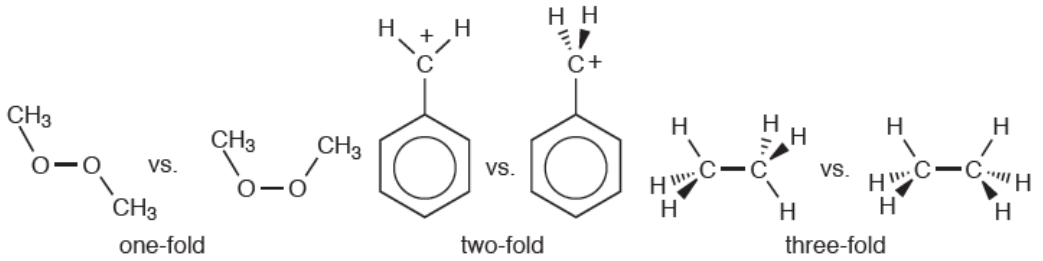
INTERPRETING CONFORMATIONAL PREFERENCES

Rotation about single bonds is periodic, retracing itself every 360°, and any function that seeks to describe the energy of internal rotation must also repeat itself every 360°. In practice, the energy function, $E^{torsion}$, is written as a combination of simpler functions, V_n, in the torsion angle Xo, each of which repeats n times in a 360° interval.

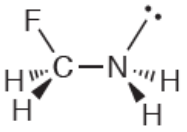
$$E^{torsion}(Xo) = V_1(Xo) + V_2(Xo) + V_3(Xo)...$$

The first three terms, V_1, V_2, and V_3 repeat every 360°, 180°, and 120°, respectively, and are referred to as one-fold, two-fold, and three-fold potentials.*

Separation into distinct n-fold potentials is a useful concept because each can be associated with a particular chemical phenomenon. For example, a one-fold potential describes the different energies of *anti* (COOC=180°) and *syn* (COOC=0°) conformers of dimethylperoxide, while a two-fold potential describes the different energies of planar and perpendicular conformers of benzyl cation. Three-fold potentials, which are more familiar to chemists, describe the difference between staggered and eclipsed conformers in molecules like ethane.

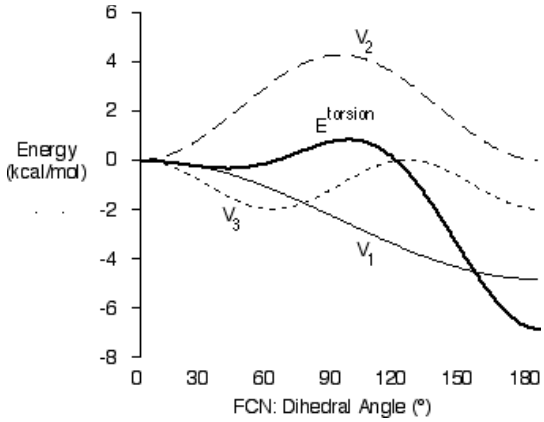


While rotation in some molecules might be adequately described using only one of the components or a combination of two components, other molecules require more complex combination. This is illustrated by fluoromethylamine.

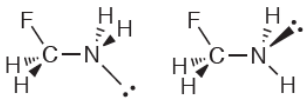


FCN: = 0°

* The torsional potential in most organic molecules will be adequately described in terms of one, two and three-fold components, and only rarely will higher-order terms be needed.



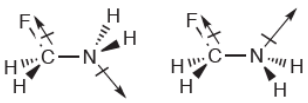
The heavy solid line describes $E^{torsion}$ for rotation about the CN bond, while the light solid line, the dashed line and the dotted line correspond to the one-fold, two-fold and three-fold components, respectively. There are two distinct minima.



FCN: = 180° FCN: ≈ 45°

The lower (global) minimum arises when the CF bond and the nitrogen lone pair are *anti*, while the higher and much more shallow minimum is close to a *gauche* structure, FCN: dihedral angle of ~45°. Also, note that one of the two energy maxima is close to an eclipsed structure, with the FCCN: dihedral angle of ~115°.

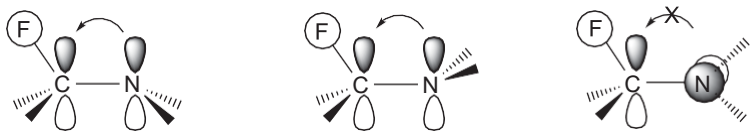
The behavior of $E^{torsion}$ becomes clear when it is resolved into its components. The one-fold term reflects a clear and very strong preference for the CF bond and the nitrogen lone pair to be *anti* and not *syn*. This preference might be electrostatic since the *anti* structure arranges the dipoles associated with the CF bond and nitrogen lone pair in opposite directions.



FCN: = 180° dipoles subtract FCN: = 0° dipoles add

The three-fold term reflects the preference for staggered over eclipsed structures and contributes less to the variation in $E^{torsion}$ than either of the one-fold or two-fold terms.

What is most interesting, and perhaps not have been easily anticipated without this type of analysis, is the large contribution made by the two-fold potential, which reflects a strong preference for a planar arrangement of FCN:. This can be attributed to stabilization resulting from donation of the lone pair orbital on nitrogen into low-energy unfilled molecular orbital associated with the CF bond, requiring that the molecule adopts either a *syn* planar or *anti* planar conformation and not a perpendicular conformation.



syn-planar *anti*-planar *perpendicular*

Note that the terms that contribute to $E^{torsion}$ are completely independent of each other, and each may be treated as one part of a larger picture. Thus, the observation that electron donation from the nitrogen lone pair into the empty orbital associated with the CF bond is optimal when the two groups are planar is independent of the observation that *cis* coplanar structure is destabilized, relative to the *anti*-structure, by dipole-dipole interactions.

CALCULATING INFRARED SPECTRA

The infrared spectrum of a molecule arises because of transitions between vibrational energy levels. Each line in an infrared spectrum is characterized by a frequency (energy) and an intensity. In one dimension (a diatomic molecule), it is common practice to assume that the frequency is proportional to the square root of the ratio of the second-derivative of the energy with respect to the internuclear distance, r, and the reduced mass (the product of the masses of the two atoms divided by their sum).

$$\text{vibrational frequency} \propto \sqrt{[(d^2E(r)/dr^2/)\text{reduced mass}]}$$

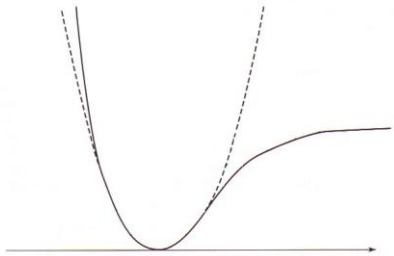
This is referred to as the ***harmonic approximation***, the origin of which can be seen by expanding the energy in a Taylor series.

$$E(r) = E(r_0) + (dE(r)/dr)\, r + (d^2E(r)/dr^2)\, r^2 + \text{higher-order terms}$$

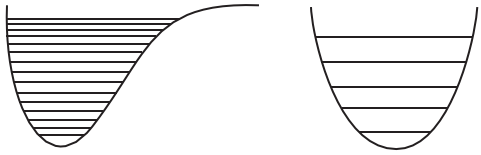
$E(r_0)$ is a constant and $dE(r)/dr$ (the gradient) is assumed to be zero. The latter implies that the underlying structure corresponds precisely to a minimum (or a maximum) on the potential energy curve. Were this not the case, the first derivative would be non-zero and the calculated frequency would be meaningless. Nearly all practical calculations ignore cubic and higher-order terms (***anharmonic*** terms), leaving only the second derivative term (the ***force constant***). This said, the frequency may be interpreted as the relative ease or difficulty of stretching the bond away from its equilibrium position, that is, the curvature of the energy surface at the minimum. Where distortion away from the equilibrium position is easy, the result is a low frequency; where it is difficult the result is a high frequency. High (reduced) mass leads to a low frequency while low mass leads to high frequency.

The expression for vibrational frequency qualitatively accounts for mass effects on reaction energies (***equilibrium isotope effects***). Even at 0K molecules vibrate, giving rise to the so-called zero-point vibrational energy (or simply ***zero-point energy***). Zero-point energy is directly proportional to the sum of the vibrational frequencies and decreases with increasing mass. For example, the zero-point energy of HCl is lowered upon replacement of hydrogen by deuterium. Thus, the measured energy (enthalpy decreases with increasing mass and the energy of DCl is smaller (more negative) than that for HCl.

Because the potential energy has been approximated by a quadratic function, calculated frequencies will almost always be larger than measured frequencies. This is because a quadratic function goes to infinity with increase in distance rather than going asymptotically to a constant (separated atoms), meaning that the potential curve will be too steep.



It is possible to extract the harmonic frequency from an experimental spectrum by measuring the spacing of the energy levels associated with the ground and excited states of a particular vibration. (The lines would be evenly spaced were the potential quadratic.) However, such an analysis is impractical for any but diatomic and very simple polyatomic molecules.



Generalization from a diatomic to a polyatomic molecule is straightforward. The energy of displacement away from the equilibrium position is expanded in the same way as before, the only difference being that a vector quantity, **x**, replaces a scalar quantity, x.

$$E(\mathbf{x}) = E(\mathbf{x_0}) + \Sigma_i(\partial E(\mathbf{x})/\partial x_i)x_i + \tfrac{1}{2}\Sigma_{ij}(\partial^2 E(\mathbf{x})/\partial x_i \partial x_j)x_i x_j + \text{higher-order terms}$$

As with the expression for a diatomic molecule, the leading term is a constant, the first derivative term is zero and cubic and higher-order terms are ignored. For a molecule with N atoms, the dimension of **x** is 3N (x,y,z Cartesian coordinates for each atom), although there are only 3N-6

(3N-5 for a linear molecule) vibrational frequencies. Six dimensions (five for a linear molecule) correspond to translation away from and rotation around the center of mass.

The first (and only computationally expensive) step involved in calculating the vibrational spectrum of a polyatomic molecule is evaluation of the full set of second energy derivatives in Cartesian coordinates. These then need to be mass weighted. Diagonal terms ($\partial^2 E(\mathbf{x})/\partial x_i^2$) are divided by the mass of the atom associated with x_i, and off-diagonal terms ($\partial^2 E(\mathbf{x})/\partial x_i \partial x_j$) are divided by the product of the square root of the masses of the atoms associated with x_i and x_j. These expressions reduce to that already provided for the one-dimensional case.

The second step involves replacing the Cartesian coordinates by a new set of coordinates $\boldsymbol{\zeta}$, such that the matrix of mass-weighted second derivatives is diagonal. δ_{ij} is the so-called Kronecker delta function which is 1 if i=j and 0 otherwise.

$$[\partial^2 E(\zeta)/\partial \zeta_i \partial \zeta_j]/(\sqrt{M_i}\sqrt{M_j}) = \delta_{ij}\, [\partial^2 E(\zeta)/\partial \zeta_i^2]/M_i$$

These new coordinates are referred to as ***normal coordinates***. While the normal coordinates for some vibrations may be described in terms of stretching of one bond or bending of one angle, more commonly they will be made up of mixtures of several bond stretches, angle bends and other motions.

The third step involves removing the six coordinates corresponding to the three translations and three rotations, leaving 3N-6 vibrational coordinates.

The intensity of a line in the infrared spectrum is proportional to the change in the dipole moment along the vibrational coordinate. It follows that where there is no change in dipole moment, for example, in a homonuclear diatomic molecule, the infrared intensity is zero.

The two major components of the earth's atmosphere, N_2 and O_2, are homonuclear diatomics and do not absorb in the infrared, that is, the intensity is zero. However, two of the four vibrational motions of CO_2, the third most common but very minor molecular component in the atmosphere, have non-zero infrared intensities. As a result, carbon dioxide absorbs radiation reflected from the earth's surface thereby trapping heat and leading to an increase in temperature (the so-called ***greenhouse effect***).

Lack of a line in the infrared spectrum does not mean that the molecule does not vibrate or that the vibrational energy for this line does not contribute to the zero-point energy. Rather, it means that absorption of radiation does not occur leading to a change in vibrational energy state. It should also be noted that a particular line that is infrared inactive might be visible in the Raman spectrum (an alternative form of vibrational spectroscopy based on reflectance rather than absorption). Here the intensity is related to the change in the polarizability rather than the change in dipole moment.

The application of quantum chemical models to infrared spectroscopy requires calculation of the second energy derivatives and first dipole moment derivatives with regard to changes in geometrical coordinate. The former completely dominates and scales as the fifth power of the size (number of basis functions). Infrared spectra may be calculated using semi-empirical molecular orbital models, for example, the PM3 models available in ***Spartan***, Hartree-Fock molecular orbital models, density functional models and MP2 models. Of the theoretical models available in ***Spartan***, semi-empirical models provide a poor account, Hartree-Fock models provide a reasonable account but density functional and MP2 models with polarization or larger basis sets perform best. Density functional models are the obvious choice for infrared spectra calculations, offering better results than Hartree- Fock models at comparable cost, and comparable results at much lower cost than MP2 models.

For two reasons, we recommend EDF2/6-31G* over the ωB97X-D/ 6-31G* model for infrared spectra calculations. First, EDF2/ 6-31G* was specifically formulated to reproduce measured infrared frequencies. Second, it is significantly (factor of two) less costly than ωB97X-D for frequency calculation and is easily applicable to the calculation of infrared spectra of organic molecules of moderate size (up to 400-500 amu). There are two major deficiencies with infrared spectra obtained directly from the EDF2/6-31G* model. The first is that calculated frequencies are almost always too large, typically by 3-5%. This can be directly traced to the harmonic approximation and to a potential energy curve that is too steep.

Other density functional models and the MP2 model show similar behavior. Vibrational frequencies obtained from Hartree-Fock models show an even larger systematic error in the same direction (frequencies too large), typically by 12-14%. Here two factors contribute. The first is the insistence on a quadratic potential, the same problem associated with density functional and MP2 models. The second is due to the fact that bond dissociation is improperly described by Hartree-Fock models, as evidenced by the fact that Hartree-Fock bond lengths are uniformly shorter than experimental distances. This suggests that the potential energy surface will be too steep and the frequency will be too large.

The second and more conspicuous deficiency is due to the fact that the lines in an infrared spectrum measured at finite temperature are broadened due primarily to rotational structure, whereas the lines in a calculated spectrum correspond to an isolated molecule at 0K and are sharp. There may be other differences, such as the absence of overtones, that is, vibrational transitions originating from excited vibrational states and, more importantly, lack of solvent. Although these are more difficult to quantify.

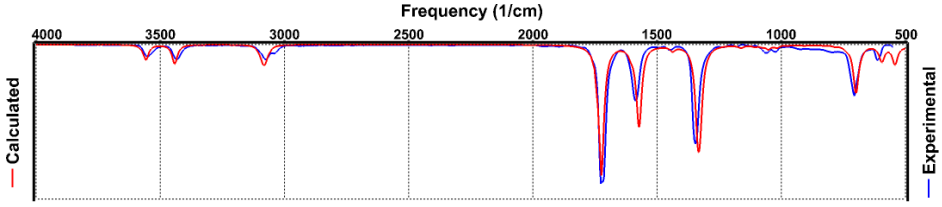
It is straightforward to correct the calculated spectrum to account for both deficiencies. First the spectrum may be uniformly scaled (multiplied by a parameter in the range of 0.95-0.97 for density functional models). Second, the calculated frequencies and intensities may be fit to a Lorentzian function with peak width and half peak height being treated as a second parameter.

Quality of Calculated Infrared Spectra

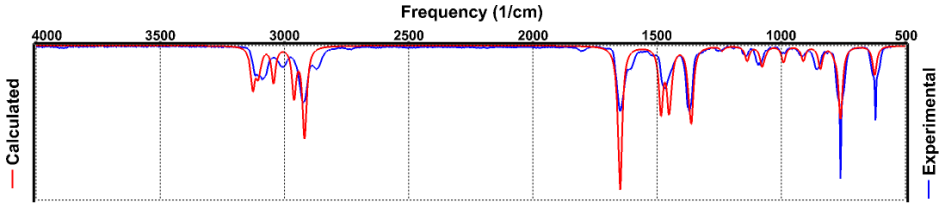
An infrared spectrometer records frequencies in the range 500- 4500 cm^{-1}. Frequencies below this range (which may require special instrumentation to measure) typically correspond to torsional motions and may depend strongly on conformation. It should be noted that the spectral region beyond ~2800 cm^{-1} is dominated by CH stretching vibrations and may be too crowded to be of value.

Calculated infrared spectra from the EDF2/6-31G* model that have been scaled to account for the systematic error in frequency and broadened to account for finite temperature are visually quite similar to the corresponding experimental spectra (taken from the NIST database). The four examples that follow, benzamide, (dimethylmethylidene) cyclopentadiene, 1,2-epoxy-*cis*-4-vinylcyclohexane and camphor are typical of the ~1000 comparisons that have been made.

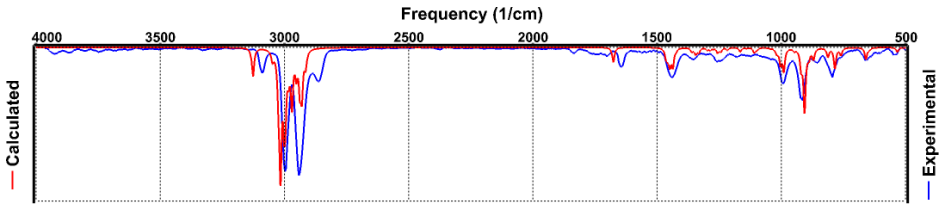
benzamide



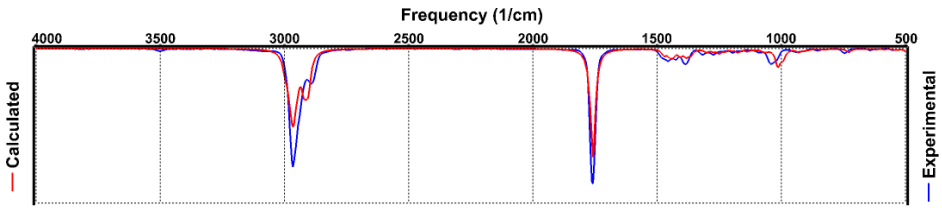
(dimethylmethylidene)cyclopentadiene



1,2-epoxy-cis-4-vinylcyclohexane



camphor



Even though scale and line-broadening parameters have been individually adjusted for these four examples, the values of the parameters are quite similar. Default parameters could have been substituted with little change.

Database of Infrared Spectra

More than 300,000 infrared spectra from the ωB97X-D/6-31G* model are available in the Spartan Spectra and Properties Database. A similar number from EDF2/6-31G* are available upon request from ***support@wavefun.com***.

Raman Spectra

Spartan is also able to calculate Raman spectra. The frequencies are identical to infrared frequencies but the intensities depend on the charge in polarizability rather than the change in dipole moment.

VCD Spectra

New in ***Spartan'26*** is the ability to calculate **Vibrational Circular Dichroism** spectra. VCD is a practical alternative to X-ray crystallography for the determination of absolute configuration. Used in conjunction with NMR protocols for conformationally flexible natural products, VCD and NMR allow exploration and conformation of constitutional as well as complete and absolute stereochemical assignment.

CALCULATING NMR SPECTRA

There are several reasons why NMR spectroscopy, in particular proton and ^{13}C NMR, is the most important analytical technique for characterizing organic molecules. The experiment is straightforward and can be carried out rapidly. It requires relatively small samples and is nondestructive. The so-called proton-decoupled ^{13}C NMR is particularly simple, comprising a single line (resonance) for each and every unique carbon. Despite its simplicity (or perhaps because of it), associating an ^{13}C spectra to a particular molecular structure can be problematic and prone to error, in particular, where alternative structures might be very similar. 2D spectra, in particular, COSY and HMBC spectra that combine chemical shifts and HH and CH coupling constants are more and more commonly employed to assist in assignment.

A routine and reliable method to predict ^{13}C chemical shifts as a function of three-dimensional structure would clearly be of value in helping to assign experimental NMR spectra of complex molecules, at the very least, either providing supporting evidence or casting doubt on a proposed assignment. One might argue that such a method already exists in the form of extensive NMR spectral databases. An exact match to an existing spectrum provides a definitive structure, while one or more close matches to entries in a database suggest what types of structures are reasonable. Of course, entirely new compounds will never give exact matches, simply because the spectrum is not in the database. Closely related are empirical relationships based primarily on connectivity and obtained from fitting previously assigned (and presumed correctly assigned) spectra. While these can achieve some degree of success, the fact that NMR chemical shifts (and of course three-bond HH and CH coupling constants) are sensitive to conformation means that molecules that appear to be very similar may give rise to entirely different spectra. Closely related to this is the fact that most organic molecules are not rigid and described by a single 3D structure. Rather, they comprise a connection of different 3D structures (conformers) and the resulting spectrum is a weighted average of energy weighted spectra of the individual conformer.

One alternative to databases and purely empirical schemes would be to calculate chemical shifts *a priori* using quantum mechanics.* In so doing, differences in structure and conformation are directly taken into account. The underlying methodology has been available for several decades for both Hartree-Fock and density functional models. However, calculations have only rarely been used to actually assist in the interpretation of spectra, and few practicing chemists seem to be aware that quantum chemical calculations are now possible (and practical) for real molecules, and how well calculations perform in accounting for chemical shifts. Those who are aware are confronted and all too often stymied, with what must seem to be an endless list of calculation methods. We believe that the full potential of quantum chemical calculations as assists to assigning NMR spectra will only be realized after a small number of alternatives or standard models are elaborated and their limits and reliability clearly defined. Stated differently, chemists need to approach quantum chemical calculations much in the same way that they now approach a spectrometer.

Underlying Theory

Application of an external magnetic field causes the nuclear spins to align either parallel or antiparallel to the field. The difference in energy (ΔE) between nuclear spin states is given by.

$$\Delta E = \gamma \hbar B_0$$

γ is the gyromagnetic ratio, a constant that depends on the magnetic moment of the nucleus, $\hbar$ is Planck's constant/2π and B_0 is the strength of the magnetic field ***at the nucleus***. What makes nuclear magnetic resonance (NMR) spectroscopy useful to chemists is that the magnetic field felt at the nucleus is different for each chemically distinct nucleus in a molecule. This is because the applied magnetic field is slightly weakened by electrons around the nucleus and the extent of this weakening depends on the detailed chemical environment. Nuclei that are well shielded by the electron cloud experience a lesser field than those that are poorly shielded and, as a result, show a smaller energy splitting. The splitting, relative to a standard, is termed a ***chemical shift***.

Two additional comments need to be made. First, not all nuclei possess non-zero magnetic moments and, therefore, give rise to an NMR signal. The magnetic moment for the proton is non-zero, although that of the dominant isotope of carbon (^{12}C) is not. Fortunately, the minor (1%) isotope ^{13}C possesses a finite magnetic moment. The fact that the dominant isotope of carbon has a zero magnetic moment is both a "curse" and a "blessing". A curse because it certainly slowed the application of carbon NMR to organic chemistry by several decades (waiting for magnet and spectrometer technology to catch up). A blessing because it results in a ^{13}C spectrum being much simpler (and easier to interpret) than a proton NMR spectrum.

* Neural networks offer another alternative, the downside being that their construction requires very large datasets. We are in the process of exploring this replacing experimental chemical shifts as reference data by values calculated from quantum chemical models.

The second comment is that proximate nuclei with finite magnetic moments will contribute to the magnetic field felt by the nuclei under investigation. Whereas nearby protons interact (couple) leading to splitting of the individual lines in the proton NMR spectrum, the very low probability (1% x 1%) that two ^{13}C nuclei will be adjacent all but eliminates carbon-carbon coupling. Proton-^{13}C coupling does occur but can be (and nearly always is) removed. The result is that the ^{13}C NMR spectrum contains only one line per unique carbon.

From the perspective of the experiment, the fact that the difference in energy between nuclear spin states (and ultimately the ability of an NMR spectrometer to distinguish chemically-different nuclei) is directly proportional to the magnetic field strength is disheartening. Magnets used in NMR spectrometers are now approaching practical and perhaps theoretical limits, and a mere 10-20% increase in field strength (translating to an equivalent increase in resolution) can mean more than doubling the cost of the spectrometer. Without a major breakthrough in magnet technology, the prognosis for greatly improved resolution over what is now possible (practical) is bleak. On the other hand, computer performance continues to double every few years (anticipated by Moore's law), and NMR spectra calculations on larger and ever more complex molecules continue to become more routine.

Correcting ^{13}C Chemical Shifts from B3LYP/6-31G*, ωB97X-D/ 6-31G* and ωB97X-V/6-31G* Density functional Models

Spartan can calculate NMR chemical shifts using Hartree-Fock models as well as a range of density functional models. We have focused on and carefully examined three density functional models, specifically B3LYP, ωB97X-D and ωB97X-V models with the 6-31G* basis set. Used directly, all three models lead to rms errors in ^{13}C chemical shifts on the order of 5-6 ppm (over a range from 0-250 ppm) and are not reduced by replacing 6-31G* by a larger more complete basis set. Errors of this magnitude are likely to be too large for reliably assigning the lines in an experimental ^{13}C spectrum, for supporting a proposed structure assignment or perhaps most important for casting doubt on one.

Empirically Corrected ^{13}C Chemical Shifts

Empirical schemes have been developed to dramatically reduce errors in ^{13}C chemical shifts, specifically for B3LYP/6-31G*, ωB97X-D/6- 31G* and ωB97X-V/6-31G* models. These utilize the same dataset for the linear regression comprising ~8000 sp^3 carbons, ~6200 sp^2 carbons (including aromatic carbons) and ~450 sp carbons.* The functional form is as follows:

$$^{13}\mathbf{C}_i = {}^{13}C_i \bullet \text{Scale} + \Sigma_k [X^{(0)}_k + X^{(1)}_k(R_{i,k} - L) + X^{(2)}_k(R_{i,k} - L)^2]$$

$^{13}\boldsymbol{C}$ is the corrected chemical shift and ^{13}C is the uncorrected chemical shift and *Scale* is a scaling factor. Summation is carried out over all bonds (1 or 2 for sp carbons, 3 for sp^2 carbons and 4 for sp^3 carbons). $X^{(0)}_k$, $X^{(1)}_k$ and $X^{(2)}_k$ are parameters that depend on the atom bonded to carbon (H, C, N, O, F, Si, P, S, Cl or Br) and $R_{i,k}$ are bond lengths to carbon. Different expressions apply to sp, sp^2 and sp^3 carbons. The equation for sp carbon involves four $X^{(0)}$parameters, four $X^{(1)}$ parameters and one $X^{(2)}$ parameter (11 parameters in total), that for sp^2 carbon involves all ten $X^{(0)}$ parameters, six $X^{(1)}$ parameters and four $X^{(2)}$ parameters (21 parameters in total), and that for sp^3 carbon involves all ten $X^{(0)}$ parameters but no $X^{(0)}$or $X^{(1)}$ parameters (11 parameters in total).**

Resulting RMS errors follow: B3LYP (2.4 ppm), ωB97X-D (2.1 ppm), and ωB97X-V (2.1 ppm), a factor of 2–3 smaller than uncorrected shifts. ^{13}C shifts noted due to changes in solvent. Assessment against the rigid molecule DELTA50[13] validation set suggests ^{13}C shift RMS errors on the order of 1.7 ppm, whereas assessment over a diverse collection of a hundred flexible natural products[9] suggests a ^{13}C RMS error of 2.1–2.2 ppm.

---

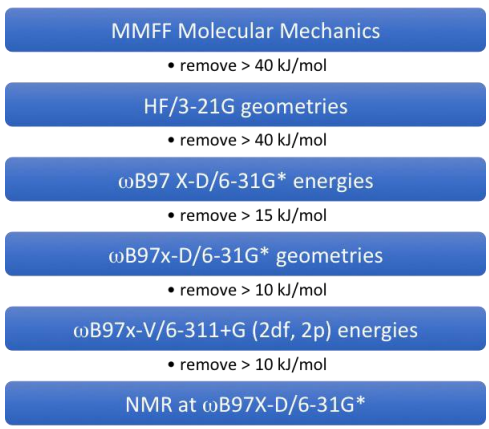
[13] Cohen, R. D.; Wood, J. S.; Lam, Y-H.; Buevich, A. V.; Sherer, E. C.; Reibarkh, M.; Williamson, R. T.; Martin, G. E. DELTA50: A Highly Accurate Database of Experimental ^{1}H and ^{13}C NMR Chemical Shifts Applied to DFT Benchmarking. Molecules. **2023**, *28* (6), 2449. https://doi.org/10.3390/molecules28062449

* Care has been taken to select either rigid molecules or molecules where a single conformer dominates the equilibrium (Boltzmann) distribution.

** Parameter files are available for review from: https://downloads-s3.wavefun.com/NMR_Correction_Parameters.zip.

CALCULATING CHEMICAL SHIFTS IN CONFORMATINOALLY FLEXIBLE MOLECULES

Spartan'18 introduced a protocol for calculating ^{13}C chemical shifts in conformationally-flexible organic molecules, targeted at the natural products community with the goal of being able to identify problematic structure assignments [*J. Nat. Prod.* **2019**, 82, 8, 2299–2306]. The protocol comprises a multi-step procedure to determine accurate Boltzmann weights starting from a systematic conformational search using MMFF molecular mechanics and terminating with a series of large basis set density functional energy calculations, and small basis set density functional chemical shift calculations that have been empirically corrected each of which is multiplied by its respective weight to produce a proper Boltzmann-averaged proton or ^{13}C spectrum. The procedure is depicted below.



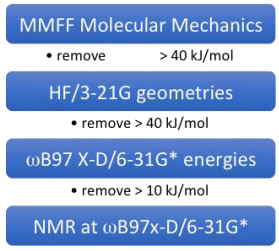
The protocol is fully automated requiring only input of a single conformer and designation of rotatable single bonds and flexible rings. Steps in the protocol may be eliminated, different quantum chemical methods employed and energy filters between steps adjusted. While the protocol has only been thoroughly assessed for ^{13}C chemical shifts, proton chemical shifts can be added to supplement (or replace) the ^{13}C data.

The performance of the protocol was first evaluated using ^{13}C chemical shifts for ~900 natural products the structures of which have been confirmed either by an X-ray crystal structure or by independent synthesis. Only a very few outliers (rms deviations > 4 ppm) were noted, almost all of which can be attributed to line misassignments rather that incorrect structures. The protocol was then applied to a second set of ~2500 natural products, the structures of which rely solely on NMR data. Here, fully10–15% fall outside the anticipated error range and furthermore cannot readily be explained by line misassignments. These problematic structure assignments perhaps warrant further examination.

A major downside of the protocol available in ***Spartan'18*** is its high computation cost, due almost entirely due to the two final two steps required for evaluation of the Boltzmann weights. While present- generation personal computers allow its application to molecules with weights <500 amu and with limited conformational flexibility, computation times for

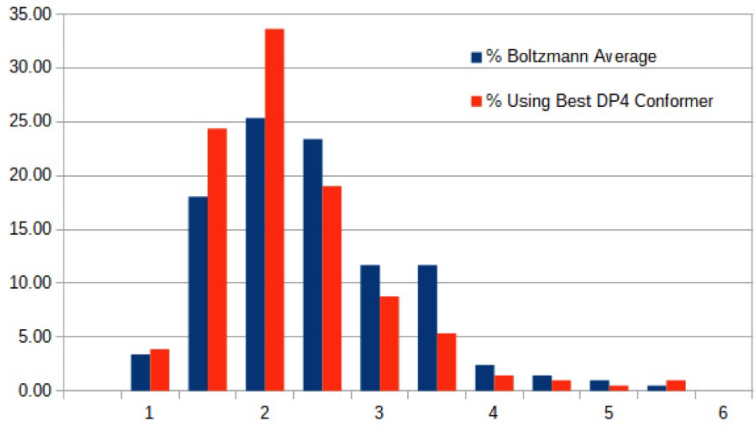
larger natural products or those with more than a few hundred accessible conformers may in some cases exceed several days.

Spartan'20 introduced a lower-cost protocol that, instead of constructing a proper Boltzmann-weighted average, selects from the set of "reasonable" low-energy conformers the one that best matches the experimental ^{13}C and/or proton chemical shifts based on its DP4 score [*J. Am. Chem. Soc.* **2010**, 132, 37, 12946–12959]. Because this does not require calculation of accurate Boltzmann weights, it relies on simpler quantum chemical models and is roughly an order of magnitude less costly than the previous protocol, greatly extending its practical range. Of course it does require an experimental spectrum! The protocol is depicted below.

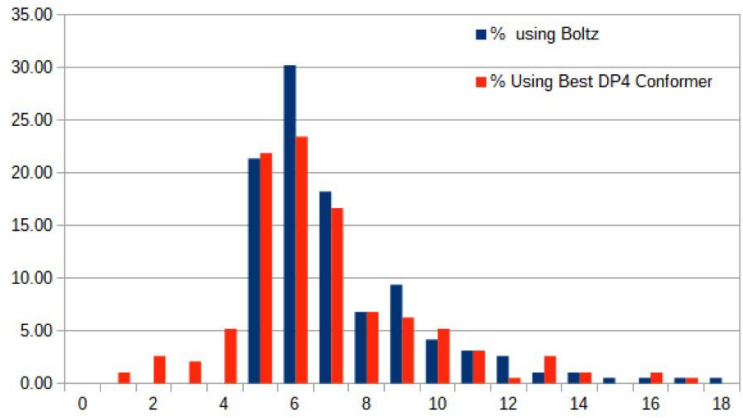


As with the previous protocol, alternative quantum chemical models and energy cutoff values may be substituted for those depicted above. In particular, the B3LYP/6-31G* density functional model may be substituted for ωB97X-D/6-31G* for both energy and NMR calculations, leading to an additional reduction in computation cost.

The two histograms below compare the performance of the original protocol based on Boltzmann averages (blue bars) and the alternative based on the best individual DP4 conformer score (red bars), for a subset of the natural products collection for which NMR structure assignments have been confirmed either by X-ray or independent synthesis. The metric is the smallest RMS deviation between calculated and experimental ^{13}C chemical shifts. The important conclusion is that the results are nearly identical. The fact that errors resulting from the new protocol are actually slightly smaller than those from the original should not be surprising as selection is based on the conformer that provides the best fit to the experimental data, irrespective of its energy.



The second pair of histograms compares the performance of the two protocols for the same natural products subset for which structure assignments are based solely on NMR, and with overall RMS errors from the original protocol of 5 ppm or greater. Errors near the lower end of this range have previously been interpreted to indicate the possibility of structure or peak misassignments while those in the middle or higher range are symptomatic of structure or peak assignments that are most likely incorrect. Aside from the fact that roughly 10% of the RMS values from the new protocol fall below 5 ppm, the two histograms are similar.



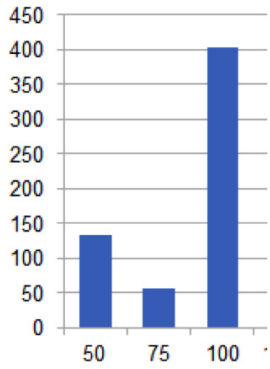
ASSIGNING STEREO AND REGIOISOMERS FROM CALCULATED CHEMICAL SHIFTS

Both protocols described in a previous topic allow calculated chemical shifts to be compared with measured values in order to choose to among possible stereoisomers. The simpler protocol introduced in ***Spartan'20*** is more practical for larger molecules with multiple degrees of conformational freedom and for which more than a few isomers need to be considered. Following the work of Goodman [*J. Am. Chem. Soc.* **2010**, 132, 37, 12946–12959], selection is based on DP4 score, using either the Boltzmann average of chemical shifts over accessible conformers for each of the isomers (original protocol) or the conformer for each isomer with the highest DP4 score (new protocol).

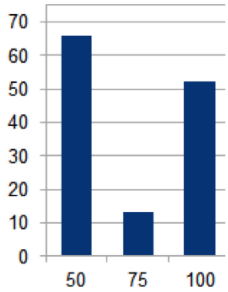
Spartan'26 provides a mechanism for isomer generation, requiring the user to designate which stereocenters are to be inverted (**Generate Isomers** under the **Geometry** menu; **Chapter 19**). Duplicate stereoisomers as well as stereoisomers that differ in absolute configuration are automatically removed. Isomer identification can be done with a single job submission, requiring input of a single conformer of one of the stereoisomers, designation of rotatable single bonds and flexible rings for this isomer (which are in turn applied to all isomers) and designation of stereocenters that are to be inverted. A more cautious approach, requiring two sequential jobs may also be followed: generate a list of stereoisomers and remove any that are high-energy or otherwise "unrealistic" and run the NMR analysis on each in turn.

As before, assessment draws on structures from the natural products literature that have been confirmed by X-ray crystallography or by independent synthesis. This involves more than 600 molecules, each of which yields two or more distinct isomers (most commonly either four or eight isomers), leading to over 4000 distinct stereoisomers.

The histogram below documents the performance of the NMR protocol introduced in ***Spartan'2***. ~70% of the molecules have DP4 scores above 75%, suggesting that the protocol successfully identifies the isomer that is observed. On the other hand, ~20% of the molecules have DP4 scores below 50%, suggesting that protocol is not successful. The remaining 10% of the molecules with DP4 scores between 50 and 75% may be interpreted to mean that while the calculations identify the correct isomer, one or more alternative isomers are competitive. These results are consistent with those obtained by Goodman based on a different and somewhat more computationally demanding procedure and applied to proton as well as ^{13}C chemical shifts for a different and smaller set of molecules.



There are at least two reasons for the failure of the calculations to correctly identify the isomer corresponding to the observed NMR spectrum. For one, it is likely that in some cases ^{13}C shifts for different isomers will be indistinguishable within the error limits of the calculations. Indeed, the majority of the "failures" involve change in chirality of just a single center. A second source of error may arise from the assumption that chemical shifts for the conformer that best fits the experimental ^{13}C data, as opposed to that from a properly Boltzmann-weighted average of chemical shifts over all conformers. This is of course what distinguishes the newer protocol in ***Spartan'20*** from both the original protocol in ***Spartan'18*** as well as from Goodman's work. It is also directly responsible for the large saving in computation cost. While it is not practical to obtain DP4 statistics over isomers for the full set of molecules using the previous protocol, we have obtained them for the 20% of the molecules with DP4 scores less than 50%. Data for this subset are provided below as a histogram.



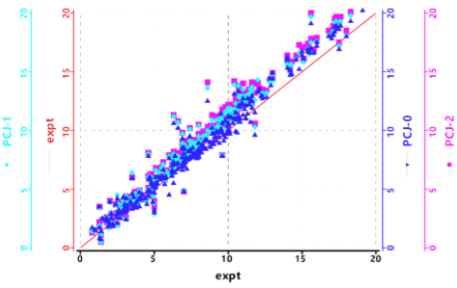
In only about a third of the cases does the "correct" isomer exhibit a DP4 score of 75% or higher, and half of the problem cases have DP4 scores below 50% are satisfactorily resolved. This suggests that the previous protocol which aimed to establish ^{13}C shifts that were Boltzmann-averaged over accessible conformers is unlikely to be more successful than the simplified procedure available in ***Spartan'20***.

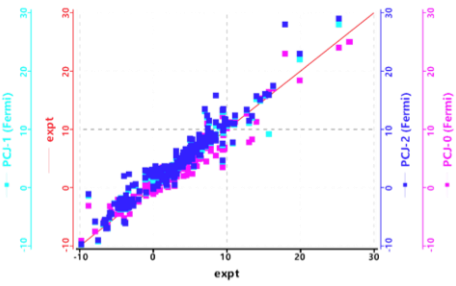
CALCULATING HH AND CH COUPLING CONSTANTS

Coupling constants may be calculated, either "exactly" or within the so-called Fermi-contact approximation. The latter approach is significantly the less demanding both in terms of computation time and memory required. 3-bond HH coupling constants may be estimated using the Karplus relationship [reference]. A parameterized variation on Karplus have long been provided in ***Spartan*** and extended to estimate both 2 and 3-bond CH coupling constants.

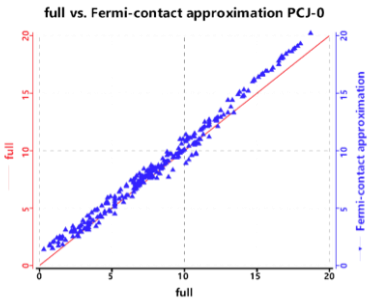
So-called "range-separated-hybrid" functionals such as ωB97X-D used for determining geometry and chemical shifts are not presently available in Spartan for coupling constant calculations. The popular B3LYP "global hybrid" functional has instead been used. In addition, the 6-31G* basis set used for chemical shift calculations has been replaced by one of a series of basis sets that have been expressly developed for coupling constant calculations. The smallest of these is PCJ-0, which is similar in construction to 6-31G* and readily applicable to molecules of comparable size. PCJ-1 and PCJ-2 basis sets are larger and therefore more limited in their practical range of application.

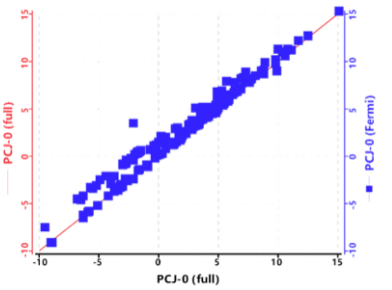
The first comparison relates coupling constants calculated within the Fermi-contact approximation and using the PCJ-0, PCJ-1 and PCJ-2 basis sets with experimental data for two sets of small molecules taken from the Reich web site [reference]. Data for 3-bond HH coupling constants with ~300 comparisons, and (combined together) 2 and 3-bond CH coupling constants with ~200 comparisons are shown below at the left and right, respectively. HH coupling constants range from roughly 0 to 20 Hz and are almost entirely positive. (The few small negative values have been excluded from the plot as have HH coupling constants larger than 20 Hz.) The overall range of CH coupling constants is similar although there are a significant number of negative values, almost entirely for 2-bond coupling. It should be noted that the sign of measured coupling constants is rarely reported and the experimental data from the Reich web site is almost entirely unsigned. We have adjusted the sign on experimental coupling constants to be in accord with the calculated values.





The second comparison examines the effect of the Fermi-contact approximation on coupling constants obtained using the PCJ-0 basis set. Results for 3-bond HH (left) and (combined together) 2 and 3-bond CH (right) coupling constants are shown below where the "full" calculations act as reference. The RMS deviations (here) are smaller than the previously mentioned RMS errors referenced to experiment.





Given that the primary objective for calculating coupling constants is to identify couplings between pairs of hydrogens that are large enough to be observed in a COSY spectrum and those between carbon and hydrogen that are large enough to be observed in an HMBC spectrum, it is likely conclude that calculations within the framework of the Fermi approximation using the PCJ-0 basis set are adequate.* As previously mentioned, current generation personal computers allow application to molecules with molecular weights on the order of 500 amu.

* It is also likely that empirically derived HH and CH coupling constants will be adequate for construction of COSY and HMBC spectra.

CALCULATING COSY AND HMBC SPECTRA

While coupling constants are rarely reported, they are of course required to determine splitting patterns in proton spectra and in providing 2D spectra, specifically HH coupling constants for COSY spectra and CH coupling constants for HMBC spectra, both of which are essential tools for assignment of structure. The "raw" experimental data are in presented in the form of contour plots, while calculated results are given in terms of a series of circles on a 2D plot corresponding to large couplings. A large degree of concurrence suggests that they are likely to be referring to the same or a very similar molecule whereas significant differences signals that something is amiss.

Two different display modes for calculated COSY and HMBC spectra are supported. The first displays all couplings ($^{3}J^{HH}$ for COSY and both $^{2}J^{CH}$ and $^{3}J^{CH}$ for HMBC) where the absolute value of the coupling constant is greater than a set magnitude. These take the form of open circles the size of which is independent of the magnitude of the coupling constant, but for HMBC the color of which depends on whether it is 2 or 3-bond. Rather than displaying the contour plot obtained directly from the COSY or HMBC experiment, assigned couplings are entered manually. Matching calculated and experimental data (that is, involving the same atoms) can be "linked" with ellipses drawn connecting the circles and crosses. The second mode, applicable only where experimental data have been provided, displays only calculated couplings for which an experimental coupling has been provided. As it is much less crowded (depending on the threshold set for displaying calculated couplings), it is likely to be the more useful of the two in trying to see if calculated and experimental HMBC spectra closely match.

MACHINE LEARNING and NEURAL NETWORK MODELS

The advent of machine learning (ML) models, in particular neural network (NN) approaches, offers the possibility of a new class of computational models that afford both high reliability and quality as well as low (computational) cost. *Seven* neural network models are implemented in the current ***Spartan'24*** release. These approaches were developed to serve four fundamental tasks:

(i) improve MMFF94[1-4] conformer energy differences.

(ii) accurately estimate ωB97X-D/6-31G*[5-7] equilibrium geometries starting from MMFF geometries.

(iii) accurately estimate electronic energies from:

ωB97X-V[8]/6-311+G(2df,2p)[11,12]//ωB97X-D/6-31G*
ωB97M-V[9]/6-311+G(2df,2p)[11,12]//ωB97X-D/6-31G*
ωB97M(2)[10]/6-311+G(2df,2p)[11,12]//ωB97X-D/6-31G*

starting from *either* ωB97X-D/6-31G* equilibrium geometries and energies, *or* from *estimated* ωB97X-D/6-31G* equilibrium geometries and ωB97X-D/6-31G* (single point) energies.

(iv) provide fast/accurate ^{1}H and ^{13}C chemical shifts, trained to reproduce the empirically corrected[13] ωB97X-D/6-31G* model.

Current machine learning (neural network) models support closed shell, uncharged molecules comprising: H, C, N, O, F, S, Cl and Br (only)[†].

With some customizations, the published SpookyNet[14] learning forcefield approach has been used to generate each neural network model. The training set for the **Corrected MMFF** neural network spanned more than ***140,000 molecules*** with more than ***1.8 million total conformers*** from ωB97X-V/6-311+G(2df,2p)[6-311G*] //MMFF calculations. The **Est. Density Functional ωB97X-D/6-31G* geometry** model required ≈***120,000 molecules***, both equilibrium geometries and another 50 distorted geometries ***(>6 million structures***) from ωB97X-D/6-31G*. Training included ≈***295,000 molecules*** from ωB97X-V/6-311+G(2df,2p)//ωB97X-D/6-31G* calculations for the **Est. ωB97X-V/6-311+G(2df,2p) energy** neural net, ≈***289,000 molecules*** from ωB97M-V/6-311+G(2df,2p)//ωB97X-D/6-31G* calculations for the **Est. ωB97M-V/6-311+G(2df,2p) energy** neural net, and ≈***289,000 molecules*** from ωB97M(2)/6-311+G(2df,2p)//ωB97X-D/6-31G* calculations for the **Est.**

ωB97M(2)/6-311+G(2df,2p) energy neural network. The NMR neural network training sets included ≈***2.2 million molecules*** with more than **44 million** calculated **^{1}H** and **33 million** calculated **^{13}C chemical shifts** from ωB97X-D/6-31G*.

Overview of Neural Network Models

1. Improving MMFF94 Molecular Mechanics (Corrected MMFF)[15]

Molecular mechanics is perhaps the only class of methods that is routinely practical for calculations of conformer energy differences for flexible molecules with several degrees of

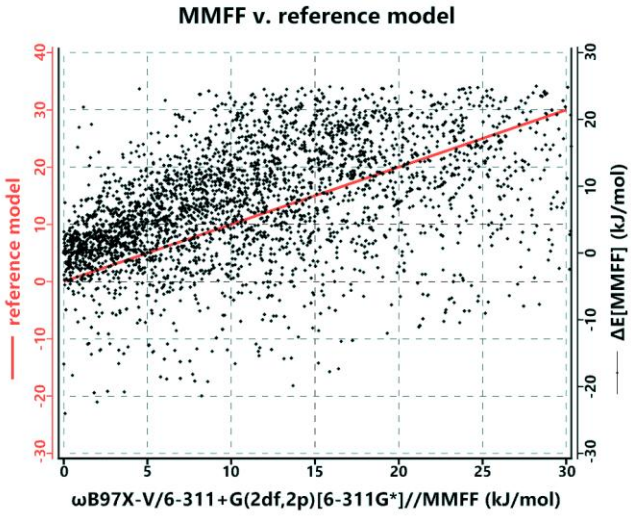
conformational freedom. All accessible conformers (or at the very least a representative sample of

† NMR shift networks available for proton and ^{13}C only (at present).

these conformers) need to be identified and ranked based on energy. Assuming three-fold rotors, two and three degrees of freedom lead to only 9 and 27 conformers, respectively, whereas a molecule with 9 degrees of freedom (not at all unusual in natural products) leads to nearly 20,000 conformers. While only a relatively few conformers are likely to seriously contribute to calculated molecular properties, all must be examined to determine *which* few.

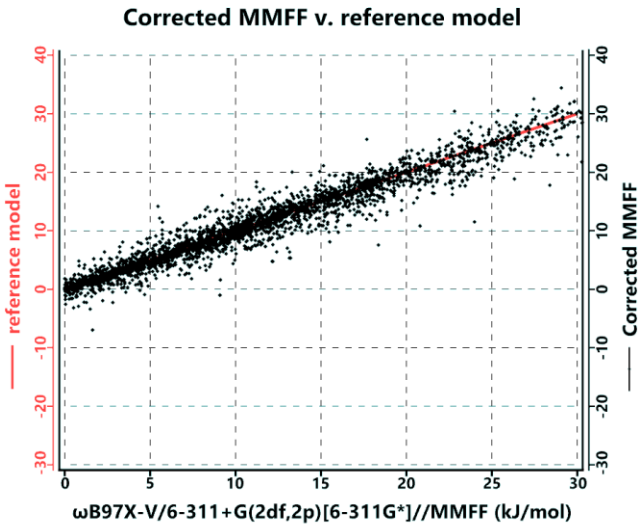
The MMFF94 approach has been shown to closely reproduce experimental conformational energy differences for small molecules with a single degree of conformational freedom. However, as suggested from the plot below, for a diverse series of moderately sized organic molecules with multiple degrees of conformational freedom, MMFF is not satisfactory. As experimental data are completely lacking, the reference energy is ωB97X-V/6-311+G(2df,2p)[6-311G*] density functional energy using MMFF equilibrium geometries, the same quantum chemical model used to train the neural net.

MMFF structures *corrected by neural network energies* provide a far better accounting with only a modest (*factor of two*) increase in computer time. The overall RMS error reduces from ***10.2* kJ/mol** for MMFF to ***1.8* kJ/mol** with the neural net correction. In conformer searching, ***Spartan's*** MMFF implementation is quite fast (typically no more than a second per conformer, and most of that is CPU overhead, not actual computational time). Adding a second or less (per conformer), to substantively improve the quality of conformational energy differences is time well spent. Ultimately this reduces the number of conformers being passed on to more accurate (and time consuming) quantum chemical approaches, used to establish an accurate Boltzmann distribution.



Comparison of Conformational Energy Differences Obtained from

MMFF and ωB97X-V/6-311+G(2df,2p)[6-311G*] Calculations



Comparison of Conformational Energy Differences Obtained from Neural Net and ωB97X-V/6-311+G(2df,2p)[6-311G*] Calculations

2. Estimating ωB97X-D/6-31G* Equilibrium Geometries[16]

Two considerations generally dictate model selection, quality of results and computation time. It is generally accepted that density functional models provide more consistent and more accurate results than Hartree-Fock molecular orbital models, but are also more time consuming, sometimes significantly so. The ωB97X-D/6-31G* density functional model is among the most used quantum chemical models for geometry calculation. It is both accurate and reliable, properly accounting for the geometries of molecules that are often poorly dealt with using Hartree-Fock models. Additionally, it has been recognized as a standout for *NMR shift prediction* among several dozen functionals surveyed[17]. While it is generally not sufficiently accurate for use in determining reaction energies, or energy differences among conformers, structures from ωB97X-D/6-31G* provide a good foundation for energy calculations with more rigorous functionals and larger (more complete) basis sets.

The neural network fit to ωB97X-D/6-31G* incorporated in ***Spartan'26*** is orders of magnitude faster than the underlying density functional model. Previously, structure optimization could take hours. With this neural network model, similar results are achieved in seconds (to just a few minutes). It should be emphasized that each one of these neural network calculations is unique and specific to the underlying computational model (and particular property, or properties) it is trained to reproduce, from a *specific functional*, and with a *specific basis set*. A neural network constructed to reproduce equilibrium geometries from the popular B3LYP-D3/6-31G* model (for example) will be distinct from that directed at ωB97X-D/6-31G* structures. Also acknowledged: *different training sets* and *different convergence criteria*, as well as different starting geometries, will ultimately lead to

different neural network definitions. Experience so far suggests that subtle differences in training sets will lead to nearly the same (or very similar) final equilibrium geometries.

To assess how well the *geometry* neural network performs, a test set comprising ***430*** natural products with molecular weights ranging from ≈200 to 700 amu (with a mean of 447 amu) was established. None of the molecules in the test set were used in training or validation. Three criteria were chosen for assessment: structure, energy, and NMR chemical shifts.

The most obvious criterion is likely structure itself: *bond lengths*, *bond angles*, and *dihedral angles* (excluding those involving hydrogen). The performance was substantive, with only 14 molecules showing bond lengths that deviate from ωB97X-D/6-31G* values by more than 0.02 Å. Similar results were observed in bond angles; only 4 and 7 molecules incorporate a bond angle that differs by more than **10°** and **5°**, respectively. Less than a quarter of the molecules show individual bond-angle deviations of more than **2°**.

Dihedral angles are perhaps the most important structural parameter as large deviations may result in molecular shapes that differ greatly from those obtained using ωB97X-D/6-31G* (the training model). Of additional concern is that further calculation of NMR chemical shifts, which are sensitive to medium to long-range interactions, might be impacted. Only two molecules show very large (>50°) deviations; 34 (8%) show deviations >20°; 30% of molecules show deviations of >10°.

The second criterion involves differences in energies resulting from use of neural network equilibrium geometries in lieu of ωB97X-D/6-31G* geometries. The RMS error for the neural network is 8 kJ/mol and the energies of only 11 molecules deviate from ωB97X-D/6-31G* by more than 20 kJ/mol, the largest of which is 42 kJ/mol. Fifty-four (54) molecules deviate by more than 10 kJ/mol.

Perhaps the most important metric is the effect of equilibrium geometry on ^{1}H and especially ^{13}C NMR chemical shifts. This is because geometry calculation is key in any effort to realize an accurate rapid procedure to calculate chemical shifts of conformationally flexible organic molecules, and to distinguish among stereoisomers that closely reproduce the results of the accurate (but time consuming) QM-based multi-step protocols available since the ***Spartan'18*** release. Only 4 and 29 molecules exhibit individual proton shift deviations greater than 1 and 0.5 ppm, respectively, more than a third of the molecules have one or more protons with chemical shifts that deviate from ωB97X-D/6-31G* values by more than 0.2 ppm.
Statistics for ^{13}C chemical shifts mimic those for proton shifts. Only one molecule exhibits an individual deviation >10 ppm and only 11 exhibit individual deviations >5 ppm. Roughly a third of the molecules based on neural net geometries show at least one individual deviation >2 ppm. The overall RMS (all carbons) and the RMS of the maximum individual ^{13}C shift errors (per molecule) from the neural net geometries are 0.8 ppm and 2.3 ppm, respectively.

3. Improved Reaction and Conformer Energy Calculations[18]

Reaction energies and conformer energies differences are among the most important quantities obtained from quantum chemical calculations. Organic chemists (synthetic chemists in particular) will benefit from calculations able to identify the dominant stereoisomer arising from a chemical reaction, and from determination of conformer energy differences, providing Boltzmann weights and with these, accurate prediction of NMR shifts of flexible molecules.

The "gold standard" CCSD(T)/CBS (**C**oupled-**C**luster theory with **S**ingles, **D**oubles, and perturbative **T**riples and a **C**omplete **B**asis **S**et) calculations[19] scale formally as the seventh power of the number of basis functions, and can take many hours (even days) to complete. Reducing computational cost while achieving CCSD(T) accuracy has been the primary motivation for development of many density functional models. In previous versions of ***Spartan***, the primary focus has been on the Head-Gordon functionals: ωB97X-V, ωB97M-V and ωB97M(2), all with the 6-311+G(2df,2p) basis set and all starting from ωB97X-D/6-31G* geometries. The CCSD(T)[20] procedure is superior to these three functionals, and while the 6-311+G(2df,2p) basis set is moderately large and diverse, the recipe developed to yield a "complete basis set" (CBS) ensures that further additions will at most have modest impact on the energy. Finally, in our case, the geometry is different from that used in CCSD(T) calculations (MP2/aug-cc-pVDZ basis set vs. ωB97X-D/6-31G*).
The two models ωB97X-V/6-311+G(2df,2p)//ωB97X-D/6-31G* and (to a lesser extent) ωB97M-V/6-311+G(2df,2p)//ωB97X-D/6-31G* are practical for routine calculations on rigid molecules (or those with only a few degrees of conformational freedom) up to molecular weights <400 amu, although they may require tens of minutes to several hours to complete. They are clearly impractical, at least routinely so, for significantly larger systems, or for molecules with multiple degrees of conformational freedom. The ωB97M(2)/6-311+G(2df,2p)//ωB97X-D/6-31G* is more computationally costly and practical ranges are even smaller.

Spartan'24 provides neural networks providing **Est. DFT energies**[18] trained to these three density functional models. These require an ωB97X-D/6-31G* geometry (or an estimated geometry from the Est. ωB97X-D/6-31G* neural network along with a single point ωB97X-D/6-31G* energy). Because of these requirements, application of these neural networks is not "instantaneous" (unlike the two previous described machine learning models, these three are not devoid of quantum chemical calculations as input). This said, the overall performance is easily an *order of magnitude faster* utilizing these neural networks to provide energies (compared to the corresponding quantum chemical energy calculations). As quality conformer energy differences are needed to establish accurate **Conformer Distributions** (as well as Boltzmann weighted NMR shifts), these energy-focused neural networks provide significant computational savings in tasks featuring conformational searching, offering the hope of an analogous machine learning option to the QM-heavy multi-step **NMR Spectrum** task (the protocol established for determining accurate ^{13}C and ^{1}H Chemical Shifts for conformationally flexible molecules)[13]. Although the routine desire for quality energies

(alone) suggests that the three **Est. Density Functional energy** models will have ample opportunity for application in conjunction with standard equilibrium geometry calculations.

4. Estimating accurate Proton and ^{13}C NMR Chemical Shifts[21]

The challenging task of assigning molecular structure for novel synthetic and/or novel natural product molecules has, in recent years, been supported by density functional calculations (typically from GIAO, Gauge-Including Atomic Orbitals) -based). Calculations have become a useful tool to aid or challenge structural assignments[22-34]. ***Spartan's*** multi-step protocol has been routinely applied in support of confirming structural assignment[35-41]. The major limiting factor to more routine application of the multi-step protocol remains the unfortunately long calculation times (many hours to days). While the original ***Spartan'20*** protocol remains an option, beginning in ***Spartan'24***, we have introduced (by default) the ***Corrected MMFF*** model as an additional filter between MMFF94 optimized structures and further DFT optimization and energy calculation. This step adds only a small fraction of overall computational time, and allows a reduction of the energy window (above "best" conformation) from 40 kJ/mol to 20 kJ/mol, as well as a reduction in the maximum total number of conformers passed on to further quantum chemical calculations (from 200 to 100). This allows for a factor of 2 to 5 faster results (compared to the ***Spartan'20*** protocol). While an improvement, the QM-heavy protocol often remains time consuming, a recent sample of conformationally flexible (mean of ≈ 280,000 conformers) natural products molecules (with a mean molecular weight of 447.8 amu) required an average of 319 minutes computational time (per isomer) to provide Boltzmann weighted Chemical Shifts (per isomer), sufficient to compare with experimental shifts and assign structure.

To be routinely applicable, improved computational time is highly desired. Two independently trained neural network models providing ^{1}H and ^{13}C chemical shifts (only), have been implemented beginning with ***Spartan'24 1.3.*** Trained on > 2.7 million equilibrium geometry calculations (over 44 million ^{1}H and 33 million ^{13}C chemical shifts) from ωB97X-D/6-31G* density functional and the Est. ωB97X-D/6-31G* machine learning model (trained to reproduce the former). These NMR chemical shift machine learning models perform consistently well when applied to organic systems outside of the formidable training set. An assessment of ***601*** molecules from the Marine Natural Product Database[42], ranging in weight from 111 to 693 amu (with an average weight of 290 amu) has been made. Results from the ^{1}H and ^{13}C chemical shift neural networks from two equilibrium geometry input sources (ωB97X-D/6-31G* density functional the est. ωB97X-D/6-31G* machine learning model) were compared with chemical shift results from direct calculation with quantum chemical (empirically corrected) GIAO density functional ωB97X-D/6-31G* calculations. The NMR neural net models performed equally well when using structures (equilibrium geometries) from ωB97X-D/6-31G* density functional and the Est. ωB97X-D/6-31G* machine learning model, with RMS errors for chemical shifts of .05 ppm for ^{1}H and .73 ppm for ^{13}C from QM geometries and .09 ppm and 1.02 ppm from machine learning geometries. Summarized in **Table 1**:

Table 1. RMS and Max. Absolute Errors for ^{1}H and ^{13}C Chemical Shifts in ppm*				
Equilibrium Geometry model:	**ωB97X-D/6-31G***		**Est. ωB97X-D/6-31G***	
signal	^{1}H	^{13}C	^{1}H	^{13}C
RMS (ppm)	0.05	0.73	0.09	1.02
Max error (ppm)	0.62	5.79	2.10	10.48

Neural Network chemical shifts compared with those from corrected ωB97X-D/6-31G (GIAO-based) density functional calculations for a series of 601 rigid marine natural product molecules.

Results are quite close to the underlying QM shifts, at essentially a few seconds per molecule. In the case of the original multi-step protocol (with no machine learning steps), calculations routinely require several hours to days to complete. The ***Spartan'26*** **QM** protocols (with no machine learning steps, or only a single Corrected MMFF option) perform on a time scale of **hours**. The default ***MLXD*** protocol (with ***five*** different machine learning models) performs on a time scale of **minutes**. A representative sampling of flexible natural products molecules (with a mean of ≈280,000 conformers) from the Natural Products Database[43] (a collection included with ***Spartan'26*** providing calculated and experimental shifts for a few thousand published natural products molecules) indicate the *est. density functional* protocol (*maximum* use of machine learning models) is on the order of **30x faster** than the *density functional* protocol (with a *single* machine learning model step) when performing the same Boltzmann weighted NMR spectrum (^{1}H and ^{13}C shifts).

Table 2 (below) provides an example of performance along with breakdown of molecular weight, conformers identified, and computational times (in minutes) for both the **Est. Density functional** protocol (machine-learning) and the **Density functional** protocol (quantum-chemical). Assessment compares times for a set of eighteen natural products molecules. The machine learning protocol performance is impressive.

Table 2. Flexible Natural Products Examples. Boltzmann Weighted NMR Spectrum task: Machine Learning (MLXD) vs. Density Functional (2024-QM) computational times.

molecule	conformers	Formula	Wt. (amu)	NN*	QM*	factor of
coodeanone A	78,732	$C_{23}H_{32}O_5$	388.5	13.6	459.8	33.8
bromophycolide R	1,417,176	$C_{27}H_{35}BrO_4$	503.5	3.3	164.8	49.9
psiguadial C	11,664	$C_{20}H_{24}O_{10}$	490.6	3.0	97.1	32.4
lythrine	2	$C_{26}H_{29}NO_5$	435.5	2.2	82.7	37.8
angustilongine A	68,024,448	$C_{43}H_{54}N_4O_5$	706.9	35.2	1277.8	36.3
ternifolide A	209,952	$C_{22}H_{28}O_7$	404.5	4.2	78.7	19.0
tonantzitlolone A	6,377,292	$C_{26}H_{40}O_7$	464.6	17.1	308.7	18.0
prunifoline D	576	$C_{21}H_{24}N_2O_4$	368.4	8.5	212.9	25.1
kopsonoline	4	$C_{19}H_{22}N_2O$	294.4	1.2	37.2	30.4
ciliatonoid A	128	$C_{26}H_{32}O_5$	424.5	7.3	161.3	22.1
epoxyresibufogenin…	1,728	$C_{25}H_{32}O_6$	428.5	3.8	102.8	27.4
achillinin B	13,122	$C_{31}H_{38}O_8$	538.6	20.7	652	31.4
lycopodine	3,888	$C_{16}H_{25}NO$	247.4	1.2	25.1	21.0
caesalappanin G	39,366	$C_{22}H_{30}O_7$	406.5	6.4	124.6	19.3
stauntonine	17,496	$C_{21}H_{28}O_7$	392.5	14.9	310.5	20.8
flueggine B	12	$C_{24}H_{28}N_2O_5$	424.5	2.4	82.6	34.5
shearinine I	2,592	$C_{37}H_{45}NO_7$	615.8	8.2	371.9	45.4
saroclide A	708,588	$C_{25}H_{32}N_2O_5$	440.5	13.5	969.5	72.1
mean	4,272,598		443.1	**9.3**	**306.7**	**≈32**

*Time in minutes. Calculations performed on a one of three different Intel systems: 12-core Intel i7 (Win 11), 10-core Intel i9 (Ubuntu 22.04 LTS), 24-core Intel i9 (Win 11). At the low end, improvement from the Neural Network protocol offers more than an order of magnitude faster results than the Quantum Chemical protocol. This dataset is available from neural network folder in the Tutorials directory "NMR-Spectrum-Timing.spartan." Average RMS against experimental ^{13}C Shifts is 1.27 ppm for the density functional calculations and 1.55 ppm for the neural network models.

From the user's perspective, the only change in the details of the **Calculations** dialog is the selection of **Est. Density Functional** instead of **Density Functional** from the "**with**" pull-down menu in the **NMR Spectrum** task. ***Spartan*** configures *all* details (which are exposed if one *clicks* on the **Details** checkbox):

Details:	Step	Theory		Level	Keep ≤	At Most
	☑ Geometry:	Molecular Mechanics		MMFF	40 kJ/mol	200 Conformers
	☑ Energy:	Molecular Mechanics		Corrected MMFF	20 kJ/mol	100 Conformers
	☑ Geometry:	Est. Density Functional		DLFFG1	10 kJ/mol	50 Conformers
	☑ Energy:	Density Functional	ωB97X-D	6-31G*	10 kJ/mol	30 Conformers
	☑ Energy:	Est. Density Functional		DLXD2XV	10 kJ/mol	30 Conformers
	☑ Energy:	Est. Density Functional		DLNMR1		

For comparison, the details of the **2024 QM-NMR** protocol for the **NMR Spectrum** task are provided below:

Details:	Step	Theory		Level	Keep ≤	At Most
	☑ Geometry:	Molecular Mechanics		MMFF	40 kJ/mol	200 Conformers
	☑ Energy:	Molecular Mechanics		Corrected MMFF	20 kJ/mol	100 Conformers
	☑ Geometry:	Hartree-Fock		3-21G	30 kJ/mol	100 Conformers
	☑ Energy:	Density Functional	ωB97X-D	6-31G*	15 kJ/mol	50 Conformers
	☑ Geometry:	Density Functional	ωB97X-D	6-31G*	10 kJ/mol	50 Conformers
	☑ Energy:	Density Functional	ωB97X-V	6-311+G(2df,2p)[6-311G*]	10 kJ/mol	30 Conformers
	☑ Energy:	Density Functional	ωB97X-D	6-31G*		

Both approaches begin with the same 2 steps: **MMFF94** is used to establish an initial conformer distribution and MMFF energies are improved by the **Corrected MMFF** model trained to ωB97X-V/6-311+G(2df,2p)[6-311G*]. The procedures fork beyond this point. The **Density Functional** protocol goes on to perform as many as 300+ additional quantum chemical calculations, some potentially quite time consuming. Final chemical shift data is obtained from density functional GIAO-based ωB97X-D/6-31G* calculations, including empirically corrected chemical shifts. Whereas in the **Est. Density Functional** protocol with only a single quantum chemical calculation (with a modest basis set), machine learning models provide the energies for Boltzmann weighting as well as ^{13}C and ^{1}H chemical shifts (one neural network per center). The combination of comparable accuracy to the QM-heavy protocol and the remarkably fast performance suggests the new ML protocol is positioned to play a substantive role in assisting practicing chemists with structure assignment of novel organic molecules.

In the case of *rigid* molecules, it is not necessary to obtain Boltzmann weighted shifts, so the only two requirements are structure (Equilibrium Geometry) and NMR (Chemical Shifts). Machine learning (neural network) models that provide both equilibrium geometries[16] and proton and ^{13}C shifts[21] are readily available. Using these models provides comparable quality results between *2 and 3 orders of magnitude faster* than QM DFT ωB97X-D/6-31G* and (corrected) GIAO NMR shifts.

Substantial gains have been made from this initial tranche of neural network models (one model improving quality of energy from MMFF input, three models improving quality of energy from single-point ωB97X-D/6-31G*, one performing geometry optimizations trained to reproduce ωB97X-D/6-31G*, and one model combining two neural networks (one per center) for providing accurate proton and ^{13}C shifts).

Assessment on a collection of 76 published natural product molecules (comprising either rigid molecules, or molecules with only one dominant conformer) suggests the new machine learning models applied to both geometry and proton and ^{13}C shift calculation averaged more than a hundred (106) times faster than ωB97X-D/6-31G* calculations. Summarized in **Table 3** (below), dataset available here.

Table 3. Performance: Calculated ^{13}C shifts for a collection of 76 published *rigid* Natural Products molecules with molecular weights ranging from 145.16 amu to 612.82 amu (average weight of 304 amu) containing H, C, N, O, F, S, Cl, and Br (only):			
Quality:		**Timing:**	
RMS Error: QM	**1.77** ppm	Avg QM geometry + NMR	**286.96** sec
RMS Error: NN	**1.94** ppm	Avg NN geometry + NMR	**3.04** sec
average Δ	**-0.17** ppm	avg. improv. factor	**106**
avg. abs. value Δ	**0.23** ppm	max. improv. factor	**840**

All calculations run on Intel® 10-Core™ i9 under Ubuntu 24.04 LTS. Comparable results on similar sized systems have been observed under Windows11 (Intel® 12-Core™ i7) and under MacOS 14.X (Apple M2 Pro 12-core). The molecule with highest timing improvement (a factor **840** in the dataset above with 10-core calculations), showed a factor of **1674** timing improvement when restricted to single core calculations. Molecules selected either have only one systematic conformer, or one clearly energetically dominant conformer.

Utilizing the *Corrected MMFF*[15] model in conformer searching, we're able to reduce the energy window (above the lowest energy conformer) by half, significantly reducing the number of conformers that must be considered for accurate Boltzmann distribution. The *Estimated DFT ωB97X-D/6-31G* geometries*[16] are often 2 or more orders of magnitude faster than QM ωB97X-D/6-31G* geometry optimizations. Single point ωB97X-D/6-31G* energies from estimated ωB97X-D/6-31G* geometries can be corrected with *Estimated DFT energies*[18], neural networks trained to *much better* QM energy models: ωB97X-V, ωB97M-V and ωB97M(2) density functionals (and the 6-311+G(2df,2p) basis set) in only seconds. Conformational energy differences obtained from the neural networks are typically within 1kJ/mol of the respective QM models (which often require several minutes to hours of computational time). The *Estimated DFT NMR Shift*[21] networks reproduce accurate empirically corrected ωB97X-D/6-31G* GIAO NMR proton and ^{13}C shifts to within 0.04 ppm for proton and 0.2 – 0.3 ppm for ^{13}C and do so in seconds.

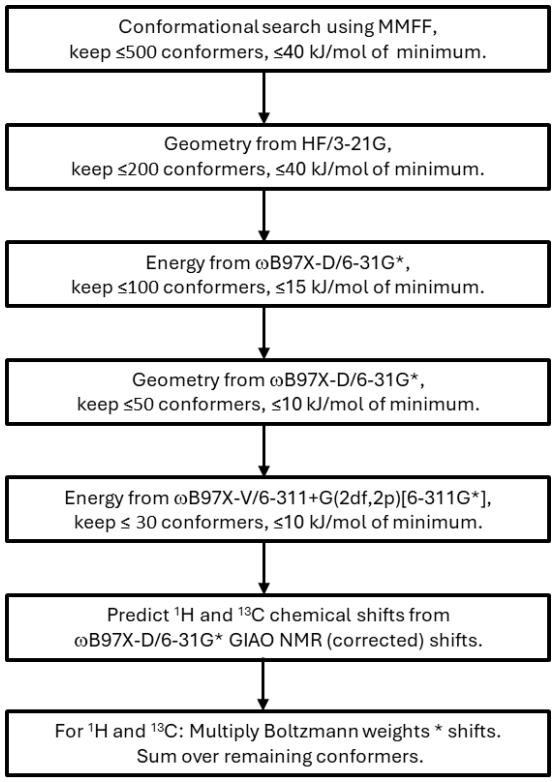
References:

1. *J. Comput. Chem.* **1996**, 33, 490.
2. *J. Comput. Chem.* **1996**, 33, 520.
3. *J. Comput. Chem.* **1996**, 33, 553.
4. *J. Comput. Chem.* **1996**, 33, 587.
5. *Phys. Chem. Chem. Phys.* **2008**, 10, 6615-6620.
6. ***Ab Initio Molecular Orbital Theory***, Wiley, New York, 1985. ISBN: 978-0-471-81241-8
7. *J. Chem. Phys.* **1989**, 90, 1007-1023.
8. *Phys. Chem. Chem. Phys.* **2014**, 16, 9904-9924.
9. *J. Chem. Phys.* **2016**, 144, 214110.
10. *J. Chem. Phys.* **2018**, 148(24), 241736.
11. *J. Chem. Phys.* **1980**, 72, 650-654.
12. *J. Comp. Chem.* **1983**, 4(3), 294-301.
13. *J. Nat. Prod.* **2019,** 82(8), 2299-2306. (Corrected-NMR-Shifts).
14. *Nat. Commun.* **2021**, 12(1) 7273. (SpookyNet).
15. *J. Comp. Chem.* **2025**, 46(1), 70016. (Corrected-MMFF-Energies).
16. *J. Chem. Inf. Model.* **2025**, 65(5), 2314-2321. (Estimated-DFT-Geometries).
17. *Molecules.* **2023**, 28(6), 2449.
18. *J. Comp. Chem.* **2025,** 46(12), 70129. (Estimated-DFT-Boltzmann-Energies).
19. *J. Chem. Theory. Comput.* **2013**, 9(5), 2151-2155.
20. *Chem. Phys. Lett.* **1997**, 281, 130-134.
21. *J. Org. Chem.* **2025**, 90(32), 11478-11485. (Estimated-DFT-NMR-Shifts).
22. *J. Am. Chem. Soc.* **1990**, 112(23), 8251–8260.
23. *J. Phys. Chem. A* **1999**, 103(36), 7357–7363.
24. *Chem. Rev.* **1999**, 99(1), 293–352.
25. *Angew. Chem.* **2004**, 116, 4585-4589.
26. *Chem. A Eur. J.* **2006**, 12, 5514–5525.
27. *Chem. Rev.* **2012**, 112, 1839–1862.
28. *Nat. Protoc.* **2014**, 9, 643–660.
29. *J. Nat. Prod.* **2015**, 78(7), 1505–1510.
30. *Chem. Eur. J.* **2016**, 22, 12246–12261.
31. *Molecules*. **2017**, 22(4), 521-526.
32. *Acc. Chem. Res.* **2020**, 53(9), 1922–1932.
33. *Org. Chem. Front.* **2021**, 8, 2019-2058.
34. *Molecules.* **2024**, 29(2), 336.
35. *J. Phys. Chem. A* **2021**, 125(28), 6029–6041.
36. *Chem. Lett.* **2022**, 51(11), 1080-1082.
37. *J. Nat. Prod.* **2023**, 86, 10, 2294–2303.
38. *J. Agric. Food Chem.* **2023**, 71(36), 13338–13345.
39. *Biomolecules.* **2024**, 14(7), 780.
40. *J. Am. Chem. Soc.* **2025**, 147(1), 498–509.
41. *J. Am. Chem. Soc.* **2025**, 147(28), 24921–24931.
42. *Nucleic Acids Res.* **2022**, 50, d665-d667.
43. *NPD.* ***Spartan'26 v.1*** **2026**, Wavefunction, Inc., Irvine, CA USA.

NMR PROTOCOLS for FLEXIBLE ORGANIC MOLECULES

Spartan'26 includes a wide range of computational options devoted to predicting accurate chemical shifts of both rigid and flexible organic molecules. Flexible molecules can be dealt with using a variety of multi-step workflows or NMR protocols. While considerable flexibility exists within these workflows (especially the QM-based options), we have distilled the options down to the four *main* published protocols, two QM (quantum chemical) and two ML (neural network-based).

The **2019 QM-NMR** protocol is based on work done by Hehre and co-workers (including substantive contributions from Masaru Hashimoto and group members at Hirosaki University), see: *J. Nat. Prod.* **2019**, *82* (8), 2299–2306. This approach has been cited dozens of times, used in conjunction with constitutional and stereochemical assignment (and reassignment) work, and follows the general workflow below:

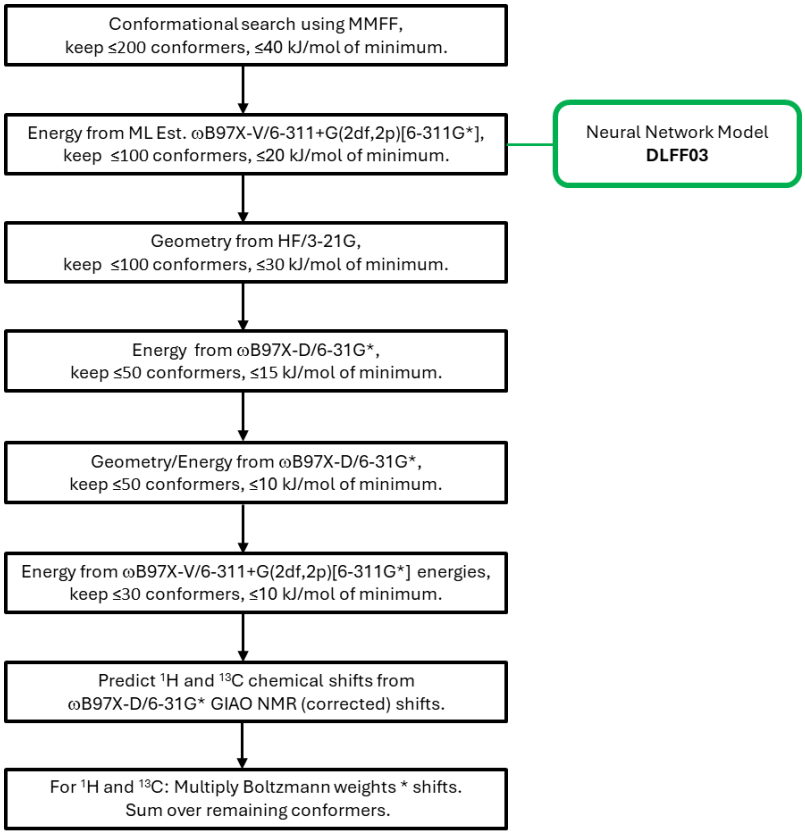


2019 QM-NMR protocol

The above approach has been assessed against >2500 published natural products from the literature with RMS errors of 2.1–2.2 ppm for ^{13}C shifts. Against 700 natural products with confirmed structures from X-ray crystallography, the RMSE is closer to 1.9 ppm. The primary limiting factor is computational time, which can exceed hours or tens of hours in the case of flexible molecules.

A **2024 QM-NMR** protocol iteration emerged once an initial ML model for improving conformational energies from MMFF was established, see: *J. Comp. Chem.* **2025**, *46* (1),

70016. Accuracy is comparable to the 2019 QM approach however performance (especially on consumer-grade hardware) has been shown to improve considerably (primarily based on improved filtering and performing *fewer* QM calculations).

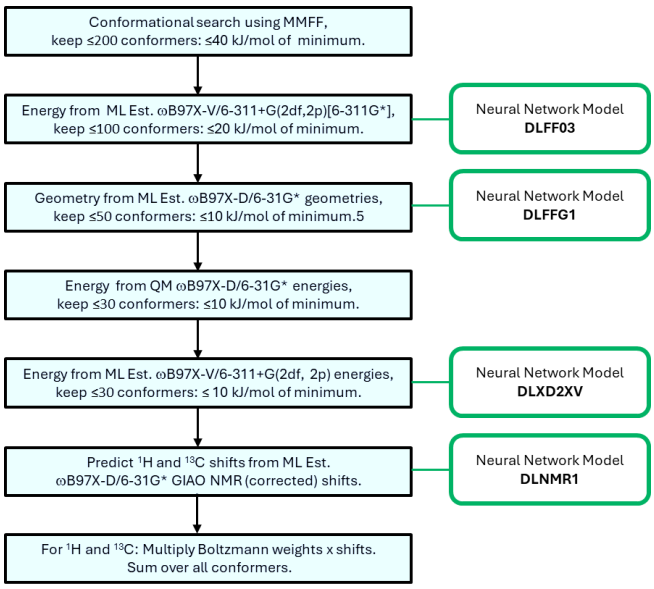


Timing and performance assessments available from ChemrRxiv 2026 preprint. The 2024 QM-NMR protocol is perhaps most noteworthy as framework including the "target" QM models that form the backbone of the machine learning-based NMR protocols MLXD, and MLHF (below).

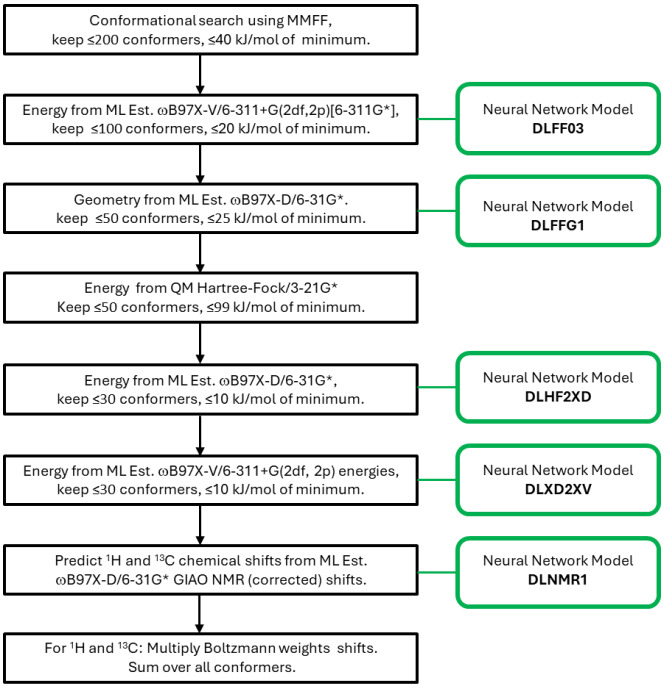
In 2025, with the development of 3 *additional* neural network approaches*, each one replacing a major step from the 2024 QM protocol, the **MLXD-NMR** protocol was made available as part of the final maintenance update to ***Spartan'24***.

* Geometries: *J. Chem. Inf. Model.* **2025**, 65(5), 2314-2321
Boltzmann Energies: *J. Comp. Chem.* **2025**, 46(12), 70129
NMR Shifts: *J. Org. Chem.* **2025**, 90(32), 11478-11485

The general form of the **2025 MLXD NMR** protocol follows:



Both the MLXD-NMR workflow and its faster cousin, MLHF-NMR perform nearly identically to the underlying QM approaches, but at tens to hundreds of times of times faster in the case of MLXD-NMR, and up to more than a thousand times faster using MLHF-NMR:



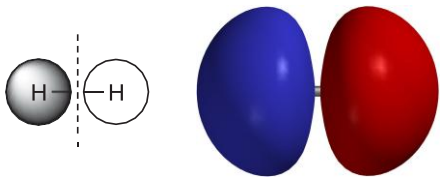
MLHF-NMR performance against experiment suggest 2.2 ppm ^{13}C shift RMSE (nearly identical to QM approaches).

ATOMIC AND MOLECULAR ORBITALS

Chemists have developed a variety of methods for describing electrons in molecules. Lewis structures are the most familiar. These drawings assign pairs of electrons either to single atoms (lone pairs) or pairs of atoms (bonds)*. The quantum mechanical equivalents are atomic and molecular orbitals which arise from solution of (approximate) Schrödinger equations for atoms and molecules, respectively. Molecular orbitals are spread throughout the entire molecule, that is, they are delocalized.** Because of this, they are typically more difficult to interpret than Lewis structures.

Orbital Surfaces

Molecular orbitals may be able to provide important clues about chemical reactivity, but before we can use this information we first need to understand the fundamentals for very simple molecules. The following figure shows two representations, a hand drawing and a ***Spartan***-generated image of an unoccupied molecular orbital of hydrogen molecule, H_2.



unoccupied molecular orbital in hydrogen

> Open ***hydrogen empty*** in the ***topics*** directory. Note that except for the colors (sign of the orbital) the two sides of the graphic are identical. The junction between red and blue regions is where the value of the orbital is zero. Close ***hydrogen empty*** when you are finished.

The familiar hand drawing shows the orbital as two circles and a dashed line. The circles identify regions where the orbital takes on a significant value, either positive (*shaded*) or negative (*unshaded*). The dashed line identifies locations where the orbital's value is exactly zero (a ***node***). The drawing is useful, but it is also limited. We only obtain information about the orbital in two dimensions, and we only learn the location of significant regions and not how the orbital builds and decays inside and outside of these regions.

---

* The present discussion is limited to molecules in which all electrons are paired. Molecules with one or more unpaired electrons (radicals, triplet states, etc.) may also be treated.

** Is it possible to localize molecular orbitals such that they correspond more closely to conventional Lewis structures. Because this is costly in terms of computation and to some extent "ill defined" (the need to specify what conditions are to be met) localization is only rarely performed.

The ***Spartan***-generated image depicts the same orbital as a surface of constant value. The surface is accurate in that it is derived from an authentic (but approximate) calculated solution to the quantum mechanical equations of electron motion (the Schrödinger equation). Equally important, the image is three-dimensional and can be manipulated and looked at from a variety of different perspectives. It is not unexpected that the orbital consists of two distinct surfaces represented by different colors. The two surfaces have the same meaning as the two circles in the orbital drawing. They identify regions where the orbital takes on a significant value, either positive (blue) or negative (red). The orbital node is not shown, but we can guess that it lies midway between the two surfaces (this follows from the fact that the orbital's value can only change from positive to negative by passing through zero).

Atomic Orbitals

Atomic orbitals (descriptions of atoms) are the fundamental building blocks from which molecular orbitals (descriptions of molecules) are assembled. The familiar atomic orbitals for the hydrogen atom are in fact exact solutions of the Schrödinger equation which can actually be solved exactly for this one electron system. They form an infinite collection (a complete set), the lowest-energy member representing the best location for the electron, and higher-energy members representing alternative locations. Orbitals for real many-electron atoms are normally (and necessarily) assumed to be similar in form to those of hydrogen atom, the only difference being that, unlike hydrogen, more than the lowest-energy atomic orbital is utilized. In practical quantum chemical calculations, atomic orbitals for many-electron atoms are made up of sums and differences of a finite collection of hydrogen-like orbitals (see the topic ***Theoretical Models***).

It is common practice to divide the full set of atomic orbitals into core and valence orbitals, and further to ignore the former. Valence orbitals for an element in the first long row of the *Periodic Table* are 2s, $2p_x$, $2p_y$ and $2p_z$, and for the second long row are 3s, $3p_x$, $3p_y$ and $3p_z$. In the case of first-row elements, a single orbital, 1s, lies underneath (is a core orbital) while in the case of second-row elements, a set of five orbitals, 1s, 2s, $2p_x$, $2p_y$ and $2p_z$, lie underneath.

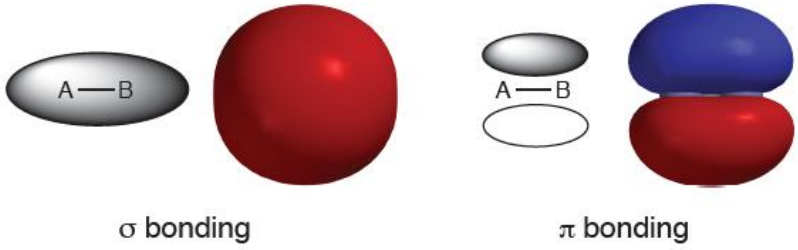
Open ***fluoride and chloride*** in the ***topics*** directory. On the top row are the four valence orbitals of fluoride anion and on the bottom row the four valence orbitals of chloride anion. You can select among them by *clicking* (left mouse button) on each in turn. First note that the three 2p orbitals in fluoride are identical except for the direction in which they point. The same is true for the three 3p orbitals in chloride. Next, note that the valence orbitals in chloride are larger than those in fluoride. Atoms further down in the *Periodic Table* are generally larger than analogous atoms further up. Close ***fluoride and chloride*** when you are finished.

Orbitals and Chemical Bonds

Although molecular orbitals and Lewis structures are both used to describe electron distributions in molecules, they are used for different purposes. Lewis structures are used to count the number of bonding and non-bonding electrons around each atom. Molecular orbitals are not useful as counting tools, but orbitals and associated orbital energies are useful tools for describing chemical bonding and reactivity. This section describes a few common orbital shapes and illustrates their use.

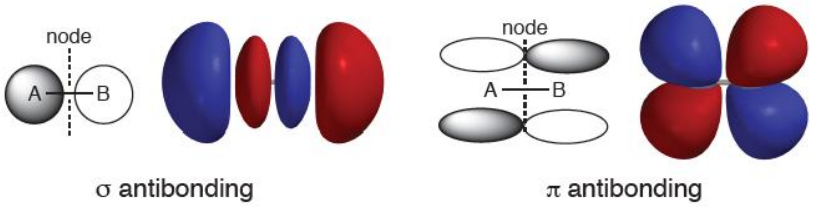
Molecular orbital surfaces can extend over varying numbers of atoms. If the orbital surface (or surfaces) is confined to a single atom or to atoms which are not close together, the orbital is regarded as non- bonding. If the orbital contains a surface that extends continuously over two neighboring atoms, the orbital is regarded as bonding with respect to these atoms. Adding electrons to such an orbital will strengthen the bond between these atoms and cause them to draw closer together, while removing electrons will have the opposite effect. Two different kinds of bonding orbitals are depicted below.

The drawing and surface on the left correspond to a σ bond while the drawing and surface on the right correspond to a π bond.



Open ***nitrogen bonding*** in the ***topics*** directory. The image on the left corresponds to the σ bonding orbital of N_2, while that on the right corresponds to one of two equivalent π bonding orbitals. Switch to a transparent or mesh model to see the underlying molecular skeleton. Note that the σ orbital is drawn in a single color (insofar as NN bonding is concerned) while the π orbital is made up of red and blue parts. This indicates a node or a break in the latter, although not involving the NN bond. Close ***nitrogen bonding*** when you are finished.

It is also possible for an orbital to contain a node that divides the region between two neighboring atoms into separate atomic regions. Such an orbital is regarded as antibonding with respect to these atoms. Adding electrons to an antibonding orbital weakens the bond and pushes the atoms apart, while removing electrons from such an orbital has the opposite effect. The following pictures show drawings and orbital surfaces for two different kinds of antibonding orbitals. As above, the left and right-hand sides correspond to σ and π type arrangements, respectively.



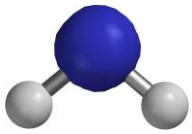
Open ***nitrogen antibonding*** in the ***topics*** directory. The image on the left corresponds to the σ antibonding (so-called σ*) orbital of N_2 while that on the right corresponds to one of the two equivalent π antibonding (so-called π*) orbitals. Switch to a mesh or transparent surface to see the underlying molecular skeleton. Note that the σ* orbital has a single node (change in color from red to blue) in the middle of the NN bond, while the π* orbital has two nodes (one in the middle of the NN bond and to the other along the bond). Close ***nitrogen antibonding*** when you are finished.

To summarize, bonds can be strengthened in two different ways, by adding electrons to bonding orbitals, or by removing electrons from antibonding orbitals. The converse also holds. Bonds can be weakened either by removing electrons from bonding orbitals or by adding electrons to antibonding orbitals.

Singlet Methylene

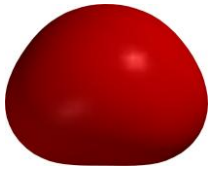
Molecular orbitals in molecules which contain many atoms are typically spread throughout the molecule (they are delocalized). Delocalized orbitals have complicated shapes and contain multiple interactions that may be bonding, non-bonding, antibonding, or any mixture of all three. Nevertheless, these shapes can usually be broken down into two-atom interactions and analyzed using the principles outlined earlier. This process is illustrated for a triatomic molecule, singlet methylene, CH_2. (Singlet refers to the fact that the eight electrons in this highly reactive molecule are organized into four pairs, and that each pair of electrons occupies a different molecular orbital. The lowest-energy state of methylene is actually a triplet with three electron pairs and two unpaired electrons.)

The lowest energy molecular orbital of singlet methylene is not very interesting in that it looks like a 1s atomic orbital on carbon. The electrons occupying this orbital restrict their motion to the immediate region of the carbon nucleus and do not significantly affect bonding. Because of this restriction, and because the orbital's energy is very low, this orbital is referred to as a ***core orbital***, and its electrons are referred to as ***core electrons***.



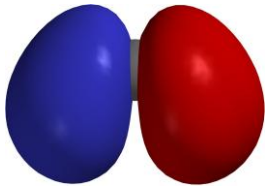
core orbital for methylene

The next orbital is much higher in energy. It consists of a single surface that is delocalized over all three atoms. This means that it is simultaneously (σ) bonding with respect to each CH atom pair.



core orbital for methylene

The next higher energy orbital is described by two surfaces, a positive (blue) surface that encloses one CH bonding region and a negative (red) surface that encloses the other CH bonding region*. Since each surface encloses a bonding region, this orbital is also (σ) bonding with respect to each CH atom pair. This reinforces the bonding character of the previous orbital. The node that separates the two surfaces passes through the carbon nucleus, but not through either of the CH bonding regions, so it does not affect bonding.



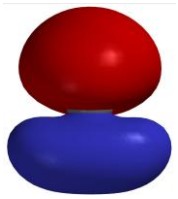
core orbital for methylene

Thus, the two CH bonds in the Lewis structure for singlet methylene are replaced by two bonding molecular orbitals.**

* While the absolute signs (colors) of a molecular orbital are arbitrary, the relative signs (colors) indicate bonding and antibonding character.

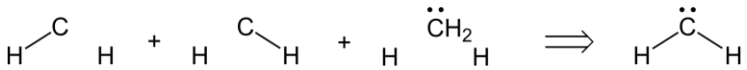
** Unlike the two Lewis structures, the two CH bonding molecular orbitals reflect the fact that the two CH bonds in methylene are identical. Specifically, the square of each of the orbitals, corresponding to the electron distribution or electron density, has the same 2-fold symmetry as the molecule. The square of either the molecular orbitals themselves or a combination of so-called degenerate (same energy) molecular orbitals must have the same symmetry as the underlying nuclear skeleton.

The highest-occupied molecular orbital (the HOMO) is also described by two orbital surfaces. One surface extends into carbon's non-bonding region opposite the two hydrogens. The other surface encompasses the two CH bonding regions. Although it is hard to track the exact path of the orbital node in this picture, it happens to pass almost exactly through the carbon. This means that this particular orbital possesses only weak CH bonding character (it is H---H bonding). It turns out that the non-bonding character of the orbital is much more important than the bonding character, in that it leads to the fact that singlet methylene is able to behave as an electron-pair donor (a nucleophile).



core orbital for methylene

In addition to two CH bonds, the Lewis structure for singlet methylene shows a lone pair on carbon.

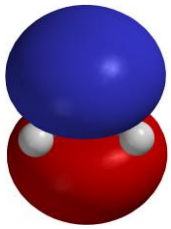


The above analysis shows that while the occupied orbitals of singlet methylene are spread over all three atoms, they are comprehensible. The orbitals divide into two groups, a single low-energy core orbital and three higher-energy valence orbitals. The latter consist of two CH bonding orbitals and a non-bonding orbital on carbon. There is no one- to-one correspondence between these orbitals and the Lewis structure. The bonding orbitals are not associated with particular bonds, and the non-bonding orbital contains bonding interactions as well.

> Open ***methylene bonding*** in the ***topics*** directory. Four images appear corresponding to the core and three valence orbitals of singlet methylene. Switch to a mesh or transparent surface to see the underlying molecular skeleton. Close ***methylene bonding*** when you are finished.

Singlet methylene also possesses unoccupied molecular orbitals. The unoccupied orbitals have higher (more positive) energies than the occupied orbitals, and these orbitals, because they are unoccupied, do not describe the electron distribution in singlet methylene.* Nevertheless, the shapes of unoccupied orbitals, in particular, the lowest-unoccupied orbital (LUMO), is worth considering because it provides valuable insight into the methylene's chemical reactivity.

The LUMO in methylene has non-bonding character and looks like a 2p atomic orbital on carbon. This suggests that singlet methylene should be able to behave as an electron-pair acceptor (an electrophile). Note, however, that were the molecule to accept electrons, these would go into non-bonding orbital; carbon would become more electron- rich, but the CH bonds would not be much affected.



LUMO of methylene

Open ***methylene LUMO*** in the ***topics*** directory and switch to a mesh or transparent surface to see the underlying skeleton. Close ***methylene LUMO*** when you are finished.

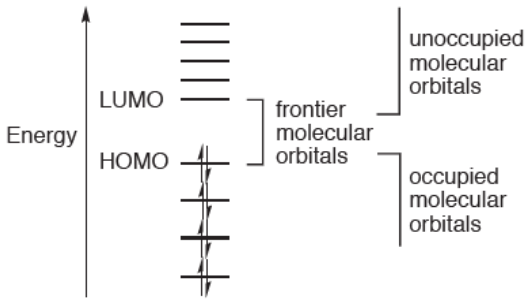
Frontier Molecular Orbitals and Chemical Reactivity

Chemical reactions often involve movement of electrons from an electron donor (base, nucleophile, reducing agent) to an electron acceptor (acid, electrophile, oxidizing agent). This electron movement between molecules can also be thought of as electron movement between molecular orbitals, and the shapes and energies of orbitals that act as electron donors or electron acceptors may provide considerable insight into chemical reactivity.

The first step in constructing a molecular orbital picture of a chemical reaction is to decide which orbitals are most likely to act as electron donors and acceptors. It is obvious that an electron-donor orbital must be drawn from the set of occupied orbitals, and an electron-acceptor orbital must be an unoccupied orbital, but there are many orbitals in each set to choose from.

Orbital energy is usually the deciding factor. The highest-energy occupied orbital (the HOMO) is most commonly assumed to be the relevant electron-donor orbital, and the lowest-energy unoccupied orbital (the LUMO) is most commonly assumed to be the relevant electron-acceptor orbital. For example, the HOMO and LUMO of singlet methylene (σ and π non-bonding orbitals, respectively) would serve as the donor and acceptor orbitals. The HOMO and LUMO are collectively referred to as the ***frontier molecular orbitals***, and most chemical reactions involve electron movement between them. In this way, the energy input required for electron movement is kept to a minimum.

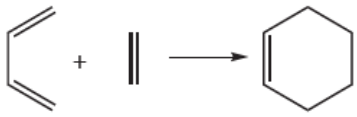
* Because Lewis structures describe electron pair bonds and non-bonding electron pairs, they may not be related to unoccupied molecular orbitals.



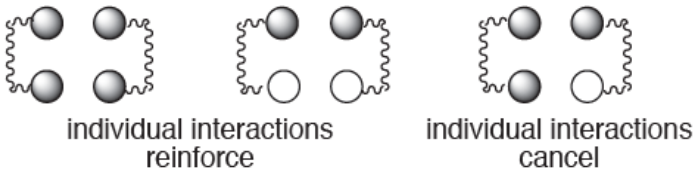
Closely related to chemical reactivity (reaction rate) is chemical selectivity. The relevant question is, where more than one combination of reagents can react, which combination will react more quickly? The answer can often be found by examining the energies of the frontier orbitals. Consider ranking the rates of a series of reagents, where chemical reaction requires electron donation from the donor's HOMO. It is reasonable to expect that the donor with the highest energy HOMO will give up its electrons most easily and be the most reactive. Electron-acceptor reagents should follow the opposite pattern. The reagent with the lowest energy LUMO should be able to accept electrons most easily and be the most reactive. For a mixture of several donor and acceptor reagents, the fastest chemical reaction would be expected to involve the reagent combination that yields the smallest HOMO-LUMO energy gap.

The Fukui-Woodward-Hoffmann Rules

In certain cases, multiple frontier orbital interactions must be considered. A good example is provided by so-called cycloaddition reactions, such as the Diels-Alder reaction between 1,3-butadiene and ethylene.



The key feature of this reaction is that the reactants combine in a way that allows two bonds to form simultaneously. This implies two different sites of satisfactory frontier orbital interaction (the two new bonds that form are sufficiently far apart that they do not interact with each other during the reaction). If we focus exclusively on the interactions of the terminal carbons in each molecule, then three different frontier orbital combinations made up of upper and lower components can be imagined.

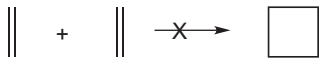


The upper orbital components are the same sign in all three combinations, meaning that their overlap is positive. In the two cases on the left, the lower orbital components also lead to positive overlap. Thus, the two interactions reinforce, and the total frontier orbital interaction is non-zero. Electron movement (leading to chemical reaction) can occur. The right-most case is different. Here the lower orbital components lead to negative overlap (the orbitals have opposite signs at the interacting sites), and the total overlap is zero. No electron movement and no chemical reaction can occur in this case.

As it happens, the frontier orbital interactions in the Diels-Alder cycloaddition shown above correspond to those found in the middle drawing, that is, the upper and lower interactions reinforce and the reaction proceeds.

> Open ***1,3-butadiene+ethylene*** in the ***topics*** directory. The image on top is the LUMO of ethylene while that on the bottom is the HOMO of 1,3-butadiene. They are properly poised to interact, but you can manipulate them independently. Close ***1,3-butadiene+ethylene*** when you are finished.

The same arguments suggest that cycloaddition of two ethylene molecules is unlikely to occur. This is because it involves a frontier orbital interaction like that found in the right drawing.



> Open ***ethylene+ethylene*** in the ***topics*** directory. The image on top corresponds to the LUMO of one ethylene while that on the bottom corresponds to the HOMO of the other ethylene. You can manipulate them independently or in concert (hold down on the **Ctrl** key while you carry out rotation and translation). Note, that in this case, the two individual atom-atom interactions cancel. Close ***ethylene+ethylene*** when you are finished.

The importance of orbital-overlap in determining why certain chemical reactions proceed easily while other seemingly similar reactions are sluggish or do not go at all was first advanced by Fukui and then beautifully elaborated by Woodward and Hoffmann, and collectively their ideas are now known as the Fukui-Woodward-Hoffmann rules.

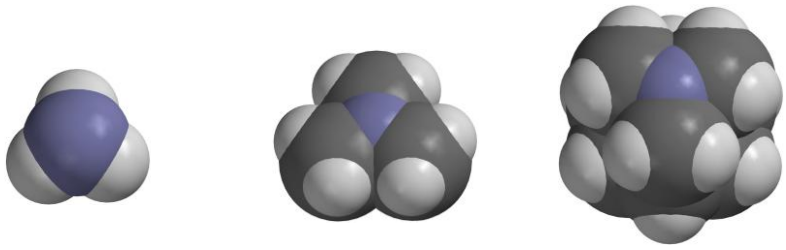
ELECTRON DENSITIES: SIZES AND SHAPES OF MOLECULES

How big is an atom or a molecule? Atoms and molecules require a certain amount of space, but how much? We know that gases can be compressed into a smaller volume but only so far, and that liquids and solids are very difficult is not nearly impossible to compress. While the individual atoms or molecules in a gas are widely separated and can be pushed into a much smaller volume, the atoms or molecules in a liquid or a solid are already close together and cannot be squeezed much further. Atoms can be talked about as having well-defined size. What is it?

Space-Filling Models

Chemists have long tried to answer the size question by using a special set of molecular models known as space-filling or CPK models. The space-filling model for an atom is simply a sphere of fixed radius. A different radius is chosen for each element in order to reproduce certain experimental observations, such as the compressibility of a gas, or the spacing between atoms in a crystal. Space-filling models for molecules consist of a set of interpenetrating atomic spheres. This reflects the idea that the chemical bonds that hold the molecule together cause the atoms to move closer together than the sum of the radii for the individual space-filling models. Interpenetration can be used as a criterion for chemical bonding. If two atomic spheres in a space-filling model strongly interpenetrate then the atoms must be bonded. Space-filling models not only show molecule size but also show which parts of the molecule are shielded and which are exposed.

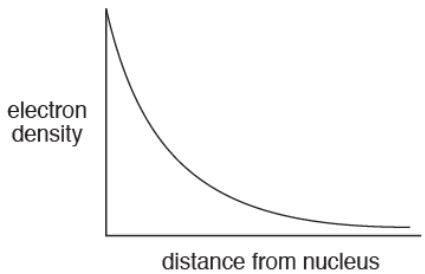
space-filling models for ammonia (left), trimethylamine (center) and 1-azaadamantane (right)



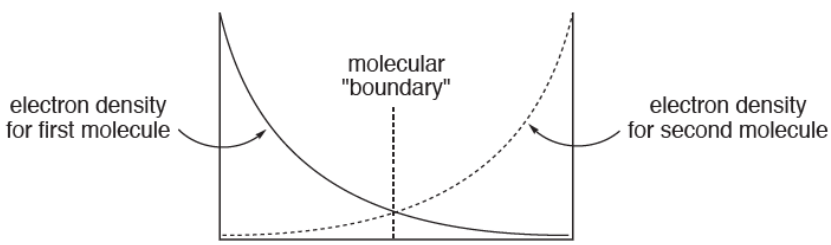
Open ***amines space filling*** in the ***topics*** directory. Space-filling models for ammonia, trimethylamine and 1-azaadamantane all appear on screen. Carbon atoms are colored dark grey, hydrogen atoms white and nitrogen blue. Note that the models clearly reveal the extent to which the nitrogen is exposed. Close ***amines space filling*** when you are finished.

Electron Density Surfaces

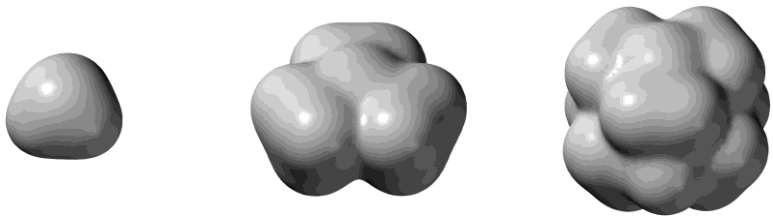
An alternative technique for portraying molecular size and shape relies on the molecule's own electron cloud. Atoms and molecules are made up of positively-charged nuclei surrounded by a negatively- charged electron cloud, and it is the size and shape of the electron cloud and not that of the nuclear skeleton that defines the size and shape of an atom or molecule. The size and shape of an electron cloud is described by the electron density (the number of electrons per unit volume). Consider a graph of electron density in the hydrogen atom as a function of distance from the nucleus.



The graph brings up a problem for chemists seeking to define atomic and molecular size, in that the electron cloud lacks a clear boundary. While electron density decays rapidly with distance from the nucleus, nowhere does it fall to zero. Therefore, when atoms and molecules rub up against each other, their electron clouds overlap and merge to a small extent.



This suggests that molecular size and shape is ill defined and that the best that can be done is to pick a value of the electron density, and to connect together all the points that have this value. The criteria for selecting this value is exactly the same as that for selecting atomic radii in space-filling models, the only difference being that only a single parameter (the value of the electron density) is involved (rather than a different radius for each element). The result is an electron density surface which, just like a space-filling model, is intended to depict overall molecular size and shape.



electron density surfaces for ammonia (left), trimethylamine (center) and 1-azaadamantane (right)

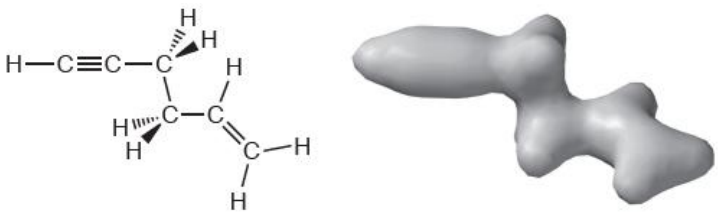
Open ***amines electron density*** in the ***topics*** directory. Electron density surfaces for ammonia, trimethylamine and 1-azaadamantane all appear on screen. Switch to a mesh or transparent surface in order to see the underlying skeletal model. *Click* on one of the surfaces and select **Mesh** or **Transparent** from the menu which appears at the bottom right of the screen. With mesh selected, change the model to **Space Filling** (**Model** menu). This allows you to see how similar the electron density representation is to that offered by a simple space-filling model. Close ***amines electron density*** when you are finished.

Both space-filling models (atomic radii) and electron density models can be chosen to yield similar molecular volumes, and both show differences in overall size among molecules. Because the electron density surfaces provide no discernible boundaries between atoms, the surfaces may appear to be less informative than space-filling models in helping to decide to what extent a particular atom is exposed. This raises an important point. Electrons are associated with a molecule as a whole and not with individual atoms. The space-filling representation of a molecule with its discernible atomic boundaries does not reflect reality.

Bond Density Surfaces

More closely representing a conventional Lewis structure is a so- called ***bond density surface***, where the boundary corresponds to a much higher value of the electron density*. As such a surface is located much closer to the atomic nuclei, it encloses a relatively small volume and does not give a correct impression of molecular size.

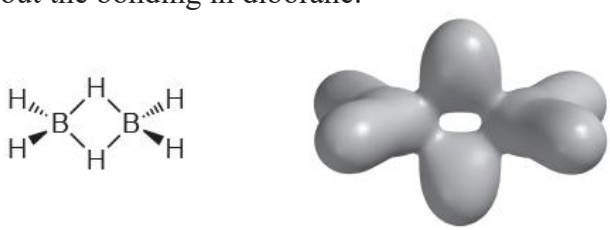
The bond density surface for hex-5-ene-1-yne clearly shows which atoms are connected, although it does not clearly distinguish single, double and triple carbon-carbon bonds.



bond density surface for hex-5-ene-1-yne

Open ***hex-5-ene-1-yne bond density*** in the ***topics*** directory, and switch to a mesh or transparent surface to see the connection between the chemical bonds in a conventional model and the electron density. Close ***hex-5-ene-1-yne bond density*** when you are finished.

The usefulness of the bond density surface is more apparent in the following model of diborane. The surface clearly shows that there is relatively little electron density between the two borons, suggesting the absence of a boron-boron bond. This is information extracted from the bond density surface model and has been obtained without reference to any preconceived ideas about the bonding in diborane.

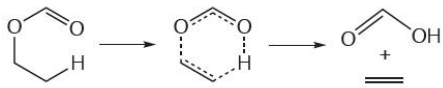


bond density surface for diborane

Open ***diborane bond density*** in the ***topics*** directory, and switch to a mesh or transparent surface to see how few electrons actually accumulate in the region between the two borons. Close ***diborane bond density*** when you are finished.

* An even higher value of the electron density leads to a surface in which only nearly spherical regions of electron density around the non-hydrogen atoms are portrayed. This serves to locate the positions of these atoms and is the basis of the X-ray diffraction experiment.

Bond density surfaces can also be informative in describing changes in bonding in moving from reactants to products through a transition state in a chemical reaction. For example, heating ethyl formate causes the molecule to fragment into two new molecules, formic acid and ethylene. A line drawing can show which bonds are affected by the overall reaction, but it cannot tell us if these changes occur all at once, sequentially, or in some other fashion.



On the other hand, the bond density surface is able to provide quantitative information.

bond density surfaces for the reactant, ethyl formate (left), pyrolysis transition state (center) and for the products, formic acid and ethylene (right)

Comparison of the bond density surface in the pyrolysis transition state with those of the reactant and the products suggests that the CO single bond of the reactant is clearly broken in the transition state and that the migrating hydrogen seems more tightly bound to oxygen (as in the product) than to carbon (as in the reactant). It can be concluded that the transition state more closely resembles the products than the reactants, and this provides an example of what chemists refer to as a late (product-like) transition state.

To see the smooth change in electron density throughout the course of the ethyl formate pyrolysis reaction, open ***pyrolysis bond density*** in the ***topics*** directory. *Click* on ▶ at the bottom left of the screen to animate the graphic (*click* on ⏸ to stop the animation). Switch to a mesh or transparent surface to follow the change in bonding. Close ***pyrolysis bond density*** when you are finished.

ELECTROSTATIC POTENTIAL MAPS: CHARGE DISTRIBUTIONS

The charge distribution in a molecule can provide critical insight into its physical and chemical properties. For example, molecules that are charged, or highly polar, tend to be water-soluble, and polar molecules may stick together in specific geometries, such as the double helix in DNA. Chemical reactions are also associated with charged sites, and the most highly-charged molecule, or the most highly-charged site in a molecule, is often the most reactive. The sign of the charge is also important. Positively-charged sites in a molecule invite attack by bases and nucleophiles, while negatively-charged sites are usually targeted by acids and electrophiles.

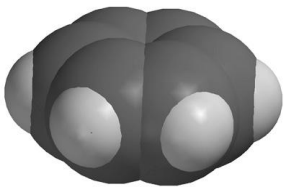
One way to describe a molecule's charge distribution is to give a numerical atomic charge for each atom. A particularly simple and familiar recipe yields so-called formal charges directly from Lewis structures. Unfortunately, formal charges are arbitrary. In fact, all methods for assigning charge are arbitrary and necessarily bias the calculated charges in one way or another. This includes methods based on quantum mechanics. The reason may be traced back to the notion that atomic boundaries are themselves ill-defined and it is impossible to decide which electrons are associated with which nuclei.

An attractive alternative for describing molecular charge distributions makes use of a quantity termed the electrostatic potential. This is the energy of interaction of a point positive charge with the nuclei and electrons of a molecule. The value of the electrostatic potential depends on the location of the point positive charge. If the point charge is placed in a region of excess positive charge (an electron- poor region), the point charge-molecule interaction is repulsive, and the electrostatic potential will be positive. Conversely, if the point charge is placed in a region of excess negative charge (an electron-rich region), the interaction is attractive and the electrostatic potential will be negative. Thus, by moving the point charge around the molecule, the molecular charge distribution can be surveyed.

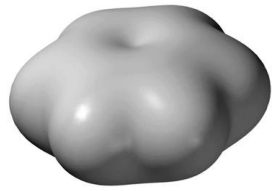
Electrostatic potentials can be depicted in various ways. For example, it is possible to make an electrostatic potential surface by finding all of the points in space where the electrostatic potential matches some particular value. This is precisely what was previously done for both molecular orbitals and electron densities. A much more useful way to show molecular charge distribution is to construct a so-called electrostatic potential map. This is done first by constructing an electron density surface corresponding to a space-filling model (see the topic ***Electron Densities: Sizes and Shapes of Molecules***). The electrostatic potential is then mapped onto this surface using different colors to represent the different values of the electrostatic potential. Mapping requires an arbitrary choice for a color scale. ***Spartan*** uses the rainbow. Red, the low energy end of the spectrum, depicts regions of most negative (least positive) electrostatic potential, and blue depicts the regions of most positive (least negative) electrostatic potential. Intermediate colors represent intermediate values of the electrostatic potential, so that potential increases in the order:

red < **orange** < **yellow** < **green** < **blue**

The connection between a molecule's electron density surface, its electrostatic potential surface, and an electrostatic potential map is illustrated below for benzene. The electron density surface defines molecular shape and size and performs the same function as a conventional space-filling model by indicating how close two benzenes can get in a liquid or in a crystal.



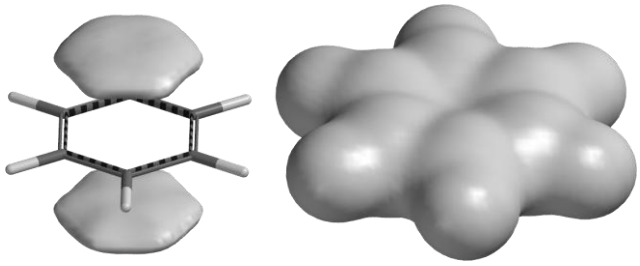
space-filling model
for benzene



electron density surface
for benzene

Open ***benzene electron density*** in the ***topics*** directory. Two different images, a space-filling model and an electron density surface, appear side by side. You can switch between the two models (in order to manipulate them) by *clicking* on each in turn. Notice how similar they are. To get an even clearer impression, switch to a mesh surface. *Click* on the graphic and select **Mesh** from the menu which appears at the bottom right of the screen. Switch to a space-filling model (**Space Filling** from the **Model** menu). The two are now superimposed. Close ***benzene electron density*** when you are finished.

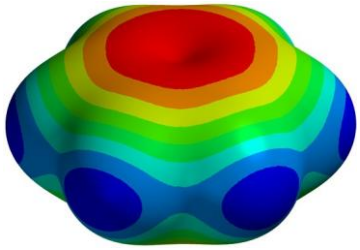
The electrostatic potential corresponding to points where the potential is negative shows two different surfaces, one above the face of the ring and the other below. Since the molecule's π electrons lie closest to these surfaces, we conclude that these electrons are responsible for the attraction of a point positive charge (or an electrophile) to the molecule. An electrostatic potential surface corresponding to points where the potential is positive has a completely different shape. It is disk-shaped and wrapped fairly tightly around the nuclei. The shape and location of this surface indicates that a point positive charge is repelled by this region, or that a point negative charge (a nucleophile) would be attracted here.



negative (left) and positive (right) electrostatic potential surfaces for benzene

Open ***benzene electrostatic potential*** in the ***topics*** directory. Two different images appear side by side. The one on the left depicts a surface of constant negative potential, while the one on the right depicts a surface of equal positive potential. You can switch between the two models (in order to manipulate them) by *clicking* on each in turn. Close ***benzene electrostatic potential*** when you are finished.

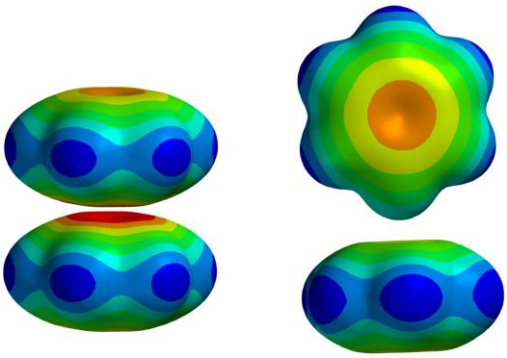
Next, combine the electron density and electrostatic potential surfaces to produce an electrostatic potential map. This simultaneously conveys both the molecule's size and shape as well as its charge distribution in a compact and easily interpretable manner. The size and shape of the map are, of course, identical to that of the electron density surface, and indicates what parts of the molecule are easily accessible to other molecules (the outside world). The colors reveal the overall charge distribution. The faces of the ring, the π system, are red (electron rich), while the plane of the molecule and especially the hydrogens are blue (electron poor).



electrostatic potential map for benzene

Open ***benzene electrostatic potential map*** in the ***topics*** directory. Manipulate the image to convince yourself that the red regions are on the π faces and the blue regions are around the edges. Close ***benzene electrostatic potential map*** when you are finished.

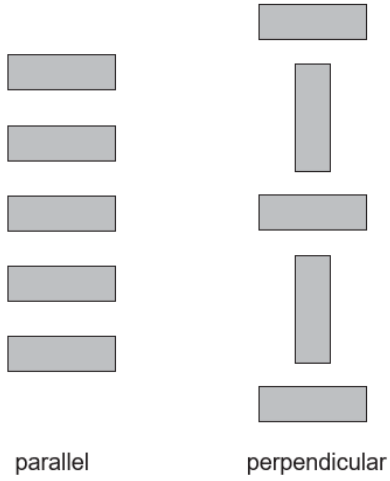
Electrostatic potential maps have made their way into mainstream general and (especially) organic chemistry textbooks as a means of displaying charge distributions in molecules. In addition, they have found application as a natural step beyond steric models for interpreting and predicting the way in which molecules fit together. A good example of this follows from the electrostatic potential map for benzene, which recall is negative on the π faces and positive around the periphery. The benzene dimer would, therefore, be expected to exhibit a perpendicular geometry, to best accommodate Coulombic interactions, instead of a parallel arrangement.



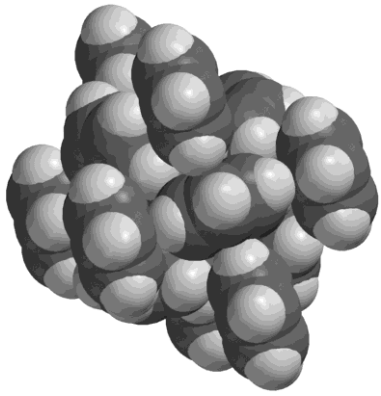
electrostatic potential maps for parallel (left) and perpendicular (right) benzene dimers

Open ***benzene dimer electrostatic potential map*** in the ***topics*** directory. Note that the parallel arrangement forces the negative region of one benzene onto the negative region of the other, while the perpendicular structure associates the negative region of one benzene with a positive region of the other. Close ***benzene dimer electrostatic potential map*** when you are finished.

Of greater interest is the structure of benzene in solid state. Intuition suggests a parallel stack. After all, benzene is flat and flat things (plates, pancakes, hamburgers) make stacks. However, Coulombs law clearly favors a perpendicular arrangement.



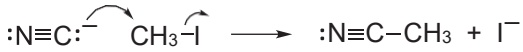
The experimental X-ray crystal structure indeed shows a perpendicular arrangement, albeit in three dimensions. There are two lessons here. Intermolecular interactions go beyond steric interactions and sometimes our simple one and two-dimensional views of the world will lead us astray.



X-ray crystal structure of benzene

Open ***benzene crystal*** in the ***topics*** directory. Manipulate in order to see the packing of benzene molecules. Close ***benzene crystal*** when you are finished.

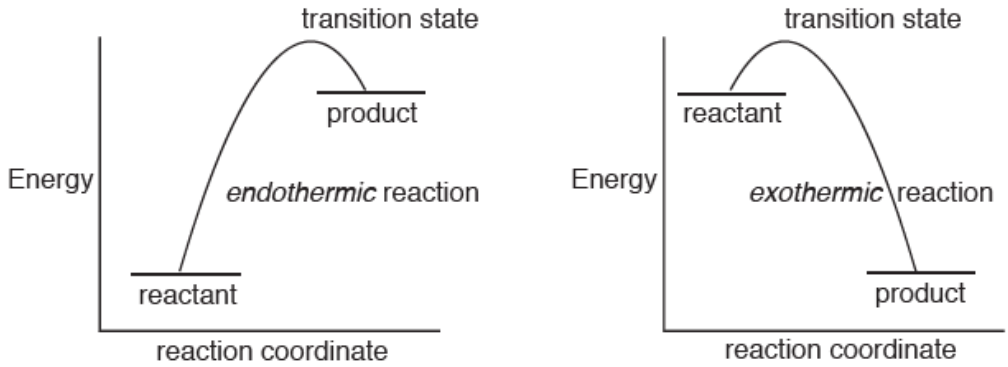
Electrostatic potential maps may also be used to describe in great detail the workings of chemical reactions. For example, a map may be used to show the transfer of negative charge during the S_n2 reaction of cyanide with methyl iodide.



Open ***Sn2 cyanide+methyl iodide*** in the ***topics*** directory. One frame of a sequence of electrostatic potential maps for the S_n2 reaction will appear. Animate by *clicking* on ▶ at the bottom left of the screen (stop the animation by *clicking* on ⏸). Note that the negative charge (red color) flows smoothly from cyanide to iodide during the reaction. Note also, that cyanide (as the reactant) is redder than iodide (as the product). Iodide is better able to carry negative charge, that is, it is the better leaving group. Switch to mesh or transparent map to see the making and breaking of bonds. Close ***Sn2 cyanide+methyl iodide*** when you are finished.

LOCAL IONIZATION POTENTIAL MAPS AND LUMO MAPS: ELECTROPHILIC AND NUCLEOPHILIC REACTIVITIES

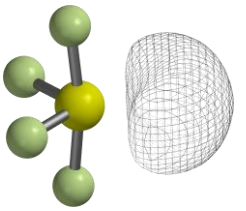
The Hammond Postulate states that the transition state in a (one-step) reaction will more closely resemble the side of the reaction that is higher in energy. Thus, the transition state of an *endothermic* reaction will more closely resemble the products, and the transition state of an *exothermic* reaction will resemble the reactants. One way to rationalize the Hammond postulate is to suggest that similarity in energy implies similarity in structure. That is, the transition state will resemble whichever reactants or products to which it is closer in energy. As seen in the reaction coordinate diagrams below this is the product in an *endothermic* reaction and the reactants in an *exothermic* reaction.



The Hammond Postulate provides a conceptual basis both for the Fukui- Woodward-Hoffmann rules (see the topic ***Atomic and Molecular Orbitals***) and for the use of graphical models. Both consider the properties of reactants as an alternative to direct calculations of transition states and reaction pathways as a way to assess chemical reactivity and selectivity. In this context, two models stand out as being particularly useful: the local ionization potential map for electrophilic reactions and (with some caveats) the LUMO map for nucleophilic reactions.

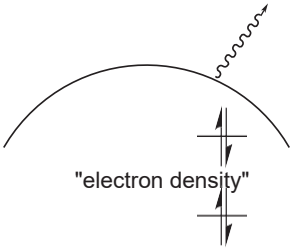
Local Ionization Potential Maps and Electrophilic Reactivity

The local ionization potential provides a measure of the relative ease of electron removal (ionization) at any location around a molecule. For example, a surface of low local ionization potential for sulfur tetrafluoride demarks the areas which are most easily ionized and is clearly recognizable as a lone pair on sulfur.

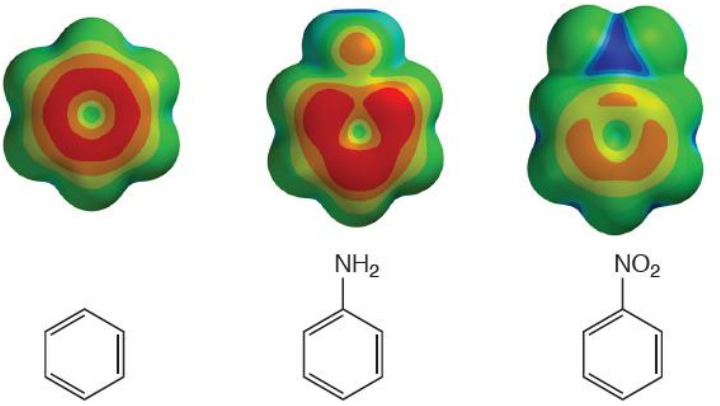


local ionization potential for sulfur tetrafluoride

While the local ionization potential by itself is not a particularly useful model, a map of the local ionization potential onto an electron density surface (a local ionization potential map) has proven to be of value in revealing regions from which electrons are most easily ionized and susceptible to electrophilic attack.



For example, they show both the positional selectivity in electrophilic aromatic substitution (NH_2 directs *ortho*/*para*, and NO_2 directs *meta*), and the fact that π-donor groups (NH_2) activate benzene while electron-withdrawing groups (NO_2) deactivate benzene.

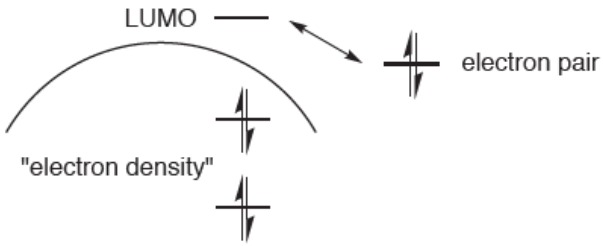


local ionization potential maps for benzene (left), aniline (center) and nitrobenzene (right)

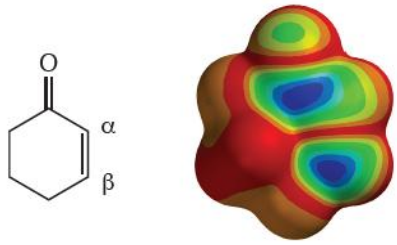
Open ***benzene, aniline, nitrobenzene local ionization potential maps*** in the ***topics*** directory. Here, the color red corresponds to regions of lowest ionization potential (most accessible to electrophiles). Note, that the π system in aniline is more accessible than the π system in benzene (the standard) and that the *ortho* and *para* positions are more accessible than *meta* positions. Note also, that the π system in nitrobenzene is less accessible than the π system in benzene and that here the *meta* positions are more accessible than the *ortho* and *para* positions, in accord with observation. Close ***benzene, aniline, nitrobenzene local ionization potential maps*** when you are finished.

LUMO Maps and Nucleophilic Reactivity

As elaborated in the topic ***Atomic and Molecular Orbitals***, the energies and shapes of the frontier molecular orbitals may anticipate chemistry of a molecule. Molecular orbital maps may also lead to informative models. The most common of these is the so-called LUMO map, in which the (absolute value) of the LUMO is mapped onto an electron density surface. This anticipates where an electron pair (a nucleophile) might attack.

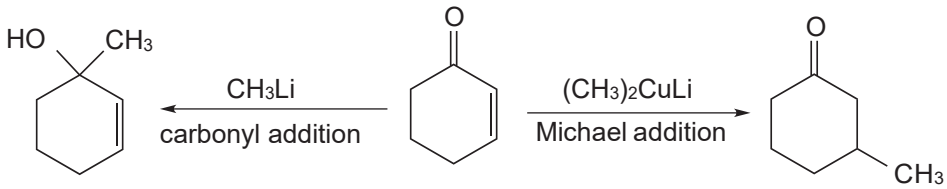


A good example is provided by the LUMO map for cyclohexenone.



LUMO map for cyclohexenone showing positions for nucleophilic attack in blue

The LUMO shows which regions most electron deficient, and hence most subject to nucleophilic attack. In this case, one such region is over the carbonyl carbon, consistent with the observation that carbonyl compounds undergo nucleophilic addition at the carbonyl carbon. Another region is over the β carbon, again consistent with the known chemistry of α,β-unsaturated carbonyl compounds, in this case conjugate or Michael addition.


HO
CH3
CH3Li
carbonyl addition
O
(CH3)2CuLi
Michael addition
O
CH3

The buildup of positive charge on the β carbon leading to possibility of Michael addition could have been anticipated from resonance arguments, although the LUMO map, like an experiment, has the advantage of showing the result ("you don't have to ask").

> Open ***cyclohexenone LUMO map*** in the ***topics*** directory and *click* on the resulting surface. Switch to a mesh or transparent surface in order to see the underlying skeletal model. On Windows, *click* on the graphic and select **Mesh** or **Transparent** from the menu which appears at the bottom right of the screen. On the Mac, position the cursor over the graphic, hold down on either the left or right button and select from the menu which appears alongside. Orient the molecule such that the two "blue regions" are positioned over the appropriate carbons and then switch back to a solid surface. Close ***cyclohexenone LUMO map*** when you are finished.

Some caution needs to be exercised. Specifically, the LUMO may not be the important molecular orbital for the chemical reaction of interest, and it may be necessary to base the map on a higher-energy unoccupied molecular orbital (LUMO+1, LUMO+2, ...).

Appendix B
Capabilities and Limitations

Molecular Mechanics Models

The molecular mechanics module calculates the energy (a combination of strain energy and intramolecular interaction energy), equilibrium geometry and vibrational frequencies, as well as provides for conformational searching*. SYBYL (Tripos, Inc.) and MMFF** (Merck Pharmaceuticals) force fields are available. There are no atom limits.

Semi-Empirical Models

The semi-empirical module calculates the heat of formation, wave function, equilibrium and transition-state geometries and vibrational frequencies. MNDO, AM1, RM1, PM3*** and PM6 models are supported. MNDO/d replaces MNDO for second-row (and heavier) main-group elements. MNDO has been parameterized for H, He, Li-F, Al-Cl, Ca, Zn, Ge, Br, Cd, Sn, I, Hg and Pb; AM1 for H, B-F, Al-Cl, Zn, Ge, Br, Sn and I; RM1 for H, C–F, P, S, Cl, Br and I; PM3 for H, Li-F, Mg-Cl, Ca, Ti-Br, Zr, Mo-Pd, Cd-I, Hf-Pt and Hg-Bi and Gd; PM6 for H–Bi except for Ce–Yb. All semi-empirical models involve a minimal valence basis set.

The preset limits for all semi-empirical calculations is 300 atoms.

Hartree-Fock Models

The Hartree-Fock module calculates the energy and wave function, equilibrium and transition-state geometries and vibrational frequencies. Both closed and open-shell (either restricted, ROHF, or unrestricted, UHF) calculations are supported. Available basis sets are enumerated later in this appendix. Hartree-Fock models have been implemented for use on multi-core processors.

Preset limits for Hartree-Fock calculations are 200 atoms and 2,000 basis functions. Calculations on molecules with up to 100 atoms are practical.

---

* MMFF is the only method available for generation of conformer libraries.

** The published parameters for MMFF have been extended to all elements in the Periodic Table. While it can be applied to any molecule, it is likely to perform poorly for molecules incorporating transition metals, lanthanides and actinides.

*** The published parameters for PM3 has been extended to transition metals. Except for Zn, Cd and Hg, these include d-type orbitals in the basis set.

Density Functional Models

The density functional module calculates the energy and wave function, equilibrium and transition-state geometries and vibrational frequencies, as well as provides for conformational searching by means of a multi-step recipe. Available functional cover a wide range drawn from several categories: GGA, GH-GGA, RSH-GGA, mGGA, GH-mGGA, RSH-mGGA, and with ***Spartan'26*** dual hybrid (DH) RSH-mGGA, where the common denominator is the Generalized Gradient Approximation (GGA), that is, depends on the electron density and its gradient. Three dozen commonly employed functionals appear in menus, separated by category, and others may be specified by name. Both closed and open-shell calculations are supported. Available basis sets are enumerated later in this appendix. Density functional models have been implemented for use on multi-core processors.

Preset limits for density functional calculations are 200 atoms and 2,000 basis functions. Calculations on molecules with up to 100 atoms are practical.

MP2 Models

The Møller-Plesset module calculates the energy and wave function and equilibrium and transition-state geometries using either the MP2 or RI-MP2 models. Vibrational frequencies are also available from both models but require numerical differentiation of analytical gradients. Both closed and open-shell calculations are supported. Available basis sets are enumerated later in this appendix. MP2 and RI-MP2 models have been implemented for use on multi-core processors.

Depending on basis set, practical considerations limit application of these techniques to molecules comprising 20-30 heavy (non- hydrogen) atoms, although program limits are set much larger.

Wave Function-Based Correlated Models

The wave function based correlated module calculates only the energy. Available from layered menus are the MP3 and MP4 Møller-Plesset methods, CCSD and CCSD(T) coupled-cluster methods and QCISD and QCISD(T) quadratic CI methods. In addition, the electronic energy from the G3, G3(MP2), G4 and G4(MP2) thermochemical recipes (see below) is available. Other methods may be specified according to their name. Available basis sets are enumerated later in this appendix.

Depending on the specific model and basis set, practical considerations limit application of these techniques to molecules comprising 10-20 heavy (non-hydrogen) atoms, although program limits are set much larger.

Thermochemical Recipes

The G3, G3(MP2), G4 and G4(MP2) recipes are available for highly- accurate heat of formation calculations which in our view represent approximations to coupled-cluster methods. Practical considerations generally limit application of the G3 and G4 recipes to molecules comprising less than 10 heavy atoms and the G3(MP2) and G4(MP2) recipes to molecules comprising less than 15 heavy atoms.

The T1 recipe has been developed to closely reproduce G3(MP2) heats of formation. It makes use of an HF/6-31G* geometry, instead of the MP2/6-31G* geometry, replaces the large basis set MP2 calculation in G3(MP2) by a dual-basis set RI-MP2 calculation, eliminates both the QCISD(T) calculation and the vibrational frequency calculation (needed to obtain zero-point energy and to correct the energy for finite temperature) and introduces an empirical correction based on the atom counts and Mulliken bond orders. T1 is 2–3 orders of magnitude faster than G3(MP2). Molecules with 30-50 heavy atoms are practical.

T1 is limited to closed-shell, uncharged molecules comprising H, C, N, O, F, Si, P, S, Cl and Br only and, because it incorporates parameters which depend explicitly on molecular geometry, is in practice limited to molecules that can be properly described in terms of a conventional valence structure. T1 is likely to be unsatisfactory for transition states.

Excited-State Models

Two classes of models are available for calculations of excited states: configuration interaction (CI) models and time-dependent density functional (TDDFT) models. The former comprises three models: CIS (configuration interaction singles), CISD (configuration interaction singles with doubles correction) and RI-CIS(D), which are analogous to the Hartree-Fock model for ground-state calculations. The latter are analogous to density functional models for ground-state calculations and support the same functionals or combinations of functionals available for geometry optimization and frequency calculations. Available basis sets are enumerated later in this appendix.

Preset limits for excited-state calculations are 200 atoms and 2,000 basis functions. Depending on the specific model and basis set, practical considerations limit applications to molecules comprising 20-30 atoms.

Basis Sets and Pseudopotentials

Available basis sets for Hartree-Fock, density functional, MP2 (RI- MP2) and wave function based correlated models for ground states and CIS, CISD (RI-CIS(D)) and time-dependent density functional (TDDFT) models for excited states include the full range of Pople basis sets: minimal STO-3G, split-valence 3-21G, 6-31G, 6-311G and polarization 6-31G*, 6-31G**, 6-311G* and 6-311G**. These may be supplemented with additional and/or higher-order polarization functions and/or with diffuse functions. Also available are the Dunning cc-pVDZ, cc-pVTZ and cc-pVQZ, and the corresponding augmented basis sets, the Weigand/Aldrich def 2 series of basis set, the G3 Large basis set, and the PCJ-0, PCJ-1 and PCJ-2 basis sets (developed for the calculation of NMR coupling constants).

Dual-basis set techniques (mixing small and large basis sets) are supported for ground-state energy calculations for Hartree-Fock, density functional and MP2 models. These can be up to an order of magnitude less costly than single-basis set calculations, depending on choice of small and large basis sets.

Pseudopotentials are available and are automatically invoked for calculations on molecules incorporating heavy atoms (for which all- electron basis sets are not available). Starting with ***Spartan'18***, this includes lanthanides and actinides pseudopotentials for which have been specifically calibrated to provide good equilibrium geometries for density functional models using the 6-31G* basis set.

Machine Learning and Neural Network Models

Spartan'26 provides seven neural net routines. The first (**Corrected MMFF**), starts from MMFF geometry and improves the quality of MMFF energies (in particular conformer energy differences) attempting to reproduce ωB97X-V/6-311+G(2df,2p)//MMFF. The second (and likely the routine to see the most utility) provides ωB97X-D/6-31G* equilibrium geometries starting from MMFF geometries. (**Est. Density Functional: ωB97X-D/6-31G***). And three additional machine learning routines estimating Total Energies from:

- **Est. Density Functional: ωB97X-V/6-311+G(2df,2p)//ωB97X-D/6-31G***
- **Est. Density Functional: ωB97M-V/6-311+G(2df,2p)//ωB97X-D/6-31G***
- **Est. Density Functional: ωB97M(2)/6-311+G(2df,2p)//ωB97X-D/6-31G***

Input is either ωB97X-D/6-31G* equilibrium geometries and energies (from QM calculations) or equilibrium geometries from the neural (**Est. Density Functional: ωB97X-D/6-31G***) and a single ωB97X-D/6-31G* energy calculation.

Two additional (proton and ^{13}C chemical shift) neural networks provide NMR shifts from either ωB97X-D/6-31G* geometries or neural network estimated ωB97X-D/6-31G* geometries.

All routines support for closed-shell, uncharged molecules comprising the elements: H, C, N, O, F, S, Cl and Br (only). There are no atom limits for the MMFF-based routines neural net. With the exception of a single-point ωB97X-D/6-31G* energy calculation (required for energies, NMR spectra and wave function related data) there are no atom limits to the second routine (Est. Density Functional ωB97X-D/6-31G* geometry). When requiring ωB97X-D/6-31G* energy or geometry as input, atom limits (≈ 200 atoms, ≈ 2000 basis functions) are the same as for standard density functional calculations. Utilized strategically, the neutral network routines can save significant computational time (orders of magnitude in some cases).

Solvent Models

Spartan supports the C-PCM and SS(V)PE models for energy, geometry and frequency

calculations with Hartree-Fock and density functional models with any basis set. C-PCM and SS(V)PE depend only on the dielectric constant of the solvent. Menus allow for specifying non-polar and polar solvents as well as water, but specific solvents (dielectric constants) may be specified. Both models alter the wave function in addition to the energy allowing graphical surfaces and property maps to be calculated in the presence of solvent.

Two additional solvent models for energy calculations (only) are supported. The SM5.4 model, which is available for H, C–F, S–Cl, Br and I, uses a parameterized procedure to estimate the aqueous solution energy. It estimates only the aqueous solvation energy. This in turn may be added to the gas-phase energy obtained from any molecular mechanics or quantum chemical model.

Q-Chem's implementation of SM8 and SM12 are available for H, C–F, S–Cl and Br, and applicable only to Hartree-Fock and density functional models with the 6-31G*, 6-31G**, 6-31+G* and 6-31+G** basis sets. These affect the wave function, meaning that graphical displays may be obtained in the presence of solvent. Both models have been parameterized for water as well as a variety of organic solvents. Common solvents may be specified.

Properties and Spectra

The properties module is automatically invoked following completion of a molecular mechanics or quantum chemical calculation. By default, it provides for text output printing, population analyses and atomic charge calculation (Mulliken, natural bond orbital and based on fits to electrostatic potentials), evaluation of thermodynamic quantities (heat capacity, enthalpy, entropy and free energy) and zero- point energy if vibrational frequencies are available, and calculation of the dipole moment.

The entropy comprises four components: translation, rotation, vibration and conformation.

$$S = S_{tr} + S_{rot} + S_{vib} + S_{conf}$$

$$S_{tr} = nR \left\{ \frac{5}{2} + \ln \left[\left(\frac{2\pi MkT}{h^2} \right)^{3/2} \left(\frac{nRT}{P} \right) \right] \right\}$$

$$S_{rot} = nR \left\{ \frac{3}{2} + \ln \left[\left(\frac{\pi V_A V_B V_C}{s} \right)^{1/2} \right] \right\} \; ; \; V_A = h^2/8\pi I_A kT$$

$$S_{vib} = nR \sum_i \left\{ (U_i e^{ui} - 1)^{-1} - \ln (1 - e^{-ui}) \right\} \; ; \; \mu_i = h\nu_i/kT$$

$$S_{conf} = \ln \left[\sum_j g_j e^{-\Delta E_j/RT} \right] 3^{Me}$$

The sum i is over vibrational modes and the sum j is over conformers, n is the number of moles, M is the molecular mass, I_A, I_B, I_C are the principal moments of inertia, s is the symmetry number, ϖ are the vibrational frequencies, E are energies relative to the lowest-energy conformer, g are the number of times the conformer appears and Me is the number of 3-fold rotors ("methyl groups"). R, k and h are the gas constant, Boltzmann's constant and Planck's constant, respectively and N_0 is Avogadro's number.

The vibrational component of the entropy is problematic as it goes to ∞ as the frequency goes to 0. In practice, the contribution is clamped at 1/2 R (the entropy of rotation).

Entropies are available for all EDF2/6-31G* (only) entries in SSPD. Note that these have not been corrected for the conformational contribution, which may range to several tens of entropy units for large flexible molecules.

Properties based on graphical displays that are available include the area, exposed area and volume of an electron density surface, the polar area of an electrostatic potential map for cutoff values of
-75, -100 and -125 kJ/mol, the maximum and minimum values of the electrostatic potential on an electrostatic potential map and the minimum value of the local ionization potential on a local ionization potential map. These require calculation of the respective surfaces and property maps, and need to be explicitly requested by *checking* the **QSAR** box inside the **Calculations** dialog (**Calculations** under the **Setup** menu; **Chapter 21**). They are either available from the **QSAR** tab of the **Molecule Properties** dialog (**Properties** under the **Display** menu; **Chapter 22**) or from the **Formula Editor** dialog (**Formulas** under the **Display** menu; **Chapter 22**).

The properties module is also responsible for calculating quantities related to infrared/Raman spectra (vibrational frequencies and infrared/Raman intensities), NMR spectra (chemical shifts and coupling constants) and UV/visible spectra (excited-state transition energies and intensities). IR spectra calculations may be carried out with molecular mechanics, semi-empirical, Hartree-Fock, density functional and MP2 and RI-MP2 models. Note, however, that MP2 and RI-MP2 infrared calculations need to be performed using numerical differentiation and are in practice limited to small molecules. NMR and Raman spectra calculations are limited to Hartree-Fock and density functional models and UV/visible spectra calculations are limited to Hartree-Fock models (CIS for excited states) and density functional models (TDDFT for excited states).

Graphical Models

The graphics module provides for data preparation associated with the display as surfaces, property maps and slices of molecular orbitals, electron densities, spin densities, electrostatic potentials, and local

ionization potentials. The sizes of electron density surfaces (and property maps based on electron density surfaces) may be chosen either using a specific value of the density or a value that encloses a specific percentage of the total number of electrons. Accessible and inaccessible regions may be distinguished for electron density surfaces and all property maps based on electron density surfaces.

In addition to its more general purpose graphics display, ***Spartan*** displays an orbital energy diagram (up to 12 highest-occupied molecular orbitals and the 2 lowest-unoccupied molecular orbitals). Molecular orbital graphics may be generated on-the-fly, given that a wave function is available.

Similarity Analysis

The similarity analysis module evaluates and quantifies the extent to which molecules are similar based either on selected atomic centers (structure) or on selected chemical function descriptors (CFD's). In addition, it quantifies the degree of fit of a molecule into a chemical environment (a pharmacophore). Similarity analyses are carried out between one or more molecules or pharmacophores in a document (the template) and one or more molecules or pharmacophores in one or more documents (the library). Molecules in the library (only) may either correspond to single conformers or to collections of diverse conformers.

Appendix C

Menus

Spartan Screen

File

New Build	Brings up a model kit for 3D molecule building or substitution.
New Sketch	Brings up the sketch pad for molecule sketching in 2D.
Build New Molecule	Adds a molecule to an existing document; brings up a model kit for molecule building or substitution.
Sketch New Molecule	Adds a molecule to an existing document; brings up the sketch pad for molecule sketching in 2D.
Delete Molecule	Deletes a molecule (or molecules) from a document.
Append Molecule(s)...	Appends molecules to an existing document.
Open...	Opens (imports) a molecule or multi-molecule document.
Close	Closes a molecule or multi-molecule document.
Open Recent Documents	Opens a molecule or multi-molecule document from a list of the 10 most recently accessed documents.
Save	Saves (exports) a molecule or multi-molecule document.

Save As...	Saves a molecule as a document under a user-specified name; creates files for use in the Spartan Reaction Database (SRD), Spartan Infrared Database (SIRD) and Spartan Spectra and Properties Database (SSPD).
Save Image As...	Saves an image as PNG, JPEG or Bitmap file
Print...	Prints on-screen display; also prints contents of output window and the **Spreadsheet**.
Access PDB Online...	Accesses the online Protein Data Bank (PDB).
Extract Ligands	Extracts ligands and pharmacophores based on protein-ligand binding from PDB files.
Embedded Data	Associates external files with a document, either at molecule or document level (or both).
Exit	Exits ***Spartan***.

Edit

Undo	Undoes previous operations.
Cut	Moves the current molecule or contents of the selection box to the clipboard.
Copy	Copies the current molecule or contents of the selection box to the clipboard.
Paste	Pastes contents of the clipboard to the screen.
Select All	Selects all atoms in the selected molecule.
Find...	Locates a text string in the output dialog or an on-screen molecular fragment.

Find Next	Locates next occurrence of a text string or molecular fragment.
Center	Centers the molecule on screen; applies to all molecules in a document.
Clear	Clears the selected molecule.

Model

Wire	Displays structure as wire-frame model.
Ball and Wire	Displays structure as ball-and-wire model.
Tube	Displays structure as tube model.
Ball and Spoke	Displays structure as ball-and-spoke model.
Space Filling	Displays structure as space-filling model.
Line	Displays structure as a line drawing.
Hide	Hides structure model from view.
Global Model	Applies model type and labels of current molecule to all molecules in the document.
Coupled	Couples motions of all molecules in the document.
Hydrogens	Toggles hydrogens on and off.
Labels	Toggles labels on and off.
Ribbons	Toggles ribbons on and off.
Ramachandran Plot	Display a Ramachandran plot for a protein that has been brought in from the Protein Data Bank (PDB).
Hydrogen Bonds	Toggles hydrogen bonds on and off.
R/S Chirality	Toggles R/S chirality labels on and off.
CFD's	Toggles chemical function descriptors (hydrophobes, hydrogen-bond donors/acceptors, positive/negative centers) on and off.

Configure...	Labels atoms, bonds, etc., provides information about polypeptides/ polynucleotides residues and designates ribbon displays. Atom labels include, among other possibilities, charges and both experimental and calculated NMR chemical shifts. Labels may also be customized.

Geometry

Measure Distance	Displays and/or sets bond distance.
Measure Angle	Displays and/or sets bond angle.
Measure Dihedral	Displays and/or sets dihedral angle.
Freeze Center	Freezes selected atomic positions.
Set Torsions	Selects bond rings for conformational searching.
Set Similarity Centers	Identifies atoms or CFD's for similarity analyses.
Constrain Distance	Constrains bond distance.
Constrain Angle	Constrains bond angle.
Constrain Dihedral	Constrains dihedral angle.
Define Point	Defines a point as a geometric mean of a set of atoms.
Define Ligand Point	Defines a ligand point as a position that is perpendicular to the centroid of a plane made by three or more atoms.
Define Plane	Defines a plane made by three or more atoms.
Generate Isomers ( )	Generates a list of regio and/or stereoisomers based on user selection
Generate Tautomers	Generates tautomers.

Define CFD	Defines the location of a CFD.
Align	Aligns molecules in a document according to selected atoms, structure or CFD's.

Build

View	Removes the model kit.
Edit Build	Brings up a 3D model kit (organic, inorganic, peptide, nucleotide, substituent or ChemDraw) with the presently selected molecule
Edit Sketch	Brings up the sketch pad with the presently selected molecule. This function is only available if the sketch has not been altered using any of the 3D model kits or has not been replaced by an entry in SSPD.
Delete	Deletes atoms, bonds, points, planes, etc. Also available at the bottom of the 3D model kits.
Make Bond	Makes bonds between free valences or atoms. Also available at the bottom of the 3D model kits.
Break Bond	Breaks a bond. Also available at the bottom of the 3D model kits.
Minimize	Performs energy minimization using MMFF molecular mechanics. Also available at the bottom of the 3D model kits.
Guess Transition State	Provides transition-state guess based on reaction database or, lacking a database entry, based on linear synchronous transit.

Setup

Calculations...	Sets up molecular mechanics and quantum chemical calculations and similarity analyses; specifies calculation of IR, Raman, NMR and UV/vis spectra.

Surfaces	Sets up generation of and displays graphical surfaces.
Submit	Submits job to the execution queue either locally or on a remote server.

Display

Output	Displays text output. This includes a summary generated on-the-fly in the interface.
Properties	Displays molecule, bond and atom properties as well as information about CFD's, geometrical constraints, graphical surfaces, plots (including spectral plots) and statistical analyses.
Orbital Energies	Displays an orbital energy diagram and allows on-the-fly generation and display of molecular orbitals.
Surfaces	Sets up generation of and displays graphical surfaces (same as entry in **Setup** menu)
Spectra	Displays IR, Raman, NMR and/or UV/visible spectra, animates vibrational modes (IR and Raman), and accesses on-line experimental spectral databases; fits calculated IR spectra to experimental spectra. A simplified and more intuitive spectra presentation mode is available.
Formulas	Constructs expressions for use in spreadsheet and database searching/mining operations.
Plots...	Creates 2D and 3D plots from the data in the spreadsheet. A simplified and more intuitive plot presentation mode is available.
Spreadsheet	Displays spreadsheet.
Similarities	Displays results of similarity analyses.
Reactions	Calculates reaction (activation) energies, enthalpies or Gibbs energies using data either

from current document or from the Spartan Spectra and Properties Database (SSPD). Database access uses energies from the ωB97X-V/6-311+G(2df,2p) density functional model with vibrational contributions (required for enthalpy and Gibbs energy calculations) from the EDF2/6-31G* density functional model.

Search

Structure Query	Add growth points to a structure defining an underlying substructure.
Reaction Query	Add curly arrows to structure query (the reactant) in order to relate it to the product of a chemical reaction or to the transition state leading to that product. Arrows may also be added using the 2D sketch pad.
Databases	Performs substructure search on Cambridge Structural Database (CSD), Spartan Spectra and Properties Database (SSPD) and Spartan Reaction Database (SRD); performs spectral search of the Spartan Infrared Database (SIRD) and the External (NIST) Infrared Database (XIRD).

Data

Expt. Chem Shifts ($^{1}H/^{13}C$)	Allows experimental proton and 13C chemical shifts to be attached to 3D models. Access from the 2D sketcher alternatively allows attachment onto 2D sketches.
HH Couplings (J^{HH})	Allows observed 3-bond HH couplings to be attached to 3D models. Access from the 2D sketcher alternatively allows attachment onto 2D sketches.

CH Couplings ($^{J_{HH}}$)	Allows observed 2 and 3-bond CH couplings to be attached to 3D models. Access from the 2D sketcher alternatively allows attachment onto 2D sketches.

Options

Preferences...	Sets various run-time and labeling preferences and specifies execution preferences and run-time queue; invokes parallel calculation; sets up icon displays; establishes url's for on-line access; sets up access to external compute servers and permits use of ***Spartan'20*** as a server.
Colors	Sets screen and model colors.
Fonts...	Sets fonts for labels and plot displays.
Graphics Fonts...	Sets fonts for graphical displays.
Monitor	Monitors and allows for killing jobs and bypassing queue to start jobs.
Calculator	Pocket calculator.
Icons	Turns on and off icons displayed under the menu bar.

Activities

Tutorials	Brings up ***Spartan'20*** tutorials as PDF documents.
Topics	Brings up selection of topics relevant to calculations performed in ***Spartan*** as PDF documents.
Look up in Wikipedia	Brings up an html page pointing to a Wikipedia page.

Help

Spartan Help	Provides information about the

	performance and timing of computational methods in ***Spartan***; provides information about using graphical models in ***Spartan***; bulletin board for FAQs about ***Spartan.***
Spartan Manual	Brings up ***Spartan*** manual (this document) as a PDF.
License Utility. . .	Provides information on licensing and maintenance. Facilitates license transfer request.
About ***Spartan***. . .	Provides program version information for citation and support.

Contextual

Main Screen

Copy	Copies selected molecule to the clipboard.
Paste	Pastes the contents of the clipboard into the selected document.
Delete Selected	Deletes selected molecule from document.
Properties	Brings up the **Molecular Properties** dialog.

Spreadsheet

Copy	Copies text of selected cell or cells to the clipboard. If leftmost cell (or cells) selected, copies molecule(s) to the clipboard
Paste	Pastes the contents of the clipboard into selected cells. If leftmost cell (or cells) selected, either pastes text or molecule(s) depending on menu choice
Add	Brings up the **Add** dialog (spreadsheet) for adding calculated quantities into the spreadsheet

Sort	Sorts the column from low to high. *Pressing* the **Shift** key prior to menu selection sorts from high to low
Format Selected	Formats selected cell(s), selected column(s) if selection is in a header cell, or entire spreadsheet if selection is header cell of leftmost column
Delete Selected	Deletes selected molecule(s) from document
Append	Appends the contents of ***Spartan*** document(s) to the spreadsheet (corresponding to the selected document)
Rename Selected Using SSPD	Rename selected molecule(s) with names in the Spartan Spectra and Properties Database (SSPD); appears only when leftmost cell(s) selected.
Properties	Brings up the **Molecular Properties** dialog.

Substituent Model Kit

Copy	Copies selected molecule(s) from a custom library onto the clipboard
Paste	Pastes molecule(s) from the clipboard into a custom library.
Delete	Deletes selected molecule(s) from the custom library.
Append	Appends the contents of ***Spartan*** document(s) to the custom library; brings up a file browser

Reactions

Copy	Copies selected text to the clipboard
Print	Prints selected text

Output Window

Copy	Copies selected text to the clipboard
Print	Prints selected text

Database

Add Columns...	Adds data columns to search results.
Copy	Copies selected molecules or data.
Paste	Pastes text to data column(s).
Format	Formats data column(s).
Delete	Deletes search results or data columns.

Appendix D

Calculation Options

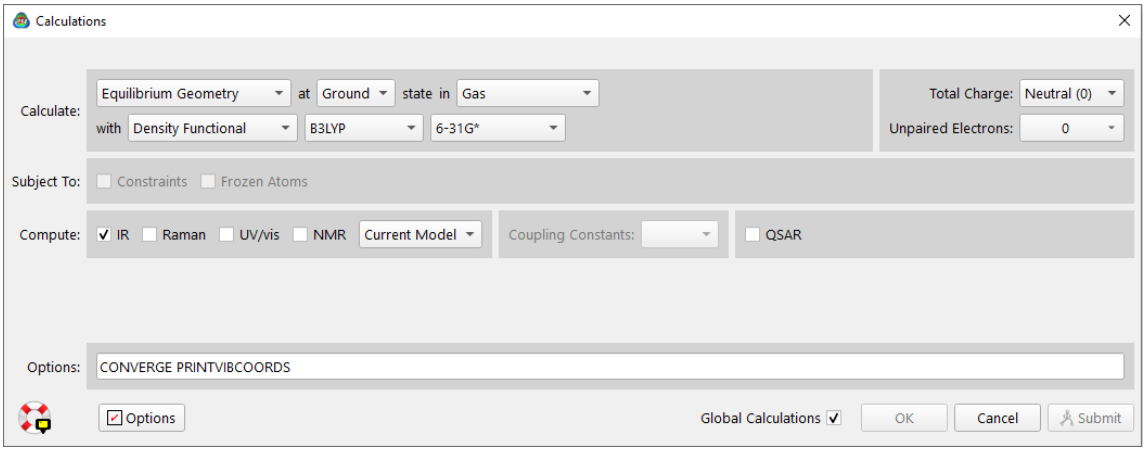
Options

While the design philosophy behind ***Spartan's*** **Calculations** dialog has historically been to streamline and simplify (where possible) the number of choices the user need make to submit a given calculation, there are times when a specific calculation may require additional user input in the form of one (or more) **Keywords**. In previous iterations of the manual a subset of available keywords was provided in this Appendix. However, based mostly on user feedback, a much longer and as complete as possible (at the time of this document) list of keywords is provided from the **Help** menu ***Spartan'26*** **Help** entry (**Chapter 27**). These are listed under the FAQ section: **Option Line Keywords**, and are broken down into the following categories:

Geometry Optimization
Energy Convergence
Conformational Searching
Energy Profile
Electrostatic Potential
Wave function and Electron Density
Properties
Intrinsic Reaction Coordinates
Excited States and UV/vis
NMR shifts
NMR couplings
Density Functional parameters

While certainly the vast majority of calculations will not require additional modification, if your calculation prompts for or is suggested to be submitted with a keyword, *check* the box next to **Options** in the **Calculations** dialog. The field should be locked for all caps, but is case insensitive. Multiple keywords should be separated by a space (not a comma or semi colon).

For example:



Above specifies a typical calculation, equilibrium geometry (and energy), followed by a frequency calculation. The input has been modified, specifying a change in the SCF convergence criteria (CONVERGE), and upon completion of the frequency calculation (**IR** check box on), to "print" the coordinates of each vibrational mode to the **Output** file (PRINTVIBCOORDS).

Over the last several release versions, we have begun to integrate a more guided **Help** system. In theory, if your job fails, you can return to the Calculations dialog and there will be a "linked" error message. *Click* on the error message. Depending on how unique the issue, there are integrated help suggestions, the often provide suggestion on use of recommended keyword. Should you encounter an error in a calculation where no additional help is referenced, please reach out to our support team via *support@wavefun.com*.

Appendix E

Units

Energies, Heats of Formation and Strain Energies, Zero-Point Energies, Enthalpies and Gibbs Energies and Entropies

Total energies from Hartree-Fock calculations are available in au, kcal/mol and kJ/mol.

Experimental heats of formation as well as those from semi-empirical calculations and from thermochemical recipes are available in kJ/ mol, au and kcal/mol.

Strain energies from molecular mechanics calculations are available in kJ/mol, au and kcal/mol.

Energies, heats of formation and strain energies corrected empirically for the effects of aqueous media are given in the same units as the corresponding gas-phase quantities.

Zero-point energies, enthalpies and Gibbs energies available in kJ/ mol, kcal/mol and au/mol. Entropies are available in J/mol•degree, kcal/mol•degree and au/mol•degree.

Orbital Energies

Orbital energies are available in eV.

Energy Conversions

	au	kcal/mol	kJ/mol	eV
1 au	-	627.5	2625	27.21
1 kcal/mol	1.593 (-3)	-	4.184	4.337 (-2)
1 kJ/mol	3.809 (-4)	2.390 (-1)	-	1.036 (-2)
1 eV	3.675 (-2)	23.06	96.49	-

a) exponent follows in parenthesis, e.g., 1.593 (-3) = 1.593×10^{-3}

Electron Densities, Spin Densities, Dipole Moments, Charges, Electrostatic Potentials and Local Ionization Potentials

Electron densities and *spin densities* are given in electrons/au^3.
Dipole moments are given in Debyes.
Atomic charges are given in electrons.
Electrostatic and *hydride potentials* are given in kJ/mol.
Local ionization potentials and *Orbital Energies* are given in eV.

Vibrational Frequencies, Infrared and Raman Intensities

Vibrational frequencies are given in wavenumbers (cm^{-1}). Intensities are given relative to the most intense peak. The simplified plot display will renormalize on-the-fly to correspond to the most intense peak that is displayed.

Chemical Shifts, Coupling Constants

Chemical shifts are given in parts-per-million (ppm) relative to appropriate standards: hydrogen (to tetramethylsilane); carbon (to tetramethylsilane); nitrogen (to nitromethane); fluorine (to fluorotrichloromethane); silicon; (to tetramethylsilane); phosphorous (to phosphoric acid). Standards have not been calculated for all models, but they are available for Hartree-Fock and several popular density functional models with the 6-31G* basis set. A full listing is provided in **Appendix L**. Chemical shifts for other models need to be adjusted using the appropriate standards calculated for the same models. Coupling constants are in Hz.

UV/visible Spectra

λ_{max} is given in wavenumbers (cm^{-1}). Intensities are given relative to the most intense peak.

Appendix F

Citation

The proper citation for ***Spartan'26*** is as follows:

Spartan'26, version 1.1, **2026**; Wavefunction, Inc.: Irvine, CA

Many have contributed to ***Spartan's*** interface over the last thirty-five years. Current team members include: Bernard Deppmeier, Andy Driessen, Michelle Fitzpatrick, Jeffrey Johnson, Thomas Hehre, Warren Hehre, Philip Klunzinger, Connor Ohlinger, William Sean Ohlinger, Pamela Ohsan, Jurgen Schnitker, Noritaka Uchida and Yuusuke Watababe. ***Spartan's*** general design and infrastructure were architected by Warren J. Hehre.

Except for **machine learning** models, **molecular mechanics** and **semi-empirical** models, and **similarity analysis**, computational methods used in ***Spartan*** have been documented in:

Epifanovsky, E.; Gilbert, A. T. B.; Feng, X.; et al. Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *J. Chem. Phys.* **2021**, *155*, 084801.

Y. Shao, Z. Gan, E. Epifanovsky, et al. Mol. Phys. **113**, 184–215 (2015).

Y. Shao, L.F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S.T. Brown, A.T.B. Gilbert, L.V. Slipchenko, S.V. Levchenko, D.P. O'Neill, R.A. DiStasio Jr., R.C. Lochan, T. Wang, G.J.O. Beran, N.A. Besley, J.M. Herbert, C.Y. Lin, T. Van Voorhis, S.H. Chien, A. Sodt, R.P. Steele, V.A. Rassolov, P.E. Maslen, P.P. Korambath, R.D. Adamson, B. Austin, J. Baker, E.F.C. Byrd, H. Dachsel, R.J. Doerksen, A. Dreuw, B.D. Dunietz, A.D. Dutoi, T.R. Furlani, S.R. Gwaltney, A. Heyden, S. Hirata, C-P. Hsu, G. Kedziora, R.Z. Khalliulin, P. Klunzinger, A.M. Lee, M.S. Lee, W.Z. Liang, I. Lotan, N. Nair, B. Peters, E.I. Proynov, P.A. Pieniazek, Y.M. Rhee, J. Ritchie, E. Rosta, C.D. Sherrill, A.C. Simmonett, J.E. Subotnik, H.L. Woodcock III, W. Zhang, A.T. Bell, A.K. Chakraborty, D.M. Chipman, F.J. Keil, A. Warshel, W.J. Hehre, H.F. Schaefer, J. Kong, A.I. Krylov, P.M.W. Gill and M. Head-Gordon, *Phys. Chem. Chem. Phys.,* **2006**, 8(27), 3172-3191.

A discussion and assessment of commonly-used calculation methods is found in: W.J. Hehre, ***A Guide to Molecular Mechanics and Quantum Chemical Calculations***, Wavefunction, Inc., Irvine, CA, 2003, available as a PDF under the **Help** menu. This is somewhat outdated as there have been many important developments since it was compiled. A guide reflecting current developments is in preparation.

Appendix G

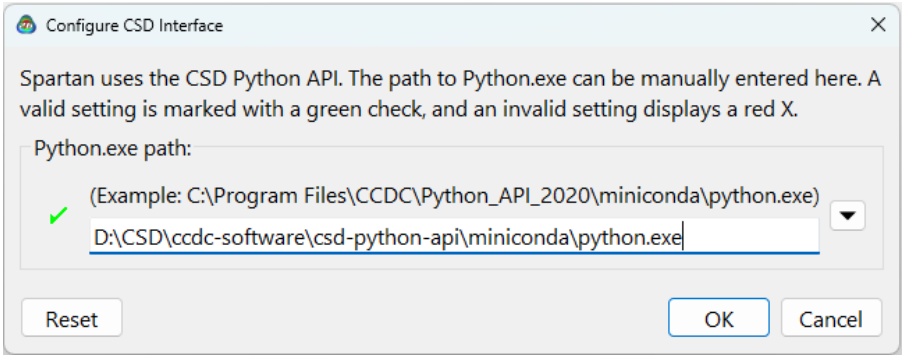
Configuring the Cambridge Structural Database

Spartan can act as an ancillary interface to the Cambridge Structural Database (**CSD**). In order to search the **CSD**, ***Spartan*** must be configured to point to the appropriate location of the **CSD** data. The **CSD** uses Python to facilitate searching, and a specific python path will need to be set. Instructions vary depending on platform (Windows, Macintosh, of Linux) and follow below.

Note: The **CSD** must be licensed in order to utilize the ***Spartan*** tools for interfacing the **CSD**. Contact admin@ccdc.cam.ac.uk for help with **CSD** licensing.

Windows:

(*Click* on **Config** near the bottom right of the **CSD** tab. This brings up the **Configure CSD Interface** dialog:

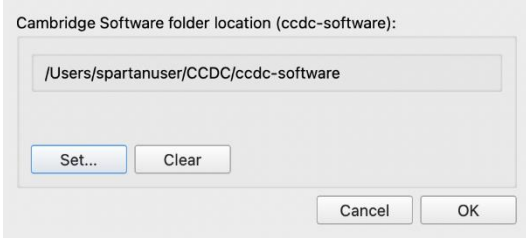


If the current (2025 at the time of this update to the manual) **CSD** System has been installed, the correct path for the python.exe will appear (similar to the above image, the exact path will depend upon the individual **CSD** installation details). The standard (full **CSD** installation) path can be restored by clicking on the **Reset** button at the bottom left of the dialog. The path may be hand edited by typing the field or may be set by using the file system (accessed by clicking on the downward triangle to the right of the path field). A valid setting is marked with a green check, and invalid setting is marked with a red X. Once enabled, you may search the **CSD** from within ***Spartan***. Contact support@wavefun.com for Spartan configuration help, or support@ccdc.cam.ac.uk for help with **CSD** installation.

Macintosh:

Click on the **Config** button near the bottom right of the **CSD** tab. This brings up the

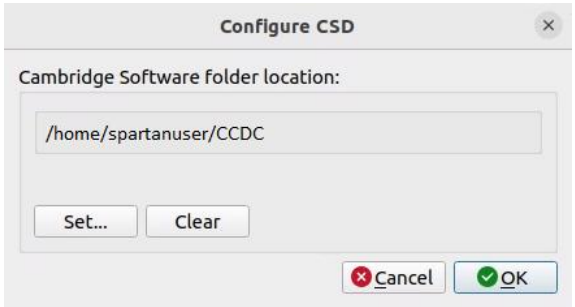
Cambridge Software folder location dialog:



If the field is blank, *click* the **Set…** button and locate the ccdc-software folder. The typical installation path is Users/currentuser/CCDC. The current license year is **2025** at the time of this document). ***Spartan*** will check look for and locate the **CSD** data (as well as a python application) necessary to interface the **CSD** from ***Spartan***.

Linux:

Installation of the current Linux release of the **CSD** creates a **CCDC** folder in the users directory, configure this via the **Config** button at the bottom right of the CSD tab:



Visit **Activities** menu, **Tutorials**, and open the last tutorial for examples of searching the **CSD** from within ***Spartan*** (**Chapter 15**).

Appendix H
Constructing Custom Databases

Beginning in ***Spartan'24*** is the computational task **SSPD Compatible** (configurable from the **Options** menu **Preferences** dialog; **Chapter 25**). If this task has been used to generate either an equilibrium geometry, or a transition state geometry, an additional option from the file menu is available: **Save to Database** . If chosen, the resulting calculation is added to a new collection, parallel to the existing SSPD (but at the same computational model), which can optionally be searched alongside the provided SSPD release. Results are *automatically* saved to the **Custom Database Directory** (**Database** tab, the **Options** menu **Preferences** dialog; **Chapter 25**).

The same approach can be used to generate custom collections utilizing *different* computational models than those available in the SSPD. In this case, the **File** menu **Save As** feature may employed to save the document as a **Spartan Database** file. In this way, other theory levels may be added on top of those provided. These may either include the geometry, energy (heat of formation, strain energy), HOMO and LUMO energies, dipole moment, atomic charges, chemical shifts and vibrational frequencies obtained from quantum chemical calculations, or (for SSPD and SIRD only) conformer libraries generated from molecular mechanics calculations. Entries for SRD (transition states) must derive from ***Spartan***'s transition state facility (**Guess Transition States** under the **Build** menu; **Chapter 20**) in order to include the original set of reaction arrows. Documents containing one or more molecules that are to be placed in a user database are first saved using **Save As** under the **File** menu with **Save as type** set to **Spartan Database**. This results in two files, a **.spentry** and a **.spindex** file for each molecule (or list). These should then be put in a single directory, the path to which must be entered into the **Paths Preferences** dialog under **Preferences...** in the **Options** menu (**Chapter 25**).

Note that it is important to specify a label ("name") for each molecule. Label specification is done either in the **Molecule Properties** dialog (see **Properties** under the **Display** menu in **Chapter 22**) or in the spreadsheet (see **Spreadsheet** under the **Display** menu in **Chapter 22**).

Appendix I
Pharmacophore Input

The following file format is used for input of a pharmacophore (illustrated for a pharmacophore with four hydrophobes and one positive ionizable center):

```
Blank line
Blank line
Blank line
ENDCART
BEGINCFD
HPHOBE|RING_AROM   8.050  -4.180  -0.400  0  2.000  1.000  1
HPHOBE|RING_AROM   9.630  -0.640  -3.590  1  2.000  1.000  1
HPHOBE|RING_AROM  10.280  -0.380   3.200  2  2.000  1.000  1
HPHOBE|RING_AROM   8.130  -2.300   4.330  3  2.000  1.000  1
POS               14.370  -1.700   0.280  4  3.000  1.000  1
ENDCFD
```

The lines between BEGINCFD and ENDCFD define the individual pharmacophore elements (one line/element):

type1|type2… (| separates types for multiple definitions), Cartesian (X,Y,Z) coordinates (in Angstroms), unique number, radius of element (in Angstroms), weight of element in fit, use the element in fit (1) or ignore the element (5)

The following element types are available:

HPHOBE	hydrophobe
RING_AROM	aromatic ring
POS	positive ionizable center
NEG	negative ionizable center
HBA_CENTER	hydrogen bond acceptor
HBD_CENTER	hydrogen bond donor
EXCL_VOL	excluded volume

Appendix J
Input of Experimental Spectra

The databases of infrared and UV/visible spectra accessible on- line from ***Spartan*** are in JCAMP format. Specification of both and examples can be found at the NIST website (NIST.gov).

The database of ^{13}C spectra is in CML (Chemical Markup Language) format. A description may be found at https://www.ch.ic.ac.uk/omf/ cml. A sample file is provided below.


```
<?xml version="1.0" encoding="ISO-8859-1"?>
<cml>
 <spectrum type="NMR">
  <metadataList>
   <metadata name="nmr:OBSERVENUCLEUS" content="13C"/>
  </metadataList>
  <peakList>
   <peak xValue="125.1399993896484" peakHeight="0.458" />
   <peak xValue="128.4499969482422" peakHeight="0.558" />
   <peak xValue="129.0099945068359" peakHeight="0.658" />
   <peak xValue="21.64999961853027" peakHeight="0.758" />
   <peak xValue="137.8399963378906" peakHeight="0.858" />
   </peakList>
  </spectrum>
</cml>Blank line
```


This contains five shifts (xValue) and five associated intensities (peakHeight).

Appendix K
Platform Dependencies

*This appendix covers variations in **Spartan** across Windows, Macintosh, and Linux operating systems.*

Accessing the Preferences… and About… dialogs

Under Windows and Linux, program **Preferences…** and the **About…** dialogs are accessed from the **Options** menu at the top right of the ***Spartan*** interface. For Macintosh users, program **Preferences…** may be accessed via the **Options** menu and, by convention, both **Preferences…** and the **About…** dialog are accessible from the ***Spartan'26*** (program) menu at the top left of the Mac interface

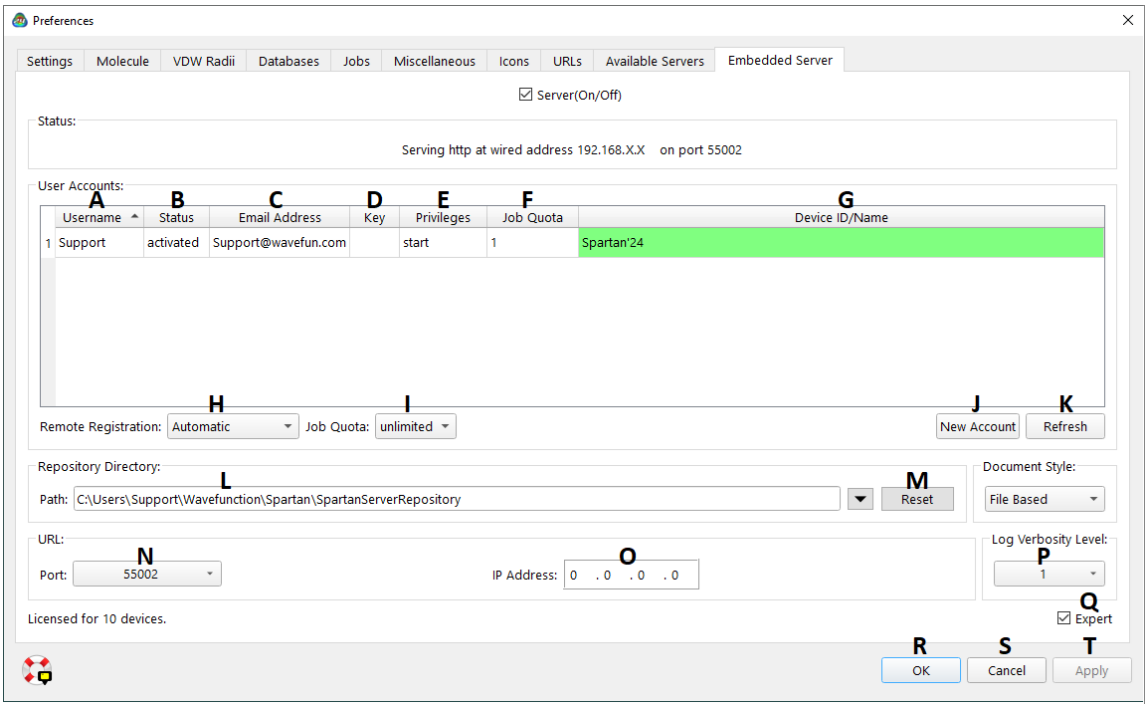
Accessing Controls for Spartan Server

Under Windows, the server controls are accessed via the **Embedded Server** tab under **Preferences…** from the **Options** menu. The server functions as part of the graphical interface and the program must remain open in order for users to remotely submit to the server.

Under Linux and Macintosh, the server controls are accessed from the **Service Control** entry under the **Options** menu. The server functions as an independent service and the graphical interface does not need to remain open in order for users to remotely submit to the server. Under Linux (only) the service control is web based and may be accessed either locally or remotely via http://machinename:8080. This requires that ports 8080 (for communication with the service controls) and 50002 (initial default for remote submission) are opened. The default username is admin, the default password is password. For https communication instructions, contact support@wavefun.com.

A description of the **Embedded Server/Service Control** dialog follows:

Settings/Options for the Embedded Server:



A Username. The Username on the account.

B Status. The account status, possible values are pending and activated. Pending means a remote registration has been received but not yet approved by the server administrator.

C E-mail Address. The email address of the account user. At present this is a secondary means of identification of the account, ***Spartan*** does not (currently) generate/send emails.

D Key. A security key can be issued for an account. If a security key is issued, it must also be provided when the account is registered (from the **Available Servers** tab).

E Privileges. Option to allow an account with the privilege to start a job irrespective of the Job Queue.

F Job Quota. Sets limits on how many jobs a user can submit.

G Device ID/Name. A unique device ID and the name of the device remote submission is received from, in the case of other ***Spartan*** licenses, the name is ***Spartan'26***, in the case of ***iSpartan***, the name of the iPad, iPhone, or iPod Touch is used. A user account can have more than one Device ID/name associated with it.

H Remote Registration. Controls the security level for new user registration, from lowest (Automatic = automatically register any new user) to intermediate (Sign Off Required = new users can register, but the server administrator must approve the account before it becomes active) to highest (Disabled = new accounts cannot be remotely registered and must be generated manually by the server administrator).

I Job Quota. This is the default job quota (see above) that will be provided to newly registered user accounts.

J New Account. Allows the server administrator to manually create a new user account (not needed if the Remote Registration is kept in either Automatic or Sign Off Required mode).

K Refresh. Refreshes the panel, required to see newly applied changes to this panel.

L Repository Directory (Path). The location (on the server) that remotely submitted jobs save to.

M Reset. If the server repository path has been changed, reset will revert to the default path location.

N Port. The default port used for communication between the client (account) and server. If this is changed from 55002, the URL path specified when client accounts verify with the server, should be modified with the addition of a colon (:) and the new port number, for example:

If using default port 55002, if the Host URL is designated as: http://opel.wavefun.com.

If the port is changed on the server from 55002 to 55000, for example, the Host URL will need to be changed to include an appended colon and port number: http://opel.wavefun.com:55000.

O IP Address. Only used if the server has a secondary Ethernet connection to specify which IP Address the server should listen on.

P Log Verbosity Level. Specifies the amount of communication logging written when jobs are remotely submitted/returned/monitored/ etc. (0 means no logging, 5 is maximum logging)

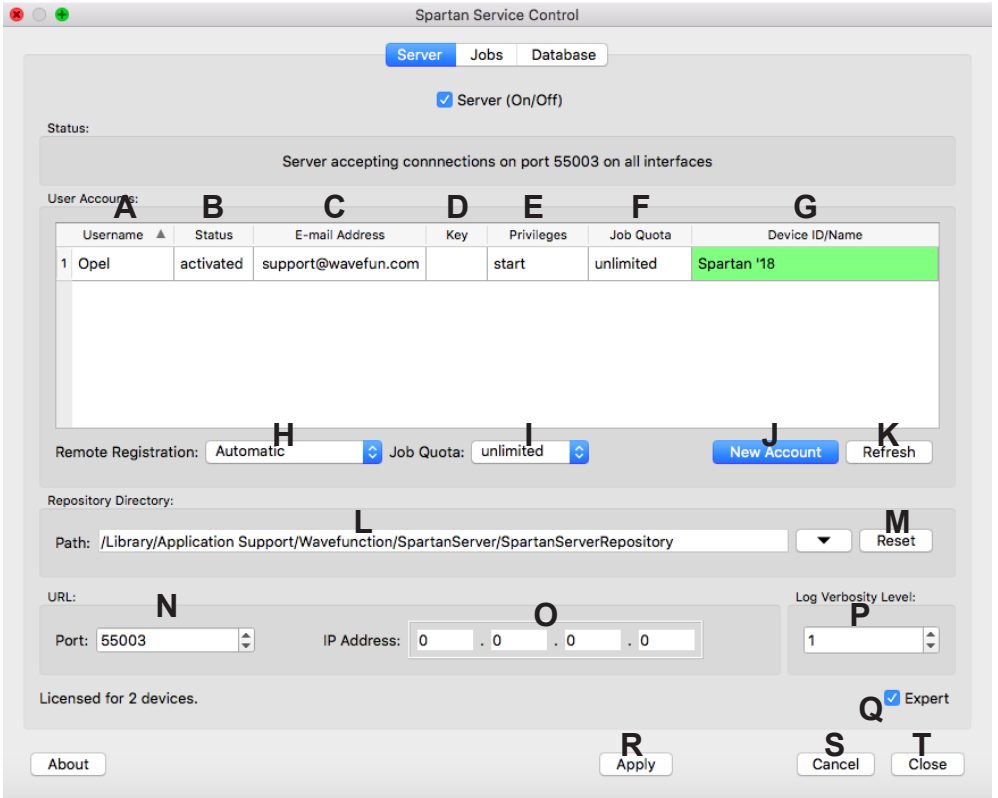
Q Expert. Off, by default, if turned on enables access to more advanced (and typically lesser required) settings for server functionality.

R OK. Saves changes and releases the Preferences dialog (Windows), Service Control dialog (Macintosh and Linux).

S Cancel. Cancels any changes and releases the Preferences dialog (Windows), Service Control dialog (Macintosh and Linux).

T Apply. Applies any changes without releasing the Preferences dialog (Windows), Service Control dialog (Macintosh and Linux).

Settings/Options for the Service Control Panel (Macintosh and Linux):



*** Server(On/Off).** Check box turns the server on or off (checked = on)

A Username. The Username on the account.

B Status. The account status, possible values are pending and activated. Pending means a remote registration has been received, but not yet approved by the server administrator.

C E-mail Address. The email address of the account user. At present this is a secondary means of identification of the account, ***Spartan*** does not (currently) generate/send emails.

D Key. A security key can be issued for an account. If a security key is issued, it must also be provided when the account is registered (from the **Available Servers** tab).

E Privileges. Option to allow an account with the privilege to start a job irrespective of the Job Queue.

F Job Quota. Sets limits on how many jobs a user can submit.

G Device ID/Name. A unique device ID and the name of the device remote submission is received from, in the case of other ***Spartan*** licenses, the name is ***Spartan'26***, in the case of ***iSpartan***, the name of the iPad, iPhone, or iPod Touch is used. A user account can have more than one Device ID/name associated with it.

H Remote Registration. Controls the security level for new user registration, from lowest (Automatic = automatically register any new user) to intermediate (Sign Off Required = new users can register, but the server administrator must approve the account before it becomes active) to highest (Disabled = new accounts cannot be remotely registered and must be generated manually by the server administrator).

I Job Quota. This is the default job quota (see above), provided to newly registered user accounts.

J New Account. Allows the server administrator to manually create a new user account (not needed if the Remote Registration is kept in either Automatic or Sign Off Required mode).

K Refresh. Refreshes the panel, required to see newly applied changes to this panel.

L Repository Directory (Path). The location (on the server) that remotely submitted jobs save to.

M Reset. If the server repository path has been changed, reset will revert to the default path location.

N Port. The default port used for communication between the client (account) and server. If this is changed from 55002, the URL path specified when client accounts verify with the server, should be modified with the addition of a colon (:) and the new port number, for example:

If using default port 55002, if the Host URL is designated as: http://opel.wavefun.com.

If the port is changed on the server from 55002 to 55000, for example, the Host URL will need to be changed to include an appended colon and port number: http://opel.wavefun.com:55000.

O IP Address. Only used if the server has a secondary Ethernet connection to specify which IP Address the server should listen on.

P Log Verbosity Level. Specifies the amount of communication logging written when jobs are remotely submitted/returned/monitored/ etc. (0 means no logging, 5 is maximum logging)

Q Expert. Off, by default, if turned on, enables access to more advanced (and typically lesser required) settings for server functionality.

R Apply. Saves changes to the Server Tab of the Service Control panel.

S Cancel. Cancels any changes and releases the Service Control panel.

T Close. Applies any changes and releases, Service Control panel.

Additional platform dependencies. While in general Windows, Macintosh, and Linux versions include identical functionality, a few important differences should be mentioned, in particular, the embedded server functionality. In order to utilize the server features, the GUI must remain opened (it can be minimized, but must be open). On both the Macintosh and Linux platforms, the server is a true service/daemon, and does not require the ***Spartan*** program nor the Service Control Panel to remain open.

Additionally, the Linux panel is a web-based tool that can be accessed both locally (on the Linux device) or remotely from a web browser on any of the client machines. A user login and password are required. The default username is **admin**, the default **password** is password. Once initially logged in, the password may be changed.

Appendix L
NMR ^{13}C Chemical Shift Standards

	Nucleus					
	H	**C**	**N**	**F**	**Si**	**P**
STO-3G	33.7521	249.4284	-99.3060	368.5390	578.6340	499.2810
3-21G(*)	33.6181	213.1943	-251.1870	279.7970	509.1390	502.3530
6-31G	33.6747	208.2044	-276.2000	222.5260	484.5230	373.9520
6-31G*	32.9027	201.7207	-160.1430	269.3660	450.1490	423.6900
6-31+G*	32.8181	201.1517	-159.5610	261.0900	446.6710	415.3570
6-31G**	32.3367	203.1517	-160.3110	269.4610	448.3970	423.1850
6-31+G**	32.2689	203.0758	-159.7060	261.0800	444.6390	415.3570
6-311G*	32.8321	195.9091	-177.3440	246.9560	395.6350	368.3180
6-311+G*	32.7799	195.8544	-184.5910	246.5820	395.6260	366.5100
6-311G**	32.4751	196.1543	-177.7210	247.0910	396.7650	366.8630
6-311+G**	32.4498	195.8412	-184.7400	246.5860	396.7780	366.6250
6-311++G**	32.4412	196.1088	-184.9250	246.8600	395.3640	365.8180
6-311+G(2d,p)	32.4000	194.5000	-182.2900	246.8690	385.7540	360.0760
PC-J0	33.8736	201.0753				
PC-J1	32.2196	194.6676				
EDF2/6-31G*	32.1340	190.3690	-110.5960	180.7560	412.4700	368.8170
EDF2/6-31+G*	32.0120	191.3710	-112.6650	175.5010	406.8510	360.9880
EDF2/6-31G**	31.7100	192.4450	-110.4530	180.8260	410.6870	368.0760
EDF2/6-31+G**	31.5980	193.1940	-112.4710	175.3460	404.6300	360.3590
EDF2/6-311G*	32.2410	184.7540	-136.9780	153.8800	337.3290	294.6500
EDF2/6-311+G*	32.2010	184.5320	-145.1610	155.2890	336.8110	293.2240
EDF2/6-311G**	31.9570	184.9030	-137.0820	154.0490	338.9200	292.8010
EDF2/6-311+G**	31.9310	184.3930	-145.1760	155.3610	337.7560	291.9270
EDF2/6-311++G**	31.9190	184.5420	-145.2370	155.3650	336.8950	292.1020
EDF2/6-311+G(2d,p)	31.8930	183.7730	-146.0210	158.3640	326.5240	288.3070
EDF2/PC-J0	32.9634	186.4313				
EDF2/PC-J1	31.6027	181.6027				
BLYP/6-31G*	31.8898	184.0004	-108.0000	134.3119	402.4730	351.3090
BLYP/6-31+G*	31.7630	185.0810	-111.5990	130.3580	397.7680	344.5740
BLYP/6-31G**	31.4860	186.2780	-107.7320	134.3010	402.2490	350.9010
BLYP/6-31+G**	31.3780	187.3770	-111.4030	130.2770	392.9890	344.0820
BLYP/6-311G*	32.0410	179.6600	-136.0360	104.7670	328.6000	270.0880
BLYP/6-311+G*	32.0070	179.0360	-145.2880	107.8570	321.1830	268.6140
BLYP/6-311G**	31.7760	178.0360	-136.1390	104.4190	331.9530	268.6970
BLYP/6-311+G**	31.7420	178.3800	-145.3260	107.3890	323.1760	267.6970
BLYP/6-311++G**	31.7310	178.7600	-145.1020	107.9910	323.2360	268.8390

BLYP/6-311+G(2d,p)	31.7130	178.2330	-146.8750	110.3870	316.6360	267.5940
BP/6-31G*	31.7540	188.0611	-97.0220	142.4206	406.3890	354.6710
BP/6-31+G*	31.6280	188.6910	-98.8150	139.9320	401.9640	348.6710
BP/6-31G**	31.3470	190.0140	-98.8200	142.4720	405.1030	353.9890
BP/6-31+G**	31.2390	190.6410	-98.6030	139.8580	397.9700	347.9100
BP/6-311G*	31.8560	182.9310	-121.4130	115.7860	341.1380	282.1140
BP/6-311+G*	31.8130	182.5960	-129.4560	116.8670	339.5860	282.5860
BP/6-311G**	31.5810	182.8640	-121.4970	115.7740	344.6450	280.5890
BP/6-311+G**	31.5450	182.2710	-129.3250	116.8340	331.3310	281.4280
BP/6-311++G**	31.5330	182.7840	-129.4050	116.8490	335.0130	280.9070
BP/6-311+G(2d,p)	31.5070	181.9580	-131.0130	121.3390	328.7120	278.3140
B3LYP/6-31G*	32.1842	189.6263	-117.7850	180.1410	413.6860	370.9310
B3LYP/6-31+G*	32.0622	190.6658	-120.1610	173.4960	408.2430	362.4460
B3LYP/6-31G**	32.7550	191.7270	-117.5850	180.3140	411.8990	369.5210
B3LYP/6-31+G**	31.6590	192.5140	-119.9760	173.4550	405.9210	361.8720
B3LYP/6-311G*	32.2872	184.2227	-143.7842	152.2110	338.7460	294.9520
B3LYP/6-311+G*	32.2467	183.9371	-152.4219	153.8600	338.3000	293.2580
B3LYP/6-311G**	31.9998	184.3793	-143.9561	152.1760	340.3220	293.0720
B3LYP/6-311+G**	31.9852	183.9396	-152.4219	153.8350	339.2570	292.0990
B3LYP/6-311++G**	31.9669	184.0443	-152.4773	153.8360	338.2340	292.2340
B3LYP/6-311+G(2d,p)	31.9778	183.9092	-153.2297	156.5940	328.2710	289.0650
B3LYP/6-311G(2d,2p)	31.8149	183.7248			329.0500	
B3LYP/PC-J0	33.0129	185.8806				
B3LYP/PC-J1	31.6464	180.8468				
ΩB97X-D/6-31G*	32.2255	195.0090	-115.3090	203.6770	419.2400	473.9430
ΩB97X-D/6-31+G*	32.1132	195.6450	-120.1610	173.4960	415.1300	362.4460
ΩB97X-D/6-31G**	31.7943	196.7768	-117.5850	180.3140	417.4570	369.5210
ΩB97X-D/6-31+G**	31.6954	197.2185	-119.9760	173.4550	412.9630	361.8720
ΩB97X-D/6-311G*	32.2896	189.4991	-143.7842	152.2110	354.7120	294.9520
ΩB97X-D/6-311+G*	32.2378	189.2949	-152.4219	153.8600	354.3290	293.2580
ΩB97X-D/6-311G**	32.0044	189.6250	-143.9561	152.1760	356.1450	293.0720
ΩB97X-D/6-311+G**	31.9772	189.2128	-152.4219	153.8350	355.5770	292.0990
ΩB97X-D/6-311++G**	31.9639	189.4118	-152.4773	153.8360	354.4180	292.2340
ΩB97X-D/6-311+G(2d,p)	31.9297	188.3941	-153.2297	156.5940	344.4360	289.0650
ΩB97X-D/6-311G(2D,2P)	31.8051	188.8165			344.8400	
ΩB97X-D/PC-J0	33.0809	192.4019				
ΩB97X-D/PC-J1	31.6687	186.1951				
ΩB97X-V/6-31G*	32.1775	197.4313	-117.7850	180.1410	424.0790	370.9310
ΩB97X-V/6-31+G*	32.0228	197.1721	-120.1610	173.4960	419.2680	362.4460
ΩB97X-V/6-31G**	31.7343	199.4499	-117.5850	180.3140	422.3330	369.5210
ΩB97X-V/6-31+G**	31.6065	198.9342	-119.9760	173.4550	417.1200	361.8720
ΩB97X-V/6-311G*	32.2128	194.0790	-143.7842	152.2110	366.9670	294.9520
ΩB97X-V/6-311+G*	32.1399	193.8788	-152.4219	153.8600	366.5780	293.2580
ΩB97X-V/6-311G**	31.9122	194.1445	-143.9561	152.1760	368.2830	293.0720
ΩB97X-V/6-311+G**	31.8812	193.7954	-152.4219	153.8350	367.8110	292.0990
ΩB97X-V/6-311++G**	31.9510	193.4655	-152.4773	153.8360	367.1550	292.2340
ΩB97X-V/6-311+G(2d,p)	31.9778	183.9092	-153.2297	156.5940	328.2710	289.0650
ΩB97XV/6-311G(2d,2p)	31.6679	192.3933			356.5920	
ΩB97XV/6-311++G(2d,2p)	31.6674	192.0270			356.2930	

Glossary

This glossary addresses nomenclature ("jargon") used throughout the manual, and more generally, in applications of molecular modeling. Some of the terms, such as **Basis Set**, will be familiar to the modeling community but foreign to the broader community, while other terms, such as **Kinetic Product**, may be foreign to the modeling community and familiar to the broader community. Terms such as **Entropy** should be familiar to all and are generally avoided. A number of terms from both categories have been defined "in the context" of their use in molecular modeling rather than more generally. For example, the term **NOE** has been defined in terms of providing information about which pairs of (non-bonded) atoms are inside a specified cutoff distance (information that is useful in restricting the limits of a conformational search) rather than the nature of the phenomenon which produces this information.

3-21G. A **Split-Valence Basis Set** in which each **Core Basis Function** is written in terms of three **Gaussians**, and each **Valence Basis Function** is split into two parts, written in terms of two and one **Gaussians**, respectively. 3-21G basis sets have been determined to yield the lowest total **Hartree-Fock Energies** for atoms. The nomenclature 3-21G implies use of the **3-21G(*)** basis set on molecules incorporating second-row (and heavier) main-group elements.

3-21G$^{(*)}$. The **3-21G Basis Set** supplemented by d-type **Gaussians** for each second-row and heavier main-group elements only. This is required in order to get good equilibrium geometries. Without the added functions, the 3-21G basis set is not recommended for use for molecules incorporating second-row (and heavier) main-group elements.

6-31G. A **Split-Valence Basis Set** in which each **Core Basis Function** is written in terms of six **Gaussians**, and each **Valence Basis Function** is split into two parts, written in terms of three and one **Gaussians**, respectively. 6-31G basis sets have been determined to yield the lowest total **Hartree-Fock Energies** for atoms. The 6-31G basis set is rarely used alone but rather is supplemented by one or more sets of **Polarization Functions** and/or **Diffuse Functions**.

6-31G*, **6-31G****. The **6-31G Basis Set** in which non-hydrogen atoms are supplemented by d-type **Gaussians** and (for 6-31G**) hydrogen atoms are supplemented by p-type **Gaussians** (**Polarization Functions**). 6-31G* and 6-31G** are **Polarization Basis Sets**.

6-31+G*, **6-31+G****. **Basis Sets** that are identical to **6-31G*** and **6-31G**** except that all non-hydrogen atoms are supplemented by diffuse s and p-type **Gaussians** (**Diffuse Functions**). 6-31+G* and 6-31+G** are supplemented **Polarization Basis Sets**.

6-311G. A **Split-Valence Basis Set** in which each **Core Basis Function** is written in terms of six **Gaussians**, and each **Valence Basis Function** is split into three parts, written in terms of three, one and one **Gaussians**, respectively. 6-311G basis sets have been determined to

yield the lowest total **MP2 Energies** for atoms. The 6-311G basis set is rarely used alone but rather is supplemented by one or more sets of **Polarization Functions** and/or **Diffuse Functions**.

6-311G*, **6-311G****. The **6-311G Basis Set** in which non-hydrogen atoms are supplemented by d-type **Gaussians** and (for 6-311G**) hydrogen atoms are supplemented by p-type **Gaussians** (**Polarization Functions**). 6-311G* and 6-311G** are **Polarization Basis Sets**.

6-311+G*, **6-311+G****. **Basis Sets** that are identical to **6-311G*** and **6-311G**** except that non-hydrogen atoms are supplemented by diffuse s and p-type **Gaussians** (**Diffuse Functions**). 6-311+G* and 6-311+G** are supplemented **Polarization Basis Sets**.

***Ab Initio* Models**. The general term used to describe methods seeking approximate solutions to the many-electron **Schrödinger Equation**, but which do not involve empirical parameters. *Ab initio* models include **Hartree-Fock Models**, **Møller-Plesset Models** and **Density Functional Models**.

Acidity. The energy required to break an XH bond into X^- and H^+, that is, to undergo **Heterolytic Bond Dissociation**. Acidities in the gas phase are very large positive numbers (producing isolated ions is highly unfavorable); they are much smaller (and of either sign) in solution.

Activation Energy. The energy of a **Transition State** above that of reactants. Activation energy is related to reaction rate by way of the **Arrhenius Equation**.

AM1. **A**ustin **M**ethod **1**. A **Semi-Empirical Model**.

Anharmonic Frequency Correction. Non-quadratic contributions to a **Vibrational Frequency**. Because calculated frequencies are harmonic, that is, assume that the potential is quadratic, they are typically ~5% larger than measured frequencies, which include anharmonic corrections. Frequencies can be corrected for anharmonicity but, due to high computational cost, this is seldom done for any but very small molecules.

Anion. An atom or molecule with a **Total Charge** of -1.

Antibonding Molecular Orbital. A **Molecular Orbital** which is antibonding between particular atomic centers. The opposite is a **Bonding Molecular Orbital**.

Arrhenius Equation. An equation governing the rate of a chemical reaction as a function of the **Activation Energy** and the temperature.

Atomic Orbital. A **Basis Function** centered on an atom. Atomic orbitals typically take on the form of the solutions to the hydrogen atom (s, p, d, f... type orbitals).

Atomic Units. The set of units which remove all of the constants from inside the **Schrödinger Equation**. The **Bohr** is the atomic unit of length and the **Hartree** is the atomic unit of energy.

Atomization Energy. The energy of dissociation of an atom or molecule into separated

nuclei and electrons. This is another term for **Total Energy**.

B3LYP Model. A **Global Hybrid Generalized Gradient Approximation (GH-GGA)** functional which improves on the **Local Density Model** by accounting explicitly for non-uniformity in electron distributions. It incorporates the **Exchange Energy** from the **Hartree-Fock Model** and involves three adjustable parameters. The B3LYP model is perhaps the most widely used density functional model and provides a solid account of equilibrium geometries (including the geometries of molecules incorporating transition metals), vibrational frequencies and reaction energies.

Basicity. The energy gained from adding a proton to an atom or molecule. Basicities in the gas phase are very large negative numbers; they are much smaller (and of either sign) in solution.

Basis Functions. Functions usually centered on atoms which are linearly combined to make up the set of **Molecular Orbitals**. Except for **Semi-Empirical Models** where basis functions are **Slater** type, basis functions are **Gaussian** type.

Basis Set. The entire collection of **Basis Functions**.
Becke-Lee-Yang-Parr Model. ***See*** **BLYP Model, B3LYP Model. Becke-Perdew Model**. *See* **BP Model**.

BLYP Model. A **Generalized Gradient Approximation (GGA)** functional which improves on the **Local Density Model** by accounting explicitly for non-uniformity in electron distributions.

Bohr. The **Atomic Unit** of length. 1 Bohr = 0.529167Å.

Boltzmann Distribution. The equilibrium distribution of products in a **Thermodynamically-Controlled Reaction**. A special (and very important) case involves the equilibrium distribution of **Conformers** of a flexible molecule.

Boltzmann Equation. The equation governing the distribution of products in **Thermodynamically-Controlled Reaction**.

Bonding Molecular Orbital. A **Molecular Orbital** which is bonding between particular atom centers. The opposite is an **Antibonding Molecular Orbital**.

Bond Order. A measure of the number of electrons shared between "bonded" atoms.

Bond Separation Reaction. An ***Isodesmic*** **Reaction** in which a molecule described in terms of a conventional valence structure is broken down into the simplest (two-heavy-atom) molecules containing the same component bonds.

Bond Density Surface. A **Density Surface** used to elucidate the bonding in molecules. The value of the density is typically taken as
0.1 electrons/**bohr**^{3}.

Born-Oppenheimer Approximation. An approximation based on the assumption that

nuclei are stationary. Applied to the **Schrödinger Equation**, it leads to the **Electronic Schrödinger Equation**.

C-PCM. The **Conductor** like **Polarizable Continuum Model** is a solvent model that depends on the shape and charge distribution of the molecule (solute) and the dielectric constant of the solvent. This is the default solvent model used in the ***Spartan'18*** version.

Cambridge Structural Database (CSD). A collection of >900,000 experimental structures for organic and organometallic compounds almost entirely from X-ray crystallography. Structures lacking carbon are excluded as are biopolymers (the latter are available from the Protein Data Bank).

Cartesian Coordinates. x, y, z spatial coordinates.

CAS Number. Chemical Abstract Service registry number; a unique substance identifier used to a large degree in the chemistry literature, see www.cas.org for more information.

Cation. An atom or molecule with a **Total Charge** of +1.

cc-pV5Z. A **Polarization Basis Set** which has been developed specifically for use with **Correlated Models**. The **Valence** has been split into five parts, and functions though g have been included. May be supplemented with **Diffuse Functions**.

cc-pVDZ. A **Polarization Basis Set** which has been developed specifically for use with **Correlated Models**. The **Valence** has been split into two parts. May be supplemented with **Diffuse Functions**.

cc-pVQZ. A **Polarization Basis Set** which has been developed specifically for use with **Correlated Models**. The **Valence** has been split into four parts, and functions though g have been included. May be supplemented with **Diffuse Functions**.

cc-pVTZ. A **Polarization Basis Set** which has been developed specifically for use with **Correlated Models**. The **Valence** has been split into three parts, and functions though f have been included. May be supplemented with **Diffuse Functions**.

^{13}C Spectrum; *see* **NMR Spectrum**

CCSD/CCSD(T) Models. **C**oupled **C**luster **S**ingles and **D**oubles optionally with a triples correction. **Correlated Models**.

Chemical Function Descriptors. A set of locations each indicating specific "chemical behavior", for example, a location indicating a possible role as a hydrogen-bond donor or acceptor. Chemical function descriptors may either be used without reference to atomic centers (the collection being termed a **Pharmacophore**) or attached to atoms.

Chemical Shift. The difference in the energies of nuclear spin states for a nucleus in a molecule caused by an external magnetic field. This is the fundamental quantity measured in **NMR** spectroscopy.

Chiral Molecule. A molecule with a non-superimposable mirror image. Most commonly (but not necessarily) a chiral molecule will incorporate one or more **R/S** centers.

CID Model. **C**onfiguration **I**nteraction **D**oubles. A limited **Configuration Interaction** scheme in which only double excitations from occupied to unoccupied molecular orbitals are considered.

CIS Model. **C**onfiguration **I**nteractions **S**ingles. A limited **Configuration Interaction** scheme in which only single excitations from occupied to unoccupied molecular orbitals are considered. This is perhaps the simplest method available to the description of **Excited States** of molecules.

CISD Model. **C**onfiguration **I**nteraction, **S**ingles and **D**oubles. A limited **Configuration Interaction** scheme in which only single and double excitations from occupied to unoccupied molecular orbitals are considered.

Closed Shell. An atom or molecule in which all electrons are paired.

Configuration Interaction. Provides an account of **Electron Correlation** by way of explicit promotion (excitation) of electrons from occupied molecular orbitals into unoccupied molecular orbitals. Full configuration interaction (all possible promotions) is not a practical method and limited schemes, for example, **CIS**, **CID** and **CISD Models**, need to be employed.

Conformation. The arrangement about single bonds and of flexible rings.

Conformational Analysis. One of several different procedures to deal with molecules that are not properly described in terms of a single **Conformer**. Determination of **Equilibrium Conformer** or establishing a **Conformer Distribution** or a **Similarity Library** are examples of conformational analysis.

Conformer; See **Conformation**.

Core. Electrons which are primarily associated with individual atoms and do not participate significantly in chemical bonding (1s electrons for first-row elements, 1s, 2s, $2p_x$, $2p_y$, $2p_z$ electrons for second-row elements, etc.).

Core Orbital. A **Molecular Orbital** that is primarily involved with the **Core** rather than with the **Valence**.

Corrected MMFF. See **DLFF03**, a machine learning approach improving MMFF energies. *J. Comp. Chem.* **2025**, *46* (1), 70016. https://doi.org/10.1002/jcc.70016

Correlated Models. Models which take implicit or explicit account of the **Correlation** of electron motions. **Møller-Plesset Models**, **Configuration Interaction Models** and **Density Functional Models** are correlated models.

Correlation. The coupling of electron motions not explicitly taken into account in **Hartree-Fock Models**.

Correlation Energy. The difference in energy between the **Hartree- Fock Energy** and the experimental energy.

Correlation Functional. A function of the **Electron Density** and perhaps as well the gradient of the **Electron Density**. The functional form derives from "exact" solution of the **Schrödinger Equation** for an idealized many-electron problem. Used in **Density Functional Models**.

COSY Spectrum; *See* **NMR Spectrum**.

Coulomb Energy. The electron-electron repulsion energy according to Coulomb's law.

Coulombic Interactions. Charge-charge interactions which follow Coulomb's law. Stabilizing when charges are of opposite sign and destabilizing when they are of the same sign.

Coupling Constant. A measure of the extent to which nuclear spins on neighboring atoms interact (couple) with each other, causing splitting of the lines in an NMR spectrum.

CPK Model. A molecular model in which atoms are represented by spheres, the radii of which correspond to **van der Waals Radii**. Intended to portray molecular size and shape. Also known as a space-filling model.

CSD; *See* **Cambridge Structural Database**.

Curtin-Hammett Principle. The idea that the reactive conformer in a **Kinetically-Controlled Reaction** is not necessarily the lowest- energy conformer. The rationale is that the energy barriers separating conformers are typically much smaller than barriers to chemical reaction, and that any reactive conformers will be replenished.

Density; *See* **Electron Density**.

Density Functional Models. Methods in which the energy is evaluated as a function of the **Electron Density**. **Electron Correlation** is taken into account explicitly by incorporating into the **Hamiltonian** terms which derive from exact solutions of idealized many-electron systems.

DEPT Spectrum; *See* **NMR Spectrum**.

Diffuse Functions. Functions added to a **Basis Set** to allow description of electron distributions far away from atomic positions. Important for descriptions of anions.

Diffusion-Controlled Reactions. Chemical reactions without **Transition States** (or energy barriers), the rates of which are determined by the speed in which molecules encounter each other and how likely these encounters are to lead to reaction.

Dipole Moment. A vector indicating the magnitude, sign and direction of the overall distribution of charge in a molecule based on the combination of nuclear and electron charges. Most commonly only the magnitude of the dipole moment is known experimentally.

DLFF03. **D**eep **L**earning **F**orce **F**ield #**3**. Under the umbrella of machine learning (ML) models, this early neural network (commonly termed "Corrected MMFF" was our first pass at implementing ML approaches into Spartan. It improves MMFF conformer energy differences to density functional quality: *J. Comp. Chem.* **2025**, *46* (1), 70016. https://doi.org/10.1002/jcc.70016

DLFFG1. **D**eep **L**earning **F**orce **F**ield **G**eometry #**1**. Commonly termed "Est. density functional", this was our third class of ML-based method, and the first to tackle geometry. Trained on more than 6 million structures, this neural network improves MMFF geometries to ωB97XD/6-31G* quality. *J. Chem. Info. Mod.* **2025**, *65* (5), 2314–2321. https://doi.org/10.1021/acs.jcim.4c01898

DLHF2XD. **D**eep **L**earning **H**artree-**F**ock/3-21G(*) to (**2**) ωB97**X-D**/6-31G*. This neural network model (our seventh ML-based method, trained on 11 million molecules) take HF/3-21G(*) energy and corrects this to ωB97XD/6-31G*. While this can be used to improve energies from, for example, the DLFFG1 model, it is most often used in conjunction with the **MLHF** or **MLXD** NMR workflows for predicting accurate ^{1}H and ^{13}C shifts of flexible organic molecules. *ChemRxiv* preprint. **2026**: https://doi.org/10.26434/chemrxiv.15001441/v1

DLXD2M2. **D**eep **L**earning ωB97**X-D**/6-31G* to (**2**) ωB97**M(2)**/6-311+G(2df,2p). see below. *J. Comp. Chem.* **2025**, *46* (13), 70129. https://doi.org/10.1002/jcc.70129

DLXD2MV. **D**eep **L**earning ωB97**X-D**/6-31G* to (**2**) ωB97**M-V**/6-311+G(2df,2p). see below. *J. Comp. Chem.* **2025**, *46* (13), 70129. https://doi.org/10.1002/jcc.70129

DLXD2XV. **D**eep **L**earning ωB97**X-D**/6-31G* to (**2**) ωB97**X-D)**/6-311+G(2df,2p). This approach, along with the above 2 options are machine learning models trained on ≈300k energy calculations (per neural network) providing "better" energies for use in Boltzmann weighting. *J. Comp. Chem.* **2025**, *46* (13), 70129. https://doi.org/10.1002/jcc.70129

DLNMR1. **D**eep **L**earning **NMR** #**1**. Technically two different neural networks (one for ^{1}H shifts and one for ^{13}C shifts), this machine learning model supports neutral organic systems comprising H, C, N, O, F, S, Cl, and Br, and can furnish accurate predicted ^{1}H and ^{13}C shifts in seconds. DLNMR1 was trained on >2.7 million empirically corrected ωB97X-D/6-31G* shifts. Definition and rigid molecule assessment: *J. Org. Chem.* **2025**, *90* (32), 11478–11485. https://pubs.acs.org/doi/10.1021/acs.joc.5c00927

Double Hybrid Functional. A class of functionals that take account of unoccupied molecular orbitals in addition to occupied molecular orbitals (the density) usually by way of an MP2-like procedure.

Doublet. A molecule with one unpaired electron**,** also known as a radical.

Dual Basis Set. A technique whereby the converged wavefunction resulting from a calculation using a small basis set is passed on for a single **SCF** cycle using a large basis set. This closely approximates the energy and wavefunction that would have resulted from a full calculation using the large basis set but with 5-10 times less computation.

EDF2 Model. A **Global Hybrid Generalized Gradient Approximation (GH-GGA)**

functional which improves on the **Local Density Model** by accounting explicitly for non-uniformity in electron distributions. EDF2 was explicitly formulated to reproduce experimental vibrational frequencies (infrared spectra), and it was for this reason (together with practical concerns) that the EDF2/6-31G* model was chosen for use in the **Spartan Spectra and Properties Database (SSPD)**.

Electron Affinity. The energy gained as a result of adding an electron to an atom or molecule. May be calculated as the (negative of) the difference in energies between the atom/molecule and it corresponding radical anion. Note that not all atoms/molecules have positive electron affinities, that is, bind an electron.

Electron Correlation. *See* **Correlation**.

Electron Density. The number of electrons per unit volume at a point in space. This is the quantity which is measured in an X-ray diffraction experiment.

Electron Density Surface. A surface of constant **Electron Density**. May be used to elucidate bonding (**Bond Density Surface**) or to characterize overall molecular size and shape.

Electronic Schrödinger Equation. The equation which results from incorporation of the **Born-Oppenheimer Approximation** to the **Schrödinger Equation**.

Electrostatic Charges. Atomic charges chosen to best match the **Electrostatic Potential** at points surrounding a molecule, subject to overall charge balance.

Electrostatic Potential. A function describing the energy of interaction of a positive point charge with the nuclei and fixed electron distribution of a molecule.

Electrostatic Potential Map. A graph that shows the value of **Electrostatic Potential** on an **Electron Density Isosurface** corresponding to a **van der Waals Surface**.

Electrostatic Potential Surface. It may be used to elucidate regions in a molecule which are particularly electron rich and subject to electrophilic attack and those which are particularly electron poor, subject to nucleophilic attack.

Energy Profile. A plot of energy vs. change in geometrical coordinate,
for example, change in bond length or in dihedral angle.

Enthalpy. The heat of a chemical reaction. Related to energy by a pressure-volume that is insignificant at normal pressures.

Equilibrium Conformer. The lowest-energy **Conformer**.

Equilibrium Geometry. A **Local Minimum** on a **Potential Energy Surface**.

Equilibrium Isotope Effect. The effect on the heat of reaction due to a change in nuclear mass of one or more atoms. The most common (and largest) isotope effects arise from substitution of deuterium for hydrogen.

Est. Density Functional. A contextual prefix indicating the model is machine learning-

based, for example: Est. Density Functional > Est. ωB97X-D/6-31G* geometries.

Exchange Functional. A function of the **Electron Density** and perhaps as well the gradient of the **Electron Density**. The functional form derives from "exact" solution of the **Schrödinger Equation** for an idealized many-electron problem. Used in **Density Functional Models**.

Exchange Energy. An "attractive" (negative) component of the electron-electron interaction energy. Arises due to an overestimation of the "repulsive" (positive) component or **Coulomb energy**.

Excited State. An electronic state for an atom or molecule which is not the lowest-energy or **Ground State**.

Force Constant. For a diatomic molecule, the second derivative of the energy with respect to change in bond length. For a polyatomic molecule, the second derivative of the energy with regard to a change in **Normal Coordinate.**

Force Field. The set of rules underlying **Molecular Mechanics Models**. Comprises terms which account for distortions from ideal bond distances and angles and for **Non-Bonded van der Waals** and **Coulombic Interactions**.

Fourier Series. An expansion of a periodic (in 360° rotation) in terms of powers of the cosine of the torsion angle.

Frontier Molecular Orbitals. The **HOMO** and **LUMO**.

Gx Models. "Recipes" which combine a series of **Hartree-Fock** and **Correlated Models** to properly account for **Atomization Energies** of atoms and molecules. Useful for providing accurate thermochemical data.

G3(MP2). A "simplified" version of the **G3 Model**. Reproduces experimental **Heats of Formation** with an RMS errors of ~8 kJ/ mol, practical for molecules comprising less than 20 atoms.

Gaussian. A function of the form $x^l y^m z^n \exp(\alpha r^2)$ where l, m, n are integers (0, 1, 2 . . .) and α is a constant. Used in the construction of **Basis Sets** for **Hartree-Fock**, **Density Functional**, **MP2** and other **Correlated Models**.

Gaussian Basis Set. A **Basis Set** made up of **Gaussian Basis Functions**.

Global Minimum. The lowest energy **Local Minimum** on a **Potential Energy Surface**.

Ground State. The lowest energy electronic state for an atom or molecule.

Hamiltonian. An operator which accounts for the kinetic and potential energy of an atom or a molecule.

Hammond Postulate. The idea that the **Transition State** for an *exothermic* reaction will more closely resemble reactants than products. This provides the basis for "modeling" properties of **Transition States** in terms of the properties of reactants.

Harmonic Frequency. A **Vibrational Frequency** which has been corrected to remove all non-quadratic (anharmonic) components. Calculated **Vibrational Frequencies** correspond to harmonic frequencies. The corrections require data on isotopically-substituted systems and are typically available only for small molecules.

Hartree. The **Atomic Unit** of energy. 1 Hartree = 627.47 kcal/mol.

Hartree-Fock Approximation. Separation of electron motions in many-electron systems into a product form of the motions of the individual electrons.

Hartree-Fock Energy. The energy resulting from **Hartree-Fock Models**.

Hartree-Fock Equations. The set of differential equations resulting from application of the **Born-Oppenheimer** and **Hartree-Fock Approximations** to the many electron **Schrödinger Equation**.

Hartree-Fock Models. Methods in which the many-electron wavefunction in written terms of a product of one-electron wavefunctions. Electrons are assigned in pairs to functions called **Molecular Orbitals**.

Hartree-Fock Wavefunction. The simplest quantum-mechanically correct representation of the many-electron wavefunction. Electrons are treated as independent particles and are assigned in pairs to functions termed **Molecular Orbitals**. Also known as **Single-Determinant Wavefunction**.

Heat of Formation. The **Enthalpy** at 298 K of a balanced chemical reaction in which a molecule is converted into the most stable form of each element, for example, the hydrogen molecule for hydrogen and graphite for carbon.

Hessian. The matrix of second energy derivatives with respect to geometrical coordinates. The Hessian together with the atomic masses lead to the **Vibrational Frequencies** of molecular systems.

Heterolytic Bond Dissociation. A process in which a bond is broken and a cation and anion result. The number of electron pairs is conserved, but a non-bonding electron pair has been substituted for a bond.

HOMO. **H**ighest **O**ccupied **M**olecular **O**rbital.

Homolytic Bond Dissociation. A process in which a bond is broken and two radicals result. The number of electron pairs is not conserved.

HOMO-LUMO Gap. The difference in energies between the **HOMO** and **LUMO**.

HOMO Map. A graph of the absolute value of the **HOMO** on an **Isodensity Surface** corresponding to a **van der Waals Surface**.

Hybrid Density Functional Models. **Density Functional Models** which incorporate the **Exchange Energy** from the **Hartree-Fock Model**. The **B3LYP Model** is a hybrid density functional model.

Hydrogen-Bond Acceptor. An electronegative element (typically, nitrogen, oxygen or halogen) with one or more lone pairs able to "accept" a hydrogen from a **Hydrogen-Bond Donor**.

Hydrogen-Bond Donor. A hydrogen bonded to an electronegative element (typically nitrogen or oxygen) that can be donated to a **Hydrogen-Bond Acceptor**.

Hydrophobe. A non-polar region of a molecule.

Hypervalent Molecule. A molecule containing one or more main- group elements in which the normal valence of eight electrons has been exceeded. Hypervalent molecules are common for second-row and heavier main-group elements but are uncommon for first-row elements.

Imaginary Frequency. A frequency which results from a negative element in the diagonal form of the **Hessian**. **Equilibrium Geometries** are characterized by all real frequencies while **Transition States** are characterized by one imaginary frequency.

Infrared Spectrum. Spectrum resulting from transitions between vibrational energy levels. Infrared intensities depend on the change in dipole moment associated with the vibration.

Internal Coordinates. A set of bond lengths, bond angles and dihedral angles (among other possible variables) describing molecular geometry. The **Z Matrix** is a set of internal coordinates.

Intrinsic Reaction Coordinate. The name given to procedures with seek to "walk" in a smooth pathway connecting reactant and product.

Ionization Potential. The energy required to completely eject an electron from an atom or molecule. May be calculated as the difference in energies between the atom/molecule and its corresponding radical cation. May be approximated from **Koopman's Theorem** as the negative of the **HOMO Energy**.

***Isodesmic* Reaction**. A chemical reaction in which the number of formal chemical bonds of each type is conserved.

Isosurface. A three-dimensional surface defined by the set of points in space where the value (isovalue) of the function is constant.

Kinetically-Controlled Reaction. Refers to a chemical reaction which has not gone all the way to completion, and the ratio of products is not related to their thermochemical stabilities, but rather inversely to the heights of the energy barriers separating reactants to products.

Kinetic Isotope Effect. The effect on the rate of reaction due to a change in nuclear mass of one or more atoms. The most common (and largest) isotope effects arise from substitution of deuterium for hydrogen.

Kinetic Product. The product of a **Kinetically-Controlled Reaction**.

kJ/mol. The unit commonly used to express chemical energy differences, replacing kcal/mol. 1 kJ/mol = 0.239 kcal/mol = 0.000381 au.

Kohn-Sham Equations. The set of equations obtained by applying the **Local Density Approximation** to a general multi-electron system. An **Exchange/Correlation Functional** which depends on the electron density has replaced the **Exchange Energy** expression used in the **Hartree-Fock Equations**. The Kohn-Sham equations become the **Roothaan-Hall Equations** if this functional is set equal to the **Hartree-Fock Exchange Energy** expression.

Koopman's Theorem. Relates the **Ionization Potential** to the negative of the **HOMO Energy.**

LCAO Approximation. **L**inear **C**ombination of **A**tomic **O**rbitals approximation. Approximates the unknown **Hartree-Fock Wavefunctions** (**Molecular Orbitals**) by linear combinations of atom-centered functions (**Atomic Orbitals**) and leads to the **Roothaan-Hall Equations**.

LACVP*, **LACVP****. **Pseudopotentials** which include not only **Valence** s and p-type orbitals (s, p and d-type for transition metals) but also the highest set of **Core** orbitals, and in which non-hydrogen atoms are supplemented by d-type **Gaussians** and (for LACVP**) hydrogen atoms are supplemented by p-type **Gaussians**.

Ligand. This term has two meanings in the context of molecular modeling. It denotes a molecular fragment that is (viewed as) non-covalently bonded to a transition metal in an inorganic or organometallic compound. For example, both carbon monoxide and benzene are said to be ligands in benzene chromium tricarbonyl. It also denotes a molecule that is "associated" (non-covalently bonded) to a protein.

Linear Synchronous Transit. The name given to procedures which estimate the geometries of transition states based on "averages" of the geometries of reactants and products, sometimes weighted by the overall thermodynamics of reaction.

LMP2 Model. An **MP2 Model** in which the **Hartree-Fock** orbitals are first localized.

Local Density Models. **Density Functional Models** which are based on the assumption that the **Electron Density** is constant (or slowly varying) throughout all space.

Local Ionization Potential. A function of the relative ease of electron removal (ionization) from a molecule.

Local Ionization Potential Map. A graph of the **Local Ionization Potential** on an **Isodensity Surface** corresponding to a **van der Waals Surface**.

Local Minimum. Any **Stationary Point** on a **Potential Energy Surface** for which all elements in the diagonal representation of the **Hessian** are positive.

LogP. An empirical measure of the extent to which a molecule partitions between water and octanol.

Lone Pair. A **Non-Bonded Molecular Orbital** which is typically associated with a single atom.

LUMO. **L**owest **U**noccupied **M**olecular **O**rbital.

LUMO Map. A graph of the absolute value of the **LUMO** on an **Isodensity Surface** corresponding to a **van der Waals Surface**.

Mechanism. The sequence of steps connecting reactants and products in an overall chemical reaction. Each step starts from an equilibrium form (reactant or intermediate) and ends in an equilibrium form (intermediate or product).

Minimal Basis Set. A **Basis Set** which contains the fewest functions needed to hold all the electrons on an atom and still maintain spherical symmetry. **STO-3G** is a minimal basis set.

Minimal Valence Basis Set. A **Minimal Basis Set** from which **Core Orbitals** have been removed. **Semi-Empirical Models** make use of a minimal valence basis set.

MM2, MM3, MM4. A series of **Molecular Mechanics Force Fields** developed by the Allinger group at the University of Georgia.

MMFF94. **M**erck **M**olecular **F**orce **F**ield. A **Molecular Mechanics Force Field** for organic molecules and biopolymers developed by Merck Pharmaceuticals.

MNDO. **M**odified **N**eglect of **D**ifferential **O**verlap. A **Semi- Empirical Model.**

MNDO/d. An extension of the **MNDO Semi-Empirical Model** in which second-row (and heavier) main-group elements are provided a set of d-type functions.

Molecular Mechanics Models. Methods for structure, conformation and strain energy calculation based on bond stretching, angle bending and torsional distortions, together with **Non-Bonded Interactions**, and parameterized to fit experimental data.

Molecular Orbital. A one-electron function made of contributions of **Basis Functions** on individual atoms (**Atomic Orbitals**) and delocalized throughout the entire molecule.

Molecular Orbital Models. Methods based on writing the many- electron solution of the **Electronic Schrödinger Equation** in terms of a product of one-electron solutions (**Molecular Orbitals**).

Møller-Plesset Energy. The energy resulting from **Møller-Plesset Models** terminated to a given order, e.g., the **MP2** energy is the energy of the second-order **Møller-Plesset Model** (or **MP2**).

Møller-Plesset Models. Methods which partially account for **Electron Correlation** by way of the perturbation theory of Møller and Plesset.

Monte-Carlo Conformation Search. A limited search of **Conformers** resulting from a random sampling of all torsional angles (including constrained torsions associated with flexible rings) while satisfying the requirements imposed by the **Boltzmann Equation**.

MP2 Energy. The energy resulting from **MP2 Models**.

MP2 Model. A **Møller-Plesset Model** terminated to be second order in the energy.

MP3 Model. A **Møller-Plesset Model** terminated to be third order in the energy.

MP4 Model. A **Møller-Plesset Model** terminated to be fourth order in the energy.

Mulliken Charges. Atomic charges obtained from a **Mulliken Population Analysis**.

Mulliken Population Analysis. A partitioning scheme in which electrons are shared equally between different **Basis Functions**.

Multiplicity. The number of unpaired electrons (number of electrons with "down" spin) +1. 1=singlet; 2=doublet; 3=triplet, etc.

Natural Charges. Atomic charges obtained from a **Natural Population Analysis.**

Natural Population Analysis. Another partitioning scheme in which electrons are shared equally between different **Basis Functions**.

NDDO Approximation. **N**eglect of **D**iatomic **D**ifferential **O**verlap approximation. The approximation underlying all present generation **Semi-Empirical Models**. It says that two **Atomic Orbitals** on different atoms do not overlap.

Neural Network. Also often call "machine learning" or artificial (or simulated) neural network(s). In the context of ***Spartan***, these refer to a growing number computational routines, trained to specific tasks (provide energy, structure, property) as a significant time improvement over quantum chemical calculations. See **Appendix A** for an extended explanation.

NIST Chemistry Web Book (National Institutes of Standards and Technology). On-line source of experimental **Infrared** and **UV/visible spectra**.

NMR Spectrum. Spectrum resulting from transitions between nuclear spin states under the influence of an external magnetic field. Both 1D spectra (most important, proton and ^{13}C) and 2D spectra (most important COSY, HMBC and HSQC) are available.

NOE (Nuclear Overhauser Effect). An NMR experiment that leads to identifying pairs of (non-bonded) atoms separated by less than a specified distance (typically 4-5Å).

Non-Bonded Interactions. Interactions between atoms which are not directly bonded. **van der Waals Interactions** and **Coulombic Interactions** are non-bonded interactions.

Non-Bonded Molecular Orbital. A molecular orbital which does not show any significant **Bonding** or **Antibonding** characteristics. A **Lone Pair** is a non-bonded molecular orbital.

Non-Local Density Functional Models. **Density Functional Models** such as the **BLYP**, **BP**, **EDF1, EDF2** and **B3LYP Models** which explicitly take into account electron inhomogeneities.

Normal Coordinates. Coordinates that lead to a diagonal **Hessian**, and which correspond to vibrational motions in molecules.

Normal Mode Analysis. The process for calculating **Normal Coordinates** leading to **Vibrational Frequencies**. This involves diagonalizing the **Hessian** and accounting for atomic masses.

Octet Rule. The notion that main-group elements prefer to be "surrounded" by eight electrons (going into s, p_x, p_y, p_z orbitals).

OD/OD(T). **Correlated Models.**

Orbital Energy. The energy of a **Molecular Orbital.**

Orbital Energy Diagram. A diagram listing **Orbital Energies** from most negative at the bottom. Usually restricted to occupied **Valence Orbitals** and the few lowest unoccupied orbitals.

Orbital Symmetry Rules; *See* **Woodward-Hoffmann Rules**.

Open Shell. An atom or molecule in which one or more electrons are unpaired.

Pharmacophore. Collection of **Chemical Function Descriptors** which characterizes a molecular environment. A pharmacophore may serve as a target for **Similarity Analysis**.

PM3. **P**arameterization **M**ethod **3**. A **Semi-Empirical Model**. **PM6**. **P**arameterization **M**ethod **6**. A **Semi-Empirical Model**.

Polar Area. The area of an **Electrostatic Potential Map** that is either more positive or more negative than some preset value (typically between 80-120 kJ/mol).

Polarization Basis Set. A **Basis Set** which contains functions of higher angular quantum number (**Polarization Functions**) than required for the **Ground State** of the atom, e.g., d-type functions for main-group elements and p-type functions for hydrogen. **6-31G***, **6-31G****, **6-311G***, **6-311G****, **cc-pVDZ**, **cc-pVTZ** and **cc-pVQZ** are polarization basis sets.

Polarization Functions. Functions of higher angular quantum than required for the **Ground State** atomic description. Added to a **Basis Set** to allow displacement of **Valence Basis Functions** away from atomic positions.

Polarization Potential. A function describing the energy of electronic relaxation of a molecular charge distribution following interaction with a point positive charge. The polarization potential may be added to the **Electrostatic Potential** to provide a more accurate account of the interaction of a point-positive charge and a molecule.

Polar Surface Area. The area of a space-filling (**CPK**) model due to electronegative elements (typically, nitrogen and oxygen) and any hydrogens that are bonded to these elements.

Pople-Nesbet Equations. The set of equations describing the best **Unrestricted Single Determinant Wavefunction** within the **LCAO Approximation**. These reduce the **Roothaan-Hall Equations** for **Closed Shell** (paired electron) systems.

Potential Energy Surface. A function of the energy of a molecule in terms of the geometrical coordinates of the atoms.

Property Map. A representation or "map" of a "property" on top of an **Isosurface**, typically an **Isodensity Surface**. **Electrostatic Potential Maps**, and **HOMO** and **LUMO Maps** and **Spin Density Maps** are useful property maps.

Proton Affinity; *See* **Basicity**.

Proton Spectrum; *see* **NMR Spectrum**

Pseudopotential. A **Basis Set** which treats only **Valence** electrons in an explicit manner, all other electrons being considered as a part of a "**Core**". **LAVCP** (and extensions including **Polarization** and/ or **Diffuse Functions**) are pseudopotentials.

Pseudo rotation. A mechanism for interconversion of *equatorial* and *axial* sites around trigonal bipyramidal centers, e.g., fluorines in phosphorous pentafluoride.
QCCD/QCCD(T) Models. Correlated Models.

QCISD/QCISD(T) Models. **Q**uadratic **CISD Models**. **Correlated Models**.

QSAR. **Q**uantitative **S**tructure **A**ctivity **R**elationships. The name given to attempts to relate measured or calculated properties to molecular structure.

Quantum Mechanics. Methods based on approximate solution of the **Schrödinger Equation**.

Ramachandran Plot. A plot for each α carbon in a polypeptide of the two dihedral angles that define the conformation of the backbone. The points will typically cluster in two regions denoting helices and sheets.

Raman Spectrum. Spectrum resulting from transitions between vibrational energy levels. Infrared intensities depend on the change in polarizability associated with the vibration.

Rate Limiting Step. The step in an overall chemical reaction (**Mechanism**) which proceeds via the highest-energy **Transition State**.

Reaction Coordinate. The coordinate that connects the **Local Minima** corresponding to

the reactant and product, and which passes through a **Transition State**.

Reaction Coordinate Diagram. A plot of energy vs. **Reaction Coordinate**.

Restricted SCF. An **SCF** procedure which restricts electrons to be paired in orbitals or permits orbitals to be singly occupied.

RM1. A **Semi-Empirical Model**.

Roothaan-Hall Equations. The set of equations describing the best **Hartree-Fock** or **Single-Determinant Wavefunction** within the **LCAO Approximation**.

R/S. Atomic chirality markers.

SCF. **S**elf-**C**onsistent **F**ield. An iterative procedure whereby a one- electron orbital is determined under the influence of a potential made up of all the other electrons. Iteration continues until self-consistency. **Hartree-Fock**, **Density Functional** and **MP2 Models** all employ SCF procedures.

Schrödinger Equation. The quantum mechanical equation which accounts for the motions of nuclei and electrons in atomic and molecular systems.

SDBS. Database of experimental infrared and NMR spectra from AIST (Advanced Industrial Science and Technology, Japan). Freely accessible on-line but automatic access is not permitted limiting its utility.

Self-Consistent Field; *See* **SCF**.

Semi-Empirical Models. **Quantum Mechanics** methods that seek approximate solutions to the many electron **Schrödinger Equation**, but which involve empirical parameters.

Similarity Analysis. A procedure to determine the extent to which a molecule is "similar" to either a standard molecule (defined in terms of a set of atoms or a set of **Chemical Functional Groups**) or by a set of unassociated **Chemical Functional Groups** (a **Pharmacophore**).

Similarity Distribution. A selection of **Conformers** based on their energies, with emphasis on maximum diversity.

Size Consistent. Methods for which the total error in the calculated energy is more or less proportional to the (molecular) size. **Hartree- Fock** and **Møller-Plesset** models are size consistent, while **Density Functional Models**, (limited) **Configuration Interaction Models** and **Semi-Empirical Models** are not size consistent.

Size Surface. A **Density Surface** used to establish overall molecular size and shape. The value of the density is typically taken as 0.002 electrons/**bohr**^{3}.

Slater. A function of the form $x^l y^m z^n \exp(-\zeta r)$ where l, m, n are integers (0, 1, 2 . . .) and ζ is a constant. Related to the exact solutions to the **Schrödinger Equation** for the hydrogen atom. Used as **Basis Functions** in **Semi-Empirical Models**.

SM5.4. An empirical solvation model for water only. Furnishes only an energy correction to be added to the (gas-phase) energy obtained from any quantum chemical model. A lower (computational) cost variation is available for use in conjunction with **Molecular Mechanics Models**.

SM8. An empirical solvation model applicable to Hartree-Fock and density functional models . Available for water and a variety of organic solvents. Furnishes a wavefunction in addition to an energy, meaning that solvated structures, infrared spectra and properties may be obtained.

SM12. An improvement over SM8 using a more general charge model which is less sensitive to the particular basis set. The SM12 training set is both larger and more diverse and SM12 covers the majority of the periodic table.

Space-Filling Model; *See* **CPK Model**.

Spartan Infrared Database (SIRD). An alternative entry point into the **Spartan Spectra and Properties Database (SSPD)** allowing searching by match to an (unknown) infrared spectrum.

Spartan Molecular Database (SMD). An early collection of molecular structures and properties for >125,000 molecules obtained from up to nine different quantum chemical models. Searchable by structure, name, formula, molecular weight and isomer. The **SMD** was the precursor to the **SSPD**.

Spartan Reaction Database (SRD). A collection of ~8,500 transition-state structures for a wide variety of simple organic and organometallic reactions obtained from up to five different quantum chemical models. Searchable by structure.

Spartan Spectra and Properties Database (SSPD). A collection of molecular structures, infrared and NMR spectra, molecular and atomic properties and QSAR descriptors for >300,000 organic molecules obtained from the ωB97X-D/6-31G* and EDF2/6-31G* density functional models based on the lowest-energy conformer assigned from the **T1** model. The wavefunction is available, as such, surfaces and property maps may be generated "on-the-fly". Searchable by structure, name, formula, molecular weight and isomer.

Spin Density. The difference in the number of electrons per unit volume of "up" spin and "down" spin at a point in space.

Spin Density Map. A graph that shows the value of the **Spin Density** on an **Isodensity Surface** corresponding to a **van der Waals Surface**.

Spin Orbital. The form of **Wavefunction** resulting from application of the **Hartree-Fock Approximation** to the **Electronic Schrödinger Equation**. Comprises a space part (**Molecular Orbital**) and one of two possible spin parts ("spin-up" and "spin-down").

Split-Valence Basis Set. A **Basis Set** in which the **Core** is represented by a single set of

Basis Functions (a **Minimal Basis Set**) and the **Valence** is represented by two or more sets of **Basis Functions**. This allows for description of aspherical atomic environments in molecules. **3-21G**, **6-31G** and **6-311G** are split-valence basis sets.

SSPD; *see* Spartan Spectra and Properties Database.

Stationary Point. A point on a **Potential Energy Surface** for which all energy first derivatives with respect to the coordinates are zero. **Local Minima** and **Transition States** are stationary points.

STO-3G. A **Minimal Basis Set**. Each atomic orbital is written in terms of a sum of three **Gaussian** functions taken as best fits to **Slater**-type (exponential) functions.

SVWN Model. (**S**later, **V**osko, **W**ilk, **N**usair) A **Density Functional Model** which involves the **Local Density Approximation**.

SYBYL. A **Molecular Mechanics Force Field** developed by Tripos, Inc.

Systematic Conformation Search. An exhaustive search of **Conformers** resulting from stepping through all torsional angles (including constrained torsions associated with flexible rings). This is the only procedure that guarantees identification of the **Global Minimum**.

T1. A combination of quantum chemical models and empirical parameters formulated to reproduce **Heats of Formation** obtained from the **G3(MP2) Model** using 2-3 orders of magnitude less computation. T1 reproduces experimental heats with an RMS error of ~8 kJ/mol (the same error as obtained from the **G3(MP2) Model**).

Tautomer. An isomer related by rearrangement of hydrogen atoms, for example vinyl alcohol is a tautomer of acetaldehyde. All tautomers are isomers but not all isomers are tautomers.

TDDFT (Time Dependent Density Functional Theory).

Theoretical Model. A "recipe" leading from the **Schrödinger Equation** to a general computational scheme. A theoretical models needs to be unique and well defined and, to the maximum extent possible, be unbiased by preconceived ideas. It should lead to **Potential Energy Surfaces** which are continuous. It is also desirable (but not required) that a theoretical model be **Size Consistent** and **Variational**. Finally, a theoretical model will be applicable for the problems of interest.

Theoretical Model Chemistry. The set of results following from application of a particular **Theoretical Model**.

Thermochemical Recipes. Combinations of quantum chemical models and (in the case of **T1**) empirical parameters to reproduce experimental (or in the case of **T1**) **G3(MP2) Heats of Formation. G3(MP2)** and **T1** are thermochemical recipes.

Thermodynamically-Controlled Reaction. A chemical reaction which has gone all the way to completion, and the ratio of different possible products is related to their thermochemical stabilities according to the **Boltzmann Equation**.

Thermodynamic Product. The product of a reaction which is under Thermodynamic Control.

Total Charge. The sum of the nuclear charges minus the total number of electrons.

Total Energy. The energy of a balanced chemical reaction in which a molecule is split into single particle fragments, that is, nuclei and electrons. Total energy is the standard way to report energy data derived from quantum chemical models (except **Semi-Empirical Models** and **Thermochemical Recipes**).

Transition State. A **Stationary Point** on a **Potential Energy Surface** in which all but one of the elements in the diagonal representation of the **Hessian** are positive, and one element is negative. Corresponds to the highest-energy point on the **Reaction Coordinate**.

Transition-State Geometry. The geometry (bond lengths and angles) of a **Transition State**.

Transition State Theory. The notion that all molecules react through a single well-defined **Transition State**.

Triplet. A molecule with two unpaired electrons.

Unrestricted SCF. An **SCF** procedure which does not restrict electrons to be paired on orbitals.

UV/visible Spectrum. The spectrum resulting between ground and excited electronic states.

Valence. Electrons which are delocalized throughout the molecule and participate in chemical bonding (2s, $2p_x$, $2p_y$, $2p_z$ for first-row elements, 3s, $3p_x$, $3p_y$, $3p_z$ for second-row elements, etc.).

Valence Orbital. A **Molecular Orbital** that is primarily involved with the **Valence** rather than the **Core**.

van der Waals Interactions. Interactions which account for short- range repulsion of non-bonded atoms as well as for weak long-range attraction.

van der Waals Radius. The radius of an atom (in a molecule), which is intended to reflect its overall size. van der Waals radii are used to define space-filling models.

van der Waals Surface. A surface formed by a set of interpreting spheres (atoms) with specific **van der Waals radii**, and which is intended to represent overall molecular size and shape.

Variational. Methods for which the calculated energy represents an upper bound to the exact (experimental) energy. **Hartree-Fock** and **Configuration Interaction Models** are variational while **Møller- Plesset Models**, **Density Functional Models** and **Semi-Empirical Models** are not variational.

VCD. Vibrational Circular Dichroism – the stereoscopic of left v. right polarized IR light, used as confirmation of absolute stereochemical configuration.

Vibrational Frequencies. The energies at which molecules vibrate. Vibrational frequencies correspond to the peaks in an infrared and Raman spectrum.

Wave function. The solution of the **Schrödinger Equation**. In the case of the hydrogen atom, a function of the coordinates which describes the motion of the electron as fully as possible. In the case of a many-electron system a function which describes the motion of the individual electrons.

Zero-Point Energy. The energy of molecular vibration at 0K, given as half the sum of the **Vibrational Frequencies** times Planck's constant.

Z Matrix. A set of internal coordinates (bond lengths, bond angles and dihedral angles only) describing molecular geometry. Of historical interest only. Not supported in ***Spartan***.

Zwitterion. A neutral valence structure which incorporates both a formal positive charge and a formal negative charge.