

# ***Spartan '14***

***For Windows, Macintosh and Linux***

## Tutorial and User's Guide

January 10, 2014



### **Wavefunction, Inc.**

18401 Von Karman Avenue, Suite 370  
Irvine, CA 92612 U.S.A.  
[www.wavefun.com](http://www.wavefun.com)

### **Wavefunction, Inc., Japan Branch Office**

3-5-2, Kouji-machi, Suite 608  
Chiyoda-ku, Tokyo, Japan 102-0083  
[japan@wavefun.com](mailto:japan@wavefun.com) • [www.wavefun.com/japan](http://www.wavefun.com/japan)

Copyright © 2014 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.

ISBN978-1-890661-42-2

Printed in the United States of America

# Acknowledgements

---

The *Spartan'14* Tutorial and User's Guide was prepared by Warren Hehre and Sean Ohlinger, who take both credit for its utility and blame for its limitations and inaccuracies. Several members of Wavefunction contributed significantly, with Philip Klunzinger, Bernard Deppmeier, Andy Driessen and Jeffrey Johnson warranting special mention. As with past manuals, sincere thanks go to Pamela Ohsan for turning a sloppy manuscript into a finished manual.

# Scope of this Guide

---

This guide provides a general reference for *Spartan*'14 for Windows, Macintosh and Linux. Following an introduction which presents *Spartan* as a tool for exploring organic, bioorganic, inorganic and organometallic chemistry by way of molecular mechanics and quantum chemical calculations, together with an array of graphical models for conveying the results of these calculations, it comprises 25 chapters grouped into four sections and several appendices.

**Section I (Operations and Overview; Chapters 1-2)** describes the operation of *Spartan*'s graphical user interface and presents a “walk through” illustrating some of its most basic components.

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

**Section III (Advanced Tutorials; Chapters 8-14)** 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, for finding reaction transition states and for assessing the “similarity” of molecules. Many of 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 IV (Features and Functions; Chapters 15-25)** 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'14*.

What this guide *does not do* is document the performance and cost (in computation time) of the different molecular mechanics and quantum chemical models available in *Spartan'14*, or 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. These topics, among several others, are touched on in **Appendix A** (also available under *Topics* from the **Activities** menu). They are covered in greater depth in *A Guide to Molecular Mechanics and Quantum Chemical Calculations* included as a PDF under the **Help** menu and available from Wavefunction as a hardbound volume. 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 citation for the program (**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**), directions for installing a network HASP (**K**) and a listing of NMR chemical shift standards (**L**). Additional materials relating to several of these appendices may be found as PDFs under the **Help** menu.

An up-to-date version of this *Tutorial and User's Guide* is available on Wavefunction's website (directly accessible from the **Help** menu).

# Table of Contents

## INTRODUCTION

---

### Section I OPERATIONS AND OVERVIEW

---

<b>Chapter 1</b>	<b>Operating Spartan .....</b>	<b>2</b>
	Starting and Quitting <i>Spartan</i> .....	2
	Menus and Icons .....	2
	File.....	3
	Edit .....	4
	Model .....	4
	Geometry .....	5
	Build.....	5
	Setup.....	5
	Display .....	6
	Search .....	6
	Options .....	7
	Activities .....	7
	Help .....	7
	Additional Icons.....	7
	Tabs.....	8
	Mouse/Keyboard Operations .....	8
	Selection.....	9
	Manipulation .....	9
	Touch-Screen Operations.....	11
	Selecting Molecules.....	11
	Stereo Displays .....	12
	Changing Colors and Setting Preferences .....	12
	Monitoring and Terminating Jobs.....	12
<b>Chapter 2</b>	<b>Walking Through <i>Spartan</i> .....</b>	<b>13</b>

### Section II ORGANIC MOLECULES

---

<b>Chapter 3</b>	<b>Building Organic Molecules in 3D .....</b>	<b>35</b>
	Organic Model Kit.....	35

	1. Acrylonitrile .....	38
	2. Cyclohexanone.....	40
	3. Limonene .....	41
	4. Nicotine.....	41
	5. Coumarin .....	42
<b>Chapter 4</b>	<b>Sketching Organic Molecules in 2D .....</b>	<b>44</b>
	Sketch Palette.....	44
	Sketching a Molecule.....	45
	Manipulating a Sketch .....	45
	1. N,N-Dimethylaniline .....	46
	2. <i>trans</i> -Stilbene.....	46
	3. Indigo .....	47
	4. 3-Cyano-4-cyclohexenyl Radical .....	48
	5. Androsterone.....	49
<b>Chapter 5</b>	<b>Spectra and Properties of Organic Molecules from the Spartan Spectra and Properties Database.....</b>	<b>52</b>
	1. Limonene .....	52
	2. Indigo .....	54
	3. <i>trans</i> -Stilbene.....	55
<b>Chapter 6</b>	<b>Graphical Models for Organic Molecules from the Spartan Spectra and Properties Database. ....</b>	<b>57</b>
	1. Nicotine.....	57
	2. N,N-Dimethylaniline .....	59
	3. Androsterone.....	60
<b>Chapter 7</b>	<b>Spectra, Properties and Graphical Models for Organic Molecules from Quantum Chemical Calculations .....</b>	<b>62</b>
	1. Acrylonitrile.....	62
	2. Cyclohexanone.....	67
	3. 3-Cyano-4-methylcyclohexenyl Radical .....	70
	4. Coumarin .....	73

## Section III    **ADVANCED TUTORIALS**

<b>Chapter 8</b>	<b>Groups of Organic Molecules .....</b>	<b>76</b>
	Dienophiles in Diels-Alder Cycloadditions.....	77
	Addition vs. Substitution .....	82
	Hydration of Carbonyl Compounds.....	84
	Acidities of Carboxylic Acids.....	87
	Positional Selectivities of Substituted Naphthalenes.....	89
	Tautomers of Nucleotide Bases .....	92
<b>Chapter 9</b>	<b>Spectra of Organic Molecules.....</b>	<b>95</b>
	Infrared Spectrum of Methyl Formate .....	96
	Searching Spartan's Infrared Spectral Database..	98
	Proton NMR Spectrum of 1-Methylindole .....	101
	<sup>13</sup> C NMR Spectrum of Caulophylline .....	105
	<sup>13</sup> C NMR Spectrum of <i>cis</i> -1,2-Dimethylcyclohexane .....	107
	Stereochemical Assignments from <sup>13</sup> C Spectra....	108
<b>Chapter 10</b>	<b>Organic Reactions.....</b>	<b>111</b>
	Ene Reaction of 1-Pentene.....	112
	S <sub>N</sub> 2 Reaction of Bromide and Methyl Chloride...	115
	Carbene Additions to Alkenes.....	120
	Stereospecific Diels-Alder Reactions .....	123
	Thermodynamic vs. Kinetic Control .....	126
	Activation Energies of Diels-Alder Reactions....	128
<b>Chapter 11</b>	<b>Medicinal Chemistry .....</b>	<b>133</b>
	Anticipating Blood-Brain Transport .....	134
	Terfenadine. A Potassium Channel Blocker?.....	137
	Morphine. Structure vs. Pharmacophore .....	139
	Gleevec. Making a Pharmacophore from PDB ...	145
<b>Chapter 12</b>	<b>Flexible Molecules.....</b>	<b>149</b>
	Internal Rotation in Dimethylperoxide .....	150
	Ethinamate .....	153
	<sup>13</sup> C Chemical Shifts Depend on Conformation....	155
	Allyl Vinyl Ether.....	157

	Gleevec. Protein-Bound vs. Free Conformer.....	159
<b>Chapter 13</b>	<b>Inorganic and Organometallic Molecules.....</b>	<b>163</b>
	Sulfur Tetrafluoride.....	164
	Benzene Chromium Tricarbonyl.....	166
	Ziegler-Natta Polymerization of Ethylene.....	168
<b>Chapter 14</b>	<b>“Dry Labs”: Using the Spartan Spectra and Properties Database.....</b>	<b>171</b>
	Reactivity of Silicon-Carbon Double Bonds .....	172
	Isomeric C <sub>5</sub> H <sub>8</sub> Dienes .....	174
	Using Infrared Spectroscopy to Identify an Unknown Ethyl Benzoate Derivative .....	176
	Infrared Spectra of Short-Lived Molecules .....	179
	Using <sup>13</sup> C NMR to Distinguish Structural Isomers.....	180
	Using <sup>13</sup> C NMR Spectra to Distinguish Stereoisomers.....	182
	Finding Stable Enols.....	183
	Electrophilic Reactivity of Polycyclic Aromatics.....	185
<b>Section IV</b>	<b>FEATURES AND FUNCTIONS</b>	
<b>Chapter 15</b>	<b>The File Menu .....</b>	<b>188</b>
	New Build.....	188
	New Sketch.....	188
	Open.....	188
	Close .....	189
	Save.....	190
	Save As .....	190
	Build New Molecule.....	190
	Sketch New Molecule.....	190
	Delete Molecule.....	191
	Append Molecule(s) .....	191
	Access PDB Online .....	191
	Print.....	192
	Embedded Data.....	192
	Start/Stop QuickTime Recording.....	193

	Exit.....	194
<b>Chapter 16</b>	<b>The Edit Menu .....</b>	<b>195</b>
	Undo .....	195
	Cut.....	195
	Copy.....	195
	Paste.....	195
	Select All.....	196
	Find .....	196
	Find Next .....	196
	Center.....	196
	Clear.....	196
<b>Chapter 17</b>	<b>The Model Menu.....</b>	<b>197</b>
	Wire.....	198
	Ball and Wire .....	198
	Tube .....	200
	Ball and Spoke.....	200
	Space Filling.....	201
	Line .....	201
	Hide.....	202
	Global Model .....	202
	Coupled.....	203
	Hydrogens.....	203
	Labels.....	203
	Ribbons .....	203
	Ramachandran Plot.....	204
	Hydrogen Bonds .....	204
	CFD's .....	204
	Configure .....	207
	Configure Labels .....	207
	Configure Objects.....	208
	Configure Ribbons .....	208
	Configure CFD's .....	209
<b>Chapter 18</b>	<b>The Geometry Menu.....</b>	<b>211</b>
	Measure Distance.....	211
	Measure Angle .....	211

Measure Dihedral.....	211
Freeze Center .....	212
Set Torsions.....	214
Set Similarity Centers .....	216
Similarity Based on Structure .....	216
Similarity Based on CFD's.....	217
Similarity Based on a Pharmacophore .....	218
Constrain Distance.....	218
Constrain Angle .....	218
Constrain Dihedral.....	218
Define Point .....	221
Define Plane.....	221
Define CFD .....	222
Align .....	222
Align Based on Structure .....	223
Align Based on CFD's .....	224
Align to a Pharmacophore.....	225
<b>Chapter 19 The Build Menu .....</b>	<b>226</b>
2D Sketch Palette.....	226
Defined Atoms.....	226
Element/Functional Group Library .....	226
Common Rings.....	227
Common Carbonyl Groups .....	227
Stereochemical Markers.....	227
Open-site Markers.....	227
Charge/Radical Markers.....	227
Drawing Tools .....	228
Making a Sketch .....	228
Manipulating a Sketch .....	229
Sketch Operations.....	229
Add an Atom, Ring or Carbonyl Group .....	229
Add Multiple Bonds.....	229
Fuse 5 and 6-Member Rings .....	229
Replace an Atom with Another .....	230
Access an Element or Functional Group from the Library.....	230

Add Stereochemical Markers (Dashes, Wedges).....	230
Add an <i>Axial</i> or <i>Equatorial</i> Marker .....	231
Assign Charges and Radical Sites.....	231
Specify “Open Site” Markers.....	232
Undo the Last Action .....	232
Clean Up a Drawing.....	232
Remove an Atom or Bond.....	233
Remove a Multiple Bond .....	233
Remove a Stereochemical Marker .....	233
Change or Remove an <i>Axial</i> or <i>Equatorial</i>	
Label from a Stereochemical Marker.....	233
Remove a Charge or Radical Marker.....	233
Remove an “Open Site” Marker .....	233
Clear the Screen .....	233
3D Model Kits .....	233
Organic Model Kit .....	234
Groups .....	236
Rings.....	237
More.....	238
Clipboard .....	239
Inorganic Model Kit.....	239
Ligands .....	242
Peptide Model Kit .....	242
Specify Amino Acid Sequence .....	244
Specify Macroscopic Structure.....	244
Terminate .....	244
Nucleotide Model Kit.....	245
Specify Base Sequence.....	247
Specify Helical Structure.....	247
Substituent Model Kit .....	248
Library Preparation.....	250
Attachment.....	251
List Generation .....	252
Accessing ChemDraw.....	253
General Molecule Building Functionality.....	254
Multiple Fragments .....	254

Rotate/Stretch Bonds.....	254
Replace Atom/Fragment.....	255
Invert Chirality .....	255
Building/Editing Menu Functions .....	256
View .....	256
Edit Build .....	256
Edit Sketch .....	257
Delete .....	257
Make Bond .....	258
Break Bond.....	258
Minimize .....	259
<b>Chapter 20 The Setup Menu .....</b>	<b>260</b>
Calculations .....	261
Calculate.....	264
Task.....	264
State .....	268
Type of Calculation.....	269
Basis Sets .....	275
Dual Basis Sets.....	276
Pseudopotentials .....	277
Solvent .....	277
Intrinsic Reaction Coordinate.....	278
Start From (Energy, Equilibrium Geometry) ....	278
Subject to.....	279
Compute .....	280
IR .....	280
Raman.....	281
NMR.....	281
UV/vis.....	282
QSAR.....	283
Total Charge .....	283
Unpaired Electrons.....	283
Print .....	284
Orbitals and Energies.....	284
Thermodynamics .....	284
Vibrational Modes .....	285

Charges and Bond Orders .....	285
Options .....	285
Similarity .....	285
Global Calculations.....	289
Surfaces.....	289
Common Surfaces and Property Maps.....	295
Surface.....	295
Property .....	297
Resolution .....	297
IsoValue.....	298
Global Surfaces .....	298
Submit.....	299
<b>Chapter 21 The Display Menu.....</b>	<b>301</b>
Output .....	301
Properties .....	302
Molecule Properties, QSAR Descriptors, Thermodynamics and Acidity/Basicity .....	303
Molecule .....	304
QSAR.....	305
Thermodynamics .....	306
Acidity/Basicity .....	306
Molecule Utilities .....	307
Atom Properties .....	308
Bond Properties .....	309
Constraint Properties.....	309
Point and Plane Properties .....	311
Surface Properties .....	311
Curve and Plot Properties.....	313
Substituent Properties.....	317
CFD Properties.....	318
Regression Properties.....	318
“Selected” Style .....	319
Orbital Energies .....	319
Surfaces.....	320
Spectra .....	320
Simplified Presentation of Spectra.....	321

Generalized Presentation of Spectra .....	323
IR, Raman .....	324
NMR .....	328
UV/vis.....	332
Formulas .....	335
Spreadsheet .....	339
From the Add Dialog.....	342
From Post ( <b>P</b> ) Buttons .....	345
From the Clipboard .....	346
User-Defined Expressions .....	347
Numerical Data .....	347
Plots .....	350
Simplified Presentation of Plots .....	350
Generalized Presentation of Plots .....	353
XY Plots.....	354
XYZ Plots .....	355
Similarities .....	357
Reactions.....	358

<b>Chapter 22    The Search Menu .....</b>	<b>361</b>
Structure Query.....	361
Reaction Query .....	362
Databases .....	365
Cambridge Structural Database (CSD) .....	366
Spartan Molecular Database (SMD) and Spartan Spectra and Properties Database (SSPD).....	370
Accessing and Processing Data from SMD and SSPD .....	373
Molecule Replacement.....	374
Substructure and Other Modes of Searching SMD and SSPD.....	376
Plotting Data from an SMD or SSPD Search.....	380
Regression Analysis of Data Resulting from an SMD or SSPD Search .....	381
Relationships Involving Chemical	

	Reactions.....	383
	Retrieving Data from an SMD or SSPD Search.....	386
	Spartan Reaction Database (SRD) .....	386
	Substructure Searching.....	387
	Retrieving Data from an SRD Search.....	389
	Spartan Infrared Database (SIRD) .....	390
	Spectra Searching .....	390
	Filters .....	392
	Retrieving Data from an SIRD Search .....	396
	External Infrared Database (XIRD).....	396
	Guess Transition State .....	397
	Identify Tautomers.....	398
	Extract Ligands.....	399
<b>Chapter 23</b>	<b>The Options Menu .....</b>	<b>402</b>
	Preferences.....	402
	Settings .....	403
	Molecule.....	407
	VDW Radii.....	409
	Databases.....	410
	Jobs.....	411
	Miscellaneous.....	412
	Icons .....	415
	URLs .....	416
	Available Servers.....	416
	Embedded Server .....	418
	Colors.....	421
	Fonts/Graphics Fonts .....	422
	Local Monitor .....	423
	Network Monitor .....	424
	Calculator.....	424
	Icons.....	424
<b>Chapter 24</b>	<b>The Activities Menu .....</b>	<b>425</b>
	Tutorials/Topics .....	425
	Lookup in Wikipedia .....	425

<b>Chapter 25</b>	<b>The Help Menu .....</b>	<b>427</b>
	Help.....	427
	About .....	428
<b>Appendix A</b>	<b>Topics .....</b>	<b>429</b>
	Potential Energy Surfaces.....	430
	Theoretical Models .....	438
	Finding and Verifying Equilibrium and Transition-State Geometries.....	448
	Total Energies and Thermodynamic Kinetic Data .....	455
	Calculating Accurate Heats of Formation.....	461
	Dealing with Conformationally Flexible Molecules .....	464
	Interpreting Conformational Preferences.....	467
	Calculating Infrared Spectra .....	470
	Calculating NMR Spectra.....	476
	Atomic and Molecular Orbitals .....	485
	Electron Densities: Sizes and Shapes of Molecules .....	496
	Electrostatic Potential Maps: Charge Distributions .....	501
	Local Ionization Potential Maps and LUMO Maps: Electrophilic and Nucleophilic Reactivities .....	507
<b>Appendix B</b>	<b>Capabilities and Limitations.....</b>	<b>511</b>
	Molecular Mechanics Models.....	511
	Semi-Empirical Models .....	511
	Hartree-Fock Models .....	511
	Density Functional Models.....	512
	Møller-Plesset Models .....	514
	Advanced Correlated Models .....	514
	Thermochemical Recipes.....	514
	Excited-State Models.....	515
	Basis Sets and Pseudopotentials .....	515
	Solvent Models .....	516
	Properties and Spectra .....	517

	Graphical Models.....	519
	Similarity Analysis.....	519
<b>Appendix C</b>	<b>Menus .....</b>	<b>521</b>
	<i>Spartan</i> Screen .....	521
	Contextual.....	528
<b>Appendix D</b>	<b>Commonly-Used Program Options.....</b>	<b>531</b>
	Conformational Search .....	531
	Similarity Analysis.....	532
	Geometry Optimization .....	532
	Molecular Mechanics Calculations.....	533
	Quantum Chemical Calculations .....	533
	Semi-Empirical Calculations .....	535
	Property and Spectra Calculations.....	535
	Miscellaneous and Printing.....	536
<b>Appendix E</b>	<b>Units .....</b>	<b>538</b>
<b>Appendix F</b>	<b>Citation .....</b>	<b>540</b>
<b>Appendix G</b>	<b>Installing the Cambridge Structural Database.....</b>	<b>541</b>
<b>Appendix H</b>	<b>Constructing Custom Databases .....</b>	<b>543</b>
<b>Appendix I</b>	<b>Pharmacophore Input .....</b>	<b>544</b>
<b>Appendix J</b>	<b>Input of Experimental Spectra .....</b>	<b>545</b>

# Introduction

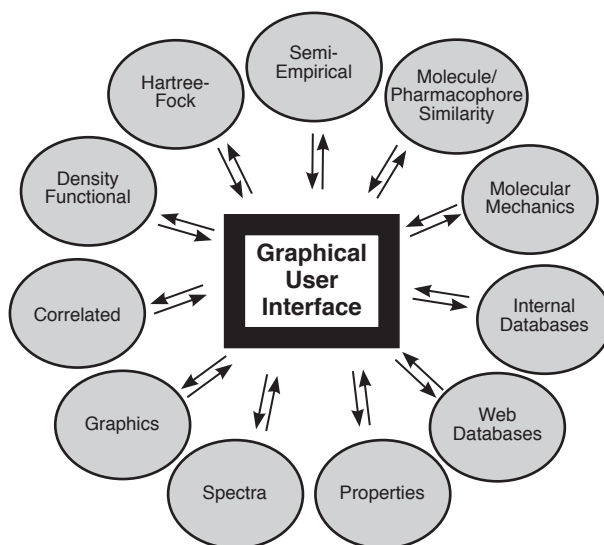
---

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 the chemical community. Quantum chemical calculations are much more time demanding, and only recently have fast enough computers become widely available to make their application routine among mainstream chemists.

Quantum chemical calculations may also be 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 still needed for systems that are too large or too complicated to be subjected to the more rigorous treatments. Finally, quantum chemical calculations can supply information to complement existing experimental data or to replace it altogether, for example, atomic charges for QSAR analyses, and intermolecular potentials for molecular mechanics and molecular dynamics calculations.

*Spartan'14* (“*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 codes supplement and extend the traditional strengths of *Spartan* as an easy to learn and use tool for molecular mechanics and quantum chemical calculations. *Spartan is intended to be utilized by chemists*, not only computational chemists who are already familiar with the capabilities of molecular mechanics and quantum chemical calculations, but also experimental chemists who may have little or no prior experience, but who want to use calculations much in the same way as experimental techniques such as NMR spectroscopy.

**Spartan** comprises a series of independent modules tightly connected via a highly functional, but simple and uncluttered graphical user interface. It 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 the interpretation of output.



**Spartan**'s interface provides the gateway to a range of modern computational methods, including molecular mechanics models, semi-empirical and Hartree-Fock molecular orbital models, and a variety of correlated models including density functional and Møller-Plesset models. 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 another (more rigorous) model.

**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 (Hartree-Fock and density functional models), NMR spectra (Hartree-Fock and density functional models) and UV/visible spectra (CIS, CIS(D) and density functional models). Experimental IR, NMR and

UV/visible spectra from public on-line databases may be automatically accessed and superimposed onto 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 proteins and nucleotides in the PDB.

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

**Spartan** accesses two different databases of information obtained from quantum chemical calculations on stable molecules. The Spartan Molecular Database (SMD) contains structures, energies and limited atomic and molecular properties for ~150,000 molecules, each from up to nine different quantum chemical models. The Spartan Spectra and Properties Database (SSPD), while limited to the EDF2/6-31G\* density functional model, contains structures, energies and a more extensive selection of atomic and molecular properties for ~250,000 molecules. More important, the entries include both IR and NMR spectra as well as the wave function. The latter allows graphical surfaces and property maps to be generated and displayed “on-the-fly”.

A small (~1,600 entry) database of calculated transition-state geometries is also provided. Finally, **Spartan** accesses the public PDB database of ~85,000 protein structures and (by separate license) the Cambridge Database of ~650,000 X-ray crystal structures.

New features in **Spartan’14** (since the previous release of **Spartan’10**) are as follows:

## Graphical User Interface

**Icons.** Icons have been revised and are now available for all menu functions. It is up to the user to choose which icons are displayed.

***Touch-Screen Operations.*** Touch-screen operations available under Windows 7 and Windows 8 are supported both in the new 2D sketch builder (see below) and in ***Spartan's*** main screen. Molecules may be manipulated (rotated, translated and scaled) without a mouse or touchpad. As a preference, “touch friendly” menus are available.

***Sketch Builder.*** This allows organic molecules to be sketched in 2D using a simple sketch palette. This is an attractive alternative to 3D building in previous versions of ***Spartan***, in particular, for Windows's touch-screen and tablet computers. It may also be used on conventional PCs and Macs with a mouse and/or touchpad. The 2D sketcher is identical in style and operation to that in ***iSpartan*** for IOS devices. 2D sketches are automatically converted to 3D structures. Cues provided in the 2D sketch (up and down “wedges”) lead to proper stereochemistry in the 3D structure. The 3D structure can be extended or modified with ***Spartan's*** (3D) model kits, but the (modified) 3D structure cannot be brought back into 2D. Seamless access to ChemDraw (which must be separately licensed) is still provided (Windows only).

***Spectra Plots.*** Simplified plots for NMR, IR, Raman and UV/visible spectra have been added. The more general spectra plot routines in previous versions of ***Spartan*** remain available as a preference. The new plot routines take advantage of touch-screen operations, and plots generated can be pasted into Word or PowerPoint (Keynote) documents.

***Plots.*** Simplified plot routines for data presentation from ***Spartan's*** spreadsheet have been added. The more general plot routines in previous versions of ***Spartan*** remain available as a preference. The new plot routines take advantage of touch-screen operations, and plots generated can be pasted into Word or PowerPoint (Keynote) documents.

## Spectroscopy

***Improved Proton NMR Chemical Shifts.*** The empirical scheme used to correct  $^{13}\text{C}$  chemical shifts from EDF2/6-31G\* density functional calculations released in ***Spartan'10***, has been extended to proton

shifts. The RMS error in proton shifts is has been reduced from 0.30 to 0.15 ppm.

***Improved  $^{19}\text{F}$  NMR Chemical Shifts.*** An empirical correction scheme has also been applied to  $^{19}\text{F}$  chemical shifts calculated from the EDF2/6-31G\* model. The RMS error of corrected shifts is 3.7 ppm.

***Improved Three-Bond HH Coupling Constants ( $^3J_{\text{HH}}$ ) and Splitting Patterns.*** The empirical procedure for estimating HH coupling constants and the resulting proton splitting patterns used in *Spartan* '10 have been completely revamped.

***Improved Infrared Spectral Matching.*** An improved scheme has been implemented using line width, scale and offset as adjustable parameters to provide an “on-the-fly” fit of a calculated infrared spectrum to an (unknown) experimental spectrum. For quantum chemical calculations performed at the EDF2/6-31G\* model, the fitted spectrum is generally a good match to the (correct) experimental spectrum. Given that infrared spectra for greater than 250,000 molecules are available in the Spartan Spectra and Properties Database, this encourages the use of theory IR databases in lieu of experimental databases to identify unknowns. IR spectral searching may be combined with substructure searching.

***NIST IR Database Search.*** The ability to match an (unknown) infrared spectrum with an entry in the NIST experimental infrared database has been provided. Each NIST entries is represented as a set of discrete frequencies and intensities (in the same way as a calculated spectrum). Matching to an unknown spectrum involves optimization of line width, scale and offset parameters, exactly the same procedure used for matching to calculated spectra.

## Databases

***Spartan Spectra and Properties Database (SSPD).*** SSPD now contains >250,000 molecules (up from ~70,000 at the time of release of *Spartan* '10). A transition from indexing by SMILES to InChi allows ions and radicals to be distinguished. SSPD now contains several thousand deprotonated and protonated forms (for neutral acids

and bases in SSPD), and gas-phase acidity and basicity can now be accessed as properties. Indexing by InChi also facilitates automatic identification and swapping of tautomers.

## Computation

***Access to Remote Compute Server.*** *Spartan* may now submit jobs to the server (see below) licensed and residing on any Windows PC or Mac. This allows calculations to be spawned off, with results returned as soon as they are available.

***Calculations on Lanthanides.*** Pseudopotentials and valence basis sets for lanthanides to be used in conjunction with all-electron 6-31G\* basis sets have been provided. Pseudopotentials have been modified from those originally proposed from Dolg, Stoll and Preuss [Theoret. Chim. Acta, **75**, 173 (1989); **85**, 441 (1993)] in order to closely reproduce equilibrium geometries for a variety of lanthanide complexes and organometallics.

***Parallel.*** Additional components to be parallelized.

## Compute Server

***Spartan'14*** may now be used as a compute server. This allows calculations to be off-loaded from any copy of ***Spartan'14*** on a Windows PC or Mac, onto any Windows PC or Mac that has been licensed for the server. All molecular mechanics and quantum chemical models are supported. It also provides IOS devices (iPad, iPhone and iPod Touch) running the ***iSpartan*** app (available from the Apple iTunes store) access the 250,000 molecules included in the Spartan Spectra and Properties Database (SSPD) and to quantum chemical calculations, for the same quantities (structure, energy, NMR and IR spectra, and heat of formation) for any molecule using the same level of theory (EDF2/6-31G\*) as in SSPD. ***iSpartan*** together with the server capability of ***Spartan'14*** puts the power of molecular modeling on mobile devices. Server capability is provided as part of the ***Spartan'14 Parallel Suite***.

Since the release of ***Spartan'10***, *Spartan* users have diligently reported program bugs via [support@wavefun.com](mailto:support@wavefun.com). All reported bugs

have been addressed. Wavefunction would like to thank our customers for their help in making *Spartan* a better program. User-feedback is greatly appreciated!

# Section I

## Operations and Overview

---

The two chapters that make up this section are intended to 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 most rudimentary operations of the interface (**Walking Through Spartan**). Together, these offer sufficient information “to get started”. Description of the full range of *Spartan*'s capabilities is deferred until **Section IV** which is intended as a reference.

# Chapter 1

## Operating *Spartan*

---

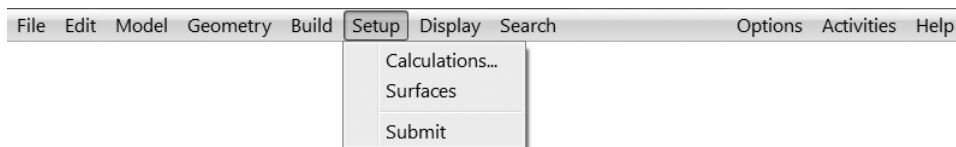
*This chapter describes the general operating features of **Spartan'14**. It should be perused prior to starting the tutorials.*

### Starting and Quitting *Spartan*

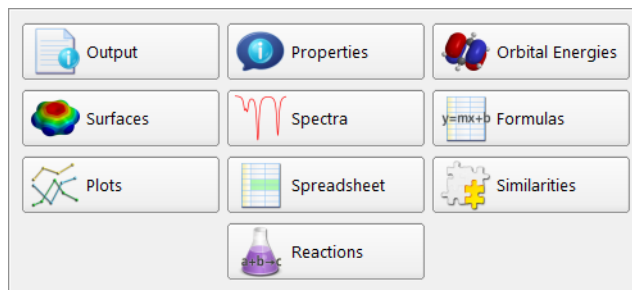
To start *Spartan* under Windows, *click* on the **Start** button, then *click* on **All Programs**, and finally *click* on **Spartan'14** (or *double click* on the *Spartan* icon on your desktop). To start under Macintosh, *double click* on the **Spartan'14** icon in the Applications Folder. To open under Linux, bring up a terminal and *type* **spartan**. To quit *Spartan*, select **Exit** from the **File** menu on Windows and **Quit Spartan 14** from the **Spartan 14** menu on Mac, or *click* the **Close** button (☒) at the top right (top left for Mac) of the *Spartan* interface.

### 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.



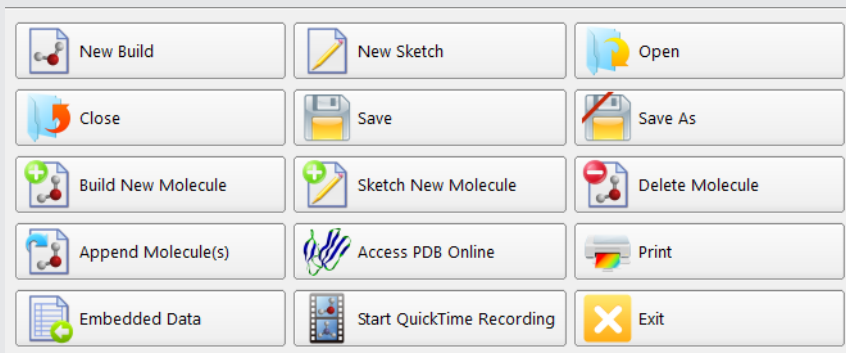
or from a list of icons presented in a palette (**Button Pad**), for example, the **Display** menu



Selection is made in the **Settings Preferences** dialog (**Preferences...** under the **Options** menu; **Chapter 23**).

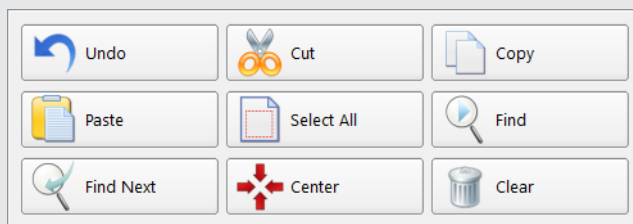
Icons for all menu functions (as shown alongside text in both **Classic List** and **Button Pad** styles) are available on screen above the menus. There are too many icons to be simultaneously displayed and the choice (beyond the default initially supplied with *Spartan*) is made in the **Icons Preferences** dialog (**Preferences...** under the **Options** menu; **Chapter 23**). Icon size is also user selectable in the **Setting Preferences** dialog (**Preferences...** under the **Options** menu; **Chapter 23**). The default settings are for 20 medium size icons.

## File



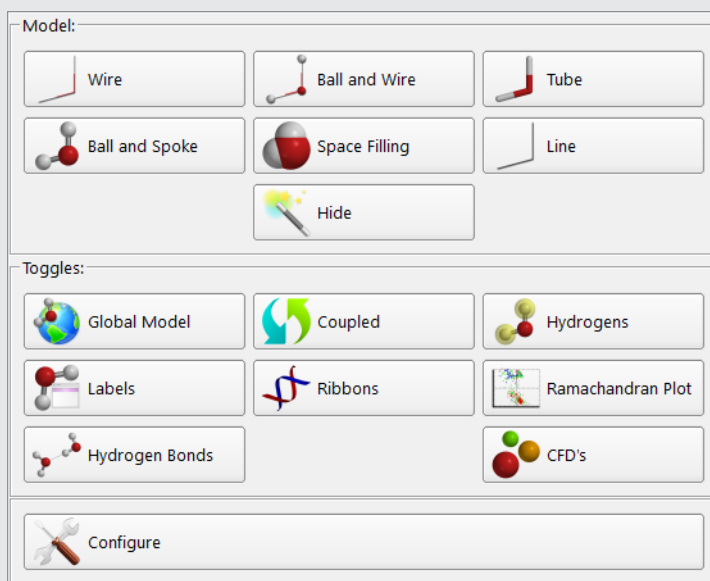
Allows you to build or sketch a new molecule, add or delete a molecule from an existing *Spartan* document, read in a molecule that you have previously saved, or a file created by another program (for example, an SD file created by ChemDraw). Allows you to print text and graphics, embed external files (for example, Word and Excel files) into *Spartan* documents, to access PDB files online and to make QuickTime movies.

## Edit



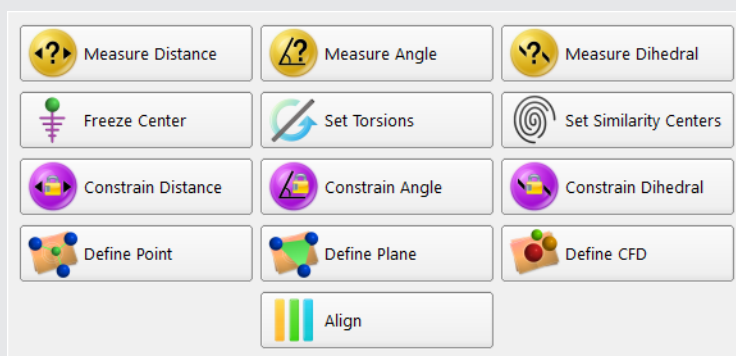
Allows you to transfer information to and from the clipboard, to undo the last operation, to find text strings and molecule fragments, to center molecules on screen, and to clear the active molecule by deleting it.

## Model



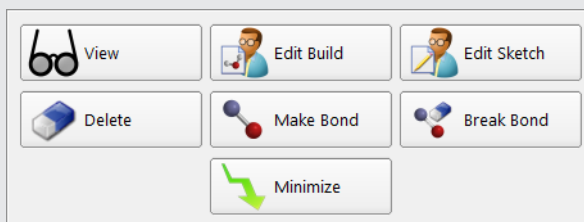
Allows you to control the style of your model, to display hydrogen bonds and chemical function descriptors, and to couple or decouple molecules in a multi-molecule file. Allows you to display of a Ramachandran plot for a protein structure brought in from PDB.

## Geometry



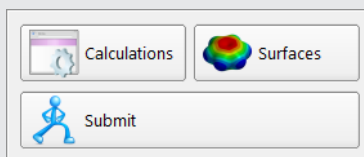
Allows you to measure and constrain bond lengths, angles and dihedrals, define points and planes, setup frozen atoms, alter default settings for conformational degrees of freedom, select centers for similarity analysis and align molecules.

## Build



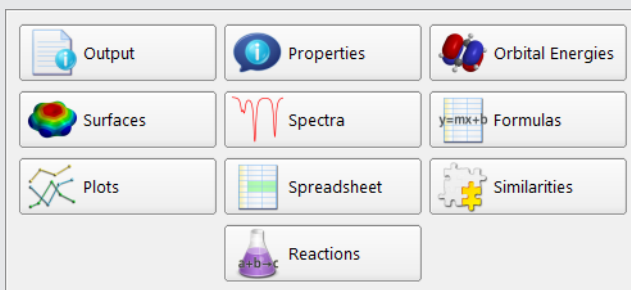
Allows you to build or sketch and edit molecules and to create lists of substituted molecules. Provides seamless access to ChemDraw™ (Windows only).

## Setup



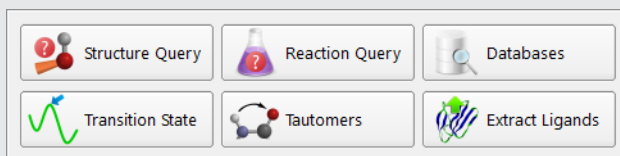
Allows you to specify the task to be performed and the theoretical model to be employed for this task, to specify graphical surfaces and property maps and to submit jobs for calculation.

## Display



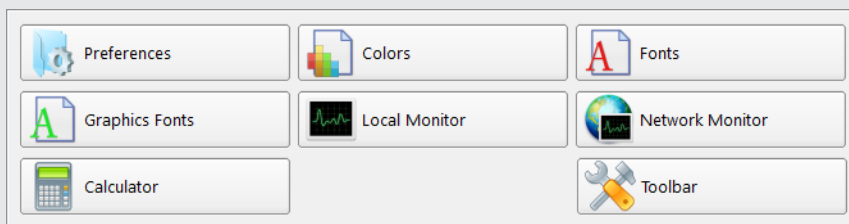
Allows you to display text output, molecular and atomic properties, QSAR descriptors, thermodynamic quantities and gas-phase acidities and basicities, orbital energy diagrams, surfaces and property maps and infrared, Raman, NMR and UV/visible spectra, as well as to access experimental IR, NMR and UV/visible spectra over the internet or from local files. 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 Spartan Molecular Database (SMD) or Spartan Spectra and Properties Database (SSPD).

## Search



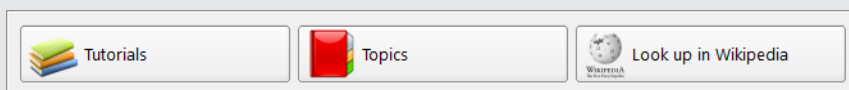
Allows you to specify queries for the Spartan Molecular Database (SMD), the Spartan Spectra and Properties Database (SSPD) and the Spartan Reaction Database (SRD). Allows you to search and mine SMD and SSPD for calculated structures, properties and spectra, and to search SRD for calculated transition-state structures. Allows you to search the Cambridge Structural Database (CSD) of experimental X-ray crystal structures, and to extract ligands from the Protein Data Bank (PDB). Allows you to match an unknown infrared spectrum to calculated spectra in the Spartan Infrared Database (SIRD) or to experimental spectra from the NIST database (XIRD). Allows you to guess a transition-state geometry based on a library of reactions and to identify tautomers.

## Options



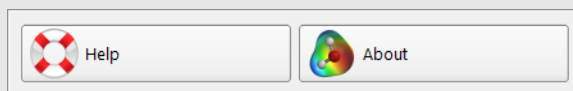
Allows you to set display standards, specify databases, assign compute servers, monitor executing jobs and specify queues and customize icons and other aspects of the graphical user interface. Allows you to setup accounts on the embedded server (*Spartan'14 Parallel Suite* only).

## Activities



Allows you to display tutorials and topics inside of *Spartan* and to search Wikipedia.

## Help

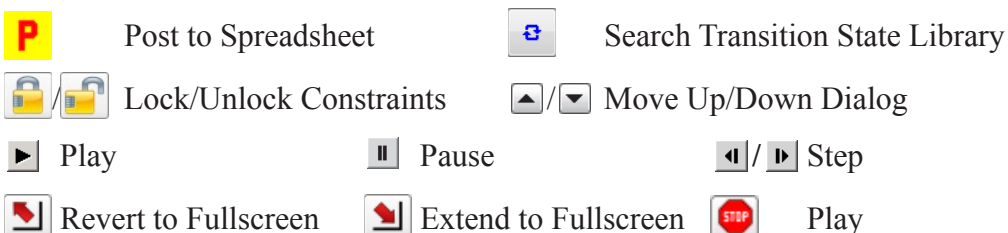


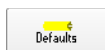
Provides access to information on *Spartan's* general operation, the *Spartan'14 Tutorial and User's Guide*, *A Guide to Molecular Mechanics and Quantum Chemical Calculations*, and a number of computational FAQ's.

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 bars at the bottom of the screen.



[Restore Default Settings](#)[Look Up in Wikipedia](#)

## Tabs

*Spartan'14* 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 read. Backward and forward step keys ( and ) at the far right provide access to tabs outside of those displayed. Documents are selected for display by *clicking* on its tab. To the left of each tab is a check box, which if checked indicates that the document will be displayed on screen even if it is not the selected document. By default, tabs are not checked. Selecting **Pin New Documents** from the **Settings Preferences** dialog (**Preferences...** under the **Options** menu; **Chapter 23**) changes the behavior and checks all tabs.

## Mouse/Keyboard Operations

The following functions are associated with the (two-button) mouse and keyboard.

keyboard	button	
	left	right
–	selection, X/Y rotate, atom/fragment exchange <sup>a</sup> , fragment insertion <sup>a</sup>	X/Y translate
<b>Shift</b>	range selection, Z rotate	scaling <sup>b</sup>
<b>Ctrl</b>	global X/Y rotate <sup>c</sup>	global X/Y translate
<b>Ctrl + Shift</b>	multiple selection, global Z rotate <sup>c</sup>	scaling <sup>b</sup>
<b>Ctrl</b> (add fragment mode)	fragment X/Y rotate chirality invert <sup>a,d</sup>	fragment X/Y translate
<b>Ctrl + Shift</b> (add fragment mode)	fragment Z rotate absolute configuration invert <sup>a,d</sup>	scaling <sup>b</sup>
<b>Ctrl + Shift</b> <sup>d</sup> (Mac & Linux)	bond rotation	bond stretching

<b>Alt</b> (Windows), <b>option</b> (Mac)	group selection <sup>e</sup> , bond rotation <sup>f</sup>	bond stretching
--	---	-----------------

- a) Requires *double clicking*.
- b) Scaling is always applied to all open molecules and all fragments. The scroll wheel will also scale molecules.
- c) Global rotations can be either molecule or screen centered. This is controlled by **Global Rotate** in the **Miscellaneous Preferences** dialog (**Preferences...** under **Options** menu; **Chapter 23**).
- d) For Macintosh, **command** key replaces **Ctrl** key for chirality inversion.
- e) In **View** mode only.
- f) In **Edit Build** mode with bond selected (red arrow visible).

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

**Selection.** *Clicking* (left button) selects objects on screen and/or of menu items. Left and right buttons together are used to define a selection box for copying to the clipboard, as well as for multiple model 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 selection apply not only to text items in lists, but to atoms and bonds in molecules as well. Together with the **Alt** key (**option** key for Mac), the left button allows for selection of an entire group (detached molecular fragment).

In **Edit Build** mode (only), *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** key leads to inversion in chirality of the atom and *double clicking* on an atom while holding down both the **Ctrl** and **Shift** keys inverts the absolute configuration of the molecule. These operations are not available in the 2D sketcher (**Edit Sketch** mode). Once an initial fragment, group or ring has been drawn, *double clicking* on the background will insert it alongside (but not bonded to) whatever fragments currently exist on screen.

**Manipulation.** The left button is used for rotation and the right button is used for translation and scaling of objects on 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.

The **Ctrl** key in conjunction with the left or right mouse buttons and (optionally) the **Shift** key, signifies a change in focus away from the default for the purpose of rotations and translations. Outside of **Edit Build/Edit Sketch** mode, the default is focus on a single molecule (the selected molecule). Use of the **Ctrl** key changes focus to the entire set of molecules on screen, meaning that rotations and translations are carried out globally. 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. Does not apply to **Edit Sketch** mode.

In **Edit Build** mode (only), moving the mouse while holding down the **Alt** key (**option** key for Mac, **Ctrl** and **Shift** keys for Linux) together with the left mouse button rotates about the selected bond and, with the right mouse button, changes the length of the selected bond. Bond rotation (only) 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. Does not apply to **Edit Sketch** mode.

Additional keys control various *Spartan* functions.

<b>3</b>	<b>3</b> selects red-cyan stereo display. Pressing again returns to non-stereo display.
<b>Page Up, Page Down Home, End</b>	Moves up ( <b>Page Up</b> ), down ( <b>Page Down</b> ), to the top ( <b>Home</b> ) and to the bottom ( <b>End</b> ) of the set of open molecules. Also, moves up and down pages in the <b>Output</b> dialog.




<b>Insert</b> ( <b>option</b> for Mac)	In <b>Edit Build/Edit Sketch</b> fragment mode only, inserts a new fragment on screen. This is accomplished by selecting the fragment from the model kit, holding down the <b>Insert</b> key and <i>clicking</i> on screen. Insertion may also be accomplished by <i>double clicking</i> on the background following selection of a fragment.
<b>Delete</b>	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 <b>Delete</b> key and <i>clicking</i> on the fragment, etc.
<b>Enter</b> ( <b>return</b> for Mac)	Required following text or data entry into spreadsheet or dialogs.



## Touch-Screen Operations

*Tapping* is equivalent to *clicking* and *double tapping* is equivalent to *double clicking*. Touch commands for range and multiple selection has not as of yet been implemented. One finger motions on screen are equivalent to left button motions (object rotation). Two finger motions are equivalent to right button motions (object translation). Two finger pinching is equivalent to scroll wheel operations (zooming).

## Selecting Molecules

While two or more molecules may be simultaneously displayed in *Spartan*'s window (see **Tabs** earlier in this chapter), 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. The previously selected molecule is deselected.

Where the molecule belongs to a document with more than a single molecule, selection from among the different molecules may be made using either the  and  buttons or the scroll bar at the bottom left of the screen. *Clicking* on  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 Preferences** dialog (**Preferences...** under the **Options** menu; **Chapter 23**). *Clicking* on  (that replaces ) stops the animation. If the spreadsheet associated with the document is open (**Spreadsheet** under the **Display** menu; **Chapter 21**), 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, *press* the **3** key. *Press* again to return to non-stereo mode. Does not apply to the 2D sketcher.

### Changing Colors and Setting Preferences

**Colors** and **Preferences...** under the **Options** menu (**Chapter 23**) allows for changing default background and graphical object colors, and for setting (and resetting) program defaults, respectively. Further control of color is available in the various **Style** dialogs accessible from **Properties** under the **Display** menu; **Chapter 21**.

### Monitoring and Terminating Jobs


Monitors list jobs that are either queued or executing on the local machine (**Local Monitor**) or on a network server (**Network Monitor**). Queued jobs may be started and executing jobs may be terminated. The current text output and molecular structure of locally executing jobs may be examined. **Local** and **Network Monitor** are accessible from the **Options** menu (**Chapter 23**).




# 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. It is a must for new **Spartan** users, and should at least be skimmed by experienced users to expose them to new features of **Spartan**'14. Specifically it shows how to: i) open molecules, ii) view different models and manipulate molecules on screen, iii) measure bond distances, angles and dihedral angles, iv) display energies, dipole moments, atomic charges and infrared, NMR and UV/visible spectra, and v) display graphical surfaces and property maps. Spreadsheet operations are not illustrated, no molecules are sketched or built and no quantum chemical calculations are performed.*

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 if it appears at the top of the screen. A file browser appears.

 Tap on  at the top of the screen. If the icon is not available, tap on **File** in the menu bar to bring up a palette of icons and then tap on .

Move to the **Tutorials** directory\*, click on **basic operations** and click on **Open** (or double click on **basic operations**).

 Tap on **basic operations** and tap on **Open** (or double tap on **basic operations**). Note that click (left mouse button) and tap (one finger) are equivalent as are double click and double tap.

---

\* For Windows, the **Tutorials** directory is found in **Program Files/Wavefunction/Spartan14** or **Program Files(x86)/Wavefunction/Spartan14**. 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 of the **Spartan14** disc image.

We shall only indicate *click/double click* throughout the text that follows.

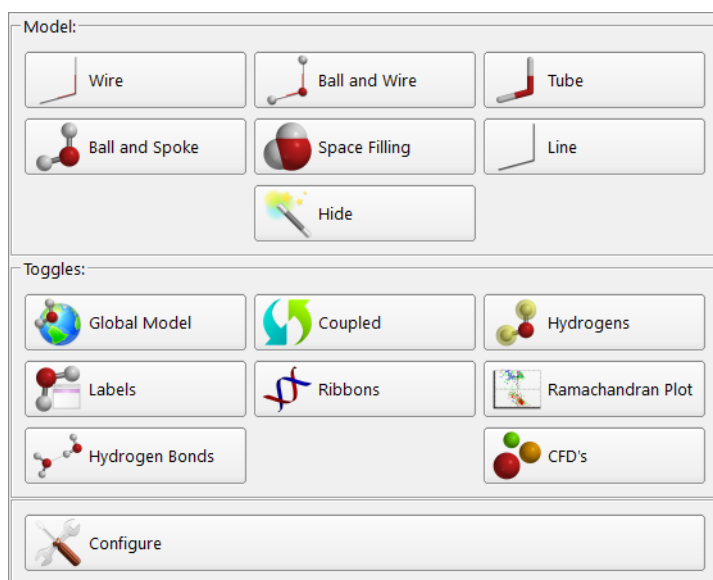
A single file containing *ethane*, *acetic acid dimer*, *propene*, *ammonia*, *hydrogen peroxide*, *acetic acid*, *water*, *cyclohexanone*, *camphor*, *3-aminobenzophenone*, *ethylene*, *benzene*, *aniline* and *cyclohexenone* will be opened. A ball-and-spoke model for the first molecule (*ethane*) will be displayed, and its name appears at the bottom right of the screen. The appearance of the name means that the molecule is included in the Spartan Spectra and Properties Database (SSPD).

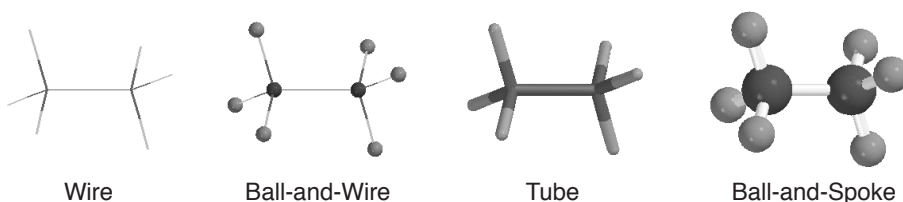
- Practice rotating (*move* the mouse while holding down the left button) and translating (*move* the mouse while holding down the right button). Use the scroll wheel to zoom in and out, or alternately move the mouse while holding down both the right button and the **Alt** key.



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

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

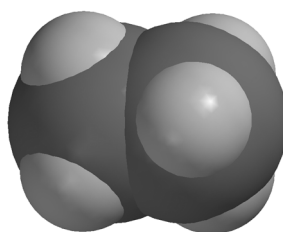




One after another, select **Wire**, **Ball and Wire**, **Tube** and finally **Ball and Spoke** from the palette of icons that results from selection of the **Model** menu. All four models for *ethane* show roughly the same information. The wire model looks the most like a conventional line formula. It 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 palette of icons that results from the selection of **Model** menu.

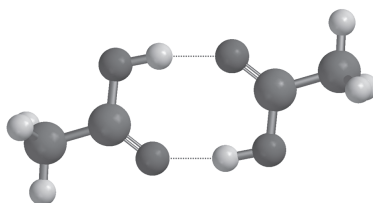


Space-Filling

The space-filling model is different from the other models in that bonds are not shown. Rather, each atom is displayed as a colored sphere that represents its approximate size. Thus, the space-filling model for a molecule provides a measure of its 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. If you make a mistake, use the backward  or forward  step keys to get to **acetic acid dimer** in the document. Switch to a space-filling model and 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

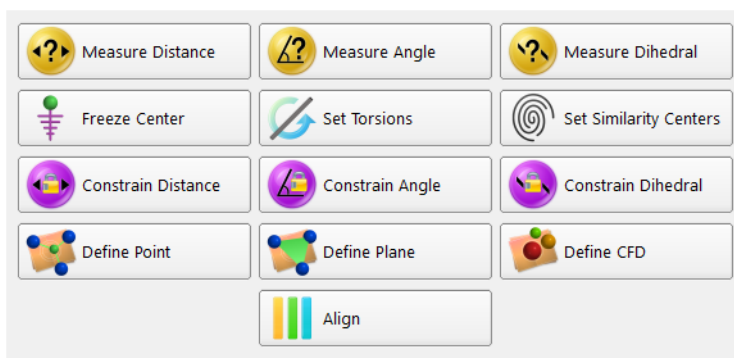
The two hydrogen bonds, that are responsible for holding the acetic acid molecules together, will be drawn.




Use the **3** key to toggle between stereo 3D and standard display. To view in 3D you will need to wear red/blue glasses.

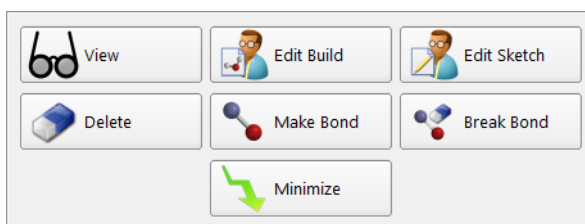





If you are on a tablet without a physical keyboard, you need to bring up a virtual keyboard. Select **Preferences...** from the **Options** menu . Click on the **Settings** tab to bring up the **Settings Preferences** dialog and check the box to the left of **On-Screen Keyboard**. This box should remain unchecked if you have a physical keyboard.




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.



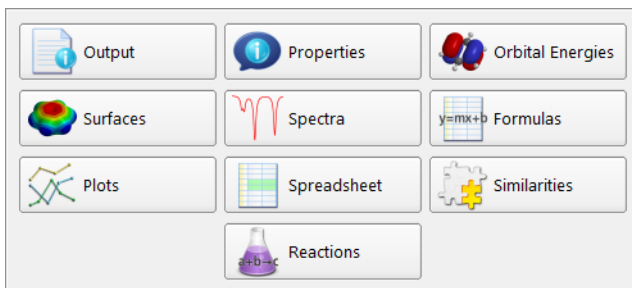
- a) **Measure Distance:** This measures the distance between two atoms. *Click* once on  to move to the next molecule, *propene*. *Click* on the **Geometry** menu and select **Measure Distance** (or *click* on the  icon if it appears at the top of the screen). *Click* on a bond or on two atoms (the 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).





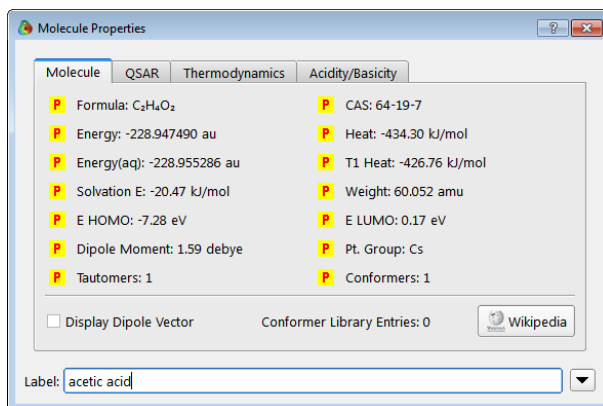
- b) **Measure Angle:** This measures the angle around a central atom. *Click* once on  to move to the next molecule, *ammonia*. *Click* on the **Geometry** menu and select **Measure Angle** (or *click* on the  icon if it appears at the top of the screen). *Click* first on H, then on N, then on another 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 four 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, **Thermodynamics** provides the enthalpy, heat capacity, Gibbs energy and zero-point energy and **Acidity/Basicity** provides information about the energies of deprotonation and protonation in the gas phase. Make certain that the **Molecule** tab is selected.

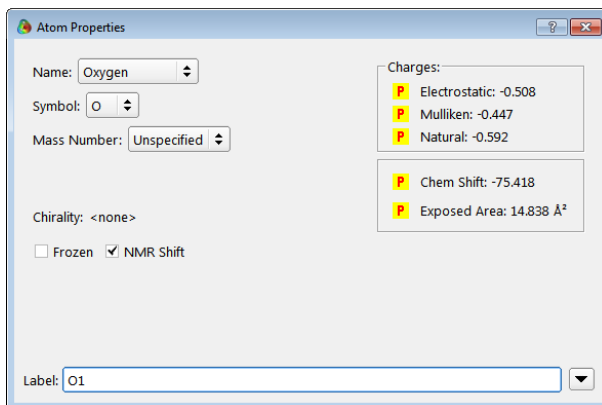


This provides the energy for acetic acid in atomic units (**Energy** in au). Also provided is an estimate of the energy in water (**Energy(aq)** in au) and the difference between the two (**Solvation E** in kJ/mol).

- 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,  $\text{+} \rightarrow$  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** near the bottom of the dialog.

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.

- c) **Atomic Charges:** To display the charge on an atom, *click* on it with the **Molecule Properties** dialog on the screen. The **Atom Properties** dialog replaces the **Molecule Properties** dialog.



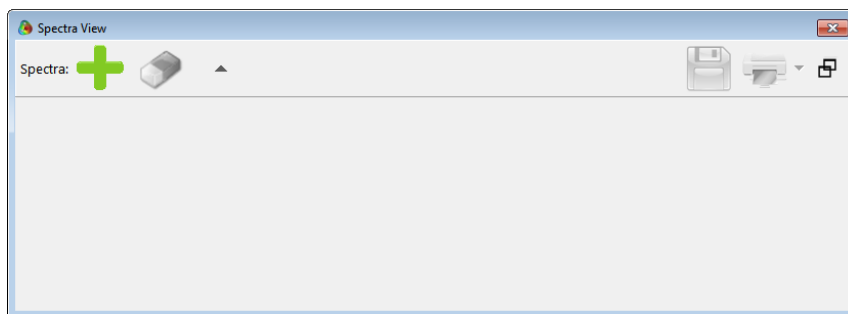
Three different atomic charges, **Electrostatic**, **Mulliken** and **Natural**, are given in units of electrons. A positive charge indicates a deficiency of electrons on an atom and a negative charge, an excess of electrons. Repeat for other atoms. Confirm that the positively-charged atom(s) lie at 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 because 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.

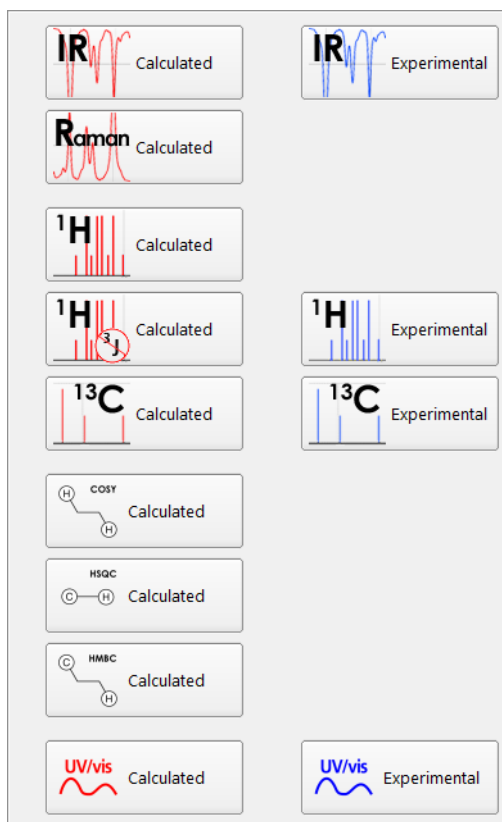
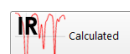
- d) **Infrared Spectra:** Molecules vibrate (stretch, bend, twist) even if they are cooled to absolute zero. This is the basis of infrared spectroscopy, where absorption of energy occurs when the frequency of a particular molecular motion matches the frequency of the light. Infrared spectroscopy is 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

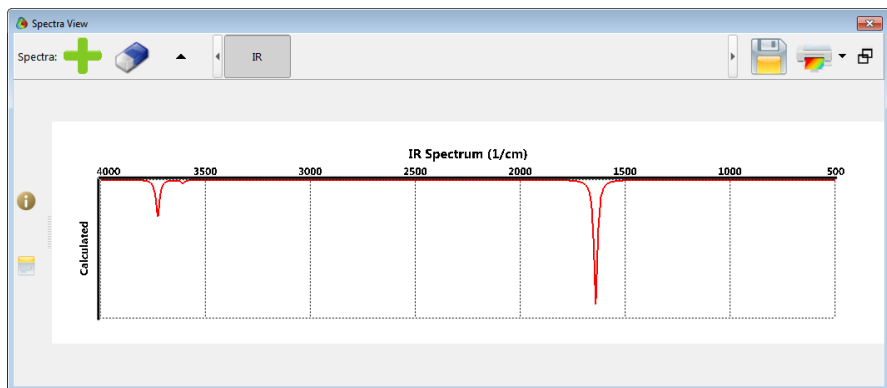
top of the screen). This leads to an empty spectra pane at the bottom of the screen.




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




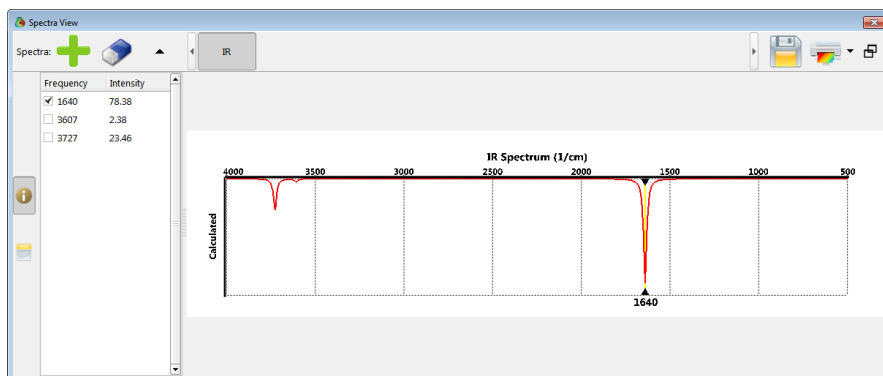
The calculated IR spectrum of water from 4000 - 500  $\text{cm}^{-1}$  appears in the pane.





There are three lines, one of moderate intensity around  $3730\text{ cm}^{-1}$ , one very weak around  $3610\text{ cm}^{-1}$  and one very strong around  $1650\text{ cm}^{-1}$ . In turn, move the cursor on the spectrum (move the mouse while holding down the left button over each of these lines). 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 increase (or decrease) the size of the plot, position the cursor inside the menu bar at the top and move the mouse up (or down) while holding down the left button. 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 **Alt** key). To reset the spectra plot to the original values, *click* on  in the bar at the top of the spectra pane.






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

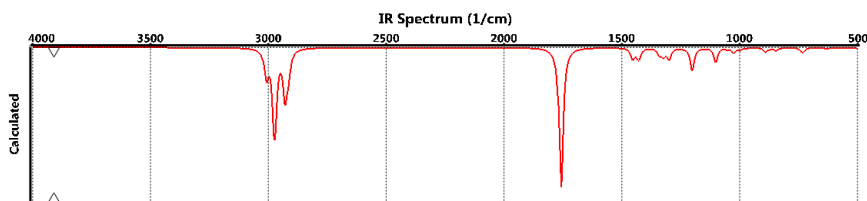


Click on each entry in the table to highlight the frequency in the spectrum and animate the vibration. Click on  again to dismiss the table and click on  to remove the spectrum.




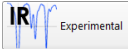
Changing the size of the spectrum dialog as well as translating and altering the visible scale are quite simple with touch screen operations. To resize the spectrum, position one finger inside the menu bar at the top of the spectra pane and move up or down. To translate the scale, move two fingers over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

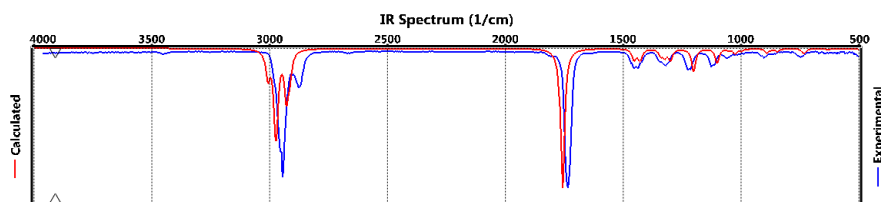
Click once on  to move to *cyclohexanone*, the next molecule in the list. The spectra pane is still on screen but should be empty. (If it is not on screen, select **Spectra** from the **Display** menu or click on  if it appears at the top of the screen to restore.) Click on  in the bar at the top of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



The spectrum obtained from quantum chemical calculations 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), but the same broadening

and scaling parameters are used for all molecules. The strongest line appears at  $1757\text{cm}^{-1}$  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. Move the cursor over this line and examine the “vibrating” model for cyclohexanone on screen above the spectrum.



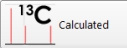
Click again on  and this time select . The experimental IR spectrum (from the public NIST database) is superimposed on top of the calculated spectrum.

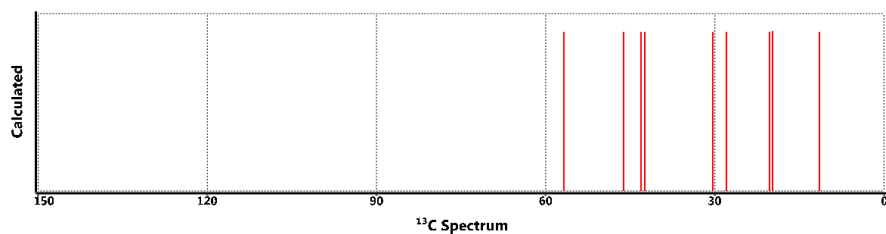


Note that the two spectra are similar but not perfectly matched.

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

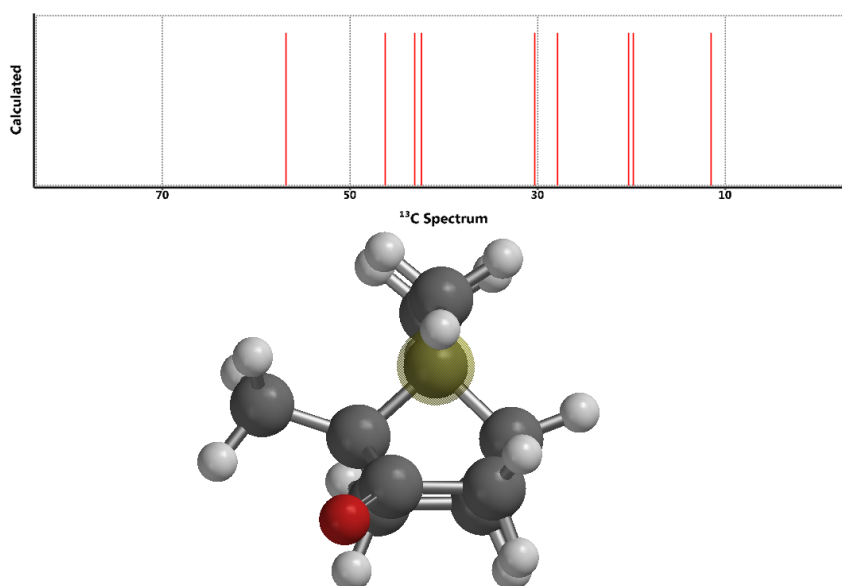
- e) **NMR Spectra:** Along with mass spectrometry, NMR spectroscopy is the most powerful tool available with which to assign molecular structure. Many nuclei exhibit NMR spectra, but proton and  $^{13}\text{C}$  are by far the most important.



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 and select . The calculated  $^{13}\text{C}$  NMR spectrum appears.

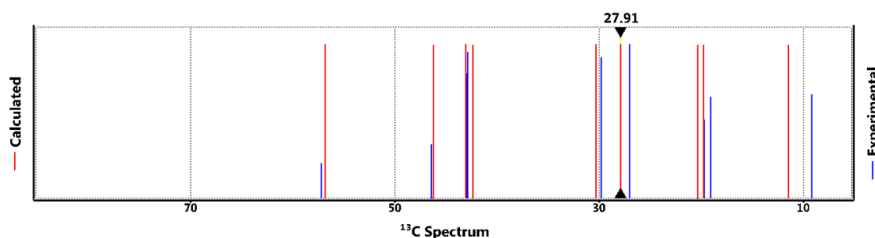


This comprises nine lines, in the range of 150 to 0 ppm (there is a tenth line corresponding to the carbonyl carbon at 217 ppm). You can zoom out to see this line by using the scroll wheel on your mouse. More instructive is to zoom in on the range from 60 to 0 ppm to get a better look at the other lines.


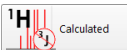
Move the mouse while holding down the left button over the spectrum. When you come to a line, the chemical shift will appear at the top of the spectrum and the atom responsible for this line will be highlighted on the model displayed above the spectrum.

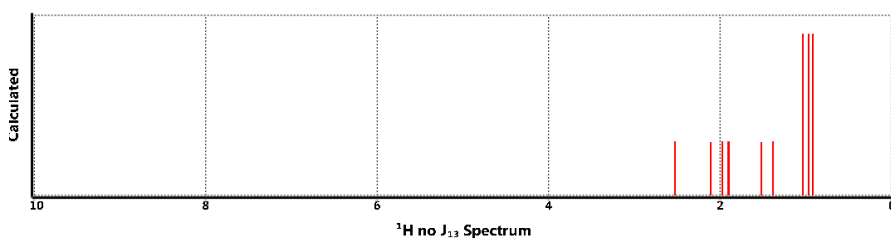


Again, *click* on  in the bar at the top of the spectra pane and select . The experimental  $^{13}\text{C}$  spectrum obtained from a public database will be superimposed on top of the calculated spectrum.


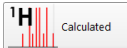


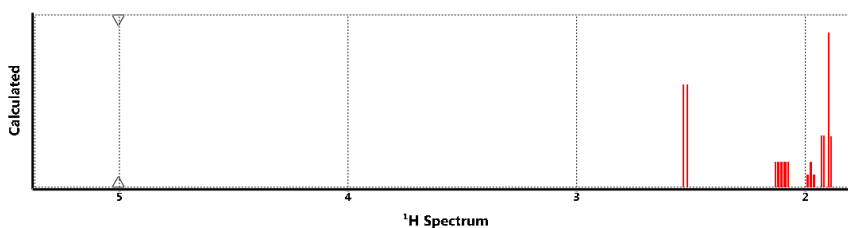
You will see that the overall agreement between calculated and experimental  $^{13}\text{C}$  spectra is quite good. As with infrared spectra (see preceding discussion of cyclohexanone), the data resulting from the quantum chemical calculations has been empirically corrected.

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




The 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  $^{13}\text{C}$  spectra.



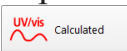
Click again on  and select  from the palette. The spectrum that appears is more complicated and much closer to what would be observed experimentally. Coupling constants have not been calculated from quantum mechanics, rather they have been estimated based on local environment.

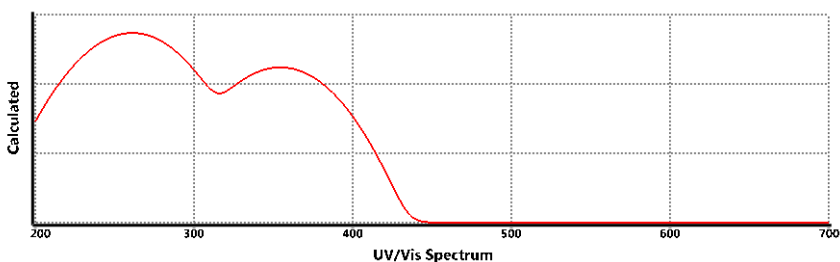



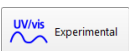
Zoom in on specific lines (scroll wheel) to see the detailed splitting patterns. For example, the two protons at  $\text{C}_3$  are both split by the proton at  $\text{C}_4$ . The doublet at 2.53 ppm shows a much larger splitting than the doublet at 1.90 ppm

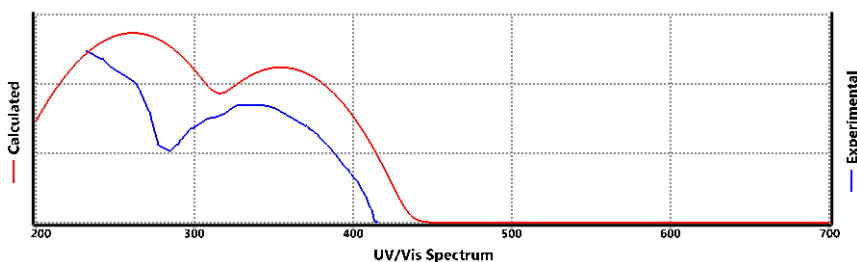
(you need to zoom way in to see that this is a doublet). This reflects the fact that the proton responsible for the line at 2.52 ppm makes a dihedral angle of  $43^\circ$  with the proton at C<sub>4</sub>, whereas the proton responsible for the line at 1.90 ppm makes a dihedral angle of  $80^\circ$ .




Finally, note that you can switch among the three calculated NMR spectra (as well as the experimental <sup>13</sup>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.

- f) **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 valuable 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-aminobenzophenone**. The spectra pane should still be on screen. *Click* on  and select . No empirical corrections have been applied to the calculated spectrum that appears.





*Click* again on  and select . The experimental UV/visible spectrum from the public NIST database will be drawn on top of the calculated spectrum.

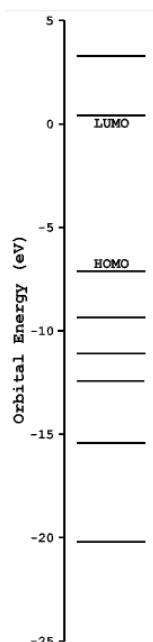


The two spectra are visually similar at least in a qualitative sense. However, calculated and experimental UV/visible spectra are likely to be sufficiently different that the theory will not often be able to account for the “color” of the molecule. Where the theory is likely to be more successful is in anticipating changes in color resulting from 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  if it appears at the top of the screen.

6. *Spartan* permits display, manipulation and query of a number of important graphical quantities resulting from quantum chemical calculations. Most important are the *electron density* (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.

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 two unoccupied molecular orbitals.



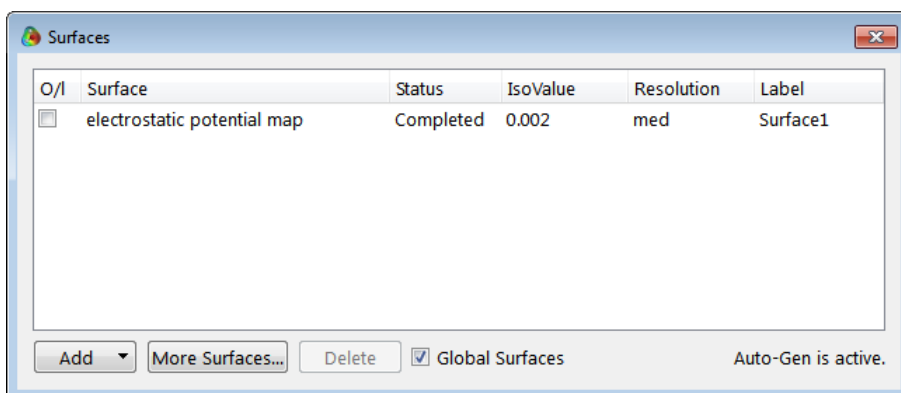
*Click* on the energy level in the diagram labeled **HOMO**. In a second, the familiar  $\pi$  bond in ethylene will appear. Note that the graphic has “blue and red” regions. These correspond to positive and negative values of the orbital (the absolute sign is arbitrary). Examine the other occupied orbitals (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.




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

*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).

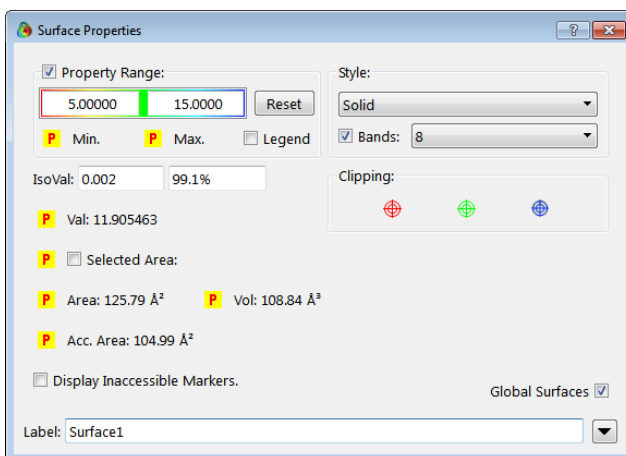
The **Surfaces** dialog appears.





Select *electrostatic potential map* inside the **Surfaces** dialog (*click* inside 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 surface is colored red in the  $\pi$  system (by convention, indicating negative potential and the fact that this region is attracted to a positive charge), and blue in the  $\sigma$  system (by convention, indicating positive potential and the fact that this region is repelled by a positive charge). *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. Remove the checkmark from the box to the left of **Bands** in the **Surface Properties** dialog to replace the series of color bands (discrete display) by a continuous display.\*

---

\* Discrete displays are the default. You can change the default to continuous displays from the **Molecule Preferences** dialog (**Preferences...** under the **Options** menu; **Chapter 23**).

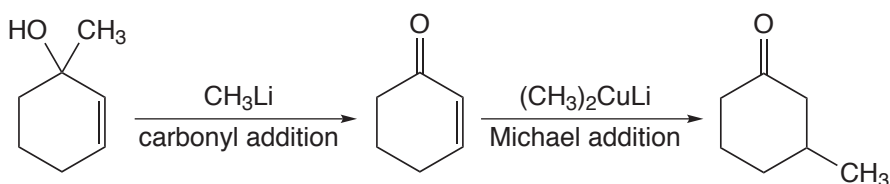



Click once on  to move to the next molecule in the list, ***aniline***, and select ***local ionization potential map*** inside the **Surfaces** dialog. By convention, red regions on a local ionization potential map indicate areas from which electron removal (ionization) is relatively easy, meaning that 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 more red than the *meta* carbons, consistent with the known directing ability of the amino substituent.

Click once on  to move to the last molecule in the list, ***cyclohexenone***, and select ***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 (select ***LUMO*** in the **Surfaces** dialog) and then turn on the LUMO map (select ***|LUMO|map*** in the dialog). Note that there are two blue regions, one directly

over the carbonyl carbon and the other over the  $\beta$  carbon. This is entirely consistent with known chemistry. Enones may either undergo carbonyl addition or conjugate (Michael) addition.



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

# Section II

## Organic Molecules

---

The first step in performing a molecular mechanics or quantum chemical calculation usually involves specifying the structure of a molecule. *Spartan* offers two ways to do this for organic molecules. One way is 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 *Spartan* provides tools for building organic, 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 organic chemists, who after all have been taught to draw molecules in 2D on paper or on a whiteboard. The advent of “touch” devices, tablets and now full-fledged PC's makes “on-screen” 2D drawing a viable alternative to 3D building.

New to *Spartan*'14 is a simple but highly functional 2D sketcher which, at present, is limited to organic molecules. Common atoms, functional groups and rings are all contained in a single palette. Conversion to a 3D structure (as would be produced by the 3D builder) 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. Sketch capability is designed for touch screens (and is modeled after the *iSpartan* app for iOS devices), but is fully functional using a conventional mouse and keyboard. Tutorials illustrating the essential features and capabilities of *Spartan*'s 2D builder are provided in **Chapter 4**.

Properties relating to 3D structure, 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 more advanced functions to tutorials in later chapters.

The tutorials in the remaining chapters may be completed starting from either the 2D sketcher or 3D builder. 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 produced. No calculations are actually required for these tutorials. Rather, results are taken from the *Spartan Spectra and Properties Database (SSPD)*. The first tutorials involving quantum chemical calculations are provided in **Chapter 7**.

Completing the full set of tutorials in this section will take perhaps one or two hours, with only a modest demand of computer time for the tutorials in the last chapter. This will not make you an “expert”, but will leave you comfortable enough to navigate throughout *Spartan*. “Advanced” tutorials grouped under the next section will build upon these skills and open you to the full possibilities of the program.

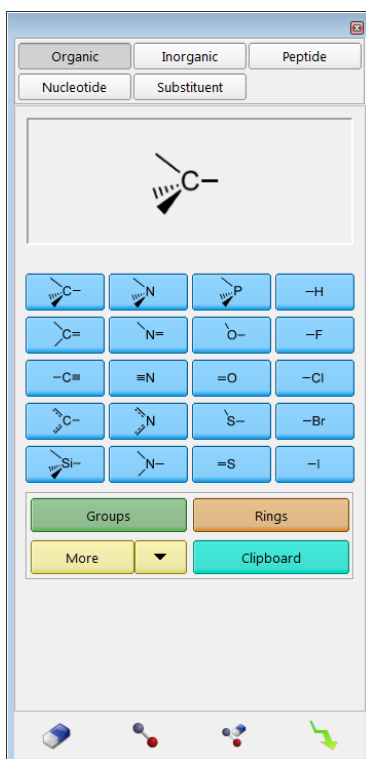
# Chapter 3

## 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

The organic model kit contains a selection of atomic fragments corresponding to elements commonly found in organic molecules.



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

C(sp <sup>3</sup> )	N(sp <sup>3</sup> )	P(sp <sup>3</sup> )	H
C(sp <sup>2</sup> )	N(sp <sup>2</sup> )	O(sp <sup>3</sup> )	F
C(sp)	N(sp)	O(sp <sup>2</sup> )	Cl
C(aromatic)	N(aromatic)	S(sp <sup>3</sup> )	Br
Si(sp <sup>3</sup> )	N(planar)	S(sp <sup>2</sup> )	I

A fragment is chosen by *clicking* on its icon, which is then displayed in a box at the top of the model kit. 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, *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<sup>2</sup> 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 directly underneath. *Clicking* on the name brings up a menu of available groups.

---

\* Previous versions of *Spartan* required only a single click to initiate building, and this option is still available. *Uncheck Double Click Start* in the **Settings Preferences** dialog (**Preferences** under the **Options** menu; **Chapter 23**). The switch to double click is for compatibility with 2D sketch mode.

	Alkenyl
	Alkynyl
	Allenyl
	Amide
	Azido
	Carbonyl
	Cyano
	Carboxylic Acid/Ester
	Isocyanato
	Nitro
	Nitroso
	Phosphine Oxide
	Sulfone
	Sulfoxide

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.

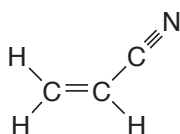
*Clicking* on the **Rings** button near the bottom of the model kit changes the focus to rings, one of which will be shown in the box at the top of the model kit and named immediately underneath. *Clicking* on the name brings up a menu of available rings.


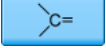


	Cyclopropane
	Cyclobutane
	Cyclopentane
	Cyclohexane
	Cycloheptane
	Benzene
	Naphthalene
	Anthracene
	Phenanthrene
	Indene
	Fluorene

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.




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. Note also, that the amide and carboxylic acid/ester groups and the cyclohexane, cycloheptane, naphthalene, phenanthrene, indene and fluorene rings have more than one different 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.

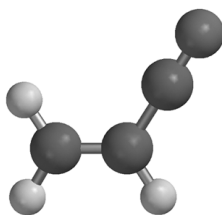
## Acrylonitrile



1. *Click* (left button) on the **File** menu and select (*click* on) **New Build**. Alternatively, *click* on  if it appears at the top of the screen. The organic model kit appears. *Click* on trigonal planar  $sp^2$  hybridized carbon  from the fragment library. 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. 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 **File** 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 name of whatever group is shown in the text box 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 now 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 also appears 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 if it appears at the top of the screen). 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 in the Spartan Spectra and Properties Database (SSPD).




ball-and-spoke model

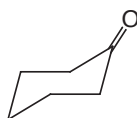
This model can be rotated, translated and zoomed by using the mouse in conjunction with keyboard functions. To rotate the model, *drag* the mouse while holding down the left button; to rotate in the plane of the screen also hold down the **Shift** key. To translate the model, *drag* the mouse with the right button depressed. To zoom the model (translation perpendicular to the screen), 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  if it appears at the top of the screen).

## Cyclohexanone

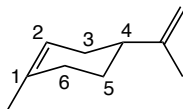




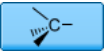


1. *Click* on the **File** menu and select **New Build** from the palette or *click* on  if it appears at the top of the screen. *Click* on the **Rings** button near the bottom of the model kit, *click* on the name of whatever ring is shown and select **Cyclohexane** in the menu. *Double click* anywhere on screen.
2. Select  $sp^2$  carbon  from the model kit. *Double click* on any carbon (**not 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 an  $sp^3$  carbon by an  $sp^2$  carbon requires that at least two free valences are available.

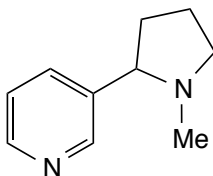
3. Select  $sp^2$  oxygen  from the model kit. *Click* on the double free valence on the  $sp^2$  carbon. You have made cyclohexanone (its name will appear at the bottom right of the screen). *Click* on  at the bottom of the model kit to produce a structure with  $C_s$  symmetry. *Click* on  to remove the model kit. The name **cyclohexanone** appears at the bottom of the screen as the molecule is in SSPD.
4. Close cyclohexanone.


## Limonene

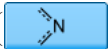

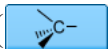




1. Select **New Build** from the **File** menu (or *click* on the  icon if it appears at the top of the screen) to bring up the organic model kit. *Click* on the **Rings** button near the bottom of the model kit, *click* on the name of whatever ring is shown and select **Cyclohexane** in the menu. *Double click* anywhere on screen.
2. *Click* on the **Groups** button near the bottom of the model kit, *click* on the name of whatever group is shown and select **Alkenyl** from the menu. *Click* on the *equatorial* free valence on C<sub>4</sub> (see figure above for numbering). You have made vinylcyclohexane.
3. *Click* on () at the bottom of the model kit. One after another *click* on the *axial* free valence on C<sub>1</sub> and then the *axial* free valence on C<sub>2</sub>. You have made 4-vinyl-1-cyclohexene.
4. Select sp<sup>3</sup> carbon () from the model kit and one after another *click* on the free valence on C<sub>1</sub> and on the free valence of the vinylic carbon attached to the ring. You have made limonene. *Click* on  at the bottom of the model kit to give a refined geometry and finally *click* on () to remove the model kit. The name will appear at the bottom of the screen as limonene is in SSPD.
5. Close limonene.

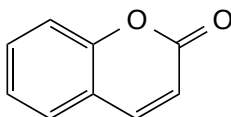
## Nicotine




1. Select **New Build** from the **File** menu (). *Click* on the **Rings** button in the model kit. *Click* on the name of whatever ring is shown in the window at the top of the model kit and select **Benzene** from the menu. *Double click* anywhere on screen.


2. Click on the name **Benzene** (the currently selected ring) in the model kit and select **Cyclopentane** from the menu. Click on one of the free valences on benzene on screen. You have made phenylcyclopentane (the name will appear at the bottom of the screen).
3. Click on aromatic nitrogen in the model kit () and double click on the appropriate (*meta*) carbon (not a free valence) in the benzene ring for the structure on screen. 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 for the structure on screen. 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 (named at the bottom of the screen). 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 **Configure** from the **Model** menu (). Under the **Labels** tab, click the radio button next to **R/S** and then click on **OK**. The R/S chirality will be displayed. The S isomer is the naturally occurring isomer of nicotine.
7. Close nicotine.

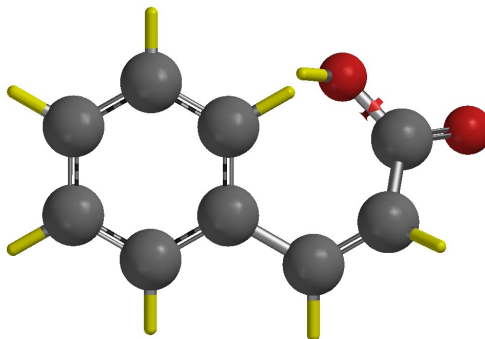
## Coumarin




1. Select **New Build** from the **File** menu (). Click on the **Rings** button in the model kit. Click on the name of whatever ring is shown in the window at the top of the model kit and select **Benzene** from the menu. Double click anywhere on screen.
2. Click on the **Groups** button, click on the name of whatever

group is shown in the window at the top of the model kit and select **Alkenyl** from the menu. *Click* on one of the free valences of benzene to make styrene.

3. *Click* on the name **Alkenyl** (the selected group) and select **Carboxylic Acid/Ester** from the menu. 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. Make a bond between the free valence on the oxygen and that on benzene adjacent (*ortho*) to the group that you added. *Click* on the connecting bond. 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. The two fragments will rotate relative to each other.



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.



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

5. Close *coumarin*.

# 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 automatically 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.

### Sketch Palette

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

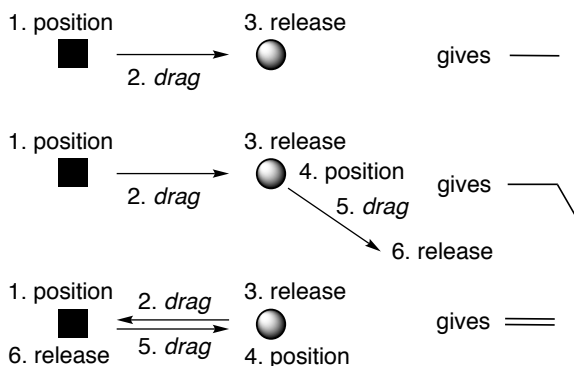


These include atoms that are most commonly found in organic molecules (H, B, C, N, O, F, Si, P, S, Cl, Br and I), the phenyl, cyclohexyl and cyclopentyl rings and the carbonyl, acid/ester and amide functional groups. A “wildcard” (immediately below H and B atoms) allows for entering additional elements and functional groups

from a library. The palette also contains stereochemical markers and charge/radical markers. Complete details are found in **Chapter 19**.

## 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). Note the difference in starting a 3D structure (*single click*) and starting a 2D sketch (*double click*). Both the 3D builder and the 2D sketcher use a *double click* to insert a new (unattached) fragment to an existing structure. 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.

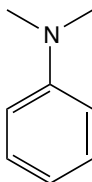
## Manipulating a Sketch








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.



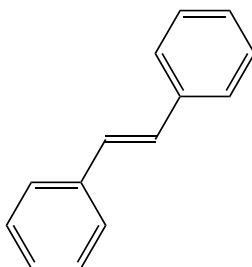
To translate the sketch, move two fingers over the screen. To rotate the sketch in the plane of the screen, “twist” two fingers on the screen. To resize the sketch, pinch (or spread) two fingers on the screen.




## N,N-Dimethylaniline

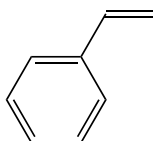





1. Select **New Sketch** from the **File** menu (or *click* on  if it appears at the top of the screen) to bring up the sketch pad. *Click* on () in the palette and *double click* on screen.
2. *Click* on () in the palette. Position the cursor over the “top” carbon on the benzene ring, *drag* it up and release. You have drawn aniline.
3. *Click* on () in the palette. Position the cursor over the nitrogen, *drag* it up and to the left and release. () 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

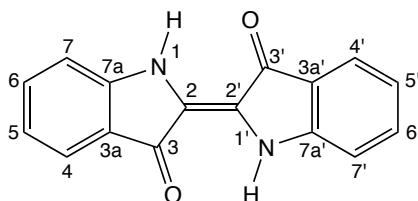





1. Select **New Sketch** from the **File** menu (  ) to bring up the sketch pad. Click on (  ) in the palette. Position the cursor on the screen, *drag* it to the right and release. Position the cursor on one end of the line (CC bond) that you have just drawn, *drag* it to the other end and release.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

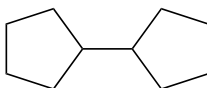



3. (  ) is still selected. Position the cursor over the right end of the double bond, *drag* it up and to the right and release. You are done.
4. Click on (  ) to clean up your drawing. Click on (  ) to convert it to a 3D structure. The name **trans-stilbene** will appear at the bottom of the screen as the molecule is in SSPD.
5. Close trans-stilbene.

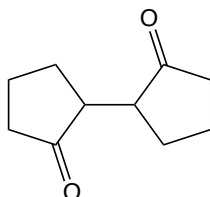
## Indigo




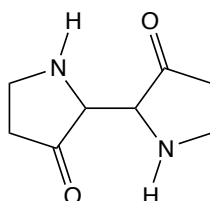
1. Select **New Sketch** from the **File** menu (  ). Click on (  ) and *double click* on screen.
2. (  ) is still selected. Position the cursor over C<sub>2</sub> (see numbering in diagram above), *drag* it to the right and release. Your sketch should appear as below.



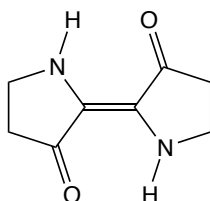
3. Select (  ) from the palette. Position the cursor above C<sub>3</sub>, *drag* it away from the ring and release. Again position the cursor above C<sub>3</sub>, drag it along the CO bond to the oxygen and release. Repeat for C<sub>3</sub>'. You are left with a drawing.






4. Select (  ) from the palette and one after another *double click* on C<sub>1</sub> and C<sub>1</sub>'.

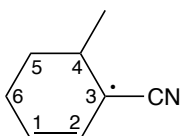


5. Select (  ) from the palette. Position the cursor above C<sub>2</sub>, *drag* it to C<sub>2</sub>' and release.















6. Select (  ) from the palette and *double click* on the bond connecting C<sub>3a</sub> and C<sub>7a</sub>. Repeat for C<sub>3a</sub>' and C<sub>7a</sub>'.
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.

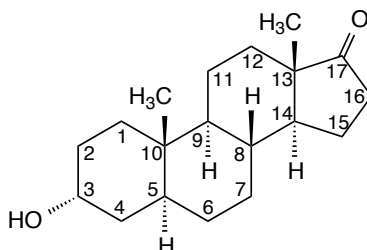
### 3-Cyano-4-methylcyclohexenyl Radical





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

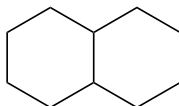
1. Select **New Sketch** from the **File** menu (). Click on () in the palette and *double click* on screen.
2. Click on () in the palette. Position the cursor over C<sub>4</sub> (see diagram above for numbering), *drag* it away from the ring and release. You have drawn methylcyclohexane.
3. () is still selected. Position the cursor over C<sub>1</sub>, *drag* to C<sub>2</sub> and release. You have drawn 4-methylcyclohexene.
4. *Double-click* on the wildcard icon that is located immediately below H and B in the palette (it will contain the name of an element or functional group). Type **CN** in the box in the dialog that appears and *click* on **OK**. **CN** now appears as the wildcard icon. Position the cursor over C<sub>3</sub>, and *drag* it away from the ring and release. You have drawn 3-cyano-4-methylcyclohexene.
5. One of these icons (, , ) will appear in the palette directly below (). If it is not () , *click* on the icon until the icon is () . *Double click* on C<sub>3</sub>. A “dot” (radical marker) will appear next to C<sub>3</sub>.
6. Click on () to clean up your drawing and *click* on () to make a 3D structure. The name will not appear at the bottom of the screen as this radical is not in SSPD.
7. Close 3-cyano-4-methylcyclohexenyl radical.

## Androsterone

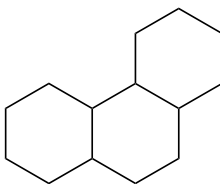








The steroid androsterone is typical of a molecule that is difficult to build but quite easy to sketch.

1. Select **New Sketch** 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 5-10 bond (see diagram above).







2. Cyclohexane is still selected. *Double click* on the 8-9 bond.



3. Click on () and *double click* on the 13-14 bond. You have now drawn the complete steroid skeleton.
4. Click on () , position the cursor over C<sub>10</sub>, *drag* up and release. Repeat for C<sub>13</sub>.
5. Click on () . Position the cursor over C<sub>3</sub>, *drag* down and to the left and release. Position the cursor over C<sub>17</sub>, *drag* up and to the right and release. Convert the single (CO) bond at C<sub>17</sub> to a double bond. Again position the cursor over C<sub>17</sub>, *drag* along the bond to oxygen and release.
6. It is necessary to explicitly specify hydrogens at C<sub>5</sub>, C<sub>8</sub>, C<sub>9</sub> and C<sub>14</sub> in order to incorporate the necessary stereochemical cues (up and down “wedges”) in your drawing. Click on () , position the cursor over C<sub>5</sub>, *drag* away from the ring and release. Repeat for C<sub>8</sub>, C<sub>9</sub> and C<sub>14</sub>.
7. Click on () . Position the cursor over C<sub>10</sub>, *drag* along the bond to the methyl group that you drew in step 4 and release. Repeat for C<sub>13</sub>. Position the cursor over C<sub>8</sub>, *drag* along the CH bond that you drew in step 6 and release. Up wedges will appear for all three centers.
8. Click on () . Position the cursor over C<sub>3</sub>, *drag* along the CO

bond that you made in step 5 and release. Position the cursor over C<sub>5</sub>, *drag* along the CH bond that you made in step 6 and release. Repeat for the CH bonds at C<sub>9</sub> and C<sub>14</sub>. Down wedges will appear for all four centers.

9. 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.
10. Androsterone incorporates seven chiral centers. To assign them as R or S, select **Configure...** from the **Model** menu or *click* on (  ) if it appears at the top of the screen. *Check* **R/S** under **Atom** in the dialog that appears. *Click* on **OK**. R/S labels will appear next to each of the chiral centers.
11. Close androsterone.

# Chapter 5

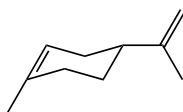
## Spectra and Properties of Organic Molecules from the Spartan Spectra and Properties Database








---

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

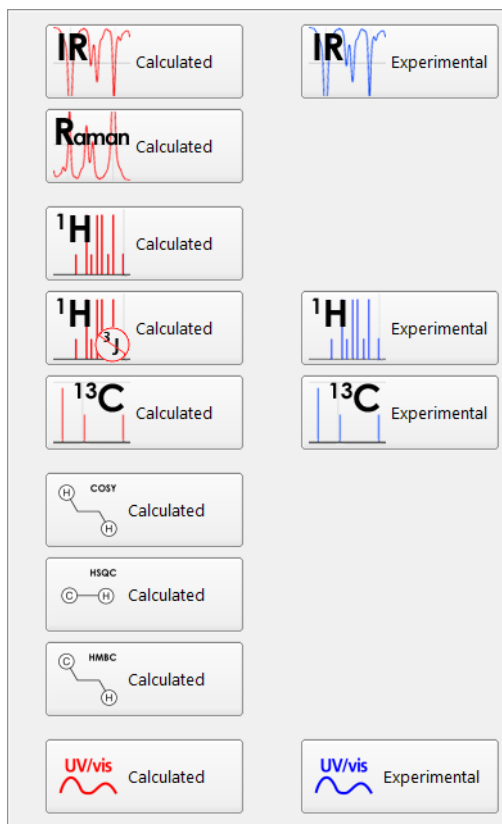
The Spartan Spectra and Properties Database (SSPD) provides atomic and molecular properties and IR and NMR spectra for more than a quarter of a million molecules, calculated using the EDF2/6-31G\* density functional model. This allows us to explore what is calculated without having to actually specify and run calculations.

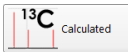
### Limonene






1. Build limonene () , minimize () and exit the builder (). Alternatively, sketch limonene () and exit the sketcher (.
2. Click on the name **limonene** at the bottom of the screen and click on **Replace** in the dialog that results. Properties and spectra for limonene are now available.
3. Select **Spectra** from the **Display** menu, or click on () if it appears at the top of the screen. Click on  at the top of the

spectra pane that results to show available calculated spectra (in red) and (possibly) available experimental spectra (in blue).



Select . The calculated  $^{13}\text{C}$  NMR spectrum appears in the spectra pane. Move the cursor horizontally over the spectrum. You will see that when you intersect a line, it will be highlighted in the spectrum and the value of the chemical shift indicated. Also, the carbon (or carbons) in the structure responsible for this line will be highlighted in the structure model (in the top part of the screen). You will see that there are ten lines in the calculated  $^{13}\text{C}$  spectrum, corresponding to ten unique carbons in limonene.

- Again, *click* on  at the top of the spectra pane, but this time, select . The experimental  $^{13}\text{C}$  spectrum of limonene will be superimposed on top of the calculated spectrum. Visual comparison will give you an idea of what you can expect from  $^{13}\text{C}$  shift calculations. To get an even better idea, shift

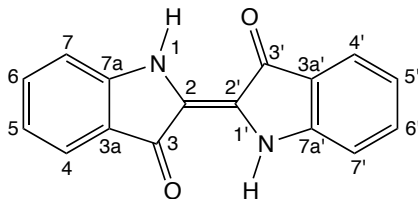
the range of the scale (initially from 150 to 0 ppm), move the cursor over the spectrum while holding down the the right button. You can also change the scale (initially from 150 ppm), that is, zoom-in or zoom-out, using the scroll wheel on your mouse. You can return to the original setting by *clicking* on () in the bar at the top of the spectra pane.










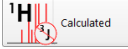



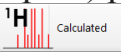
Move one finger over the spectrum to select a line. Move two fingers over the spectrum to shift the range of the scale and pinch two fingers to zoom in and out.


## 5. Close limonene.

### Indigo

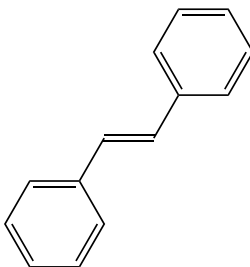









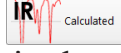
1. Build indigo () , minimize () and exit the builder () . Alternatively, sketch indigo () and exit the sketcher () .
2. *Click* on the name **indigo** at the bottom of the screen and *click* on **Replace** in the dialog that results. Properties and spectra for indigo are now available. If the name does 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 four unique hydrogens. To see a “more familiar” (and more complex) proton spectrum, *click* again on , but this time select . The same four lines appear, but all are split (as in a real proton spectrum). The lines at 6.85 and 7.80 ppm are doublets due to C<sub>7</sub> and C<sub>4</sub>, respectively (and split by C<sub>6</sub> and C<sub>5</sub>, respectively). The lines centering at 6.97 and 7.38 are quartets (doublet of doublets), due to C<sub>5</sub> and C<sub>6</sub>, respectively (and split by C<sub>4</sub> and C<sub>6</sub> and C<sub>5</sub> and C<sub>7</sub>, respectively).


4. The experimental proton NMR for indigo is not available from the on-line database. However, the experimental <sup>13</sup>C spectrum is available. You can if you wish compare it to the corresponding calculated spectrum. Either is accessed by *clicking* on  at the top of the spectra pane followed by selecting the appropriate entry from the resulting palette.
5. Close indigo.

### *trans*-Stilbene





1. Build *trans*-stilbene () , minimize () and exit the builder (). Alternatively, sketch *trans*-stilbene () and exit the sketcher ().
2. *Click* on the name ***trans-stilbene*** at the bottom of the screen, and then *click* on **Replace** in the dialog that results. Data from SSPD is now available.
3. Select **Spectra** from the **Display** menu (). *Click* on  in the bar at the top of the spectra pane and select . The calculated IR spectrum of *trans*-stilbene appears in the spectra

pane. You may find it valuable to increase the size of the spectrum. Position the cursor inside the bar at the top of the spectra pane and *drag* it up.

4. Move the cursor 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 of the lines of moderate intensity in the vicinity of  $1500\text{ cm}^{-1}$  (at  $1446$ ,  $1494$  and  $1609\text{ cm}^{-1}$ ). 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 right button). You can return to the original settings by *clicking* on () in the bar at the top of the spectra pane.



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.

5. Click on  from the bar at the top of the spectra pane and select . The experimental IR spectrum of *trans*-stilbene obtained from the public NIST database will be superimposed onto the calculated spectrum. Note that the two spectra are similar although the experimental spectrum exhibits a number of (small) lines not found in the calculated spectrum.
6. Close *trans*-stilbene.

# Chapter 6

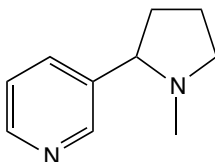
## Graphical Models for Organic Molecules from the Spartan Spectra and Properties Database






---

*The tutorials in this chapter illustrate three 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, atomic and molecular properties and IR and NMR spectra, but also the wave function. This allows graphical models to be constructed and displayed “on-the-fly”.

### Nicotine

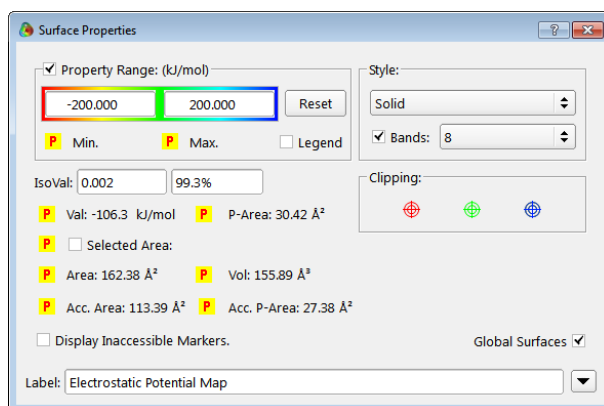


1. Build nicotine () , minimize () and exit the builder (). Alternatively, sketch nicotine () and exit the sketcher (.
2. Click on the name **nicotine** at the bottom of the screen, and click on **Replace** in the dialog that results. Your structure will be replaced by that in SSPD making the wave function available.
3. Select **Surfaces** from either the **Setup** or **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 mapped). A 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). Which is the more negative, the nitrogen in the pyridine ring or that 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. This will bring up the **Surface Properties** dialog.



*Check* the box to the left of **Legend** to display the property range on screen. (*Uncheck* the box to remove the legend.) To translate the legend, *click* on the legend to select, then hold down the

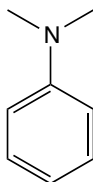
right mouse button and move the mouse. The legend is useful when making qualitative comparisons of property values. Turn the map such that you can clearly see the pyridine nitrogen and *click* on the area that is “most red”. An arrow marks the point on the surface and the value of the potential is shown 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 very 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 both inside the **Surface Properties** dialog and at the bottom right of the screen. Select **Transparent** or **Mesh** from this menu. You now can “see through” the map to the underlying molecular structure.
6. Close nicotine and any open dialogs.

## N,N-Dimethylaniline

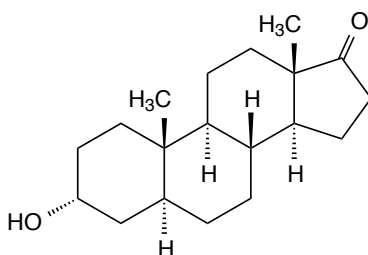









1. Build N,N-dimethylaniline () , minimize () and exit the builder () . Alternatively, sketch N,N-dimethylaniline () and exit the sketcher () .
2. *Click* on **N,N-dimethylaniline** at the bottom of the screen and *click* on **Replace** in the dialog that results. Your structure will be replaced by that in SSPD.
3. Select **Surfaces** from either the **Setup** or **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 is automatic and 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. 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.

## Androsterone



1. Build androsterone () , minimize () and exit the builder () . It is much easier to sketch the molecule () and exit the sketcher () .
2. Click on the name **androsterone** at the bottom right of the screen and click on **Replace** in the dialog that results. Your structure will be replaced by that in SSPD. 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 () .
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  $\pi^*$  orbital localized

on the carbonyl group. It is clearly more concentrated on carbon than on oxygen, but you cannot easily tell which side of the steroid skeleton (toward or away from the two methyl groups) it is concentrated. For this, a LUMO map is a much better graphical model.

4. Select **Surfaces** from the **Display** menu (). Click on **Add** and select **|LUMO| map** from the menu. The graphics calculation will run automatically and will only require 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. Note that you can easily see that the LUMO is more concentrated on the face away from the methyl groups.
5. Close androsterone and any open dialogs.

# Chapter 7

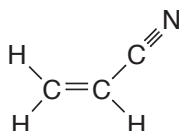
## Spectra, Properties and Graphical Models for Organic Molecules from Quantum Chemical Calculations

---







*The tutorials in this chapter provide the earliest, and some of the simplest examples, of specifying and performing quantum chemical calculations. As in the previous two chapters, they refer back to tutorials in **Chapters 3 and 4**.*

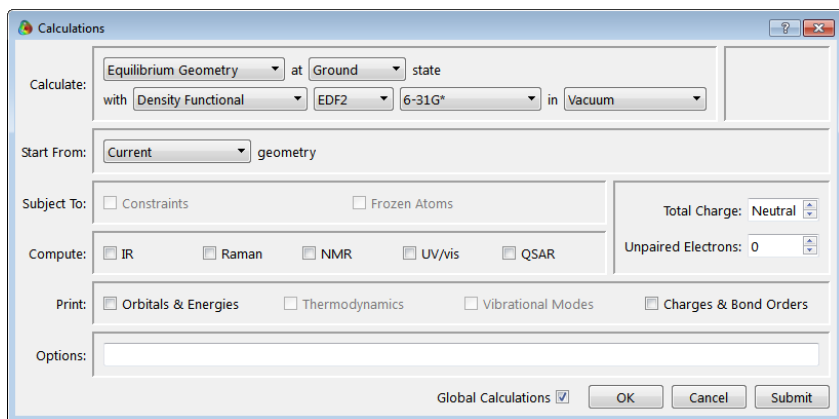
1 min

### Acrylonitrile

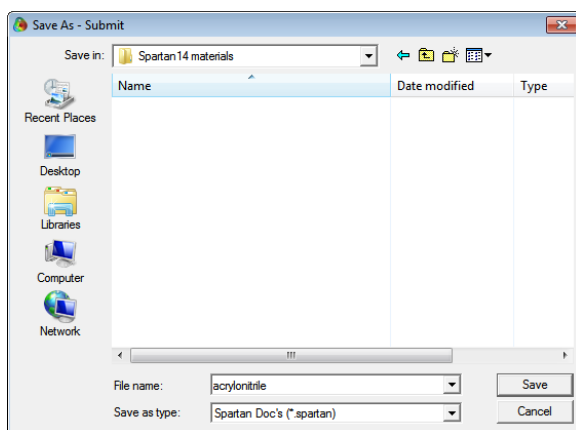


We return to acrylonitrile for the first tutorial that actually involves quantum chemical calculations. The molecule is very small and the calculations will require only a minute or two.

1. Build acrylonitrile () , minimize () and exit the builder () . Alternatively, sketch nicotine () and exit the sketcher () .
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 which appears.

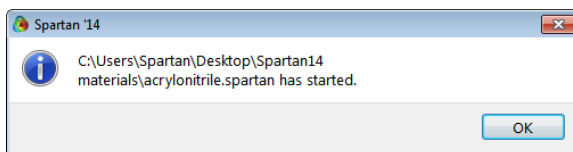


- Select **Equilibrium Geometry** from the top menu to the right of **Calculate**. This specifies optimization of equilibrium geometry. **Ground** and not **First Excited** should appear in the menu to the left of **State**.
- Select **Density Functional**, **EDF2**, **6-31G\*** and **Vacuum** from the four bottom menus to the right of **Calculate**. This requests that the EDF2/6-31G\* density functional model is to be used for this calculation (the same model as that in SSPD).
- Click* on **Submit** at the bottom of the dialog. A file browser appears.



Because the model is in SSPD (even though we will not use the data), the name *acrylonitrile* will be presented to you in the box to the right of **File name:**. Either use it or type in whatever

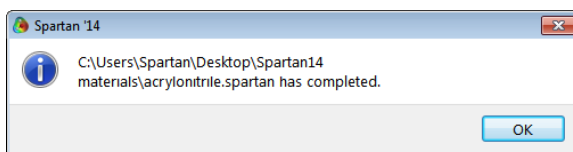
name you like and then *click* on **Save**. You will be notified that the calculation has been submitted.




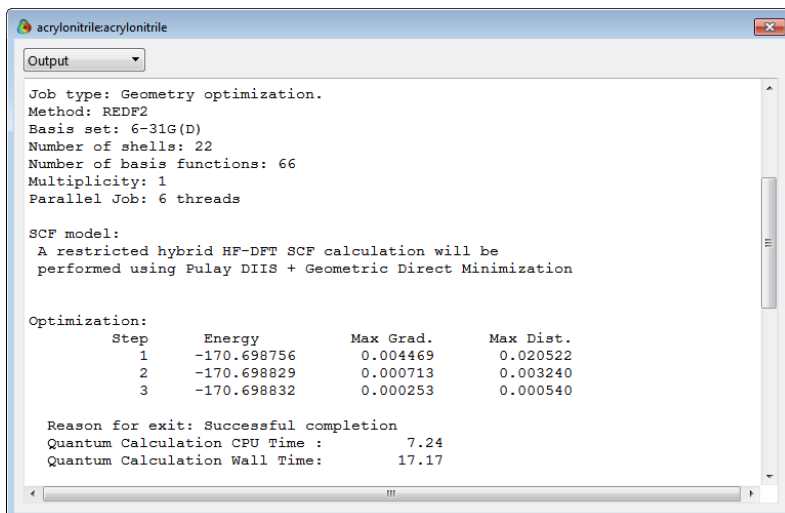
*Click* on **OK** to remove the message from the screen.

After a molecule has been submitted, and until the calculation has completed, you are not permitted to modify any dialogs or other information associated with it.

3. You will be notified when the calculation has completed (less than a minute).




*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. A window containing text output for the job appears.



You can scan the output from the calculation by using the scroll bar at the right of the window or by *clicking* (left button) on or inside the output window and using the scroll wheel on your mouse. The information at the top of the dialog includes the task, basis set, number of electrons, number of unpaired electrons and the charge, as well as further details of the calculation. Below this is the symmetry point group of the molecule that was maintained during the optimization.

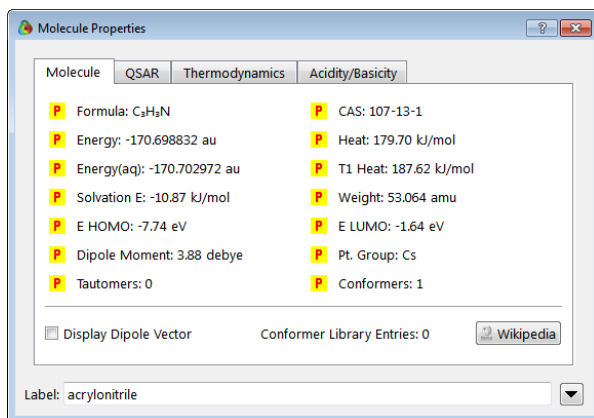
Eventually, a series of lines appear, under the heading **Optimization**. These tell the history of the optimization process. Each line (or **Step**) provides results for a particular geometry. Ideally, the energy will monotonically approach a minimum value for an optimized geometry. If the geometry was not optimized satisfactorily an error message, such as: **Optimization has exceeded N steps – Stop**, will be displayed following the last optimization cycle. If this were the case, you would have been notified that the job had failed.

Near the end of the output is the final total energy\* (-170.69883 atomic units for acrylonitrile with the EDF2/6-31G\* model), and the computation time. *Click* on  at the top of the output dialog to close it.

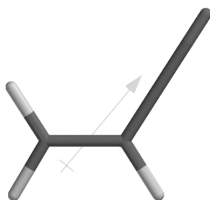
You may examine the total energy and dipole moment among other calculated properties without having to go through the output. Select **Properties** from the **Display** menu to bring up the **Molecule Properties** dialog (make certain that the **Molecule** tab and not the **QSAR**, **Thermodynamics**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

---

\* See **Calculations...** (**Setup** menu; **Chapter 20**) for a discussion of how total energy relates to heat of formation and strain energy.

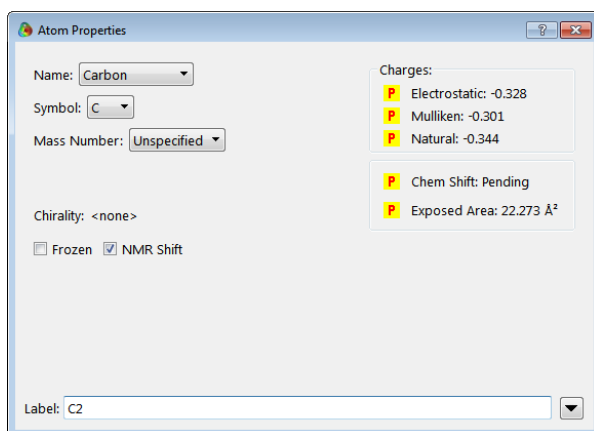


To see the dipole moment vector (indicating the sign and direction of the dipole moment), *check* the box to the right of **Display Dipole Vector**. Wire, ball-and-wire or tube models are best for this display.





*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.



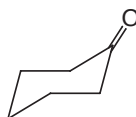
Among other things, this provides three different sets of atomic charges: **Electrostatic**, **Mulliken** and **Natural**. To obtain the charge on another atom, simply *click* on it. Inspect all the atomic

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.






4. Select **Surfaces** from either the **Setup** or **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 onto which the value of the electrostatic potential is mapped). The graphics calculation will run automatically (without needing to resubmit the job) 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 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 and the hydrogens (the most electropositive atoms) are blue.
5. Close *acrylonitrile* and any open dialogs.



1 min

## Cyclohexanone

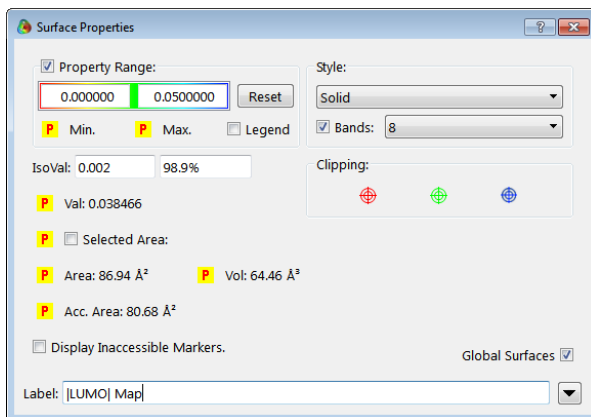


Cyclohexanone provides another example of a quantum chemical calculation on a very simple molecule. It also provides an opportunity to further illustrate graphical models for elucidating the stereochemistry of organic reactions.

1. Build cyclohexanone () , minimize () and exit the builder (). Alternatively, sketch cyclohexanone () and exit the sketcher (.

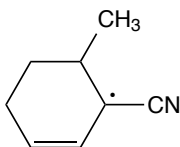
2. Select **Calculations...** from the **Setup** menu (). Specify **Equilibrium Geometry** from the top menu to the right of **Calculate**, and **Density Functional, EDF2, 6-31G\*** and **Vacuum** from the four bottom menus to the right of **Calculate**. Click on **Submit** and accept the name *cyclohexanone*. Wait until the calculation completes (2-3 minutes) 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 **Setup** or **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  $\pi^*$  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 on which face of the carbonyl carbon is the LUMO more concentrated.
5. Uncheck the box to the left of **LUMO** in the **Surfaces** dialog (to turn off the display of the LUMO, it operates in toggle mode). 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 more blue) 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 these two faces. Select **Properties** from the **Display** menu (1) and *click* anywhere on the graphic to bring up the **Surface Properties** dialog.









- Check* the box to the left of **Legend** to display the property range on screen. (Remove the checkmark to remove the legend.) 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 *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 now see the *equatorial* face of the carbonyl carbon, and *click* on the region of maximum blue on this face. Do these values support your qualitative conclusions from viewing the image?
7. Close *cyclohexanone* and any open dialogs.

### 3-Cyano-4-methylcyclohexenyl Radical



3-cyano-4-methylcyclohexene radical will be used to illustrate how different models may be combined to achieve a desired result with minimum computation cost. MMFF molecular mechanics will be employed to establish equilibrium conformation, the semi-empirical AM1 model to establish equilibrium geometry and the Hartree-Fock 3-21G model to provide a basis for graphical calculations.

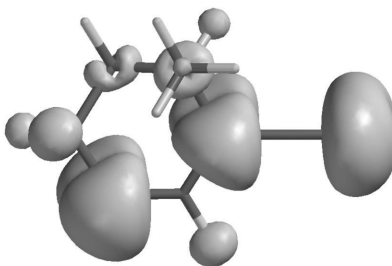
1. Build 3-cyano-4-methylcyclohexenyl radical () , minimize () and exit the builder (). Alternatively, sketch the radical () and exit the sketcher ().
2. Select **Calculations...** from the **Setup** menu (). Specify **Energy** from the top menu to the right of **Calculate**, **Hartree-Fock** and **3-21G** and **Vacuum** from the three bottom menus. Next, specify **MMFF Conformer** from the **Start from** menu and **AM1** from the menu which then appears to the right. You have requested that a Hartree-Fock energy is to be obtained using an AM1 geometry, which in turn is to be based on the best conformation according to MMFF.

This molecule has one unpaired electron. If you sketched it (which requires that you explicitly specify the radical center) the number of unpaired electrons will be correct. However, if you built the molecule, you will need to 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 SSPD.) Wait a few minutes for the calculations to complete before proceeding.

3. Select **Surfaces** from the **Setup** or **Display** menu (). Click on

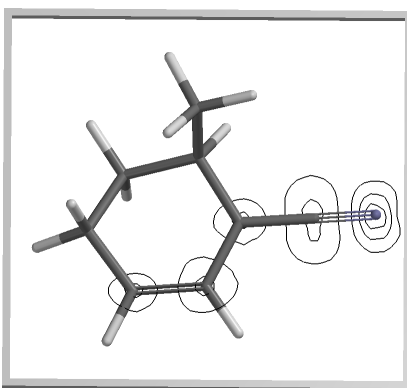
**Add** and select **spin density** from the menu. *Click* on **Add** and select **spin density map** from the menu. You have requested two different representations of spin distribution. The first presents spin density as a surface of constant value, while the second uses color to map the value of the 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  $\alpha$  spin that is, the orbital that contains the unpaired electron) from the menu.

4. The three graphics you requested will run automatically 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.



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 *press* both the **Shift** key and right button and move the mouse up and down (or use the scroll wheel). This will 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.



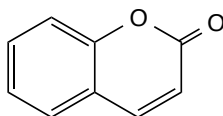
Tap on the graphic to select the graphic and on the frame to select the frame. Use one finger to rotate and two fingers to translate the graphic or frame. Pinch two fingers to zoom the graphic or frame.

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









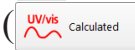
---

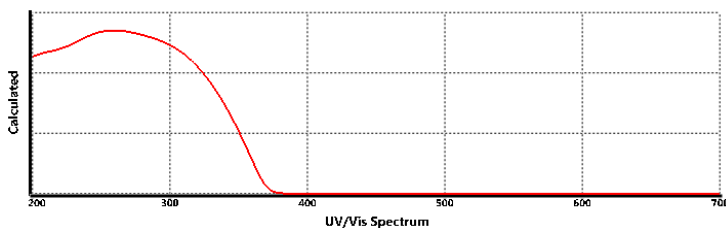
\* You can change the display style from **Contours** to **Solid** or **Transparent** using the **Style** menu at the bottom right of the screen. This will appear only when the slice is selected.



## Coumarin



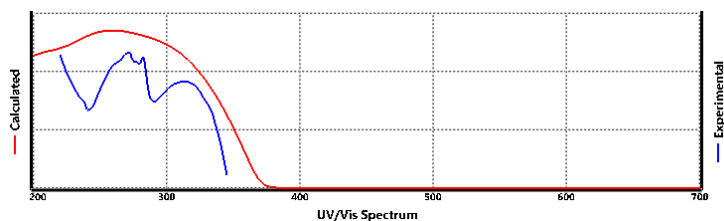
We conclude this chapter with coumarin to illustrate the calculation of a UV/visible spectrum. We will start from the equilibrium structure from SSPD.

1. Build coumarin () , minimize () and exit the builder () . Alternatively, sketch coumarin () and exit the sketcher () .
2. The name ***coumarin*** appears at the bottom of the screen. *Click* on it and then *click* on **Replace** inside the dialog that appears.
3. Select **Calculations...** from the **Setup** menu () . Change **Equilibrium Geometry** from the top menu to the right of **Calculate** to **Energy**. 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 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 public NIST database will be superimposed onto the calculated spectrum.



6. Close *coumarin*.

# Section III

## Advanced Tutorials

---

The tutorials in the chapters in this section build upon and extend the tutorials provided in the previous section. **Groups of Organic Molecules** introduces spreadsheets and associated statistics and plot capabilities. **Spectra of Organic Molecules** provides “real” examples of calculated IR and NMR Spectra. **Organic Reactions** describes and illustrates the procedures involved in finding and verifying a reaction transition state. **Medicinal Chemistry** illustrates examples of the interaction 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. **Inorganic and Organometallic Molecules** illustrates applications to “non-organic” molecules, in particular, molecules incorporating transition metals.

The tutorials in the last chapter “**Dry Labs**”. **Using the Spartan Spectra and Properties Database** further illustrate a variety of important features. They differ from earlier tutorials in that they do not require calculations.

Rough estimates of computer time are provided for each tutorial, assuming a single processor (“core”) on a current-generation Intel or AMD chip. In most cases, times will be reduced by nearly a factor of two by parallel processing using two cores and by nearly a factor of four by parallel processing using four cores. Parallel processing capability is available only with the *Spartan’14 Parallel Suite*.

Icons are provided for all menu entries as they are requested and without further clarification (“if it appears at the top of the screen”). Icon displays may be customized according to the needs of an individual user (**Icons** tab under **Preferences** in the **Options** menu; **Chapter 23**).

# 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 single molecules, 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 based either on their structure or chemical functionality. Calculations may be performed either on individual molecules or, just as simply, on the complete group of molecules. The results of calculations on a group can be examined and analyzed individually or altogether to seek out trends.

Associated with a group (or an individual molecule) is a spreadsheet. This allows convenient access to virtually any calculated quantity that can be given a 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 for further analysis. The Excel worksheet can then be embedded in the **Spartan** document (see **Embedded Data** under the **File** menu; **Chapter 15**).

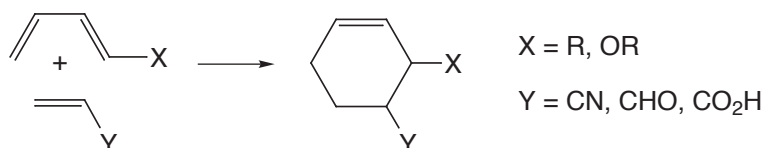
The tutorials in this chapter introduce and illustrate a number of the basic group operations available in **Spartan**. These include building a group from scratch and processing groups resulting from both

a conformational search and from 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. The use of the **Reactions** dialog is illustrated both as it operates on molecules in a group and on molecules derived from substitution of grouped molecules.

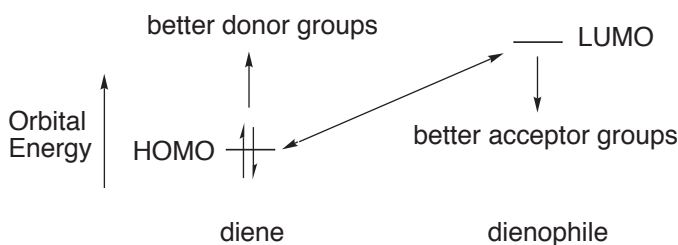
2 mins

## Dienophiles in Diels-Alder Cycloadditions










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

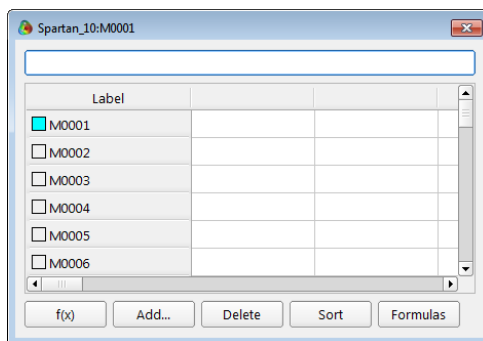




The rate of these reactions generally increases with increasing  $\pi$ -donor ability of the diene substituent, and with increasing  $\pi$ -acceptor ability of the dienophile substituent. This can be rationalized by noting that donor groups raise the energy of the highest-occupied molecular orbital (the 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 enhanced stabilization.



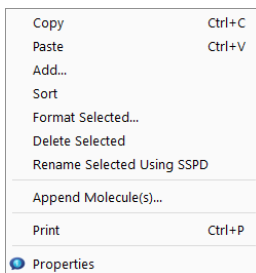
To test this hypothesis, you will examine the extent to which experimental relative rates of Diels-Alder cycloadditions involving cyclopentadiene and a variety of cyanoethylenes correlate with dienophile LUMO energies.

1. Build () and minimize () acrylonitrile,  $\text{H}_2\text{C}=\text{C}(\text{H})\text{CN}$ . 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 **Build New Molecule** (not **New Build**) from the **File** menu (). This requests that a molecule that has yet to be built is to be appended onto the end of a group of molecules in which the currently selected molecule (acrylonitrile) is a member. The screen will be cleared. *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 (**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 are all done (six molecules in total), *click* on  to remove the model kit.
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 ().

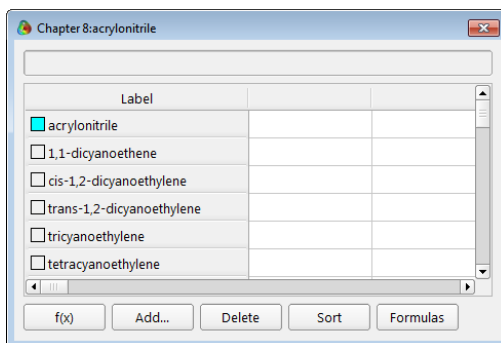



To select an individual molecule in the group, *click* on its label (**M0001**, ...) in the left hand column, or use the  and  keys at the bottom left of the screen. *Left click* on the header cell (“**Label**”) to select the entire group, then *right click* and select

**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.\*



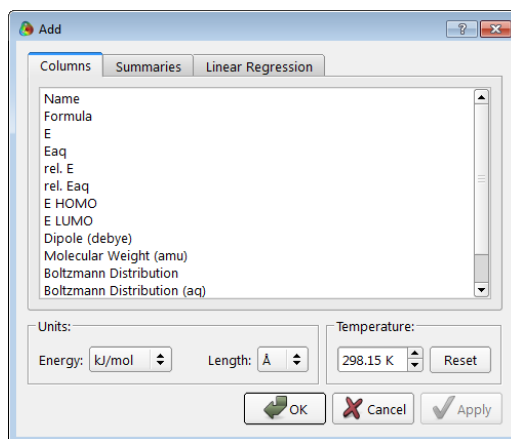
5. Select **Calculations...** from the **Setup** menu () and specify **Energy** using the **Hartree-Fock 3-21G** model in **Vacuum**, and **AM1** from the **Start from** menu. A 3-21G energy calculation will be preceded by calculation of geometry using the AM1 model. Make certain that **Global Calculations** at the bottom of the dialog is *checked* to ensure that the calculations are to be applied to all the molecules.
6. Click on the **Submit** button at the bottom of the **Calculations** dialog. Name it **Diels-Alder dienophiles**. After it completes, enter the following experimental relative rates into the spreadsheet.


\* Two databases are provided with *Spartan*, the Spartan Molecular Database (SMD) and the Spartan Spectra and Properties Database (SSPD). Selection of which occurs at the time of molecule replacement. If SMD is selected instead of SSPD, the menu will say **Rename Selected Using SMD**.

Experimental relative rates for Diels-Alder cycloadditions of cyclopentadiene*			
dienophile	log <sub>10</sub> (relative rate)	dienophile	log <sub>10</sub> (relative rate)
acrylonitrile	0	1,1-dicyanoethylene	4.64
<i>trans</i> -1,2-dicyanoethylene	1.89	tricyanoethylene	5.66
<i>cis</i> -1,2-dicyanoethylene	1.94	tetracyanoethylene	7.61
* J. Sauer, H. Weist and A. Mielert, <i>Chem. Ber.</i> , <b>97</b> , 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** key (**return** key on Mac) 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. Select **E LUMO** from the list in the dialog that results and **eV** from the **Energy** menu. *Click* on **OK**.



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).

Diels-Alder dienophiles:acrylonitrile

Label	Log(rate)	E LUMO (eV)
<input checked="" type="checkbox"/> acrylonitrile	0.000	2.66
<input type="checkbox"/> 1,1-dicyanoethene	4.64	1.20
<input type="checkbox"/> cis-1,2-dicyanoethylene	1.94	0.94
<input type="checkbox"/> trans-1,2-dicyanoethylene	1.89	0.86
<input type="checkbox"/> tricyanoethylene	5.66	-0.35
<input type="checkbox"/> tetracyanoethylene	7.61	-1.48

f(x) Add... Delete Sort Formulas

8. Select **Plots...** from the **Display** menu () . This leads to an empty plot pane at the right of the screen. Click on  in the bar at the 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.

X Axis:

- Molecule
- Log (rate)
- E LUMO (eV)

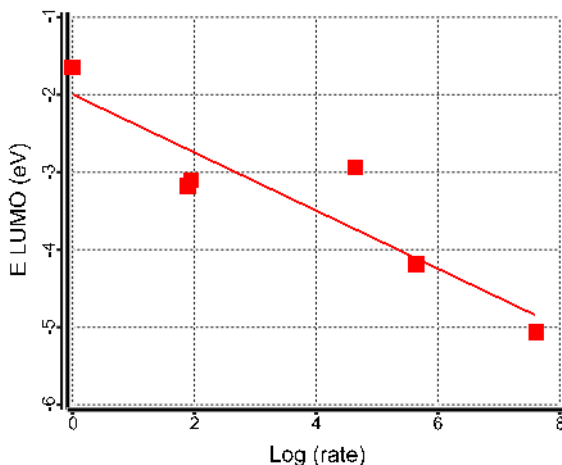
Y Axes:

- Log (rate)
- E LUMO (eV)

☐ Properties

Cancel Add

Click on **Add**.

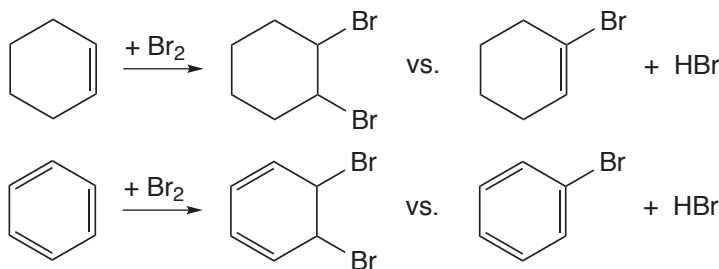


- Results for all six dienophiles calculated using the EDF2/6-31G\* model are available in the Spartan Spectra and Properties Database (SSPD). *Click* on the name of whichever molecule is selected at the bottom of the screen. *Click* on **Replace** at the bottom of the dialog that results and then *click* on **All**. The LUMO energies you have just calculated will be replaced by values from SSPD, and the plot will be updated. Is the correlation between reaction rates and LUMO energies better or worse than that previously noted?
- Close *Diels-Alder dienophiles* and any open dialogs.

no calculations







## Addition vs. Substitution

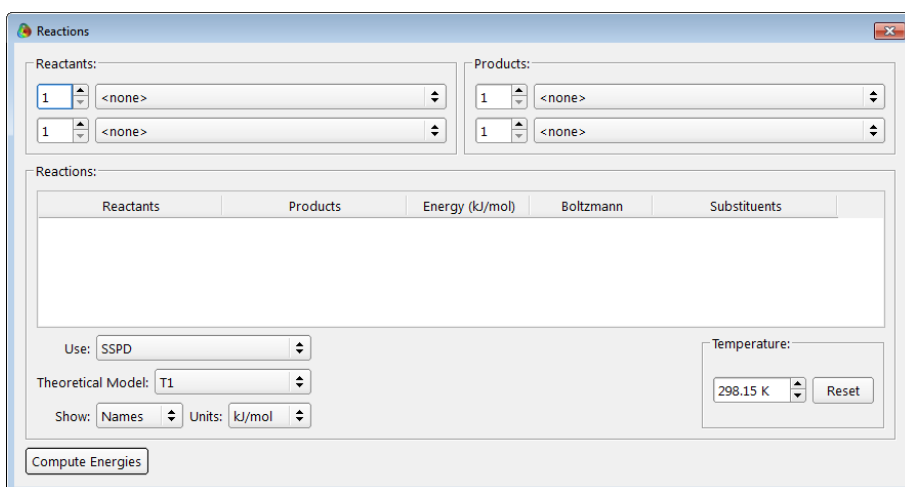
Alkenes normally undergo addition reactions whereas aromatic compounds normally undergo substitution reactions. For example, bromine reacts with cyclohexene to give *trans*-1,2-dibromocyclohexane (the addition product) not 1-bromocyclohexene, whereas it reacts with benzene to give bromobenzene (the substitution 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, heats of formation from the T1 thermochemical recipe will be employed. No calculations are involved.

- One after another, build or sketch cyclohexene, *trans*-1,2-dibromocyclohexane, 1-bromocyclohexene, benzene, *trans*-

- 5,6-dibromo-1,3-cyclohexadiene, bromobenzene, bromine ( $\text{Br}_2$ ) and hydrogen bromide (eight molecules in total). Put all in the same document. Use **New Build** () or **New Sketch** () for the first molecule and **Build New Molecule** () or **Sketch New Molecule** () for 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 **Use** and **T1** from the menu to the right of **Theoretical Model** at the bottom of the dialog.\* Compute the energy for  $\text{Br}_2$  addition to cyclohexene: select *cyclohexene* and *bromine* as **Reactants** and *trans-1,2-dibromocyclohexane* (and *<none>*) as **Products** and *click* on the **Compute Energies** button at the bottom left of the dialog. Repeat for the corresponding substitution energy (same reactants but the products are 1-bromocyclohexene and hydrogen bromide) and for both addition and substitution

\* The T1 model was developed to closely reproduce heats of formation obtained from the G3(MP2) model, which in turn closely reproduces experimental heats. The T1 heat of formation is provided as a property for the molecule in SSPD.

reactions of benzene (reactants are *benzene* and *bromine* and products are *trans-5,6-dibromomo-1-3-cyclohexadiene* and *<none>* for addition and *bromobenzene* and *hydrogen bromide* for substitution).

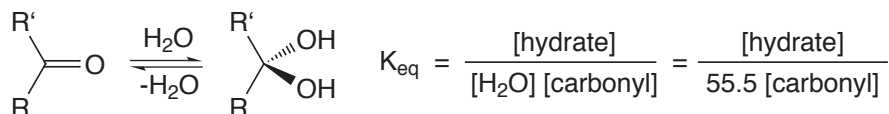
Are all reactions thermodynamically favorable (*exothermic*)? Identify any reactions that are not and provide a rationale as to why they are not. Why is there a change in preferred reaction in moving from the alkene to the arene?

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.

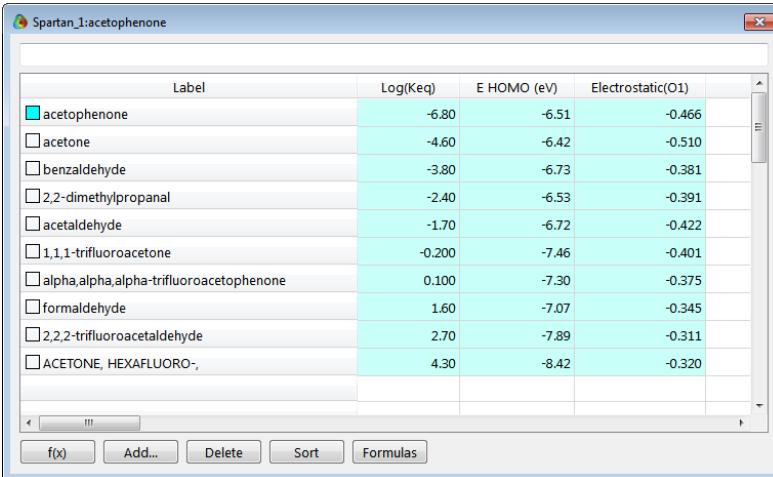
1. Build or sketch all the compounds listed in the next page. If you build, start with **New Build** from the **File** menu () for the first molecule and **Build New Molecule** from the **File** menu () for each successive molecule. If you sketch, start with **New Sketch** () for the first molecule and **Sketch New Molecule** () for each successive molecule, both from the **File** menu.
2. Click on the name of whichever molecule is selected at the bottom of the screen. Click on **Replace** in the dialog that results and then on **All**. Data from the Spartan Spectra and Properties Database (SSPD) will be used.

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 <sub>3</sub>	-6.8	CF <sub>3</sub> COCH <sub>3</sub>	-0.2
CH <sub>3</sub> COCH <sub>3</sub>	-4.6	PhCOCF <sub>3</sub>	0.1
PhCHO	-3.8	H <sub>2</sub> CO	1.6
<i>t</i> -BuCHO	-2.4	CF <sub>3</sub> CHO	2.7
CH <sub>3</sub> CHO	-1.7	CF <sub>3</sub> COCF <sub>3</sub>	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 **LogKeq** and then *click* on **Sort** at the bottom of the spreadsheet.

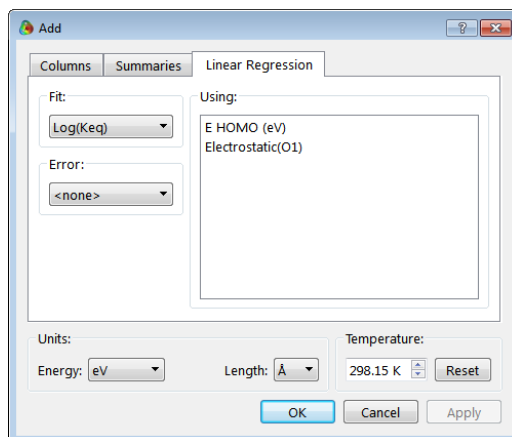


Label	Log(Keq)	E HOMO (eV)	Electrostatic(O1)
<input checked="" type="checkbox"/> acetophenone	-6.80	-6.51	-0.466
<input type="checkbox"/> acetone	-4.60	-6.42	-0.510
<input type="checkbox"/> benzaldehyde	-3.80	-6.73	-0.381
<input type="checkbox"/> 2,2-dimethylpropanal	-2.40	-6.53	-0.391
<input type="checkbox"/> acetaldehyde	-1.70	-6.72	-0.422
<input type="checkbox"/> 1,1,1-trifluoroacetone	-0.200	-7.46	-0.401
<input type="checkbox"/> alpha,alpha,alpha-trifluoroacetophenone	0.100	-7.30	-0.375
<input type="checkbox"/> formaldehyde	1.60	-7.07	-0.345
<input type="checkbox"/> 2,2,2-trifluoroacetaldehyde	2.70	-7.89	-0.311
<input type="checkbox"/> ACETONE, HEXAFLUORO-	4.30	-8.42	-0.320

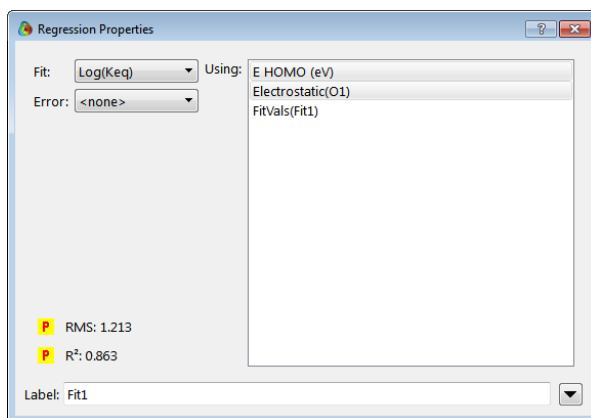
4. Select **Properties** from the **Display** menu (  ) to bring up the **Molecule Properties** dialog. *Click* on **P** to the left **E HOMO**.
5. *Click* on the oxygen atom for whatever compound is displayed. *Click* on **P** to the left of **Electrostatic** (under **Charges**) in the **Atom Properties** 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. This brings up the **Add Linear Regression** dialog.



Select **Log(Keq)** from the **Fit** menu and **EHOMO (eV)** and **Electrostatic (Ox)\*** from the **Using** list. Click on **OK**. A new row (**Fit1**) will be added at the bottom of the spreadsheet. With a **Properties** dialog on screen, click anywhere on the 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

---

\* Numbering depends on the order that atoms were introduced during building, but the reference is to the carbonyl oxygen.

better the fit.

7. To see the correlation between measured and fit rates, 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. Close the document and any open dialogs.

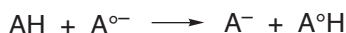
no calculations

## Acidities of Carboxylic Acids

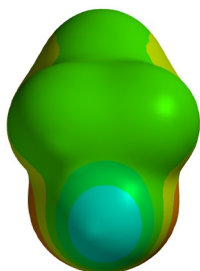
Acid strength is among the most important molecular properties. It is readily available from calculation, either in terms of absolute deprotonation energy,



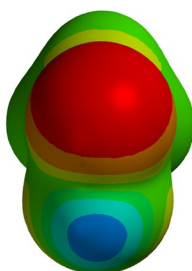
or, more commonly, as the deprotonation energy relative to that of some standard acid ( $\text{A}^\circ\text{H}$ ).



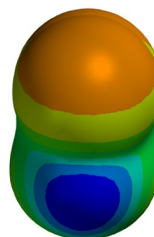
The energy calculations apply strictly to gas-phase acidities and may or may not reflect acidities in solution. This prompts a search for alternative descriptions. One possibility is the value of the electrostatic potential in the vicinity of the acidic hydrogen in the neutral acid. Electrostatic potential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



ethanol










acetic acid







nitric acid

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.

1. One after the other, build or sketch trichloroacetic, dichloroacetic, chloroacetic, formic, benzoic, acetic and pivalic acids. Put all molecules into the same document. Use **New Build** () or **New Sketch** () both from the **File** menu for the first molecule and use **Build New Molecule** () or **Sketch New Molecule** () both 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, *click* on **Replace** in the dialog that results and finally on **All**. Structures obtained from EDF2/6-31G\* 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 next available data column, *type* **pK<sub>a</sub>** and *press* the **Enter** key (**return** key on Mac). Enter the experimental pK<sub>a</sub>'s (given on the next page) into the appropriate cells under this column. You need to *press* the **Enter** (**return**) key following each entry. Alternatively, use the  key to move to the next cell.

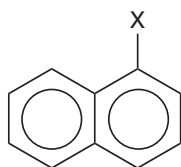
acid	pK <sub>a</sub>	acid	pK <sub>a</sub>
trichloroacetic (Cl <sub>3</sub> CCO <sub>2</sub> H)	0.7	benzoic (C <sub>6</sub> H <sub>5</sub> CO <sub>2</sub> H)	4.19
dichloroacetic (Cl <sub>2</sub> CHCO <sub>2</sub> H)	1.48	acetic (CH <sub>3</sub> CO <sub>2</sub> H)	4.75
chloroacetic (ClCH <sub>2</sub> CO <sub>2</sub> H)	2.85	pivalic ((CH <sub>3</sub> ) <sub>3</sub> CCO <sub>2</sub> H)	5.03
formic (HCO <sub>2</sub> H)	3.75		
Experimental data from: E.P. Sargeant and B. Dempsey, <i>Ionization Constants of Organic Acids in Aqueous Solution</i> , IUPAC no. 23, Pergamon Press, 1979.			

4. Arrange the seven molecules on screen such that you can clearly see the acidic hydrogen on each. 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 (.
5. Select **Surfaces** from either the **Setup** or **Display** menus () and then 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 **P** to the right of **Max** inside the **Surface Properties** dialog. The maximum value of the electrostatic potential (corresponding to the acidic proton) will be pasted to the spreadsheet. Leave the dialog on screen.
6. Plot experimental  $\text{pK}_a$  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 **Add**. Does there appear to be a correlation between  $\text{pK}_a$  and the maximum value of the electrostatic potential?
7. Close the document and any open dialogs.

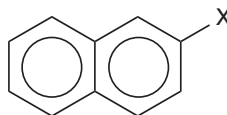
no calculations

## 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



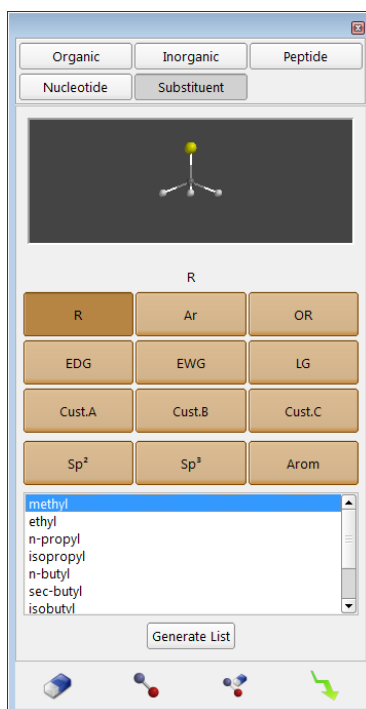
2-substituted naphthalene


In this tutorial you will use T1 results\* 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.

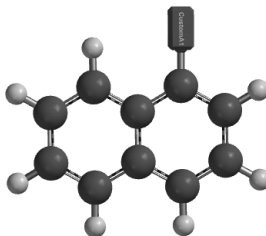
1. One after the other build methane, ammonia, water, hydrogen fluoride, hydrogen cyanide and formic acid. Put all in the same list (use **New Build** () to enter the builder for the first molecule and **Build New Molecule** () for each successive molecule). Select **Spreadsheet** from the **Display** 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 without saving.
2. Select **New Build** from the **File** menu () and build naphthalene. Then, select **Substituent** from the menu at the top of the organic model kit to bring up the substituent model kit.

---

\* The T1 model was developed to closely reproduce heats of formation obtained from the G3(MP2) model, which in turn closely reproduces experimental heats. The T1 heat of formation is provided as a property for the molecule in SSPD.



*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.

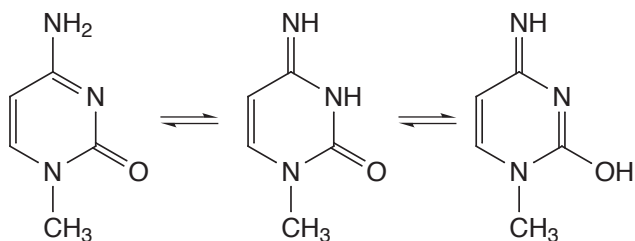


3. Select **Build New Molecule** from the **File** menu ( ). Select **Organic** from the tab at the top of the model kit and again build naphthalene. Select **Substituent** from the menu to bring the substituent model kit. Make certain that **Cust. A** is selected (if not, *click* on it). *Click* on 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 **Use** and **T1** from the menu to the right of **Theoretical Model**. Select **1-substituted naphthalenes** as **Reactants** and **2-substituted naphthalenes** as **Products**, and *click* on **Compute Energies** at the bottom of the dialog. Reactions will be written (and reaction energies computed) for all six substituents.
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. As a case in point, consider the nucleotide base cytosine (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 or alternatively one of its tautomers. If you build, minimize  and then move to view mode . If you sketch, you simply need to move to view mode .
2. Note that the word **Tautomer** appears at the bottom right of the screen, indicating that tautomers exist. Select **Identify Tautomers** from the **Search** 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 dialog that results.
3. Select **Calculations...** from the **Setup** menu () with the list of tautomers. (It is a good idea to close the molecule you used to get the tautomers.) Specify calculation of **Equilibrium Geometry** using the **3-21G Hartree-Fock** model. Submit the job with the name *cytosine tautomers*.
4. After the calculations have completed, select **Spreadsheet** from the **Display** menu (). *Click* on the line corresponding to the structure for 1-methylcytosine that you originally built. *Click* on **Add** at the bottom of the spreadsheet, select **rel.E** and **Boltzmann Distribution** from the available quantities and **kJ/**

---

\* 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.

**mol** from the **Energy** menu and *click* on **OK**. Are either of the alternatives close in energy to the normal form of cytosine?

To get a better estimate of relative tautomer energies, perform energy calculations using the EDF2/6-31G\* density functional model.

5. Make a copy of *cytosine tautomers*. Name it *cytosine tautomers EDF2*. Select **Calculations...** from the **Setup** menu () with this copy. Specify **Energy** from the top menu to the right of **Calculate** and **Density Functional, EDF2** and **6-31G\*** from the three bottom menus. Submit the job.
6. When the calculations have completed (several minutes), bring up the spreadsheet and again examine relative tautomer energies and Boltzmann populations. Are either of the alternative structures competitive?
7. Close all documents and any remaining open dialogs.

# Chapter 9

## Spectra of Organic Molecules

---

*This chapter illustrates applications involving the calculation of IR and proton and  $^{13}\text{C}$  NMR spectra. It also illustrates the use of calculated IR spectra to identify an experimental unknown.*

In addition to equilibrium geometries and conformations, reaction energies and conformational energy differences and diverse molecular properties, calculations are able to account for molecular spectra. These include, infrared and Raman spectra, NMR spectra and UV/visible spectra. The first two arise from the transitions between ground and excited vibrational states, the third from transitions between nuclear spin states and the fourth from transitions between ground and excited electronic states. Most commonly, experimental spectra are used to provide clues to the structure of an unknown molecule, that is, features in the spectrum are taken as evidence for features in the molecular structure. 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 a spectrum was measured. Lack of similarity suggests that the two molecules are not the same.

The first tutorial details the steps needed to calculate the infrared spectrum of methyl formate and to relate the spectrum to the underlying molecular structure. The second tutorial illustrates the use of the Spartan Infrared Database to identify a molecule based on its infrared spectrum. The remaining tutorials deal with NMR spectroscopy. The first details the steps involved in calculating and displaying a proton and COSY spectra for 1-methylindole, and the second the  $^{13}\text{C}$  spectrum of caulophylline, both of which are rigid molecules. The third tutorial considers the  $^{13}\text{C}$  NMR of *cis*-1,2-dimethylcyclohexane, a molecule that can flex between two equivalent conformers, while the fourth tutorial examines the dependence of carbon chemical shifts on stereochemistry.

## 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). This is in the infrared region of the electromagnetic spectrum. Frequency increases with increasing force constant or stiffness of the bond and decreases with increasing masses of the atoms involved in the bond.

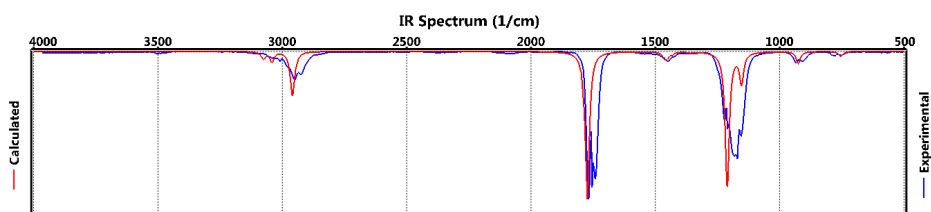
$$\text{frequency} \propto \sqrt{\frac{\text{force constant}}{\text{mass}}}$$




In order to generalize this expression to a polyatomic molecule, it is necessary to change into a coordinate system that leads to a diagonal matrix of second energy derivatives (so-called ***normal coordinates***). The normal coordinates for a polyatomic molecule will typically involve the motions of several (and perhaps all) of the atoms.

The degree (or intensity) of absorption of infrared radiation by a diatomic molecule is proportional to the change in the dipole moment with change in bond length. Note that there is no change in the dipole moment of a homonuclear diatomic molecule with change in bond length, meaning that the intensity of absorption is zero. Homonuclear diatomic molecules such as N<sub>2</sub> and O<sub>2</sub> 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. Note that some of the normal coordinates in a polyatomic molecule may not lead to a change in dipole moment, for example, the symmetric stretch in carbon dioxide where both CO bonds are simultaneously moved.

This tutorial illustrates the steps required to calculate and display an infrared spectrum and to compare it with an experimental spectrum. You will see how the ease or difficulty of molecular motion and atomic masses affect frequency.

1. Build methyl formate and *click* on . Select **Calculations...** from the **Setup** menu () and request an equilibrium geometry using the EDF2/6-31G\* density functional model. *Check IR* to the right of **Compute** and *click* on **Submit**. Accept the name.
2. After the calculation has completed (several minutes), select **Spectra** from the **Display** menu (). *Click* on the  in the bar at the top of the spectra pane and select  Calculated from the palette of icons. *Click* on the  again and select  Experimental from the icon palette. Calculated (in red) and experimental (in blue) infrared spectra are now superimposed.



3. Move the mouse while holding down the right button to shift the cursor across the spectrum. Position it over the intense line in the (calculated) spectrum at  $1774\text{ cm}^{-1}$ . Note that the molecular model (on screen above the spectra pane) vibrates. Describe the motion. Position it over the intense line at  $1211\text{ cm}^{-1}$  and describe its motion.
4. Select **Save As** from the **File** menu () to make a copy of methyl formate; name it *methyl formate d3*. 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 **Standard** 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 (). Compare the frequencies of the undeuterated and deuterated forms of methyl formate, and identify which change the most and which change the least. Explain your results.
5. Close all documents and any open dialogs.



## Searching *Spartan*'s Infrared Spectral Database

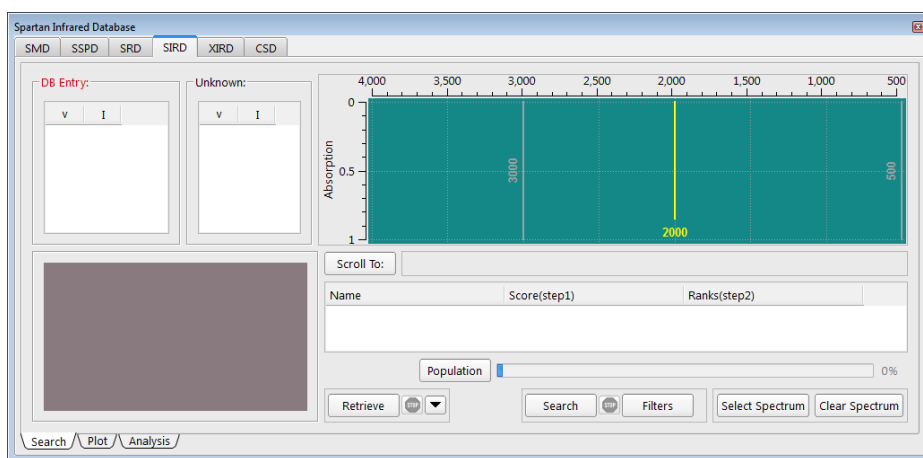
Pattern matching (“fingerprinting”) a measured infrared spectrum to experimental spectra contained in a library (database) has been common practice for many years, but has not routinely been extended to comparisons of measured spectra to databases of calculated spectra. One reason is that the results of calculations (a set of vibrational frequencies and intensities to go along with these frequencies) do not look like experimental infrared spectra, at least, they do not look like 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 (one value for all peaks) makes the two spectra visually quite similar. This parameter loosely corresponds to temperature, consistent with the fact that an infrared spectrum measured at low temperature comprises a series of sharp lines (rather than bands).

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, the curvature of the surface. Practical frequency calculations replace the real potential at the energy minimum by a quadratic function. This leads to a surface that is too steep (and an energy second derivative that is too large), 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.

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

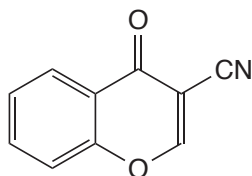
This tutorial illustrates the way in which searches of the Spartan Infrared Database (SIRD) are setup and performed and the results examined. You may choose from a small selection of measured infrared spectra (or supply your own spectra as a .dx file).

1. You need to have a *Spartan* document open in order to access the database even for a search that does not involve a structure. Select **New Build** 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.

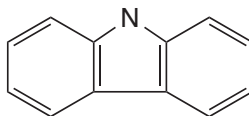


Click on **Select Spectrum** at the bottom right of the dialog. Navigate to the *spectra of organic molecules* subdirectory under the *tutorials* directory.\*

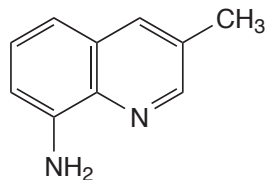
Select one of the following molecules and click on **Open**.



chromone-3-carbonitrile



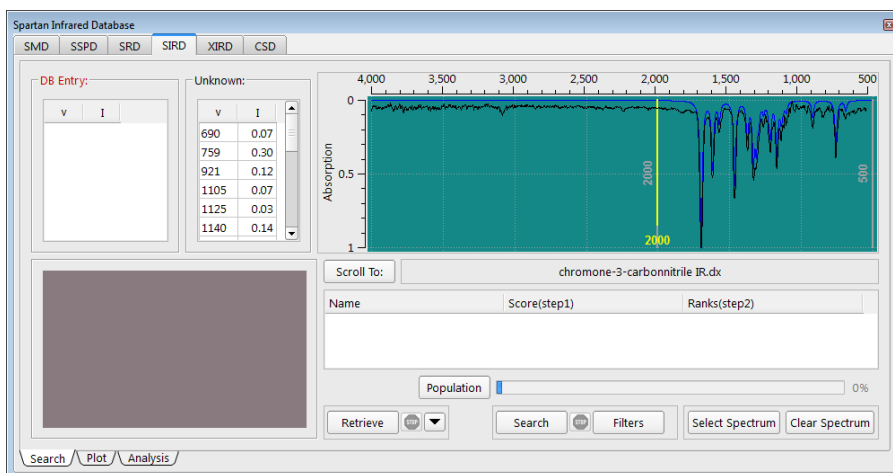
dibenzopyrrole



8-amino-2-methylquinoline

\* For Windows, this directory is found in *Program Files/Wavefunction/Spartan14* or *Program Files(x86)/Wavefunction/Spartan14*. 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 *Spartan14* 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 of the dialog 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.

10 mins

## Proton NMR Spectrum of 1-Methylindole

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. <sup>13</sup>C NMR is now more dominant, but proton NMR remains an essential tool in the chemist's arsenal.

NMR is based on 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 ( $\Delta E$ ) depends on the nucleus and on the strength of the applied magnetic field, by way of a simple relationship.

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

$\gamma$  is the gyromagnetic ratio (a constant for a given type of nucleus),  $\hbar$  is Planck's constant divided by  $2\pi$  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

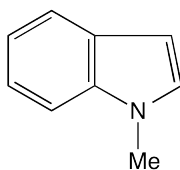
---

\* This will not always be the case. The obvious exception occurs where the query is not in the database. Problems may also occur when the (query) infrared spectrum lacks sufficient structure to distinguish it from similar molecules.








second magnetic field. The absorption of energy as a function of field strength (a *resonance*) between the states can then be detected.

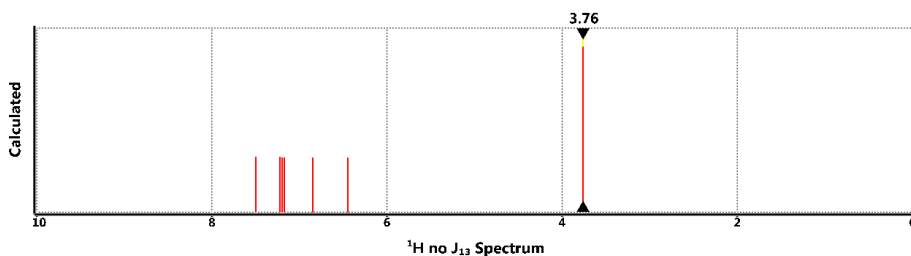
The key to the utility of the magnetic resonance experiment is that the energy at which a nucleus resonates depends on its location in the molecule, and is different for each (chemically) distinct nucleus. The reason is that the applied magnetic field is weakened by electrons around the nucleus. Nuclei 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  $^{13}\text{C}$  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 nuclei add and subtract to the external magnetic field. This leads to a splitting of lines, the splitting pattern depending on the number of neighboring protons and their geometry. 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.





In this tutorial, you will calculate the proton NMR spectrum of 1-methylindole and compare it with the experimental proton spectrum in the absence of three-bond HH coupling. You will also make a COSY plot which identifies pairs of protons that are separated by three bonds. This requires HH coupling constants. These are not calculated but rather evaluated empirically based on the three-dimensional geometry of the molecule.

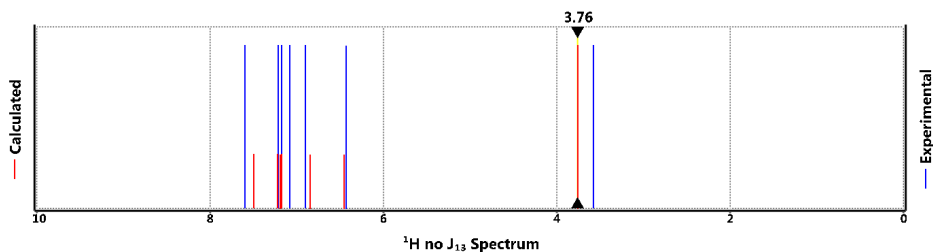
1. Build or sketch 1-methylindole. If you build, *click* on  and then on . If you sketch, *click* on . Select **Calculations...** from the **Setup** menu (). Specify calculation of equilibrium geometry using the EDF2/6-31G\* density functional model. *Check* **NMR** to the right of **Compute** and *click* on **Submit**. Accept the name *1-methylindole*. The calculation will require several 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 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 proton shift appears at the top of the spectrum.

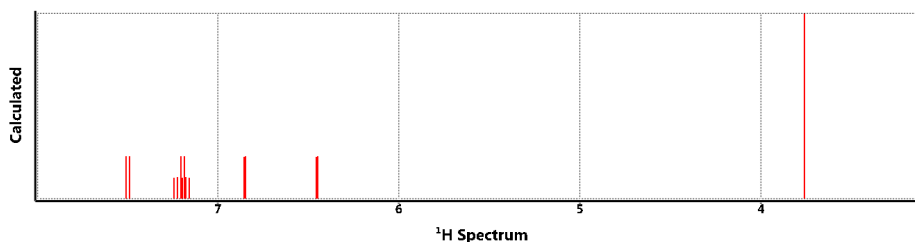
3. *Click* again on  and this time select  (experimental proton NMR spectrum with HH coupling constants set to 0). The experimental spectrum will be retrieved from the public NMR database and displayed on top of the calculated spectrum. Proton-proton coupling has been eliminated (all coupling constants are assumed to be 0).

\* EDF2/6-31G\* is the theoretical model that is supported in the Spartan Spectra and Properties Database (SSPD). You can, if you like, avoid doing any calculations and simply retrieve it. In this case, *click* on the name at the bottom of the screen and *click* on **Replace** in the dialog that results.





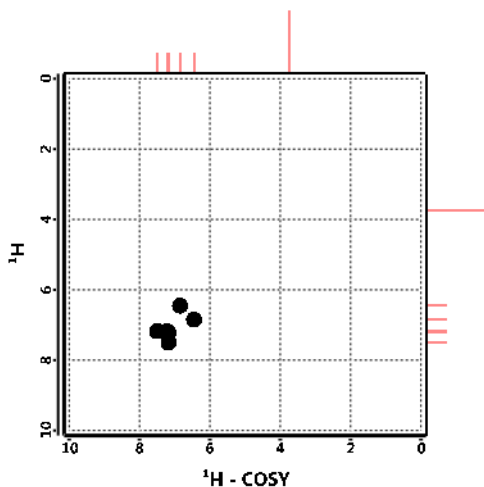
The comparison gives you an idea of the level of agreement that can be expected between calculated and experimental proton spectra.

- Click again on  and select  Calculated (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 both the right button and **Shift** key). You will see that lines due to protons at C<sub>2</sub>, C<sub>3</sub>, C<sub>4</sub> and C<sub>7</sub> are doublets, while those due to protons at C<sub>5</sub> and C<sub>6</sub> are quartets (doublet of doublets).

- Click again on  and this time select  Calculated (calculated COSY spectrum).



The plot is very small. To increase the size, position the cursor inside the bar at the top of the spectra dialog and move the mouse up while holding down the left button.

The circles indicate interactions between neighboring protons. Note that coupling constants are not obtained from quantum chemical calculations, but rather are based on geometrical relationship between protons in the calculated structure.

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

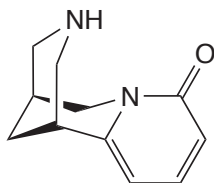
100 mins






### <sup>13</sup>C Spectrum of Caulophylline

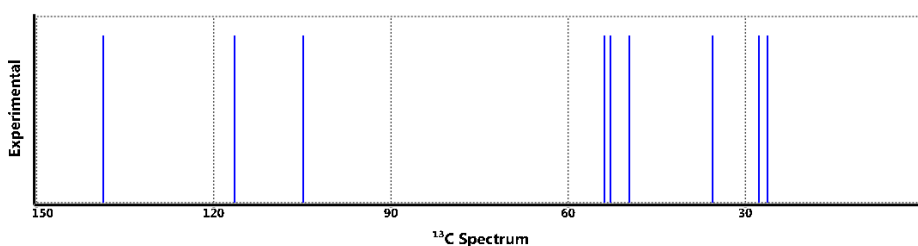
There are several reasons why NMR spectroscopy, in particular, <sup>13</sup>C NMR, is one of the most important routine analytical techniques available for characterizing organic molecules. The experiment, which is straightforward and can be carried out rapidly, requires relatively small samples and is non-destructive. The resulting (proton decoupled) spectrum is quite simple, comprising but a single line for each and every unique carbon. However, assigning <sup>13</sup>C spectra is by no means trivial, even for molecules that might appear to be quite simple. 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. Where two or more carbons in a molecule



reside in what appears to be similar environments, it may be very difficult to distinguish them.

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



1. Either build and minimize or sketch caulophylline and *click* on . Select **Calculations...** from the **Setup** menu () and specify calculation of equilibrium geometry with the EDF2/6-31G\* density functional model. *Check* **NMR** to the right of **Compute**. *Click* on **Submit** and accept the name *caulophylline*. The job will require upwards of one hour to complete\*.
2. When the calculation is done (or after you have accessed information from SSPD), select **Spectra** from the **Display** menu (). *Click* on  in the bar at the top of the spectra pane and select  (experimental  $^{13}\text{C}$  spectrum). The experimental  $^{13}\text{C}$  spectrum is drawn.



3. Use the calculated spectrum to associate the individual lines in the experimental  $^{13}\text{C}$  spectrum with specific carbons in the structure of caulophylline. *Click* on  and this time select  (calculated  $^{13}\text{C}$  spectrum). The calculated spectrum (in

\* Caulophylline is available in the Spartan Spectra and Properties Database (SSPD). If you decide to use this instead of doing the calculations, *click* on the name at the bottom of the screen and *click* on **Replace** at the bottom of the dialog that results.

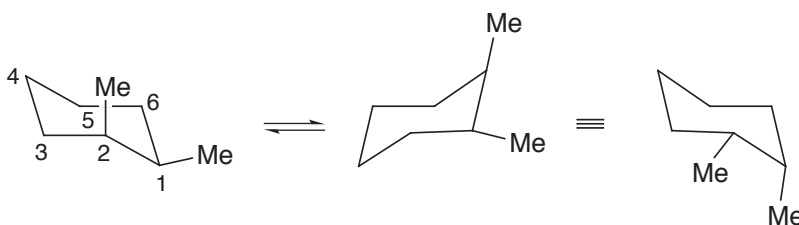
red) will be superimposed on top of the experimental spectrum (in blue). This gives an impression of the performance of the quantum chemical calculations.\* It also allows you to assign the lines in the (calculated) spectrum to individual carbons and by inference to assign the lines in the experimental spectrum. 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.

4. Close *caulophylline* and any open dialogs.

15 mins

### <sup>13</sup>C Spectrum of *cis*-1,2-Dimethylcyclohexane

At normal temperatures, the NMR spectrum of a conformationally-flexible molecule represents a (Boltzmann-weighted) average of the NMR spectra of all accessible conformers. Only when the temperature is lowered will the spectrum reveal its components, and its complexity. A simple example is *cis*-1,2-dimethylcyclohexane, a molecule with two equivalent conformers.





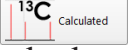


The room temperature <sup>13</sup>C NMR spectrum contains only four lines at 34.9, 31.9, 24.2 and 16.4 ppm relative to TMS, corresponding to an equal weighting of C<sub>1</sub> and C<sub>2</sub>, C<sub>3</sub> and C<sub>6</sub>, and C<sub>4</sub> and C<sub>5</sub> and the two methyl carbons, respectively. Only when the sample is cooled does the spectrum reveal eight distinct lines.

In this tutorial, you will calculate the <sup>13</sup>C NMR spectrum of *cis*-1,2-dimethylcyclohexane, and then average chemical shifts for equivalent

\* <sup>13</sup>C chemical shifts from EDF2/6-31G\* calculations have been corrected to account for local environment.

carbons to yield a room temperature spectrum.

1. Either build or sketch *cis*-1,2-dimethylcyclohexane and *click* on . Select **Calculations...** from the **Setup** menu () and specify calculation of equilibrium geometry with the EDF2/6-31G\* density functional model. *Check NMR* to the right of **Compute**. *Click* on **Submit** and accept the name *cis*-1,2-dimethylcyclohexane. The calculation will take about 10 minutes.\*
2. When the calculation is done (or after you have accessed the information from SSPD), select **Spectra** from the **Display** menu () , *click* on  in the bar at the top of the spectra pane and select  (calculated  $^{13}\text{C}$  spectrum). How many lines are in the calculated  $^{13}\text{C}$  spectrum? Average the calculated chemical shifts in order to mimic what is seen at room temperature. Are the resulting  $^{13}\text{C}$  shifts consistent with what is actually observed? Elaborate.
3. Close *cis*-1,2-dimethylcyclohexane and any open dialogs.

no calculations

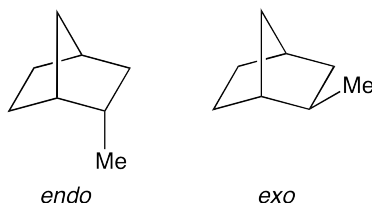
## Stereochemical Assignments from $^{13}\text{C}$ Spectra

NMR spectroscopy, in particular  $^{13}\text{C}$  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 severely limits its application. It is now practical to routinely calculate the NMR spectra of organic molecules. The availability of a “virtual NMR spectrometer” offers 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.


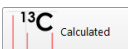


\* *cis*-1,2-dimethylcyclohexane is available in the Spartan Spectra and Properties Database (SSPD). If you decide to use this instead of doing the calculations, *click* on the name at the bottom of the screen and *click* on **Replace** at the bottom of the dialog that results.

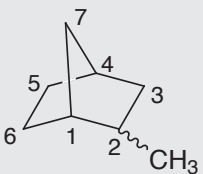
In this tutorial, you will obtain  $^{13}\text{C}$  chemical shifts for *endo* and *exo* stereoisomers of 2-methylnorbornane, and compare these to experimental shifts. The EDF2/6-31G\* density functional model will be employed. You will establish the extent to which the calculations are able to reproduce differences in chemical shifts as a function of stereochemistry.

1. Build or sketch *endo* and *exo* stereoisomers of 2-methylnorbornane.



Place both in the same document. Use **Build New Molecule** (🔍) or **Sketch New Molecule** (📐) from the **File** menu (instead of **New Build** or **New Sketch**) for the second stereoisomer.

2. Both stereoisomers are available in SSPD so there is no need to perform any calculations. *Click* on the name of whichever molecule is selected at the bottom of the screen. *Click* on **Replace** in the dialog that results and then on **All**.
3. Select **Spectra** from the **Display** menu (📊), *click* on  in the bar at the top of the spectra pane and select . You can switch between the two stereoisomers using the  and  keys at the bottom of the screen. Compare the  $^{13}\text{C}$  chemical shifts with experimental values, paying particular attention to differences between *endo* and *exo* stereoisomers.

Experimental $^{13}\text{C}$ NMR Data		<i>endo</i>	<i>exo</i>	$\Delta$
	C <sub>1</sub>	43.5	42.2	-1.3
	C <sub>2</sub>	36.8	34.6	-2.2
	C <sub>3</sub>	40.2	40.7	0.5
	C <sub>4</sub>	37.3	38.2	0.9
	C <sub>5</sub>	30.3	30.6	0.3
	C <sub>6</sub>	29.0	22.4	-6.6
	C <sub>7</sub>	35.0	38.9	3.9
	CH <sub>3</sub>	22.3	17.4	-4.9

Can the calculations be used to distinguish between the two isomers?

4. Close the document and any open dialogs.

# Chapter 10

## Organic Reactions

---

*This chapter outlines and illustrates strategies for locating and verifying transition states for organic 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. 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. However, it is often difficult to designate appropriate valence structures for transition states, let alone specify detailed geometries. While there is a complete absence of experimental data for the structures of transition states, calculated transition-state geometries are now commonplace. **Spartan** provides both a library of calculated transition-state geometries and a facility for automatically matching an entry in this library with the reaction at hand.\* This library is also available as the Spartan Reaction Database (SRD), and may 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 first four tutorials in this chapter illustrate **Spartan's** automatic procedure for guessing transition-state geometries based on its library of organic reactions. They also illustrate the use of vibrational analysis (infrared spectroscopy) to verify that a particular structure corresponds to a transition state and to show the motion connecting it

---

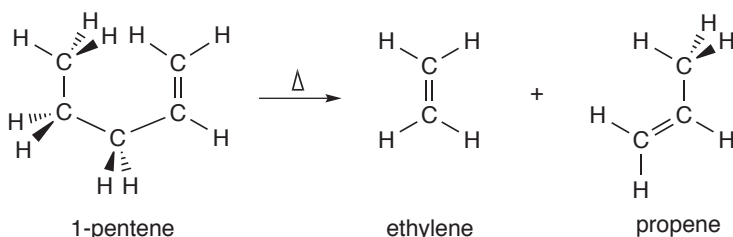
\* Where a reaction is unknown to **Spartan's** library, a fallback technique which averages reactant and product geometries (similar to the so-called linear synchronous transit method) is invoked.

to reactants and products. The fifth 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.




An example of a transition state calculation for an organometallic reaction is provided in **Chapter 13**.

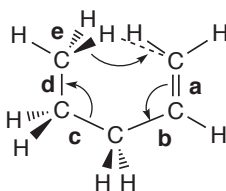
10 mins

## 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 3-21G Hartree Fock calculations, and examine the reaction coordinate for evidence of concerted motion.

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 (it will be marked 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 move the mouse up and down. Minimize and *click* on .
2. Select **Guess Transition State** from the **Search** 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. (You will need to select  to continue.) Alternatively, hold down the **Delete** key as you *click* on an arrow. 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. In this case, select **Undo** from the **Edit** menu () to return to the model with the arrows and modify accordingly.



3. Select **Calculations...** from the **Setup** menu () and specify calculation of transition-state geometry using the 3-21G Hartree-Fock model. Select **Transition State Geometry** from the top menu to the right of **Calculate**, and choose **Hartree-Fock**, **3-21G** and **Vacuum** from the three bottom menus. Finally, *check* **IR** to the right of **Compute**. This will allow you to confirm that you have found 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 () and *click* on  in the bar at the top of the spectra pane and *click* on  in the palette that results. *Click* on  at the left of the spectra pane to bring up a list of frequencies and intensities.


Frequency	Intensity
<input checked="" type="checkbox"/> i1690	0.50
<input type="checkbox"/> 185	0.85
<input type="checkbox"/> 390	5.01
<input type="checkbox"/> 436	2.84
<input type="checkbox"/> 488	3.19
<input type="checkbox"/> 504	0.46
<input type="checkbox"/> 553	10.62
<input type="checkbox"/> 633	4.34
<input type="checkbox"/> 821	6.91




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.

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.

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

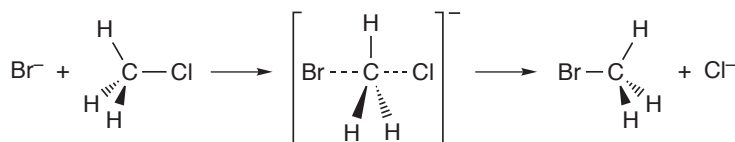
- Click on  at the left of the spectra pane. Controls in the dialog that result allow for changing both the amplitude of vibration (**Amp**) and the number of steps that make up the motion (**Steps**). The latter serves as a speed control. Change the amplitude to 0.3. Type **0.3** in the box to the right of **Amp** and press the **Enter** key (**return** key on Mac). Next, click on **Make List** at the bottom of the dialog. This will give rise to a group of structures that follow the reaction coordinate down from the transition state both toward reactant and product. You are done with *ene reaction 1-pentene*, so close it.
- Select **Calculations...** from the **Setup** menu () and specify calculation of **Energy** using the **3-21G Hartree-Fock** model (the same level of calculation used to obtain the transition state and calculate the frequencies). Make certain that **Global Calculations** is checked. Select **Surfaces** from either the **Setup**

or **Display** menu () and specify evaluation of two surfaces: a bond density surface and a bond density surface onto which the electrostatic potential has been mapped. Before you do so, make certain that **Global Surfaces** is checked. *Click* on **More Surfaces...**<sup>\*</sup>, select **density (bond)** for **Surface** and **none** for **Property** and *click* on **Apply**. Select **density (bond)** for **surface** and **potential** for **Property** and *click* on **OK**.

7. Submit for calculation<sup>\*\*</sup>. Name it *ene reaction 1-pentene sequence*. Once the job has completed, enter the **Surfaces** dialog and examine the surfaces that you have calculated. For each, 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. Also pay attention to the value of the potential on the migrating atom. This reflects its charge. Is it best described as a proton (blue), hydrogen atom (green) or hydride anion (red)?
8. Close *ene reaction 1-pentene sequence* and any open dialogs.

5mins



## S<sub>N</sub>2 Reaction of Bromide and Methyl Chloride




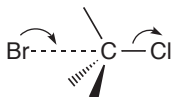
1. The S<sub>N</sub>2 reaction passes through a transition state in which carbon is in a trigonal bipyramid geometry and the entering and leaving groups are colinear. To build it, first construct methyl chloride. Then select bromine from the palette of icons in the model kit. Hold down the **Insert** key (**option** key on Mac) and *click* on screen opposite chlorine. Alternatively, *double click* on screen. Two detached fragments, methyl chloride and hydrogen


<sup>\*</sup> You need to use **More Surfaces** rather than **Add** as the electrostatic potential is mapped on the bond density surface.





<sup>\*\*</sup> 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.



bromide, appear on screen. *Click* on  or hold down the **Delete** key and then *click* on the free valence on bromine. You are left with methyl chloride and bromine atom (bromide). Manipulate the two such that bromide is poised to attack methyl chloride from the backside (as in the transition state above). (Translations and rotations normally refer to both fragments, but can be made to refer to a single fragment by first *clicking* on the fragment and then holding down on the **Ctrl** key while carrying out the manipulations.) **Do not minimize.** *Click* on .















2. Select **Guess Transition State** from the **Search** menu (). *Click* on bromide and *click* on carbon and again on bromide. A dotted line will be drawn from bromine to carbon, together with an arrow from bromine to the center of this line. Next, *click* on the CCl bond and then *click* two times (not *double click*) on the chlorine. A second arrow from the carbon-chlorine bond to the chlorine will be drawn.




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

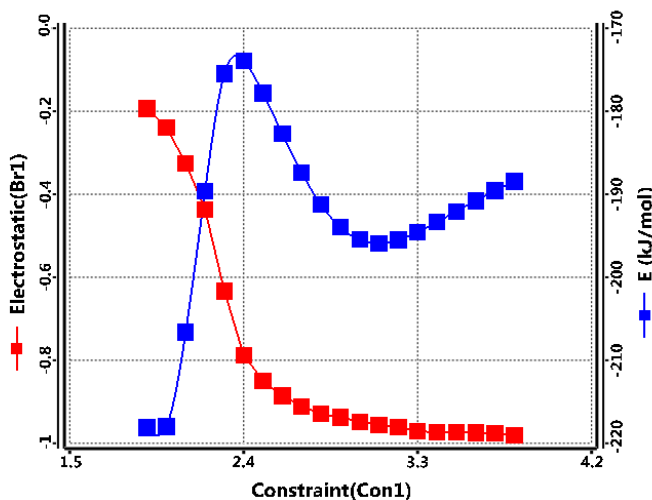
3. Select **Measure Distance** from the **Geometry** menu () and then *click* on the CBr bond. Replace the current CBr distance in the box at the bottom right of the screen by **3.8** (3.8 Å) and *press* the **Enter** key (**return** key on Mac). You have made a complex representing the reactant.
4. 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  indicating 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.

Constraint(Con1) =   3.800 Å to 1.900 Å Steps: 20 ☒ Profile


- Leave the value **3.8** (3.8Å) in the leftmost box alone, but change the number in the box to the right of **to 1.9** (1.9Å) and *press* the **Enter (return)** key. Change the number in the box to the right of **Steps** from **10** (the default) to **20**. 20 Calculations with CBr bond lengths constrained from 3.8Å (the starting point) to 1.9Å (the ending point) will be performed. The transition state should have a CBr distance in between these values. *Click* on .
5. 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. You need to change **Total Charge** to **Anion**.
  6. 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 this file. *Click* on **YES**. You can close the first document *bromide+methyl chloride*.
  7. Align the molecules which make up the sequence. *Click* on , select **Structure** from the **Align by** 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 *press* the **Align by** button at the bottom right of the screen. *Click* on .
  8. Select **Spreadsheet** from the **Display** menu (). Use the  and  keys at the bottom left of the screen to step through and *click* on **Add...** Select **E** from among the quantities listed at the top of the dialog, **kJ/mol** from the **Energy** menu, and *click* on **OK**. Next, enter the (constrained) CBr 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 spreadsheet and select **Plots** from the **Display** menu () to bring up the **Plots** dialog. *Click* on  in the bar at the top of the dialog and select **Constraint (Con1)** (the distance at

which the CBr bond has been constrained) from the **X Axis** menu, and both **E (kJ/mol)** and **Electrostatic (Br1)** from the **Y Axes** list in the dialog that results. *Click* on **Add**. By default, the data points are fit to a least squares line. This is not appropriate in this case. *Click* on  in the bar at the top of the plots plane, select **Point-to-Point** (instead of **Least Squares**) and **Electrostatic (Br1)**, then *click* the down arrow to the right of the y-axis field and select **E(kJ/mol)** and again, choose **Point-to-Point**, then *click* on **Done**.






One plot gives the energy as the reaction proceeds and the other gives charge on bromine. Are the two related? Explain.



8.  $S_N2$  reactions involving charged species normally need to be carried out in highly-polar media, for example, water. *Spartan* provides two solvation models in which the solvent is represented by a reaction field (**Appendix B**). The SM5.4 model only alters the energy and not the charges. The more recent SM8 model does affect charges but is available only for Hartree-Fock and density functional models.

Add aqueous phase data (based on the SM5.4 model) to the spreadsheet. *Click* on an empty column header, *click* on **Add...**, select **Eaq** from the list of available quantities (**kJ/mol** from the **Energy** menu), and *click* on **OK**. *Click* on  in the bar at the top of the **Plots** dialog and select **Constraint (Con1)** from the

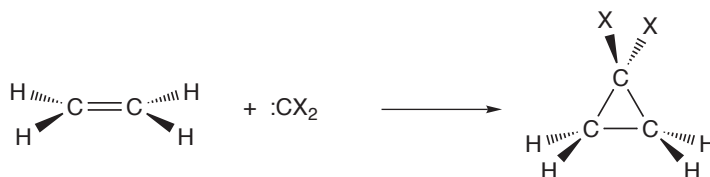
**X Axis** and **Eaq(kJ/mol)** from the **Y Axes** list. *Click* on **Add**. A second plot appears (you can go between this and the first plot using tabs at the top of the plots pane). Is the “aqueous” energy curve markedly different from the gas-phase curve? Close the **Plots** dialog when you are done.

9. Make certain that *bromide+methyl chloride.Prof.M0001* is selected. Select **Calculations...** from the **Setup** menu () , specify **Energy (Semi-Empirical** and **AM1** should be selected) and *click* on **OK**.
10. Select **Surfaces** from either the **Setup** or **Display** menu (). *Click* on **More Surfaces....** \* Select **density (bond)** from the **Surface** menu and **none** from the **Property** menu and *click* on **Apply**. Select **density (bond)** from the **Surface** menu, but this time **potential** from the **Property** menu. *Click* on **OK**. Do not close the **Surfaces** dialog.
11. Submit the job. When completed, *check* the box to the left of the bond density surface inside the **Surfaces** dialog (the first one you specified) to turn it on. *Click* on  at the bottom left of the screen to animate the display. Note, that bonds are smoothly broken and formed during the course of reaction. *Click* on  at the bottom of the screen when you are done.
12. Turn display of the bond density off (*uncheck* the box) inside the **Surfaces** dialog, and then turn display of the electrostatic potential mapped onto the bond density on. *Click* on . Relate the migration of negative charge during reaction as indicated by colors in the electrostatic potential map to the plot you constructed in step 8. Recall, that colors near red indicate maximum negative potential.
13. Close all documents and any open dialogs.

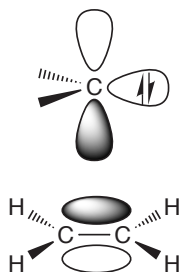
---

\* You need to use **More Surfaces** rather than **Add** as the electrostatic potential is mapped on the bond density surface.

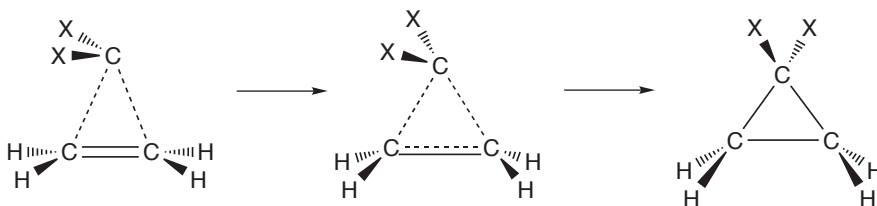
## Carbene Additions to Alkenes



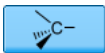
Singlet carbenes add to alkenes to yield cyclopropanes. Since a singlet carbene possesses both a high-energy occupied molecular orbital in the plane of the molecule, and a low-energy, out-of-plane unoccupied molecular orbital, this reaction presents an interesting dilemma. Clearly it would be more advantageous for the low-lying vacant orbital on the carbene, and not the high-lying filled orbital, to interact with the olefin  $\pi$  system during its approach.






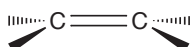
However, this leads to a product with an incorrect geometry. The carbene must twist by  $90^\circ$  during the course of reaction.




In this tutorial, you will use the Hartree-Fock 3-21G model to find the transition state and to analyze the motion of the fragments.

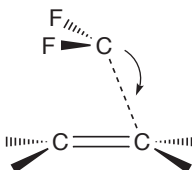
1. Build ethylene using the organic model kit.
2. Select  from the model kit. Hold down the **Insert** key (option key on Mac) and then *click* anywhere on screen. Alternatively,

*double click* on a blank area of the screen. Next, select  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.\*

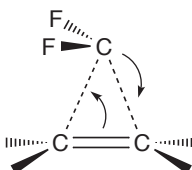


Translations and rotations normally refer to the complete set of fragments, but if you *click* on a fragment (not on a free valence) to select it, and then hold down the **Ctrl** key they will refer to an individual fragment.

3. Select **Guess Transition State** from the **Search** menu (). *Click* on the carbon on the  $CF_2$  fragment and then *click* on the  $CF_2$  carbon and on the closer of the carbons on the ethylene fragment. 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.





---

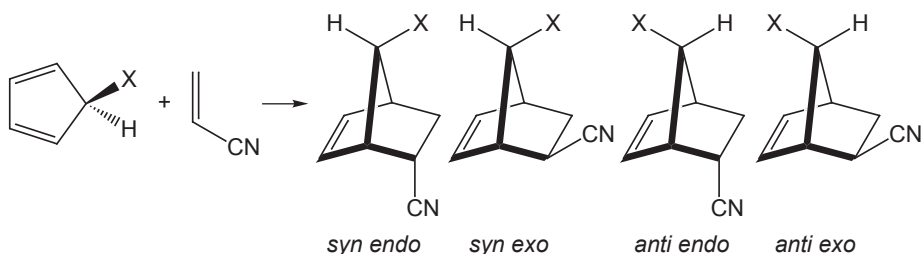
\* Proper orientation of the two fragments is not crucial in this case, but is primarily to allow you to associate the arrows with the intended reaction. Proper orientation is, however, essential where different stereochemical outcomes are possible.

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


4. Specify **Calculations...** from the **Setup** menu (). Specify calculation of a **Transition State Geometry** using the **3-21G Hartree-Fock** 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 corresponding 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  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 list of calculated frequencies and infrared intensities. Note that the frequency value at the top of the list is preceded by an “i”. This designates it as imaginary. *Click* on this frequency.

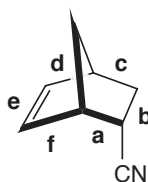
Does the animation show that the carbene reorients as it approaches the double bond? Turn the animation off by again *clicking* on the imaginary frequency.
6. Select **Properties** from the **Display** menu () and, in turn, *click* on each of the four hydrogens in the transition state. Change the value in the **Mass Number** menu in the **Atom Properties** dialog (**Properties** under the **Display** menu) from **Standard** 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. Note that they are uniformly smaller than those for the undeuterated system, and that the largest changes are for vibrational motions where hydrogens are involved.
7. Close *difluorocarbene+ethylene* and any open dialogs.


## Stereospecific Diels-Alder Reactions



Diels-Alder cycloaddition of 5-substituted cyclopentadienes with acrylonitrile can lead to four stereoproducts, 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 (consistent with sterics), 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 AM1 calculations to obtain both *syn* and *anti* transition states for *endo* addition of both 5-methylcyclopentadiene and 5-fluorocyclopentadiene and acrylonitrile, and then use Hartree-Fock 3-21G calculations to estimate relative activation energies. All four transition states are likely to be very similar to the transition state for the parent cycloaddition (cyclopentadiene and acrylonitrile). You will take advantage of this, and obtain a transition state for the parent reaction, and then to use it as a starting point for locating transition states for the substituted systems.

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










2. Select **Guess Transition State** from the **Search** 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.







*Click* on  at the bottom right to produce a guess at the transition state.

3. Select **Calculations...** from the **Setup** menu () and specify calculation of **Transition State Geometry** using the **Semi-Empirical AM1** model. *Check IR* under **Compute** to specify calculation of vibrational frequencies. This will allow you to verify that you have indeed located a transition state that is consistent with a Diels-Alder reaction. *Click* on **Submit** and name it *cyclopentadiene+acrylonitrile*.
4. When the calculation completes, examine the resulting structure and calculated frequencies. Select **Spectra** from the **Display** menu (). *Click* on the  in the bar at the top of the spectra pane and select  **Calculated**. The “infrared spectrum” calculated 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. Verify that the frequency value at the top of the list is preceded by an “i”. This designates it as imaginary. *Click* on this frequency. Does it correspond to the expected motion along the reaction coordinate? Make certain that there are no other imaginary frequencies.
5. Place the transition-state structure onto the clipboard. Right *click* on the background and select **Copy** from the resulting contextual menu. Alternatively, select **Copy** from the **Edit** menu (). Close *cyclopentadiene+ acrylonitrile*.
6. Select **New Build** from the **Build** menu () to bring up the organic model kit. *Click* on **Clipboard** (at the bottom of the

model kit) and then *click* anywhere on screen. The transition-state structure will appear. Add a fluorine to the methylene group on the cyclopentadiene fragment of the transition state either *syn* or *anti*. **Do not minimize**. Your starting structure should provide an excellent guess at the transition state for the substituted system.

7. Select **Build New Molecule** (not **New Build**) from the **File** menu (). The screen will clear. *Click* on **Clipboard** and then *click* anywhere on screen. Add fluorine to the other methylene group position. Repeat the process two more times (starting with **Build New Molecule**), to add a methyl group (sp<sup>3</sup> carbon) to both *syn* and *anti* on methylene. When you are done (four substituted transition states in total), *click* on .

Bring up the spreadsheet and replace the molecule identifiers in the leftmost column (*M0001*, etc.) but more descriptive names (*F syn*, etc.). *Click* inside each cell and type.

8. Select **Calculations...** from the **Setup** menu () and specify calculation of **Transition State Geometry** using the **Semi-Empirical AM1** model. *Check IR* to the right of **Compute**. Make certain that **Global Calculations** is checked. *Click* on **Submit** and name it **Diels-Alder stereochemistry**. When it has completed, verify that all four structures correspond to transition states, that is, have one and only one imaginary frequency.
9. Make a copy of **Diels-Alder stereochemistry** by selecting **Save As** from the **File** menu (). Name it **Diels-Alder stereochemistry Hartree-Fock**. Select **Calculations...** from the **Setup** menu () with this copy and specify a **3-21G Hartree-Fock Energy** calculation. Make certain that you remove the checkmark on **IR** and that **Global Calculations** is checked before you *click* on **Submit**.
10. When the calculation completes, select **Reactions** from the **Display** menu (). Select *syn* and *anti* F transition states for **Reactants** and **Products**, respectively, select **Current Document** from the **Use** menu, **kJ/mol** from the **Units** menu and *click* on the **Compute Energies** button at the bottom of the dialog. Use

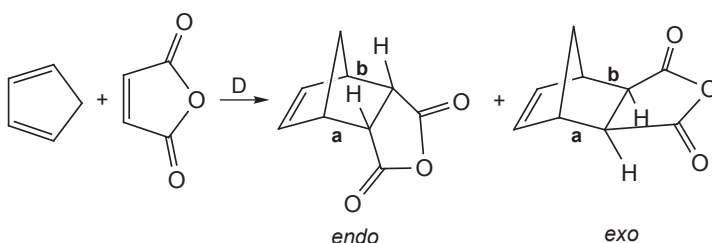
the results to identify the more stable (thermodynamic) product. Repeat for the *syn* and *anti* Me transition states.

11. 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 easily 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 3-21G Hartree-Fock 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. Click on .
2. Select **Constrain Distance** from the **Geometry** menu () and click on bond **a** in the figure above. Click on the icon  at the bottom right of the screen (it will then turn to ). Check the box to the left of **Profile** at the bottom right of the screen. This will give rise to three additional check boxes. Change the value in the leftmost box to 1.5 Å by typing **1.5** inside the box and then pressing the **Enter** key (return key on Mac). Type **2.7** (2.7 Å)

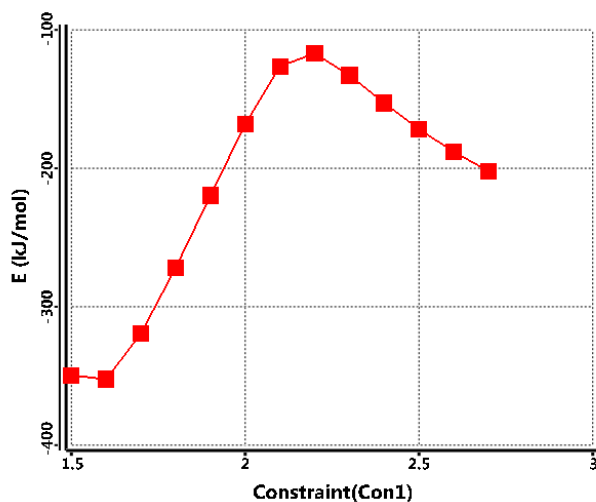
into the box to the right of **to** and *press* the **Enter (return)** key. Type **13** into the box to the right of **Steps** and *press* the **Enter (return)** key. You have specified that bond **a** will be constrained first to 1.5Å, then to 1.6Å, then to 1.7Å, etc. and finally to 2.7Å. Click on .

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, the job will give rise to a new document *cyclopentadiene+maleic anhydride endo.Prof.M0001*. This contains 13 calculations corresponding to the 13 steps that make up the energy profile. You will be prompted as to whether you want to open this file. Click on **Yes**.\*
6. Select **Spreadsheet** from the **Display** menu () and *click* on **Add...** at the bottom of the spreadsheet. Select **E** from among the entries, **kJ/mol** from the **Energy** menu, and *click* on **OK**. Next, *click* on  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 left. Click on . Select **Plots** from the **Display** menu () and *click* on  to bring up the **Add Plots** dialog. Select **Constraint (BondX)**\*\* from among the items in the **X Axis** menu and **E (kJ/mol)** from the **Y Axes** list and *click* on **Add**. By default, the data points are fit to a least squares line. This is not appropriate in this case. Click on  in the bar at the top of the plots plane, select **Point-to-Point** and *click* on **Done**.

---

\* To avoid confusion, it is a good idea to close the original file *cyclopentadiene+maleic anhydride endo*.

\*\* Bonds are numbered in the order they were formed upon initial construction of the molecule.



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* (🌤️) and make a copy (📄). Name it *cyclopentadiene +maleic anhydride endo Hartree-Fock*. Select **Calculations** from the **Setup** menu (⚙️) and specify a **Hartree-Fock 3-21G Energy** calculation in the dialog that results. *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.

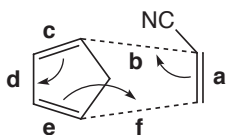
1 min


## 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 of calculations on the 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. Before you proceed with the next step, copy cyclopentadiene onto the clipboard. *Right click* and select **Copy** from the contextual menu that appears or select **Copy** from the **Edit** menu ()
2. Select **Calculations...** from the **Setup** menu () . Specify an **Equilibrium Geometry** calculation using the **Hartree-Fock 3-21G** model. Submit the job with the name *cyclopentadiene*.
3. Select **New Build** from the **File** menu () . Click on **Clipboard**. The structure of cyclopentadiene will appear at the top of the model kit. Click anywhere on screen and cyclopentadiene will appear.
4. 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. Both cyclopentadiene and ethylene will appear on screen, but 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.
5. Orient the two molecules such that they are poised for a Diels-Alder reaction leading to an *endo* product.

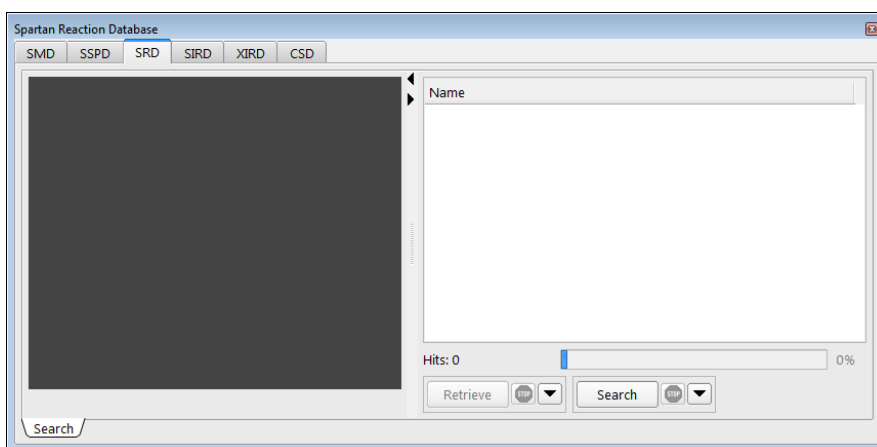



Select **Reaction Query** from the **Search** menu () .<sup>\*</sup> Click on the bond marked **a** in the above figure and then click on the two carbons that when connected will lead to the bond marked **b**. A dotted line will be drawn between these carbons and a curly

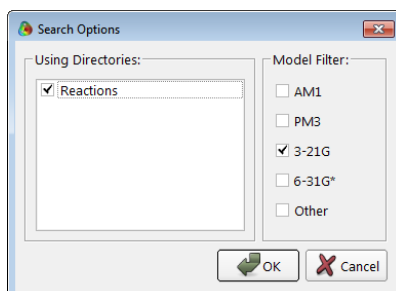
<sup>\*</sup> **Reaction Query** is equivalent to **Guess Transition State** for the purpose of defining curly arrows. It lacks the ability to automatically guess a transition state.

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, while holding down the **Shift** key, *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.

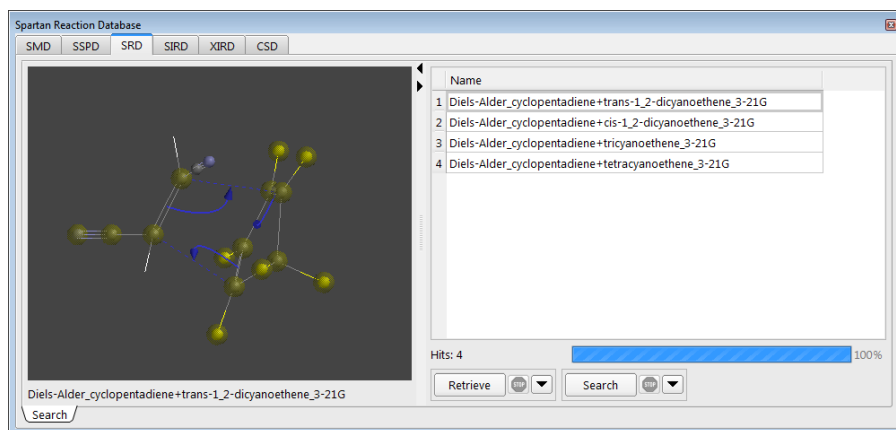



Before you begin the search, *click* on  to the right the **Search** button at the bottom of the dialog. This brings up the **Search Options** dialog.



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.



- 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. Open a spreadsheet for this document.

- Open **Diels-Alder dienophiles** that you created in an earlier tutorial\* (**Chapter 8**) and bring up a spreadsheet for this document. Select all entries in this document. *Click* inside the

\* If you have not completed this tutorial, a copy of the document **Diels-Alder dienophiles** may be found in the *organic reactions* sub-directory (*tutorials* directory). For Windows, this directory is found in **Program Files/Wavefunction/Spartan14** or **Program Files(x86)/Wavefunction/Spartan14**. 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'14** disc image.

lefthand cell of the first entry, hold down the **Shift** key and *click* on the last entry. *Right click* and select **Copy** from the contextual menu that results or select **Copy** from the **Edit** menu (). Move the cursor over the leftmost cell of the first empty column of the new (unnamed) document, *right click* and select **Paste** from the contextual menu that results. Alternatively, select **Paste** from the **Edit** menu (). The contents of *Diels-Alder dienophiles* (the dienophiles and cyclopentadiene) will be copied into the new document. Close *Diels-Alder dienophiles*. Close the spreadsheet.

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 dienophile for **Reactants** and the appropriate transition state and *none* for **Products**. Select **Current Document** under **Use** at the bottom of the dialog and *click* on **Compute Energies**. Is the ordering calculated activation energies consistent with the experimental relative rates (available in the tutorial *Diels-Alder dienophiles* in **Chapter 8**)?
11. Close all documents and any open dialogs.

# Chapter 11

## Medicinal Chemistry

---

*This chapter illustrates applications of **Spartan** to problems of relevance to medicinal chemists.*

Medicinal chemistry is a diverse field, the common thread being its concerns with the 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), not something that can be directly calculated. The polar surface area (PSA) is first examined and then refined 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 on-line Protein Data Bank (PDB). 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 into biopolymers. The focus is not the molecule itself, but rather its 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.

## Anticipating Blood-Brain Transport



To be effective, a drug must be capable of transportation to its target. This includes intestinal absorption and passage across 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  $\log (C_{\text{brain}}/C_{\text{blood}})^*$ .

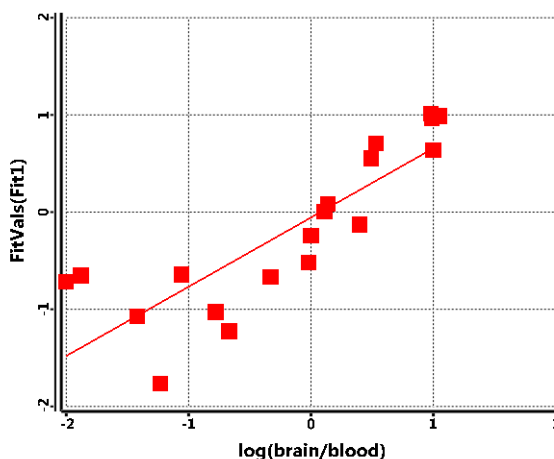
1. Open **blood brain transport** from the **medicinal chemistry** sub-directory (**tutorials** directory).\*\* This comprises a list of drugs/drug candidates for which experimental data relating to the ratio of concentrations in the brain and in the blood ( $C_{\text{brain}}/C_{\text{blood}}$ ) are 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 the  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 (most likely the **Add Columns** dialog). This leads to the **Add Linear Regression** dialog. Select **log(brain/blood)** from the **Fit** menu and *click* on **PSA** from the quantities in the box below **Using**. *Click* on **OK**. Select **Properties** from the **Display** menu () and then *click* on the cell labeled **Fit1** under **Label** in the spreadsheet. This brings up the **Regression Properties** dialog in which RMSD and  $R^2$  values are reported.

\* 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/Spartan14** or **Program Files(x86)/Wavefunction/Spartan14**. 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'14** disc image.

The smaller RMSD and the closer to unity  $R^2$ , the better the fit. Leave the **Regression Properties** dialog on screen.





4. Make a plot of  $\log(\text{brain/blood})$  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  **$\log(\text{brain/blood})$**  from the **X Axis** menu, **Fit Vals (Fit 1)** from the **Y Axis** list and click on **OK**. A least-squares line is drawn through the points.

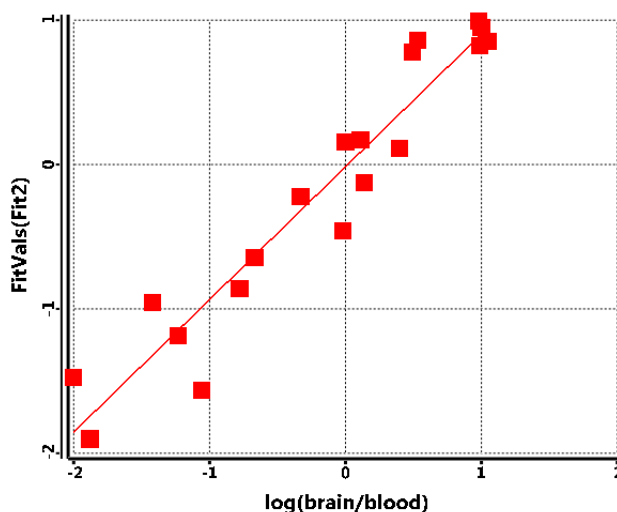


Next, you will consider an alternative definition of polar surface area based on electrostatic potential maps.

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). The overall size and shape is that of the electron density and corresponds 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  button to the left of **P-Area** inside 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 below **Using**. Click on **OK**. 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)** (instead of **Fit Vals (Fit 1)**) from the **Y Axis** list and click on **OK**. Which property provides better correlation with transport across the blood/brain barrier, PSA or polar area?



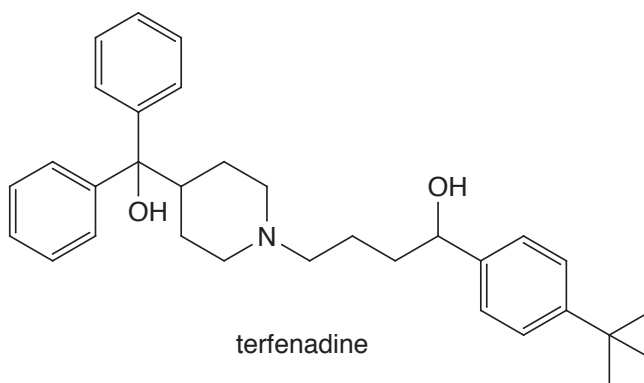
8. Close *blood brain transport* and any open dialogs.

---

\* This value may be changed in the **Settings Preferences** dialog (**Preferences** under the **Options** menu; **Chapter 23**).

## 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.




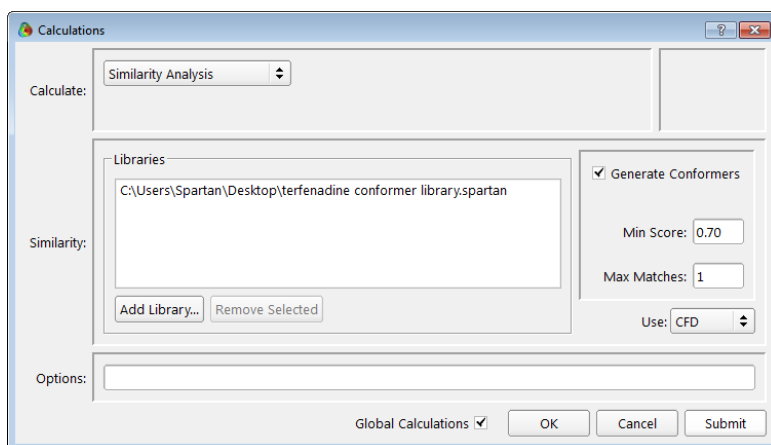
1. Open *terfenadine conformer library* from the *medicinal chemistry* sub-directory (*tutorials* directory).\*\* This contains a selection of several hundred conformers of terfenadine (the number is reported in the **Molecule Properties** dialog) obtained from a conformer library calculation. It has been supplied in order to save computer time. Close *terfenadine conformer library*.
2. Open *potassium channel blocker pharmacophore* from the *medicinal chemistry* sub-directory (*tutorials* directory). This is a five-point pharmacophore comprising four hydrophobes (blue spheres) and one positive ionizable center (red sphere).
3. Select **Set Similarity Centers** from the **Geometry** menu (⌘) and then select **CFD** from the menu at the bottom right of the

\* S. Ekins *et al*, J. Pharmacology and Experimental Therapeutics, **301**, 427 (2002).


\*\* For Windows, this directory is found in *Program Files/Wavefunction/Spartan14* or *Program Files(x86)/Wavefunction/Spartan14*. 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'14* disc image.

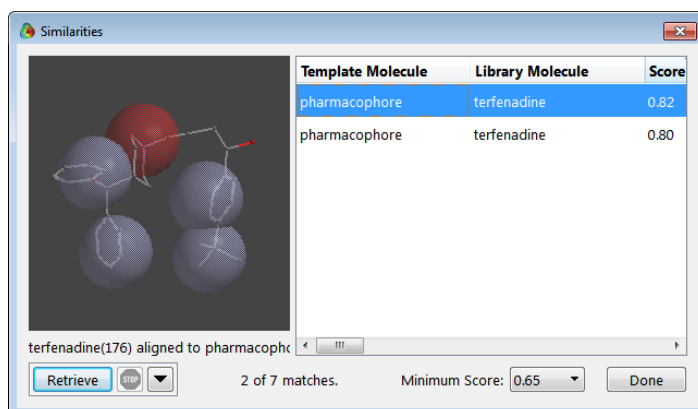
screen. In turn, *click* on each of the (five) pharmacophore elements. In response, each will be surrounded by a violet circle indicating that it is to be used in the similarity analysis.

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 form of the **Calculations** dialog.



*Click* on **Add Library** to bring up a file browser. Locate **terfenadine conformer 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, select **Similarities...** from the **Display** menu () to bring up the **Similarities** dialog.



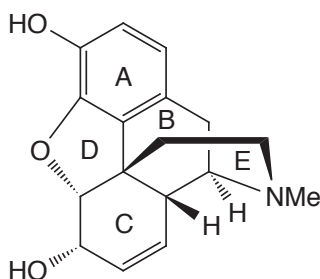
One or more hits will appear in the box at the right of the dialog, together with a similarity score (limiting on 1.0 which means a perfect match). If there is more than one hit, *click* on that with the largest score. A composite graphic of terfenadine and the pharmacophore will appear in a window at 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 the nitrogen, while the four hydrophobes are above the 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, if you wish, retrieve your 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.
9. Close all open documents and dialogs.

5 mins

## Morphine. Structure vs. Pharmacophore




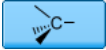




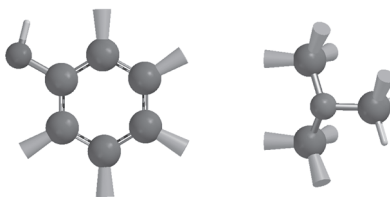
morphine

Three elements 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. Loss of any of these results in significant loss 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.

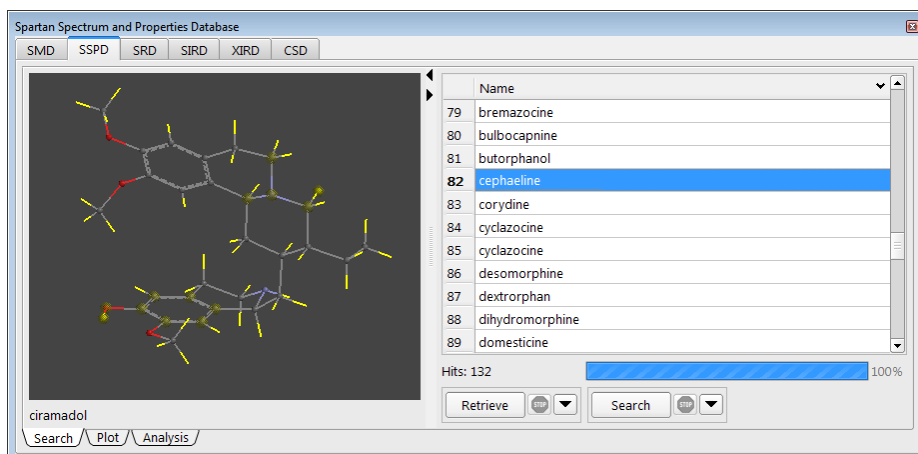
This tutorial comprises three parts. In the first part, you will identify molecules in the Spartan Spectra and Properties Database (SSPD) that incorporate the three required elements. In practice, you will look for molecules that incorporate a nitrogen center substituted by a methyl group and two other ( $sp^3$ ) carbons and a substituted phenol. You will then select a single hit and generate a conformer library for similarity analysis. (In a comprehensive investigation, you would likely use several hits or all hits; the restriction here is simply in order to save computer time.)

1. Select **New Build** 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 () , hold down the **Insert** key (**option** key on Mac) and *click* anywhere on screen. Alternatively, *double click* on a blank region on screen following selection of  $sp^3$  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.



Anything can be grown from these positions (including hydrogen) but any 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. The search will result in several hits. Find and *click* on **cephaeline** from among those hits listed inside in the scroll box at the right of the **SSPD** dialog.

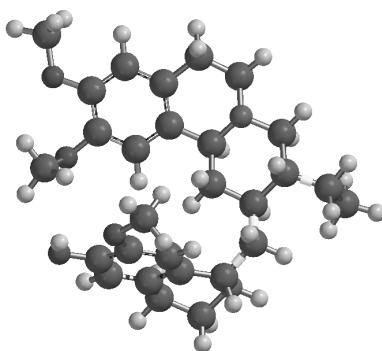


Its structure will be displayed in the window at the left of the dialog\*. You can manipulate the model in the usual way by positioning the cursor inside this window. *Click* on to the right of the **Retrieve** button at the bottom of the **SSPD** dialog to bring up the **Retrieve Options** dialog. *Check* **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.


\* *Right click* on **Name** at the top of the list to sort in alphabetical order.

Close the molecule used for the search (phenol and trimethylamine) without saving. Save the molecule you retrieved under the name *cephaeline*.

5. To shorten computer time, remove some of the degrees of conformational freedom from *cephaeline*. Select **Set Torsions** from the **Geometry** menu () . Your model will be augmented with yellow cylinders to designate flexible bonds. Remove the markers from the bonds connecting the three methoxy groups to the aromatic rings by *double clicking* on each in turn. These bonds will no longer be included in the conformational search. Your model should now appear as follows with only three single bonds selected.





If more are selected, *double click* on the appropriate torsion markers to remove them.

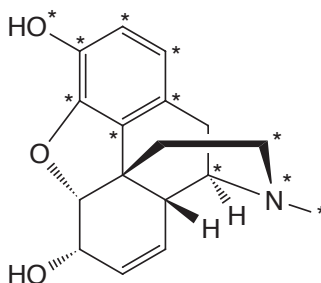
6. Select **Calculations...** from the **Setup** menu () . Select **Conformer Library** from the top left menu to the right of **Calculate**. No further information is required. *Click* on **Submit** at the bottom of the dialog. Name it *cephaeline library*. You can continue with the tutorial while you 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 the tutorial, you will identify the key structural components in morphine and perform a similarity analysis based on structure on the conformer library for cephalexin.

7. You can either sketch morphine (much easier than building it) or to save time, open it: *morphine* from the *medicinal chemistry*

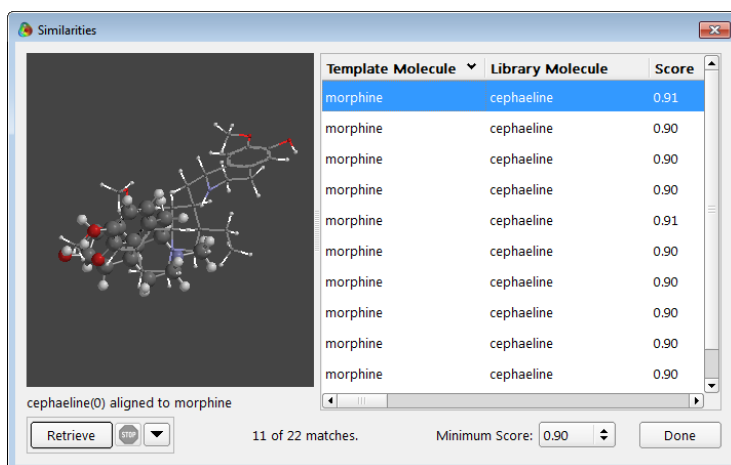
sub-directory (*tutorials* directory).<sup>\*</sup> If you sketch it, *click* on  when you are done. The name should appear at the bottom of the screen as morphine is in SSPD.

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. A red circle will be drawn around each. If you make a mistake and select the wrong atom, *click* again 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 () .

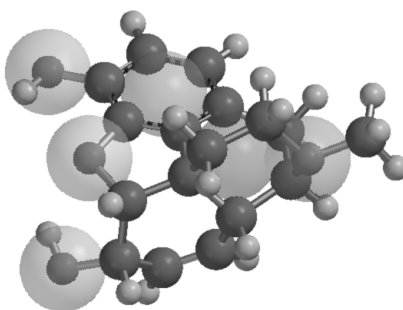
<sup>\*</sup> For Windows, this directory is found in *Program Files/Wavefunction/Spartan14* or *Program Files(x86)/Wavefunction/Spartan14*. It needs to be copied to another location available to the user prior to opening it in *Spartan*. For Linux, this is found in the directory where *Spartan* was installed. For Macintosh, this is located at the top level of the *Spartan'14* disc image.



- One or more 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 cephaline 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.



Repeat the similarity analysis using CFD's rather than structure.

- 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.



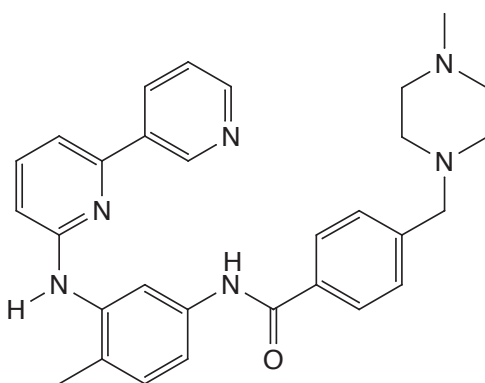
- Designate three of the CFD's as similarity centers. Select (*click* on) the CFD over the phenolic oxygen. A violet circle will surround this center indicating that it is to be used in the similarity analysis. (If you make a mistake and select the wrong CFD, *click* on the circle and it will disappear.) Select (*click* on)

the CFD at nitrogen and the CFD at the middle of the phenol ring.

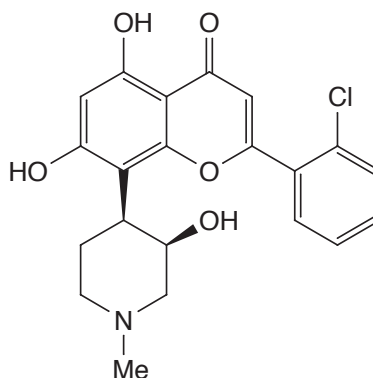
14. Select **Calculations...** from the **Setup** menu (). This should already designate **Similarity Analysis** and the appropriate library (*cephaeline*). Change the entry in the **Use** menu to **CFD**. Click on **Submit**.
15. When the analysis is complete, select **Similarities...** from the **Display** 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






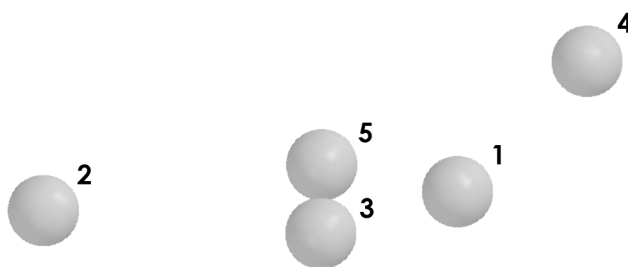
flavopiridol

In this tutorial, you will start from a protein structure from PDB in which the so-called protein kinase, gleevec, is found\*. Instead of abstracting the molecular structure of gleevec, you will abstract its “footprint”, that is, the locations from which it interacts with its 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

\* B. Nager *et. al.*, Cell, **112**, 859 (2003): PDB identification 1opj.





protein (excluded volumes), constitutes a structural pharmacophore, that is, a template for scrutiny of other possible guests. In this tutorial, you will 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 database or open **1opj** from the **medicinal chemistry** sub-directory (**tutorials** directory)\*. Select **Extract Ligands** from the **Search** menu () , *click* on the model for gleevec (it will be designated as STI1, 2, 3, or 4) inside the protein structure (displayed as a set of transparent spheres and look much like a space-filling model) to select it, and *click* on the **Exact Ligands** button. 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**.
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**.
3. The model that now appears comprises five blue spheres representing non-steric contacts between gleevec and its protein host. Turn on labels. *Click* on the **Labels** tab, *check* **CFD Labels** and *click* on **OK**.




\* For Windows, this directory is found in **Program Files/Wavefunction/Spartan14** or **Program Files(x86)/Wavefunction/Spartan14**. 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'14** disc image.

These have been assigned based on the observation that each is within a predefined distance of (one or more) complementary groups from amino acids in the protein host. Because hydrogen positions are not established (in the experimental X-ray structure) it may be impossible to say whether a particular site on the guest is acting as a (hydrogen-bond) donor or acceptor. In fact, it may not even be possible to say whether a particular site is protonated or deprotonated (although pH may dictate this). This being the case, 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 above figure for numbering.)

4. You will use only three elements of this pharmacophore (together with all 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 (). Then select **Set Similarity Centers** from the **Geometry** menu (). Make certain that **CFD** (not **Structure**) is selected in the lower right hand corner of the screen. *Click* on the pharmacophore element designated as **1** (see figure on previous page). A violet circle will surround the element designating that it is to be used in the similarity analysis and the **CFD Properties** dialog will identify the descriptor. If you make a mistake and select the wrong element, *click* on the circle and the designation will be removed. Inside the **CFD Properties** dialog, remove checkmarks from **HBD** and – **Ionizable** (you are left with **HBA** and + **Ionizable**). Repeat the process for the pharmacophore element **3**, designating it as **HBD** only and **5**, designating it as **HBA** only. When you are done, only three of the five pharmacophore elements should be marked. *Click* on .
5. 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*.

6. 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.
7. Close any open documents and dialogs.

# Chapter 12

## Flexible Molecules

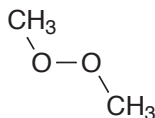
---

*This chapter describes issues associated with 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 (have only one 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 right combination of conformers can be important to properly describing molecular properties and chemical behavior.

The five tutorials in this chapter illustrate a variety of different approaches for dealing with flexible molecules. The first tutorial deals with a molecule with only one 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 deals with 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 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 the same question, but this time with regard to a proper conformation for binding to a protein.

## Internal Rotation in Dimethylperoxide



Quantum chemical calculations, in particular, Hartree-Fock molecular orbital calculations, density functional calculations and MP2 calculations, may be called on to furnish data to parameterize empirical energy functions for use in molecular mechanics and/or molecular dynamics calculations. Most important are data relating to torsional motions, for it is here that experimental data are most scarce.




Note that the 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  $360^\circ$ . One suitable choice is a truncated Fourier series.

$$E^{\text{torsion}}(\omega) = k^{\text{torsion1}} (1 - \cos(\omega - \omega^{\text{eq}})) + k^{\text{torsion2}} (1 - \cos 2(\omega - \omega^{\text{eq}})) + k^{\text{torsion3}} (1 - \cos 3(\omega - \omega^{\text{eq}}))$$






Here,  $\omega^{\text{eq}}$  is the ideal dihedral angle and  $k^{\text{torsion1}}$ ,  $k^{\text{torsion2}}$  and  $k^{\text{torsion3}}$  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 generate a potential energy function for rotation about the oxygen-oxygen bond in dimethylperoxide, first using Hartree-Fock 6-31G\* calculations, and then using EDF2/6-31G\* density functional calculations. You will fit your data to truncated Fourier series.


1. Build dimethylperoxide. If the molecule is not already in an *anti* conformation, select **Measure Dihedral** from the **Geometry** menu (🔍) and set the COOC dihedral angle to **180** ( $180^\circ$ ) by *typing 180* in the box at the lower right of the screen and *pressing* the **Enter** key (return key on Mac). ***Do not minimize.***



2. Select **Constrain Dihedral** from the **Geometry** menu (). Select the COOC torsion, 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.

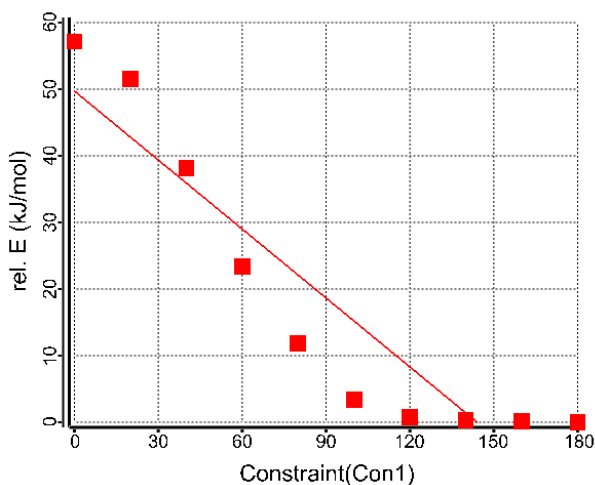
Constraint(Con1) =   180.00 ° to 0.00 ° Steps: 10 ☒ Profile


- Leave **180** ( $180^\circ$ ) in the original (leftmost) box alone, but change the contents of the box to the right of **to** to **0** ( $0^\circ$ ). You need to *press* the **Enter (return)** key after you type in the value. **Steps** should be **10**. If it is not, *type 10* and *press* the **Enter (return)** key. What you have specified is that the dihedral angle will be constrained first to  $180^\circ$ , then to  $160^\circ$ \*, etc. and finally to  $0^\circ$ . *Click* on .
4. Select **Calculations...** from the **Setup** menu () and specify **Energy Profile** from the top menu to the right of **Calculate**, and **Hartree-Fock** and **6-31G\*** from the two bottom menus. *Click* on **Submit** and accept the name *dimethylperoxide*.
  5. When the calculations have completed, they will go into a file named *dimethylperoxide.Prof.M0001*. A prompt will ask you if you want to open this file. *Click* on **OK**. 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  $180^\circ$  conformer, and the COOC dihedral angles. First *click* on the label (**M0001**) for the top entry in the spreadsheet (the  $180^\circ$  conformer), then *click* on the header cell for the leftmost blank column, and finally, *click* on **Add...** at the bottom of the spreadsheet. Select **rel. E** from the quantities in the dialog that results, **kJ/mol** from the **Energy** menu and *click* on **OK**. To enter the dihedral angle constraints, select **Constrain Dihedral** from the **Geometry** menu () , *click*

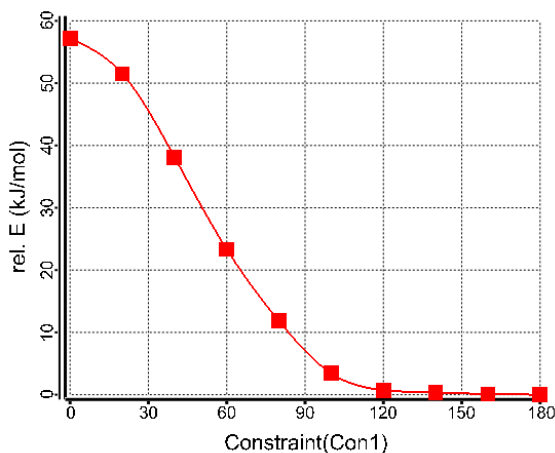
\* The difference between constraint values is given by: (final-initial)/(steps-1).

on the constraint marker attached to dimethylperoxide and *click* on **P** at the bottom of the screen (to the right of the value of the dihedral angle). *Click* on .


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 **rel. E(kJ/mol)** from the **Y Axes** list.



By default, the data points have been fit to a least square line (which is not appropriate in this particular case). To fit the points to a Fourier series, *click* on  at the top of the plot pane and select **Fourier** to the right of **Curve** in the resulting dialog.



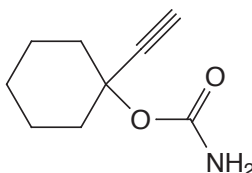
7. Use energies from EDF2/6-31G\* density functional

calculations, together with the Hartree-Fock geometries to provide a better fitting function. First, make a copy of *dimethylperoxide.Prof.M0001*. Name it *dimethylperoxide density functional*. Select **Calculations...** from the **Setup** menu () and specify calculation of **Energy** using the **EDF2/6-31G\* 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. 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** using the **MMFF** molecular mechanics model. *Click* on **Submit**, and accept the name *ethinamate*.
2. When the job completes (a few seconds), all low-energy conformers will be placed in a new document *ethinamate.Conf.M0001*. You will be asked whether you want to open this document. *Click* on **OK**. Select **Spreadsheet** from the

**Display** menu () , and size the spreadsheet such that all rows (corresponding to different conformers) are visible. *Click* on the top row of the spreadsheet (corresponding to the lowest-energy conformer according to the MMFF model). *Click* on **Add** at the bottom of the spreadsheet. Select **rel. E** and **Boltzmann Distribution** from the list at the top of the dialog that results, **kJ/mol** from the **Energy** menu at the bottom, and *click* on **OK**. Describe the lowest-energy conformer. Does the carbamate group prefer to be *equatorial* or *axial*? Does one conformer dominate the Boltzmann distribution or are two or more conformers needed to account for 90%? Have you found the expected number of conformers? If not, describe any “missing” conformers and suggest why they are missing.

3. See if the results of the molecular mechanics calculations (identity of the lowest-energy conformer and makeup of the Boltzmann distribution) maintain if you move to a better theoretical model.

Make a copy of *ethinamate.Conf.M0001* (**Save As** from the **File** menu or *click* on ). Select **Calculations...** from the **Setup** menu () with this copy and specify calculation of **Equilibrium Geometry** using the **EDF2/6-31G\* Density Functional** model. Make certain that **Global Calculations** at the bottom of the dialog is checked. (This applies the calculation model to all the molecules in the document and not just the selected molecule.) *Click* on **Submit**.

4. The calculations will require at least 15 minutes or more (depending on your computer). When completed, bring up the spreadsheet and identify the best conformer. Is it the same as that assigned from the MMFF calculations? If not, is the *equatorial* or *axial* preference for the carbamate group the same? Is the Boltzmann distribution similar in the sense that one conformer dominates or two or more conformers contribute significantly.
5. Close any open documents and dialogs.

## <sup>13</sup>C Chemical Shifts Depend on Conformation

The time for nuclear spin relaxation is very long relative to time required for equilibration among conformers, at least at normal temperatures. This means that each of the carbon chemical shifts in the NMR spectrum of a flexible molecule, <sup>13</sup>C, will be a weighted average of the shifts NMR of the individual conformers, <sup>13</sup>C<sub>i</sub>.

$$^{13}\text{C} = \sum_i \omega_i ^{13}\text{C}_i$$

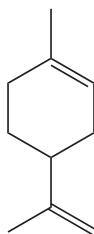
The weight,  $\omega_i$ , is given by the Boltzmann equation, and depends on its energy,  $\epsilon_i$  (relative to that of the lowest-energy conformer,  $\epsilon_0$ ) and on temperature,  $T$ . Summation is over all conformers (including the lowest-energy conformer),  $g_i$  is the number of times that conformer  $i$  appears in the overall distribution and  $k$  is the Boltzmann constant.


$$\omega_i = g_i \exp - [\epsilon_i - \epsilon_0/kT] / \sum_j \{ \exp - [\epsilon_j - \epsilon_0/kT] \}$$


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).



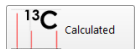




In this tutorial, you will see how well the <sup>13</sup>C spectrum of the “best” (lowest-energy) conformer of limonene reproduces the experimental room temperature NMR. More generally, you will assess the sensitivity of <sup>13</sup>C chemical shifts to change in conformation.

1. Build or sketch limonene. Click on .

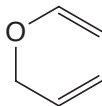


2. Select **Calculations...** from the **Setup** menu () and request a **Conformer Distribution** using the **MMFF** molecular mechanics model. Click on **Submit** and accept the name *limonene*.
3. The job will complete in a few seconds and you will be asked

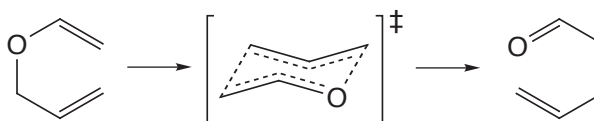
whether or not you wish to open *limonene.Conf.M0001*, the document containing the full list of conformers. *Click* on **OK**. Select **Spreadsheet** from the **Display** menu () with this list. *Click* on the top row of the spreadsheet to select the lowest energy conformer. *Click* on **Add** at the bottom of the spreadsheet, select both **rel. E** and **Boltzmann Distribution** from the list at the top of the dialog that results and **kJ/mol** from the **Energy** menu at the bottom of the dialog and *click* on **OK**.

4. Delete any conformers that are more than 10 kJ/mol higher in energy than the best conformer. You can see that these do not contribute significantly to the Boltzmann distribution. Select the conformer to be deleted and *click* on **Delete** at the bottom of the spreadsheet.
5. Select **Calculations...** from the **Setup** menu and specify an **Energy** calculation using the **EDF2/6-31G\* Density Functional** model. We are simplifying the calculations by using MMFF geometries, but several conformers are involved and it will require several minutes. *Check* **NMR** to the right of **Calculate**. *Click* on **Submit**.
6. When the calculation completes, select **Spectra** from **Display** menu () , *click* on  in the bar at the top of the spectra pane and select . *Click* again on  and this time select the calculated <sup>13</sup>C spectrum  for each of the low-energy conformers with the actual experimental spectrum. You can move from one conformer to another using the “step keys” ( and ) at the bottom left of the screen. While the calculated spectrum changes from one conformer to another, there is only one experimental spectrum. Is there significant variation as you move from one conformer to another? Does the spectrum from the lowest-energy conformer adequately reproduce the experimental spectrum?
4. Close any open documents and dialogs.







## Allyl Vinyl Ether



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 obtain relative conformer energies using single-point 6-31G\* Hartree-Fock calculations based on 3-21G Hartree-Fock equilibrium geometries.

1. Either build  and minimize  or sketch  allyl vinyl ether. Click on .
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.M0001*\*\*. A prompt will ask you if you want to open this file. Click on **OK**\*\*\*. Select **Spreadsheet** from the **Display** menu (). Size


\* By default, only conformers within 40 kJ/mol of the global minimum will be kept. This can be changed (see **Conformational Search** in **Appendix D**).



\*\* **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 21**).


\*\*\* To avoid confusion, it is a good idea to close the original file *allyl vinyl ether*.

the spreadsheet such that all rows (corresponding to different conformers) are visible at one time. *Click* on **Add...** at the bottom of the spreadsheet. Select **E** from the list at the top of the dialog which appears, **kJ/mol** from the **Energy** menu and *click* on **OK**. 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. A message will appear at the bottom left of the screen.

Select atoms.

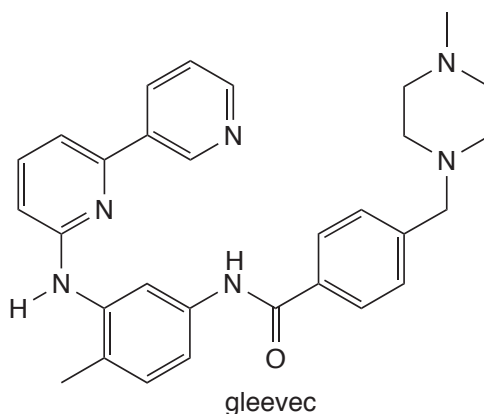
*Click* on oxygen and on the two carbons of the vinyl group bonded to oxygen to designate them as alignment centers. Each will be marked by a red circle. If you make a mistake, *click* on the circle and it will disappear. When you are done, *click* on the **Align by** button at the bottom right of the screen. *Click* on .

4. To obtain a better estimate of the energy required to twist allyl vinyl ether into a conformer suitable for Claisen rearrangement, perform 6-31G\* energy calculations using 3-21G geometries. Select **Save As** from the **File** menu () to make a copy of *allyl vinyl ether.M0001*. Name it *allyl vinyl ether Hartree-Fock*. Using the copy, delete all conformers except the global minimum and the lowest-energy conformer that is 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 calculation of an energy using the **Hartree-Fock 6-31G\*** model. Also, specify **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. What is 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 mins

### Gleevec. Protein Bound vs. Free Conformer






Gleevec (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 for gleevec docked in protein kinases have appeared in the literature and are available in the PDB. In this tutorial, you will examine one of them\* 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).

\* B. Nager *et al*, Cell, **112**, 859 (2003); PDB identification 1opj.


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  $\alpha$  helices, while blue regions correspond to parts that are  $\beta$  sheets (the remaining regions are green). Note that these assignments are taken from the PDB file. A Ramachandran plot (the  $\phi$  and  $\psi$  torsional angles connecting the individual amino acids) allows you to see the clustering of  $\alpha$  helices and  $\beta$  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.

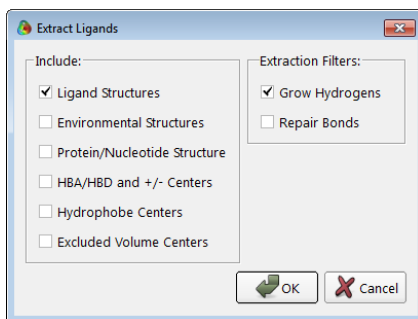
The colors on the ribbon display may be modified, by selecting **Configure...** from the **Model** menu () and *clicking* on the **Ribbons** tab. Coloring **By Residue** will give each amino acid its own unique color.

The protein you have brought in 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 **Search** 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.

---

\* For Windows, this directory is found in *Program Files/Wavefunction/Spartan14* or *Program Files(x86)/Wavefunction/Spartan14*. 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 in the *Spartan'14* disc image.




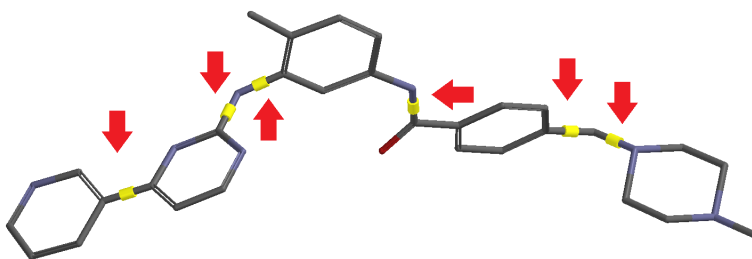
*Check **Ligand Structures** under **Include** and **Grow Hydrogens** under **Extraction Filters** inside the dialog (if other options are selected, remove them) and *click* on **OK**.*



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, select **Copy** from the **Edit** menu ( ).
4. Select **Build New Molecule** (not **New Build**) 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 and then 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 remain unchanged, 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 a starting point for a conformational search. First, remove some degrees of conformational freedom in order to shorten the computer time for the conformational search. With the second copy, select **Set Torsions** from the **Geometry** menu () . Your model will be augmented with yellow cylinders and circles to designate flexible bonds and rings, respectively. Remove the marker from the bond connecting the phenyl ring to the carbonyl end of the amide group, by *double clicking* on it. Also, remove the (two) circles on the piperazine ring by *double clicking* on each 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 **Rel. E** from the scroll box in the dialog that appears and **kJ/mol** from the **Energy** menu. 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?
9. Close *gleevec protein bound vs. free conformer* and any open dialogs.

# Chapter 13

## Inorganic and Organometallic Molecules

---

*This chapter shows how to construct inorganic and organometallic molecules using **Spartan**'s inorganic model kit.*


Many molecules are made up of a relatively few elements and obey conventional valence rules. They may be easily built using the organic model kit. However, others cannot be assembled with this model kit either because they incorporate other elements, or do not conform to normal valence rules or involve ligands. Important among these are inorganic and organometallic compounds involving transition metals. They may be constructed using the inorganic model kit.

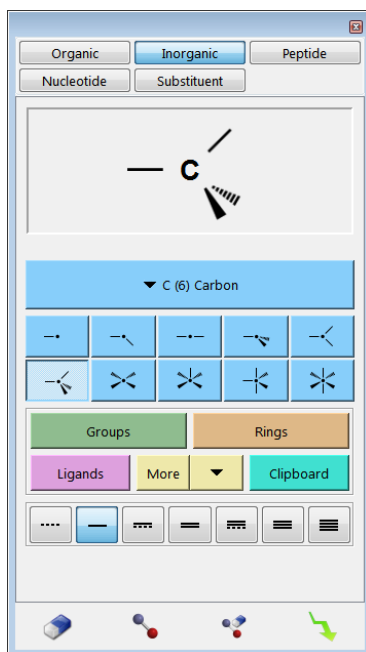
Transition-metal inorganic and organometallic compounds may also require different quantum chemical methods from those that have proven to be satisfactory for organic molecules. In particular, Hartree-Fock and MP2 models are known to produce poor results where transition metals are involved. The PM3 and PM6 semi-empirical models have been parameterized for most transition metals, and generally provide a good account of equilibrium geometries. They would not be expected to properly account for the energies of reactions involving transition metals as they generally perform poorly for reactions involving molecules without metals. Density functional models are successful for calculation of geometries of molecules incorporating both transition metals and lanthanides. It might be hoped that density functional models would also provide a solid account of thermochemistry. There are very few experimental thermochemical data with which to build a case.

## Sulfur Tetrafluoride



Sulfur tetrafluoride cannot be constructed using *Spartan's* organic model kit. 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 model kit.

1. Bring up the inorganic model kit by *clicking* on  and then selecting **Inorganic** from the menu at the top of the model kit.

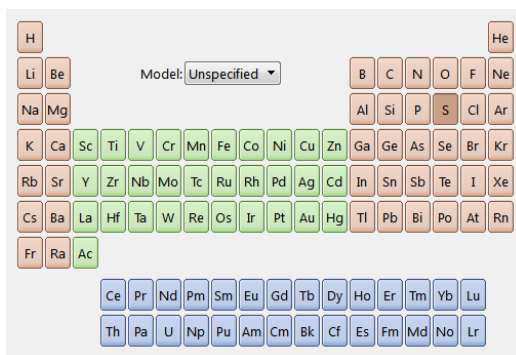



The inorganic model kit comprises an atom bar (*clicking* on which bring up the *Periodic Table*\*) followed by a selection of atomic hybrids. Buttons access menus of groups, rings and






\* 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 from the **Model** menu that appears in the center of the *Periodic Table*.

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. One after the other, *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  to remove the model kit.
6. Select **Calculations...** from the **Setup** menu (). Specify calculation of **Equilibrium Geometry**\* using the **Hartree-Fock 3-21G** model\*\* and *click on* **OK**.

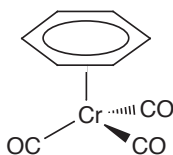
\* It should be noted that were an incorrect geometry specified at the outset, optimization would lead to the correct structure, as long as the starting geometry possessed no symmetry ( $C_1$  point group). Thus, square planar  $SF_4$  in  $D_{4h}$  symmetry would remain square planar, while an almost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

\*\* The 3-21G basis set is extended by a set of d-type valence functions for sulfur (and other second-row elements).

7. Select **Surfaces** from the **Setup** menu, *click* on **Add** and select **HOMO** from the menu that results. Leave the dialog on screen.
8. 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?
9. *Check* the box to the left of **HOMO** inside the **Surfaces** dialog. Is the highest-occupied molecular orbital in sulfur tetrafluoride consistent with the notion that sulfur is surrounded by six electron pairs? Elaborate.
10. Remove *sulfur tetrafluoride* and any open dialogs.


10 mins

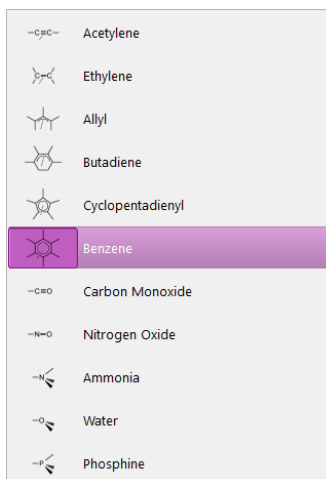
## Benzene Chromium Tricarbonyl



benzene chromium tricarbonyl

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

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 **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, 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.
4. Select **Build 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. Select **Calculations...** from the **Setup** menu (). Specify calculation of **Equilibrium Geometry** with the **Semi-Empirical PM3** model. Make certain that **Global Calculations** (at the bottom of the dialog) is checked. Click on **OK**.
6. Select **Surfaces** from either the **Setup** or **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. Submit the job. Name it *benzene chromium tricarbonyl*. When completed select **Spreadsheet** from the **Display** menu () and check the box to the left of the label for both entries. This allows the two molecules to be displayed simultaneously on screen. By default, **Coupled (Model menu)** is checked. Remove the checkmark by selecting it. The two molecules may now be

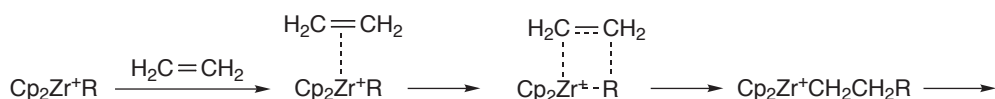
moved independently. Orient each molecule so that you can clearly see the benzene face (exposed face in the case of the organometallic).

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  $\text{Cr}(\text{CO})_3$  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. EDF2/6-31G\* results for both benzene and benzene chromium tricarbonyl are available from 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 SSPD is selected and click on **Replace**. Click on **All** in the dialog that results. The graphical calculations do not need to be specified, but will require a few seconds to complete. When they are, examine the electrostatic potential maps. Are they qualitatively similar to those from the PM3 calculations?
10. Close all documents and any open dialogs.

90 mins

## Ziegler-Natta Polymerization of Ethylene


Ziegler-Natta polymerization involves a metallocene. This complexes an olefin, which then inserts into the metal-alkyl bond.

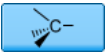


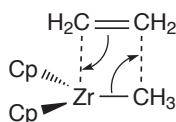
In this tutorial, you will use PM3 calculations to obtain a transition state for insertion of ethylene into  $\text{Cp}_2\text{ZrCH}_3^+$  and, estimate the activation energy using density functional calculations.

1. Select **New Build** from the **File** menu () and then **Inorganic**








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

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 and *double click* on screen. Select **Cyclopentadienyl** from the **Ligands** menu and *click* on two of the 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  $\text{sp}^3$  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 ( $\text{Cp}_2\text{ZrCH}_3$  and ethylene) as shown below:



(To move the fragments independently, hold down the **Ctrl** key.)

4. Select **Guess Transition State** from the **Search** 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 has completed, select **Spectra** from the **Display** menu (). *Click* on the  in the bar at the top of the spectra pane and select  (calculated IR Spectrum). The “infrared spectrum” calculated 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. Note that the frequency value at the top of the list is preceded by an “i”. This designates it as imaginary. *Click* on this frequency and examine the vibrational motion. Would you describe the process as concerted or occurring in discrete steps?

7. Perform EDF2/6-31G\* density functional energy calculations using PM3 geometries to obtain a better estimate for the energy barrier for ethylene insertion. **Spartan** will automatically make use of a pseudopotential for Zr instead of the all-electron 6-31G\* basis set. 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 **EDF2/6-31G\* Density Functional** model. Remove the checkmark on **IR** (to the right of **Compute**). **Total Charge** should still be set to **Cation**. *Click* on **Submit**.
8. While you are waiting for the energy calculation to complete, build both ethylene and  $\text{Cp}_2\text{ZrCH}_3^+$  (name them **ethylene density functional** and **Cp2ZrMe cation**, respectively). For  $\text{Cp}_2\text{ZrCH}_3^+$ , start with three-coordinate trigonal Zr, and then add two cyclopentadienyl ligands and a four-coordinate tetrahedral carbon. For each, select **Calculations...** from the **Setup** menu () and specify calculation of **Energy** using the **EDF2/6-31G\* Density Functional** model. Select **PM3** from the **Start from** menu to designate use of a PM3 geometry. For **Cp2ZrMe cation density functional** (only), set **Total Charge** to **Cation**. You need to be certain that **Global Calculations** at the bottom of the **Calculations** dialog is not checked.
9. Submit both jobs. When they have completed, calculate an activation energy for the insertion reaction.
10. Close all documents and any open 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 obtained from a single theoretical model, specifically the EDF2/6-31G\* density functional model. SSPD comprises over 250,000 entries, each of which provides a wide variety of molecular properties, including a selection of QSAR descriptors and, with the introduction of *Spartan*'14, estimated gas-phase acidities and basicities for selected compounds. SSPD entries also include both the infrared and NMR spectra and the wave function, allowing “on-the-fly” generation of molecular orbitals, electron densities and other graphical surfaces and the electrostatic potential and other property maps. In short, SSPD provides a rich source of diverse information for a wide selection of molecules.

The restriction of SSPD to a single theoretical model is a deliberate attempt to focus attention away from the characteristics of a particular model (and the question “which model is best”) and onto the chemistry at hand. While the choice EDF2/6-31G\* can be seen as a compromise between accuracy and practicality, in fact, EDF2/6-31G\* provides excellent results for a variety of properties that can be accurately measured, including molecular geometry and infrared and NMR spectra. Furthermore, the T1 heat of formation, a byproduct of using the T1 recipe to establish best conformer, is included in


the SSPD entry. We believe that over the wide variety of molecules considered, T1 provides as reliable an indicator of internal energy as any (calculated or experimental) measure presently available.

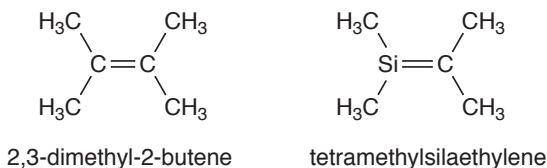
Collections such as the Spartan Spectra and Properties Database may be used to eliminate the need to perform calculations that have already been done. Taking this to the extreme, a database might serve as an exclusive source of information, doing away altogether with the need for quantum chemical calculations. From the instructional perspective (teaching about molecular models and molecular modeling) this allows use of more complex examples than might otherwise be practical. The tutorials in this chapter fall into this category, and can be thought as the ultimate in “dry labs”.

To complete the tutorials in this chapter, SSPD needs to be installed on *Spartan*. Either the sample version included with all copies of *Spartan* and which comprises ~6,000 molecules or the full version supplied as part of the *Spartan'14 Parallel Suite* may be used for the tutorials in this chapter.

## 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 and put into the same document. Start the second molecule with **Build New Molecule** from the **File** menu () instead of **New Build**.




2. Click on the name of whichever molecule is selected at the bottom




of the screen, *click* on **Replace** in the resulting menu 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 is automatic (you don't need to do anything) but will take a few seconds.
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. Select **Coupled** from the **Model** menu () to uncouple the motions of the molecules. Position the two molecules side by side on screen.
5. Inside the **Surfaces** dialog, select **local ionization potential map**. Compare local ionization potential maps for the olefin and silaolefin, recognizing that the more red 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 the local ionization potential maps off by again selecting it. Select **|LUMO|map** to turn the LUMO maps on. Here, the more blue 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? Speculate why molecules which incorporate a silicon-carbon double bond are chemically problematic. Close the (unnamed document) when you are finished.

## 6 - 8 (requires access to the Cambridge Structural Database)

Identify molecules in the Cambridge Structural Database (CSD) that incorporate a carbon-silicon double bond.

6. Build silaethylene,  $\text{H}_2\text{Si}=\text{CH}_2$ . Select **Structure Query** from the **Search** menu (). *Click* on all four free valences in silaethylene. They will be marked by orange cones.

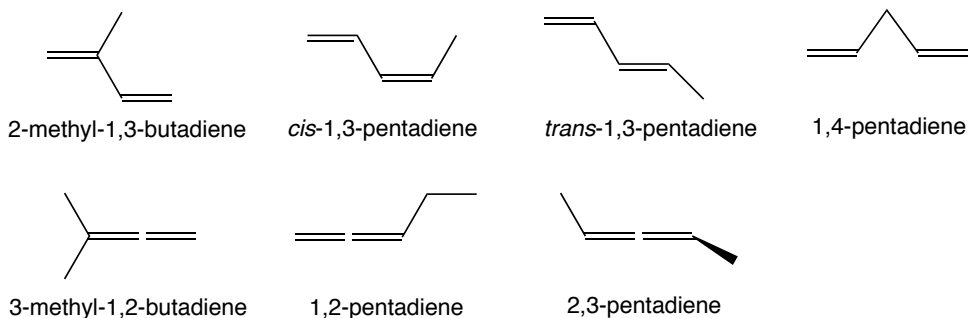
7. Bring up the **Cambridge Structural Database (CSD)** dialog by first selecting **Databases** from the **Search** menu () and then *clicking* on the **CSD** tab. *Click* on the **Search** button at the bottom of the dialog. In a few seconds, hits will begin to appear in the box at the right of the dialog. Examine each in turn (in the window at the left of the dialog) by *clicking* on its CSD reference code (**REFCODE**). Identify a structure that appears to be representative for closer scrutiny. Before you retrieve it, *click* on  to the right of the **Retrieve** button at the bottom of the dialog. Make certain that all five boxes under **Retrieval Filters** in the **Retrieve Options** dialog are *checked*. *Click* on **OK**. *Click* on the selected **REFCODE** and then *click* on the **Retrieve** button. Dismiss the **CSD** dialog.
8. Select **Minimize** from the **Build** menu (). Except for hydrogens, all atoms in the structure you have retrieved will be fixed and energy minimization will only change hydrogen positions. This is to compensate for the fact that hydrogens are generally not well located in X-ray crystal structure determinations.



Change the display to a space-filling model. Pay particular attention to the incorporated silicon-carbon double bond. Is it accessible in all of the compounds or is it encumbered by the surrounding groups? What, if anything, do your observations suggest about why this particular compound can be (has been) made?

9. Close all documents and any open dialogs.

### Isomeric C<sub>5</sub>H<sub>8</sub> 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 (1,1-dimethylallene) 1,2-pentadiene (ethyl allene) and 2,3-pentadiene (1,3-dimethylallene).








1. Build or sketch all seven  $C_5H_8$  dienes (place 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. Make certain that “*Update molecule names when replaced*” is checked. 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 **Molecule** tab in the resulting dialog is selected. Click on the **P** 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 should provide a more accurate account of relative stability than available from the EDF2/6-31G\* 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 molecules in SSPD (~2% 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. Remove the **Databases** window by again selecting **Databases** (). Bring up the spreadsheet (**Spreadsheet** from the **Display** menu (). Note that not all the experimental heats of formation are available. In fact, experimental heats of formation exist for only around 2,000-3,000 molecules. According to the T1 model, are any of the non-diene isomers more stable than the best diene?
4. Close all documents and open dialogs.


### Using Infrared Spectroscopy to Identify an Unknown Ethyl Benzoate Derivative

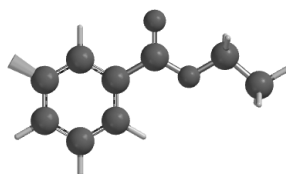
Unlike NMR spectra (at least  $^{13}C$  spectra) infrared (IR) spectra are rich in detail. Some chemists would view this as a curse, in that the complexity of an infrared spectrum masks its connection with the three-dimensional molecular structure. Others might see complexity as a blessing, providing a detailed pattern or fingerprint with which to identify a molecule. In fact, this difference 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'14** 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). 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\* and *click* on *unknown ethyl benzoate*. Make certain that **All Files** is selected in the menu to the right of the file name in the **Open** dialog. 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

\* For Windows, this directory is found in **Program Files/Wavefunction/Spartan14** or **Program Files(x86)/Wavefunction/Spartan14**. 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'14** disc image.

**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 could carry out your search with nothing other than the unknown spectrum (see the optional part of this tutorial), or you could 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. What you have done is to restrict 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 right of the **SIRD** dialog, in numerical order starting from the best. 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?

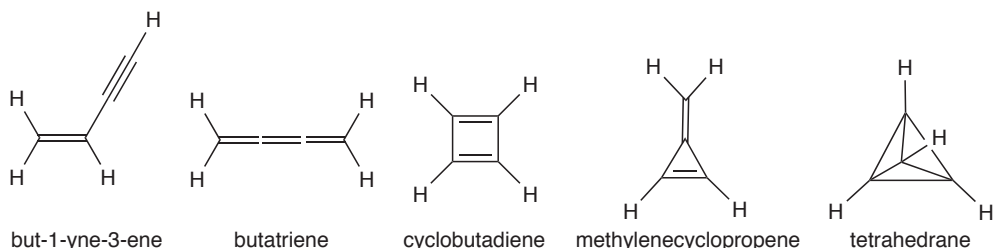
Be patient. A search of the full database will require significantly more computer time than the search of the 6,000 molecule sample database supplied with *Spartan*, simply because it contains many more molecules.


5. Close all documents and any open dialogs.


## Infrared Spectra of Short-Lived Molecules



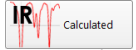
While infrared spectroscopy is by no means as formidable a tool as NMR for elucidating the structures of molecules, it does offer significant advantages which can make it the method of choice in some instances. It is more sensitive, meaning that it can be used to detect transient species, for example, 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! The problem 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  $3080\text{ cm}^{-1}$ . Possible structures include but-1-yne-3-ene, butatriene, singlet-1,3-butadiene, 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 left 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. *Click* on the **Retrieve** button. Remove the **Databases** dialog by again selecting **Databases** from the **Search** menu (  ).

2. Select **Spectra** from the **Display** menu (  ) and *click* on  at the top of the resulting dialog. *Click* on  (calculated IR spectrum). Move through the list to bring up the spectra for the other isomers. The calculated infrared spectrum of which if any of the  $C_4H_4$  isomers best fits the unknown infrared spectrum?
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 reduces to 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.

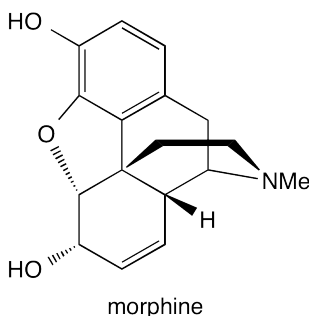
NMR assignment problems can be reduced but not altogether eliminated by requiring that both proton and  $^{13}C$  NMR spectra are consistent with a particular structure. This is routine practice. In principal, all ambiguity can be eliminated by cross-correlating the results of the  $^{13}C$  NMR with those of proton NMR using a variety of so-called 2D spectra, most important, the HMBC spectrum. However, 2D NMR experiments are not yet the norm, if for no other reason because they require significant instrument time on a spectrometer



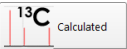


that may cost upward of \$1,000,000.

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 costing \$1000 or less raises the possibility for another tool to assist with spectral assignments. Whether this turns into reality ultimately depends on the ability of the calculations to obtain NMR spectra that are sufficiently accurate to distinguish among the different isomers.

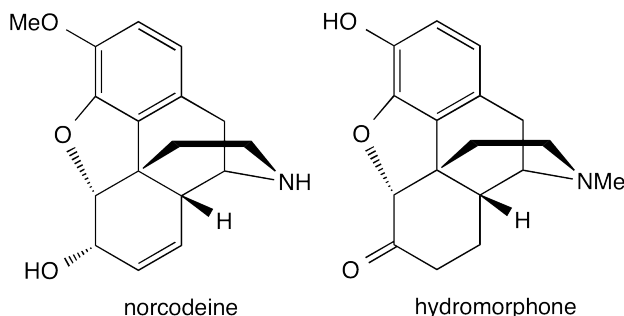
In this tutorial you will first compare calculated and experimental  $^{13}\text{C}$  spectra of morphine. You will then examine calculated spectra for two isomers, norcodeine and hydromorphone. The overall objective is to provide you with an impression of the ability of the calculations to reproduce a  $^{13}\text{C}$  spectrum for a known compound and more importantly, to challenge structure assignments based on NMR spectroscopy.


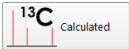
1. Sketch morphine. *Click* on its name at the bottom of the screen and *click* on **Replace**. If the name fails to appear, then you have made a mistake. In this case, select **Edit Sketch** from the **Build** menu () and correct your sketch.



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  $^{13}\text{C}$  spectra are now superimposed. Focus your attention on the four “most isolated” resonances, specifically those with calculated chemical shifts of 19.8 ppm, 60.0 ppm, 66.0 ppm and 93.8 ppm.
3. Sketch norcodeine and hydromorphone and retrieve both

molecules from the database.



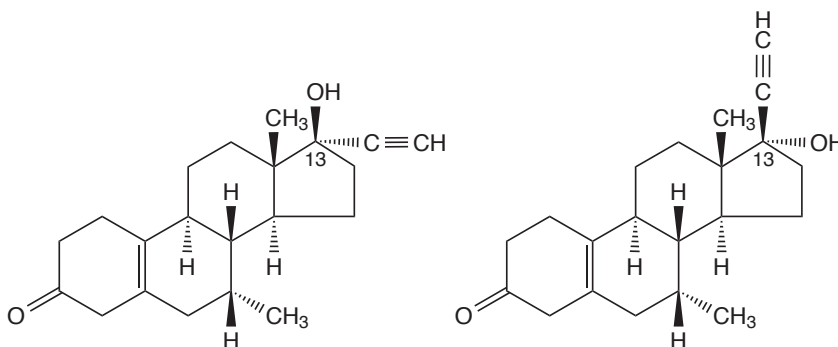
4. Select **Spectra** from the **Display** menu (🎵) and click on  and select  from the palette. You only need to do this once if you put norcodeine and hydromorphone in the same document. Does either spectrum fit the experimental  $^{13}\text{C}$  for morphine (again focus on the four isolated lines)?
5. Close all documents and any open dialogs.



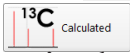
### Using $^{13}\text{C}$ NMR Spectra to Distinguish Stereoisomers

As commented 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 also noted, practical considerations (time on an expensive instrument) often make 2D experiments (in particular HMBC) the exception rather than the normal. Without the additional experiments, distinguishing stereoisomers based on their  $^{13}\text{C}$  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  $^{13}\text{C}$  NMR spectra good enough to be able to clearly be able to distinguish between stereoisomers? This tutorial provides an illustration.

The hydroxyl CO bond in tibolone, shown below on the left, is *gauche* to the methyl group at  $\text{C}_{17}$  (the alkyne is *anti*), whereas it is *anti* (and the alkyne is *gauche*) in the stereoisomer shown on the right. The

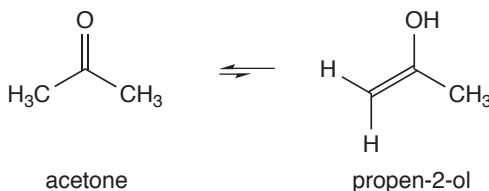
experimental  $^{13}\text{C}$  spectrum of a molecule that has been assigned as tibolone shows a line at 12.7 ppm attributed to the methyl group at  $\text{C}_{13}$  and lines at 88.9 and 74.8 ppm attributed to the internal and external alkyne carbons, respectively.



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. Select **Spectra** from the **Display** menu () , *click* on  in the bar at the top of the spectra pane and select  from the palette. In turn, *check* on the relevant atoms in the structure models for both stereoisomers and record the calculated chemical shifts. Do the calculations support or refute the experimental assignment, or are they ambiguous?
3. Close the document and any open dialogs.

## Finding Stable Enols

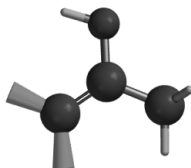
The equilibrium between ketones and their corresponding alcohols (enols) almost always greatly favors the former, for example, between acetone and propen-2-ol favors acetone.



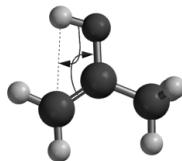
In this tutorial you will “mine” SSPD to see if you can find an enol that is actually *more stable* than its keto form. You will consider

enols related to propen-2-ol by substitution at C<sub>1</sub>.

1. Build or sketch propen-2-ol. Select **Structure Query** (🔍) from the **Search** menu and *click* on both free valences at C<sub>1</sub>. Orange cones will appear designating these as growth points.



2. Define the relationship between propen-2-ol and acetone using curly arrows.\* Select **Reaction Query** from the **Search** menu (🔍). *Click* on the OH bond, hold down the **Shift** key and one after another, *click* on the OH hydrogen and on C<sub>1</sub>. An arrow will be drawn from the OH bond to the space between the (OH) hydrogen and C<sub>1</sub>. *Click* on the (double) bond connecting C<sub>1</sub> and C<sub>2</sub> and then *click* on the (single) bond connecting C<sub>2</sub> to oxygen. A second arrow will be drawn between these two bonds.



3. Select **Formulas** from the **Display** menu (📄). *Click* on the **H(T1)** button under **Energies** and select **P-R (reaction)** from the **Applied to:** menu under **Context** (both near the bottom of the **Formula Editor** dialog). A formula for the query appears in the box at the center of the **Formula Editor** dialog, preceded by the name of the query **F0**. Replace this name by **enol to keto**, *click* on the **Enter** button (to the right of the query name and formula). You have now defined the query as the difference between the T1 heats of formation of the keto and enol forms of (substituted)propen-2-ol. Leave the **Formula Editor** dialog open.
4. Select **Databases** from the **Search** menu (🔍) and *click* on the **SSPD** tab at the top of the dialog. *Drag* the title **enol to keto**

---

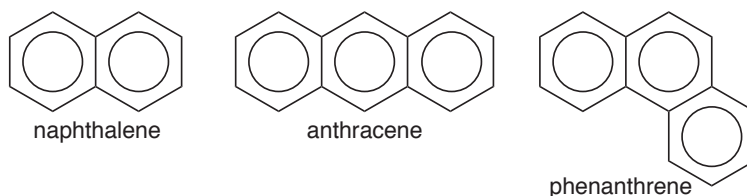
\* Organic chemists interpret curly arrows in terms of the relationship between reactants and transition state but they also connect the reactants to the products. It is in this context that they are used here.


from the formula editor onto an empty column header at the top right of the **SSPD** dialog, and *click* on the **Search** button. In a few seconds, hits (molecules in SSPD formed from propen-2-ol by single or double substitution at C<sub>1</sub>) will begin to appear at the right of the dialog. Reaction energies for any entries where T1 data exists both for actual search query (the substituted enol) and for its associated keto form will appear in the column designated **enol to keto**. **Pending** signifies that an entry exists for the search query (the enol) but not for the keto form.



- After the search has completed, position the cursor over the **enol to keto** column, hold down the right mouse button and select **Sort enol to keto** from the menu that appears. Molecules where the equilibrium most favors the keto form now appear at the top. Scroll down the list to identify molecules where the enol form is preferred.
- Close all documents and any open dialogs.

## Electrophilic Reactivity of Polycyclic Aromatics

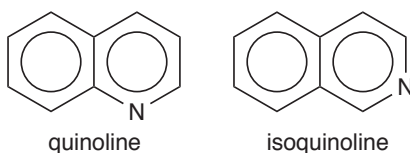
Naphthalene, anthracene and phenanthrene undergo electrophilic substitution just as does benzene. Are they likely to be more or less reactive? Local ionization potential maps should be able to provide an answer. The lower the potential on the aromatic ring, the more likely it is to give up an electron, that is, the more susceptible it will be to electrophilic attack.



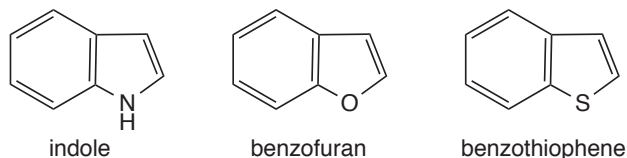
- Build or sketch benzene, naphthalene, anthracene and phenanthrene and put into a single document. *Click* on the name of whichever aromatic is selected at the bottom of the screen, *click* on **Replace** and finally *click* on **All** from the dialog that results.
- Select **Surfaces** from the **Setup** or **Display** menu (). Make

certain that **Global Surfaces** is *checked* in the dialog. Select **local ionization potential map** from the **Add** menu. When the graphics calculations have completed (they are done without you having to take further action, but they will take several seconds), select **Spreadsheet** from the **Display** menu () and *check* the box to the left of each of the four entries in the spreadsheet. The four molecules will now be simultaneously displayed, but they will be coupled. To move them independently, de-select **Coupled** from the **Geometry** menu (). Position the four molecules on screen such that the  $\pi$  system is visible, and *check local ionization potential map* inside the **Surfaces** dialog. Which of the four molecules is indicated by its local ionization potential map to be most susceptible to electrophilic attack? Which positions on naphthalene, anthracene and phenanthrene are most susceptible?

3. Compare local ionization potential maps for naphthalene, quinoline and isoquinoline. Focusing on the carbon ring, which of the three is likely to be most susceptible to electrophilic attack? Which is likely to be least susceptible?



4. Compare the local ionization potential map for naphthalene with those of indole, benzofuran and benzothiophene. Focusing on the carbon ring, which of the four compounds is most susceptible to electrophilic attack? Which is the least susceptible?



5. Close all documents and any open dialogs.

# Section IV

## Features and Functions

---

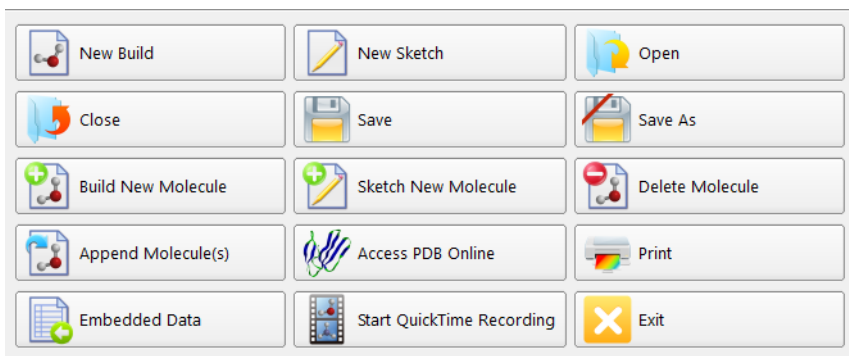
This section describes the functions available under the menus incorporated into *Spartan'14*, and is intended to serve as a general reference to the program. The coverage follows the order of the menus presented in *Spartan*'s user interface: **File**, **Edit**, **Model**, **Geometry**, **Build**, **Setup**, **Display**, **Search**, **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, while those under the **Search** menu access 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'14* as a server.

The chapters in this section provide only limited commentary about the performance and requirements of different computational methods and the utility of the different graphical models. Additional information is provided under *Topics* accessed from the **Activities** menu and to *A Guide to Molecular Mechanics and Quantum Chemical Calculations*, available as a PDF under the **Help** menu.

# Chapter 15

## 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 nucleotide structures. They allow text and on-screen graphics to be printed, text and other files to be embedded into **Spartan** documents and QuickTime movies to be created.*



### New Build ()

Brings up a model kit and clears the screen. Model kits are discussed in **Chapter 19**.

### New Sketch ()

Brings up the 2D sketch pane and clears the screen. The 2D sketch pane is discussed in **Chapter 19**.

### 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) including 2D sketch files, supported are files containing 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), and files containing 1D “strings”: SMILES (.smi). Finally, infrared, NMR and UV/visible spectra may be input (discussion of file formats is provided in **Appendix J**).

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).

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 21**). Additionally, hydrogen atoms that are missing in the input structure may be provided using **Grow Hydrogens** in the **Molecule Utilities** 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.

---

\* SDF files may contain information in addition to geometry, for example, the name of the molecule or its CAS number. Each quantity appears in a line following <xyz> where xyz is a character string denoting the quantity, for example, **CAS** to denote the CAS number.

<CAS>

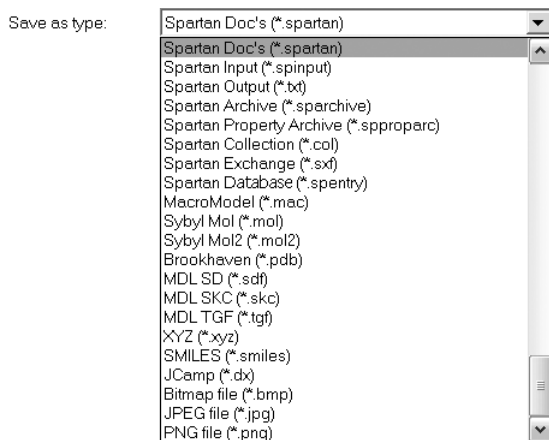
22090-26-2

This information may be accessed in the spreadsheet by typing the character string into a header cell, for example, typing **CAS** will result in the CAS number.

**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 be 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**). Support is also provided for writing QuickTime (movie) files (see discussion later in this chapter). Selection is made under the **Save as type** menu in the **Save As** dialog.



**Build New Molecule** ()

Brings up a model kit and clears the screen. **Build New Molecule** differs from **New Build**, in that the resulting molecule is appended to the end of the document associated with the molecule (or sketch) that is presently selected.

**Sketch New Molecule** ()

Brings up the 2D sketch pane and clears the screen. The menu bar is still accessible, but only the **View** () and **Sketch New Molecule** () icons are available. **Sketch New Molecule** differs from **New Sketch** in that the resulting sketch is appended to the end of the

document associated with the molecule (or sketch) that is presently selected.

### Delete Molecule ( )

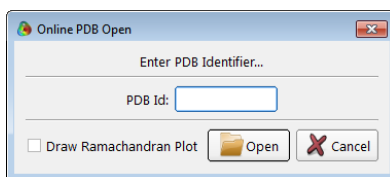
Deletes the selected molecule(s) from a document. Deleting the last molecule leads to an empty document.

### 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.\*

### Access PDB Online... ( )

Provides access to the online **Protein Data Bank (PDB)\*\*** comprising nearly 90,000 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

---

\* Alternatively, molecules may be appended onto 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** keys in the usual manner.

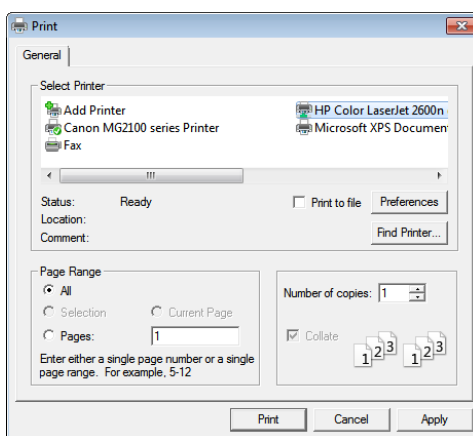
\*\* The web address is <http://www.rcsb.org>.

than one entry, all structures will be returned in a single document.

A ribbon model of the protein or nucleotide 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 17**).

## Print ()

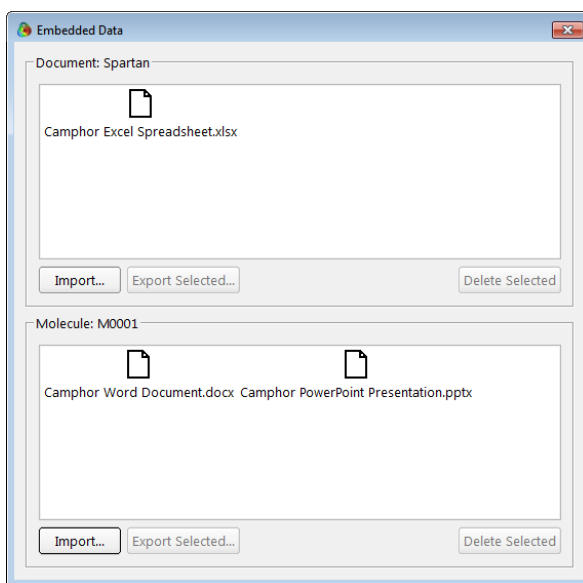
Selection leads to a dialog in order to designate a printer, specify print layout and number of copies. It also allows printing to a file.



The contents of the spreadsheet (**Spreadsheet** under the **Display** menu; **Chapter 21**) may be printed using **Print** from the contextual menu. The results of a reaction energy calculation (**Reactions...** under the **Display** menu; **Chapter 21**), may be printed using **Print** from the contextual menu.

## Embedded Data ()

This allows non-native files, for example, Word, Excel or PowerPoint files, to be embedded in a *Spartan* document. These may be associated either with an individual molecule or with all the molecules (or both). Selection results in the **Embedded Data** dialog, the upper half of which relates to document level files and the lower half of which relates to molecule level files.



A file may be imported (at the document or molecule level) by *clicking* on the appropriate **Import** button. This leads to a file browser from which a file may be selected. Once embedded in a *Spartan* document, a non-native file may be exported by first selecting (*clicking* on) it and then *clicking* on **Export Selected**. It may be deleted by first selecting it and then *clicking* on **Delete Selected**.

Note that *applications cannot be launched* from the **Embedded Data** dialog. They need to be copied outside prior to launching.

### Start/Stop QuickTime Recording ( )

This allows QuickTime movies to be created. To start making a movie, select **Start QuickTime Recording**. Any motions of all molecules in *Spartan*'s main screen will be captured. Dialogs (including the builders) will not be captured. Note the use of “tumbling” (see **Settings Preferences** under **Preferences** in the **Options** menu; **Chapter 23**) in making QuickTime movies. To stop making a movie, select **Stop QuickTime Recording** (which has replaced **Start QuickTime Recording** in the **File** menu) and supply the requested file name.

Note that the QuickTime entry in the **File** menu will only appear if QuickTime is actually installed on your machine.

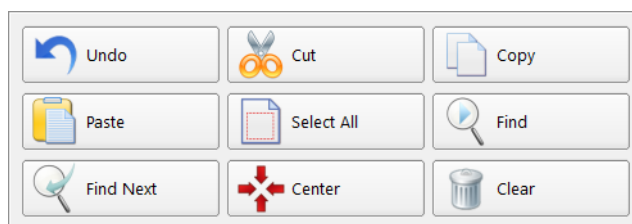
## Exit (✖)

Exits *Spartan*, that is, clears the screen and closes all open documents. A prompt for a name is provided for each document that has not previously been saved.

# Chapter 16

## 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<sup>®</sup> and PowerPoint<sup>®</sup>.
- (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<sup>®</sup>.

- (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 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*, but 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 19** and for moving data in and out of the spreadsheet is provided in **Chapter 21**.

### 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).

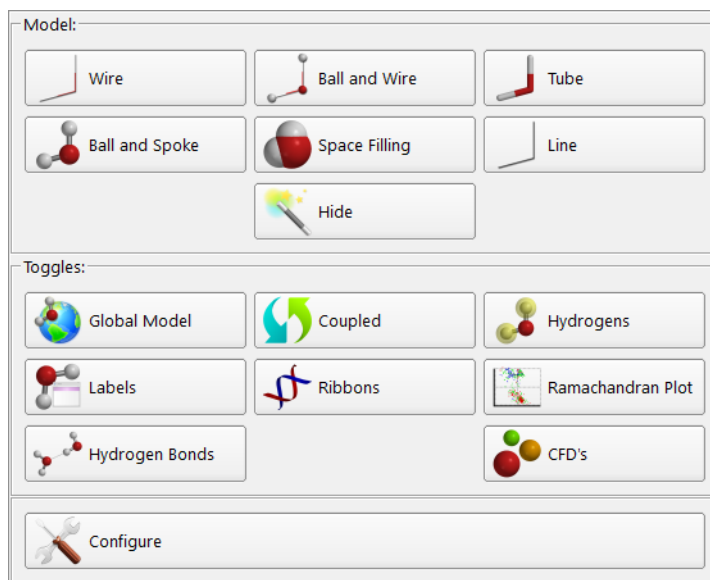
### Clear ()

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 17

## The Model Menu

*Structure models available under the **Model** menu include wire, ball-and-wire, tube, ball-and-spoke, space-filling (CPK) and line models, with or without hydrogens, with or without hydrogen bonds indicated, with or without chemical function descriptors (CFD's) shown and with or without atom labels, as well as ribbon displays for polypeptides and polynucleotides, with or without labels and with or without hydrogen bonds indicated. It allows drawing a Ramachandran plot for a protein structure retrieved from the Protein Data Bank (PDB). The menu also provides for configuring atom labels to display element name, R/S chirality, mass number, charge or chemical shift, and for specifying color coding and display style for ribbon labels, as well as turning a variety of other labels on and off. Finally, it allows model style to be applied globally (to all molecules in a document) and models to be manipulated in concert.*



Only one model style **Wire**, **Ball and Wire**, **Tube**, **Ball and Spoke**, **Space Filling**, **Line** or **Hide**) may be selected. The selected model is

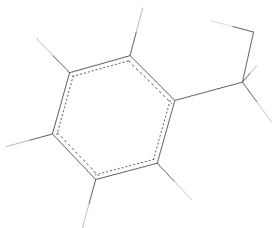
designated by a check mark ✓ in front of its entry in the menu. **Global Model, Coupled, Hydrogens, Labels, Ribbons, Hydrogen Bonds** and **CFD's** operate as toggle switches. A ✓ in front of the entry in the menu indicates that it is turned on.

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 Preferences** dialog (**Preferences...** under the **Options** menu; **Chapter 23**). Both structure models and graphics may be presented in 3D stereo. This is also controlled from the **Settings Preferences** dialog as well as from the **3** key. 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 **Table 17-1**.

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 23**). Colors of individually selected atoms may be set using the **Atom Style** dialog (**Properties** under the **Display** menu; **Chapter 21**). All models use the same color scheme for atoms, and provide for the same mechanism of changing colors globally or individually.

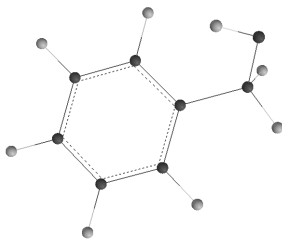
### Ball and Wire()

This represents atoms by small balls and bonds by wires.

**Table 17-1: Default Atom Colors**

main group		main group (con't)	
Hydrogen	white	Indium	orange
Lithium	orange	Tin	gray
Beryllium	green	Antimony	orange
Boron	orange	Tellurium	orange
Carbon	gray	Iodine	violet
Nitrogen	blue-gray	<b>transition metals</b>	
Oxygen	red	Scandium-Zinc	green
Fluorine	pale yellow	Yttrium-Cadmium	green
Sodium	yellow	Lanthanum-Mercury	green
Magnesium	purple	<b>lanthanides</b>	
Aluminum	magenta	Cerium-Lutetium	violet
Silicon	gray	<b>actinides</b>	
Phosphorus	orange	Thorium-Lawrencium	dark blue
Sulfur	yellow	<b>noble gases</b>	
Chlorine	green	Helium	dark orange
Potassium	red	Neon	dark orange
Calcium	red	Argon	dark orange
Gallium	orange	Krypton	dark orange
Germanium	gray	Xenon	orange
Arsenic	dark orange		
Selenium	red orange		
Bromine	dark red		
Rubidium	red		
Strontium	red		

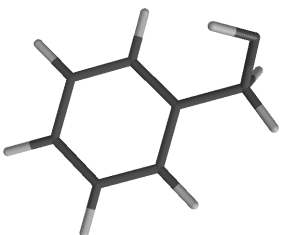
Ball-and-Wire Model



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.

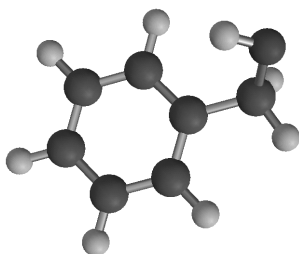


Tube Model

Tubes may either be solid cylinders or be split to represent multiple bonds depending on whether **Split Tubes** in the **Settings Preferences** dialog (**Preferences...** under the **Options** menu; **Chapter 23**) is turned off or on. As with wire models, bonds are drawn in two colors.

### 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

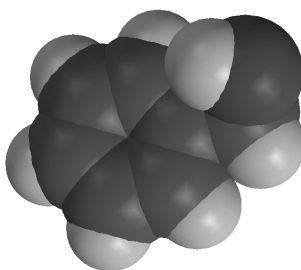
Spokes may either be cylinders or be split to represent multiple bonds depending on whether **Split Tubes** in the **Settings Preferences** dialog

(**Preferences...** under the **Options** menu; **Chapter 23**) is turned off or on. Bond (spoke) color is gray by default but it may be changed using the **Set Colors** dialog (**Colors** under the **Options** menu; **Chapter 23**). Colors of individually selected bonds may be set using the **Bond Style** dialog (accessible from **Properties** under the **Display** menu; **Chapter 21**).

## 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 21**) correspond to a space-filling model.

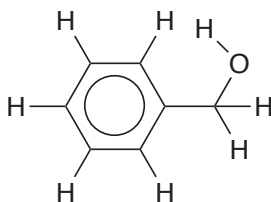
## Line ()

A line model is a 2D representation (similar to that provided from drawing programs such as ChemDraw and MarvinSketch). It designates carbons as the termini or the intersection of lines and non-carbons by their elemental symbols without any use of color.

---

\* Default values for van der Waals radii may be changed from the **VDW Radii** dialog accessible from **Preferences** under the **Options** menu (**Chapter 23**). 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 21**).



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 21**).

### Hide (🔍)

This removes the 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 interesting 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 has already been 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 21**).

### Global Model (🌐)

If *checked* (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 Preferences** dialog (**Preferences...** under the **Options** menu; **Chapter 23**). **Global Model** acts in a toggle manner, switching between global and local display. **Global Model** is normally on.

## Coupled ()

If *checked* (turned on), this signifies that all molecules in a document selected for simultaneous display will be moved together. **Coupled** is turned on following molecule alignment (see **Align** under the **Geometry** menu; **Chapter 18**). **Coupled** acts in a toggle manner, that is, repeated selection couples and decouples the molecules.

## Hydrogens ()

If *checked*, this signifies that hydrogens are to be included in the model. Note that 2D SD 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 22**) or using **Grow Hydrogens** in the **Molecule Utilities** dialog (accessible from **Properties** under the **Display** menu; **Chapter 21**). **Hydrogens** acts in a toggle manner, that is, repeated selection turns the display of hydrogens on and off.

## Labels ()

If *checked*, 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 *checked*, this signifies that ribbons are to be displayed along with the selected model. (If only ribbons are desired, for example, in proteins, select **Hide** for the model.) **Ribbons** acts in a toggle manner, that is, repeated selection turns display of ribbons on and off.

## Ramachandran Plot ()

If *checked*, this draws a Ramachandran plot for a protein input from the Protein Data Bank (see **Access PDB Online** under the **File** menu; **Chapter 15**). **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  $\alpha$ -helices, blue for  $\beta$ -sheets, green otherwise) is not based on the actual 3D geometry but rather on assignments in the PDB file.

## Hydrogen Bonds ()

If *checked*, 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 2.1 Å and making an X–H--Y (X,Y = N, O) angle of  $>120^\circ$ . 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.

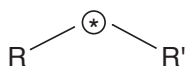
## CFD's ()

If *checked*, this signifies that **Chemical Function Descriptors (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).

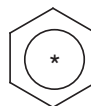
*Spartan* uses seven different kinds of CFD, the first six of which may be thought of as attributes of a molecule.

CFD	chemical language
1. hydrophobe	sterically-crowded region
2. aromatic	aromatic $\pi$ 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 21**). 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.

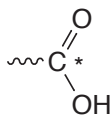


hydrogen-bond acceptor

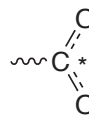


aromatic

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.

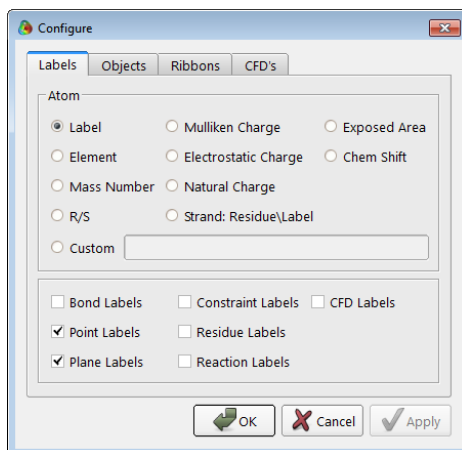
Calculated electrostatic potentials may offer significant advantage over CFD's and other qualitative descriptions when there are no heteroatoms and hydrogen bonding is not possible. The  $\pi$  system of benzene and other aromatics provide the most common examples. The same conditions that make the ring 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 **Search** menu; **Chapter 22**).

## Configure... ()

This selects the types of labels attached to atoms, ribbons and CFD's.

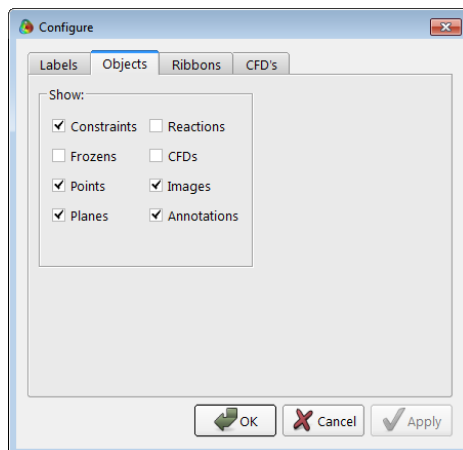
### Configure Labels



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 21**), **Element**, **Mass Number**, **R/S** (chirality), **Mulliken Charge**, **Electrostatic Charge**, **Natural Charge**, **Strand: Res\Label** (polypeptides and polynucleotides), and **Exposed Area** (of an atom in a space-filling model) and **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 21**). In addition, **Bond Labels**, **Point Labels**, **Plane Labels**, **Constraint Labels**, **Residue Labels**, **Reaction Labels** and/or **CFD Labels** may be provided. Default settings (for a new molecule) are made in the **Molecule Preferences** dialog (**Preferences** under the **Options** menu; **Chapter 23**).

## Configure Objects

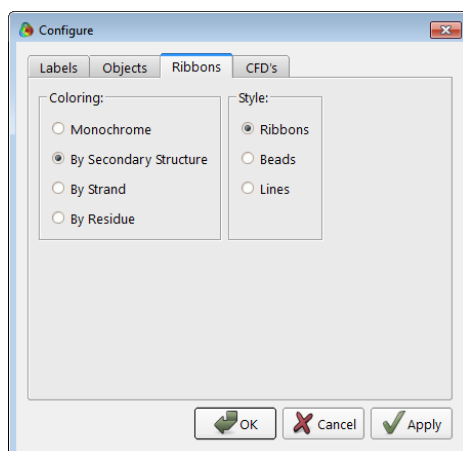
Clicking on the **Objects** tab leads to the **Configure Objects** dialog.



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.

## Configure Ribbons

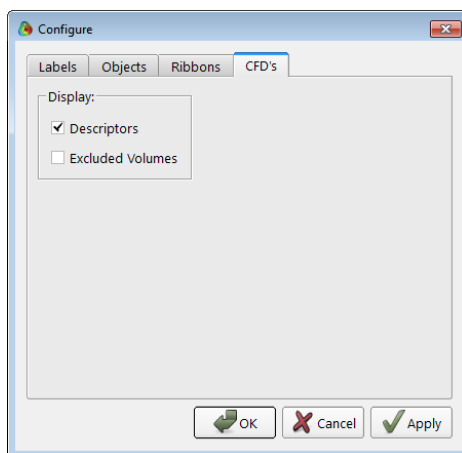
Clicking on the **Ribbons** tab leads to the **Configure Ribbons** dialog.




Ribbon coloring may be selected from among the following: **Monochrome**, **By Secondary Structure**, **By Strand** or **By Residue**. Ribbon style may be selected from among the following: **Ribbons**, **Beads** or **Lines**.

## Configure CFD's

*Clicking* on the **CFD's** tab leads to the **Configure CFDs** dialog.



This allows for turning CFD descriptors on (identifying the type of CFD) identifiers and CFD's corresponding to excluded volumes.

The **Configure** dialog is removed from the screen with all selections maintained by *clicking* on **OK**. *Clicking* on **Cancel** or on  removes the dialog but selections are lost. *Clicking* on **Apply** maintains the selections but leaves the **Configure** dialog on screen. Note, that **Labels** (from the **Model** menu) will be turned on following either *clicking* on **OK** or on **Apply**.

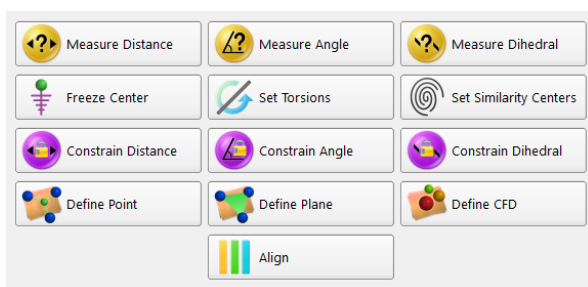
**Table 17-1: Default Atom Colors**

main group		main group (con't)	
Hydrogen	white	Indium	orange
Lithium	orange	Tin	gray
Beryllium	green	Antimony	orange
Boron	orange	Tellurium	orange
Carbon	gray	Iodine	violet
Nitrogen	blue-gray	<b>transition metals</b>	
Oxygen	red	Scandium-Zinc	green
Fluorine	pale yellow	Yttrium-Cadmium	green
Sodium	yellow	Lanthanum-Mercury	green
Magnesium	purple	<b>lanthanides</b>	
Aluminum	magenta	Cerium-Lutetium	violet
Silicon	gray	<b>actinides</b>	
Phosphorus	orange	Thorium-Lawrencium	dark blue
Sulfur	yellow	<b>noble gases</b>	
Chlorine	green	Helium	dark orange
Potassium	red	Neon	dark orange
Calcium	red	Argon	dark orange
Gallium	orange	Krypton	dark orange
Germanium	gray	Xenon	orange
Arsenic	dark orange		
Selenium	red orange		
Bromine	dark red		
Rubidium	red		
Strontium	red		

# Chapter 18

## 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 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 in a list.*



**Measure Distance** (↔?)

**Measure Angle** (∠?)

**Measure Dihedral** (∩?)

**Measure Distance** displays the distance (in Ångstroms) between two atoms, whether or not they are bonded. Selection results in a message at the bottom left of the screen.


Select two atoms, a bond, ...

*Clicking* on two atoms displays the distance at the bottom right of the screen.

Distance(C1,C2) = 1.531 Å **P**

Alternatively, *clicking* on a bond displays its length.

Length(Bond1) = 1.531 Å **P**

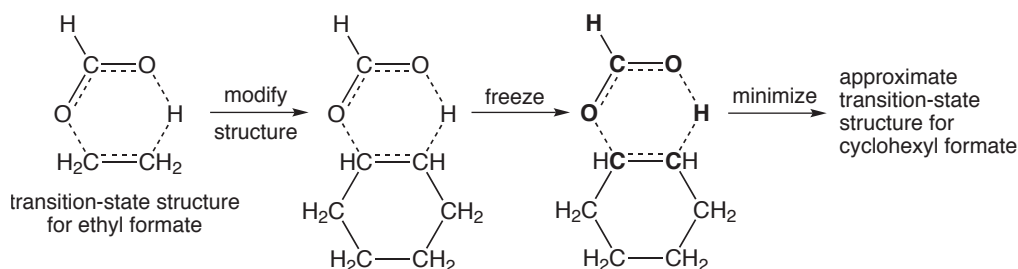
**Measure Distance** may also be used to alter the distance between atoms (as long as both are not incorporated into the same ring), by altering the contents of the box to the right of **Distance (A,B) =** or **Length (A)=**, and then *pressing* the **Enter** key (return key on Mac). The distance (length) may be entered into the spreadsheet by *clicking* on  to the right of its display (see **Spreadsheet** under the **Display** menu; **Chapter 21**). 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 21**) or into any of the dialogs associated with the Spartan Molecular Database or Spartan Spectra and Properties Database (see **Databases** under the **Search** menu; **Chapter 22**).

Angle and dihedral angle queries 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 20**).

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 figure below) 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 or free valence\*, freezes it; *clicking* again thaws it. Buttons at the bottom right of the screen allow for freezing all atoms (**Freeze All**), freezing all heavy (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 22**). Hydrogen positions are more often than not poorly located in small molecule X-ray structures, and XH bond lengths are commonly as much as 0.1 to 0.2 Å shorter than they should be. Structures incorporating such bond lengths are clearly inappropriate for energy and property calculations and may also be problematic as starting geometries in quantum chemical calculations. Protein crystal structures always 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 minimization using either **Minimize** from the **Build** menu (see discussion in **Chapter 19**) or from **Calculations...** under the **Setup** menu (see discussion in **Chapter 20**).

Frozen atoms are indicated by magenta colored markers (  ). Whether or not these are included with the model (outside of freeze center mode) is controlled from the **Molecule Preferences** dialog under **Preferences...** in the **Options** menu (**Chapter 23**).

\* The bond distance in this case is that appropriate for hydrogen being added to the free valence.

## 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 20**.) Selection of rule set is made in the **Settings Preferences** dialog (**Preferences...** under the **Options** menu; **Chapter 23**).

**Set Torsions** allows these defaults to be altered. Selection results in rotatable bonds each being marked by a gold cylinder, and flexible rings each 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, followed by *pressing* the **Enter** key (**return** key on Mac). 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.

◀ 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 have not been removed.

The total number of conformers (excluding ring conformers) that will be examined in a systematic search (see discussion in **Chapter 20**) 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 21**), into the **Formulas** dialog (see discussion in **Chapter 22**), or into any of the dialogs associated with the Spartan Molecular Database or Spartan Spectra and Properties Database (see discussion in **Chapter 22**). The number of conformers is also available under the **Molecule** tab in the **Molecule Properties** dialog (**Properties** under the **Display** menu; **Chapter 21**). This can also be dragged into the same dialog as well as posted to the spreadsheet.

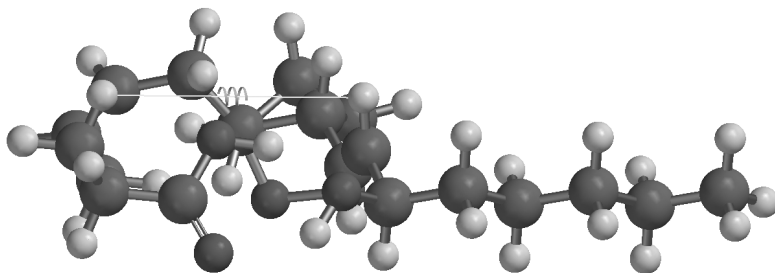
**Set Torsions** is also used to specify non-bonded distances that need to be kept below a threshold value. These follow from NOE (Nuclear Overhauser Effect) measurements and may be referred to as NOE conditions or simply NOEs. NOEs are specified by *clicking* on two (non-bonded) atoms while holding down the **Shift** key. In response, a message appears at the bottom of the screen.



*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.

---

\* A limit of 1000 conformers is enforced.



Once set, NOEs are used in conformational searching (conformer distribution only; see discussion under **Calculations...** under the **Setup** menu; **Chapter 20**) 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.

### Set Similarity Centers (🌀)

This specifies the set of atomic centers, or **Chemical Function Descriptors (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 20**).

#### Similarity Based on Structure

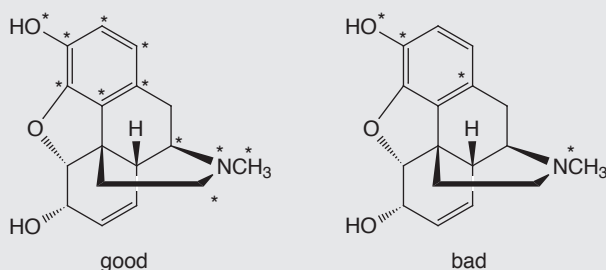
Selection of **structure** from the **Similarity by** menu at the bottom right of the screen results in a message at the bottom left of the screen.

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 trialkylamine components (left) will be much faster and much 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 few 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.

Select CFD's.

*Clicking* on a CFD designates it as a similarity center and marks it with a violet circle. *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:

$$i!j!k! \dots$$

(! 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).

Selection of similarity centers either needs to be repeated for the remaining molecules in the list or the selections may be designated as global, that is, applying to all molecules in the list. The latter is applicable where the molecules in the list are conformers or are closely related at least insofar as the alignment centers are concerned. Global designation is accomplished by *checking* the **Global Similarity Settings** at the bottom of the screen.

**Constrain Distance** ()

**Constrain Angle** ()

**Constrain Dihedral** ()


These introduce one or more geometrical 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 20**). They also allow for setting a range of constraints needed for generation of energy profiles. 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, and
- (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.

Constraint(C1,C2) =  


*Clicking* on  changes it to  and shows the current distance.

Constraint(C1,C2) =  

This (constraint) distance can now be changed by altering the contents of the box and then *pressing* the **Enter** key (**return** key on Mac). 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. In this case, *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 21**) or into any of the dialogs associated with the Spartan Molecular Database or Spartan Spectra

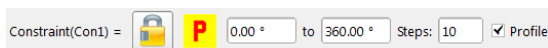
\* Note, however, that it may be problematic to carry out a constrained geometry optimization starting from a structure that is very different from that satisfying one or more constraints.

and Properties Database (see **Databases** under the **Search** menu; **Chapter 22**).

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 edit build fragment mode () , but need to be requested for methods specified in the **Calculations** dialog (**Calculations...** from the **Setup** menu; **Chapter 20**).

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 20**).



The leftmost box sets the initial value of the constraint, the middle box to the right of **to** sets the final value, and the rightmost box to the right of **Steps:** sets the number of steps. 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.


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 23**).

## 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.

Selecting **Define Point** (or *clicking* on ) while holding down on the **Shift** key, followed by *clicking* on the appropriate atoms, leads to a **ligand point**. This is 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). 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 19**) 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 23**).

## 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 23**).

### Define CFD ()

This defines the position of a new CFD. The **CFD Properties** dialog (accessible from **Properties** under the **Display** menu; **Chapter 21**) allows CFD size and type to be selected. 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 23**).

### Align ()

This aligns the selected molecule to all other molecules in the same document based either on structure or on CFD's. If the selected molecule is a pharmacophore, aligns the pharmacophore to all molecules in the same document (based on CFD's). Choice of whether alignment is based on 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 and/or pharmacophores in a document are reoriented to best coincide with the selected molecule (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 in detail in **Chapter 20**.

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 Structure

Selection of **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 using spreadsheet functions (see **Spreadsheet** under the

**Display** menu; **Chapter 21**). 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 17**). 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 21**) 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 21**). 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 17**). 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 a Pharmacophore**

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 list of available properties (see **Spreadsheet** under the **Display** menu; **Chapter 21**). A score of 0 is assigned to molecules that cannot be aligned to the selected molecule.

# Chapter 19

## The Build Menu

*The **Build** menu provides a sketch palette for drawing organic 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, 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.*

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 molecules can either be rendered as 2D sketches and later brought into 3D, or directly constructed from 3D fragments. Inorganic and organometallic molecules as well as 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.








**Defined Atoms.** H, B, C, N, O, F, Si, P, Cl, Br and I only.


**Element/Functional Group Library.** A “wildcard” icon which appears below the H and B icons allows any atom to be specified

by its element symbol, for example, **Na** or **Al**, as well as a variety of standard groups, for example, **Ph**, **OCH<sub>3</sub>**, **OMe**, **NO<sub>2</sub>**, or **TMS**. About 60 functional group symbols are presently recognized, a list of which may be found under the **Help** menu. The meaning of this icon can be changed by *double-clicking* the icon and entering a new symbol. Case is not important, for example, **sn**, **sN**, **Sn** and **SN** are all interpreted as **Sn**.




**Common Rings.** Three icons facilitate the rapid addition of benzene () , cyclohexane () and cyclopentane () rings to drawings.




**Common Carbonyl Groups.** Three icons facilitate the rapid addition of carbonyl () , carboxylic acid/ester () and amide () groups to drawings.

**Stereochemical Markers.** Wedges and dashes, represented by  and , can be used to designate in-out stereochemistry. 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

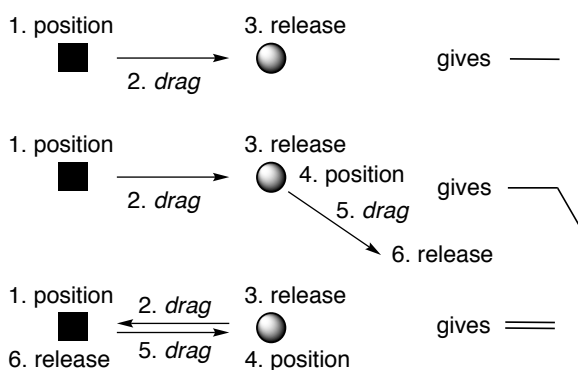
produce a 3D model of water,  $\text{H}_2\text{O}$ . However, adding the appropriate marker will result in 3D models of  $\text{H}_3\text{O}^+$  () ,  $\text{HO}^-$  () , or HO radical () , respectively. Only one charge/radical marker can be assigned to an atom.

Only one charge/radical marker is displayed on the palette, but *clicking* on the icon will cause each marker to appear in turn.

**Drawing Tools.**  undoes the most recent drawing operation.  removes or modifies parts of a drawing.  deletes an entire drawing (a warning is provided).  tries to improve the readability of a drawing by applying various “clean up” procedures.

## 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 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.



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 the sketch, pinch (or spread) two fingers on the 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/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).

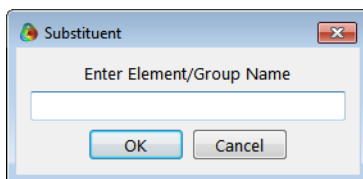
**Fuse 5 and 6-Member Rings.** *Click* on an icon for the first ring and *double click* on the screen. Next, *click* on the icon for the second ring and double tap the bond that the rings will share. This will create a drawing with a fused bicyclic ring system. Note that the (*cis* or *trans*) stereochemistry of the ring juncture is ill-defined.

This technique can also be used to add 5 and 6-member rings to an existing bond 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* 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 or Functional Group from the Library.** *Click* on the icon underneath H and B (this icon will initially be labeled with the name of the group that was previously selected and will change each time you select a new element or group). A dialog results.




To define an element, type the one or two-letter standard chemical symbol for the element and *click* return. To define a group, type the atoms in the group (for example, **OCH3** or **NO2**) or a standard abbreviation for the group (for example, **OMe**, **Ph**, or **TMS**) and *click* on **OK**. Case is not important, for example, **sn**, **sN**, **Sn** and **SN** are all interpreted as **Sn**. A listing of available groups is provided under the **Help** menu. If the element/group is not available, it will be marked as unknown.

The element/group 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 and type the symbol for the new group.



**Add Stereochemical Markers (Dashes and Wedges).** 3D information can be added to a drawing by replacing **single** bonds with stereochemical markers: dashes or wedge. 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.

**Add an Axial or Equatorial Marker** (6-member rings only). 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* marker to a stereochemical marker, *click* on either  or , then *double click* on the stereochemical marker. *ax* will appear on top of the stereochemical marker. To replace *ax* with *eq*, *double click* on the stereochemical marker again. To remove the marker, *double click* on the marker again.

Although it is possible to produce a drawing in which several bonds are marked as *axial* or *equatorial*, **only one** marker is used when converting a 2D ring drawing into a 3D model.


**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 conventional bonding rules. A neutral carbon atom is assumed to form four bonds, nitrogen three bonds, oxygen two bonds, and so on. Analogous logic is used for atoms that carry a formal charge or unpaired electron marker.


Charge and unpaired electron markers are carried over into specification of quantum chemical calculations (**Calculations...** from the **Setup** menu; **Chapter 20**). 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 22**.

**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 an unsatisfactory 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 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 affected.

**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.

A 3D structure is obtained from the 2D sketch by *clicking* on .

### 3D Model Kits

*Spartan* provides five different model kits for assembling a variety of molecular systems: an organic model kit for most organic molecules, an inorganic model kit for organic molecules not well represented in terms of an uncharged (non-zwitterionic) valence structure, as well as inorganic and organometallic molecules, model kits for polypeptides

and polynucleotides and a model kit for attaching substituents (lists of chemical groups) to previously-built molecules. The organic and inorganic model kits utilize atomic fragments, functional groups and rings (and ligands in the inorganic model kit), while the peptide model kit uses the set of natural amino acids as building blocks, and the nucleotide model kit the set of nucleotide bases.

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).

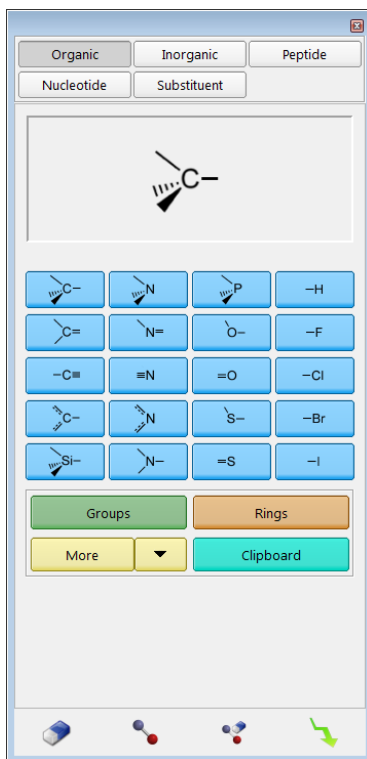
3D molecule construction in *Spartan* proceeds much in the same manner as a chemist would assemble a structure from a model kit, that is, pieces are taken from the kit one at a time and added sequentially to the molecule under construction.

## Organic Model Kit

The organic model kit contains a suite of molecule building/editing tools specifically designed to construct organic molecules.

---

\* Seamless access to ChemDraw is available for the Windows version only. All versions are able to read ChemDraw files. ChemDraw is not included with *Spartan* and must be obtained from CambridgeSoft at <http://www.cambridgesoft.com>.



In the center of the model kit are a selection of atomic fragments, which from left to right and then top to bottom, correspond to:

C(sp <sup>3</sup> )	N(sp <sup>3</sup> )	P(sp <sup>3</sup> )	H
C(sp <sup>2</sup> )	N(sp <sup>2</sup> )	O(sp <sup>3</sup> )	F
C(sp)	N(sp)	O(sp <sup>2</sup> )	Cl
C(aromatic)	N(aromatic)	S(sp <sup>3</sup> )	Br
Si(sp <sup>3</sup> )	N(planar)	S(sp <sup>2</sup> )	I

A fragment is chosen by *clicking* on its icon, which is then displayed at the top of the model kit. 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, *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,  $\text{sp}^2$  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 model kit, other elements may be substituted using atom replacement feature available in the inorganic model kit (see **General Molecule Building Functionality** later in this chapter). Similarly, bond types may be altered in the inorganic model kit. 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 21**). Note that bond type (or the fact that bonds may be present) is not utilized in quantum chemical calculations, and only atom identities and coordinates are required. They are, however, utilized in molecular mechanics calculations.

Menus inside the 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 model kit 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 model kit.












	Alkenyl
	Alkynyl
	Allenyl
	Amide
	Azido
	Carbonyl
	Cyano
	Carboxylic Acid/Ester
	Isocyanato
	Nitro
	Nitroso
	Phosphine Oxide
	Sulfone
	Sulfoxide

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 model kit). The marked position circulates among the possible positions with repeated *clicking* on the icon.

## 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 model kit.

	Cyclopropane
	Cyclobutane
	Cyclopentane
	Cyclohexane
	Cycloheptane
	Benzene
	Naphthalene
	Anthracene
	Phenanthrene
	Indene
	Fluorene

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 different 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, the menu to the right of **More** will be empty. It can be populated, by *clicking* on  to the far right. This brings up a file browser that has been set to point toward a directory containing several *Spartan* documents of common rings.

nitrogen heterocycles  
oxygen heterocycles  
sulfur heterocycles  
mixed heterocycles

saturated nitrogen rings  
saturated oxygen rings  
saturated sulfur rings  
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 model kit 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 model kit. 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.

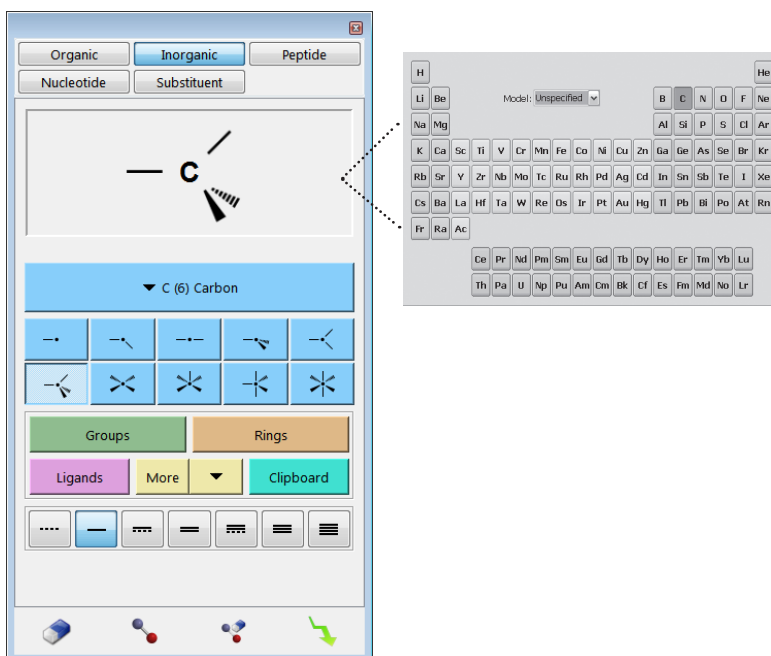
An empty clipboard will be signaled by:



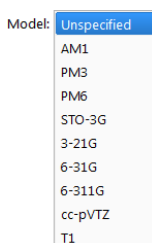
## Inorganic Model Kit

*Spartan*'s inorganic model kit allows construction of a much wider class of molecules (including inorganic and organometallic species)

than possible with the organic model kit. Structures that violate conventional bonding rules may be constructed, as this model kit purposefully provides no checking. The inorganic model kit is reached by selecting **Inorganic** from among the tabs at the top of the model kit.\*



Atoms may be selected by *clicking* on the atom button near the center of the model kit. 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 **Model** menu inside the *Periodic Table* contains a selection of theoretical models (basis sets except for semi-empirical models).



\* Tabs may require too much vertical space on computers or tablets with very small screens. Alternative **Builder Selection Methods** are available in the **Miscellaneous Preferences** dialog (**Preferences** from the **Options** menu; **Chapter 23**).

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 effected 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 append onto an existing molecular fragment. To initiate building, *click* anywhere on screen. To add alongside of an existing structure, hold down the **Insert** key (**option** key on Mac) and *click* anywhere on screen, or *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 effected 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 a menu near the bottom of the model kit. 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 model kit, and the user is free to construct any arrangement of atoms.

Menus under **Groups**, **Rings** and **More** are the same as those

---

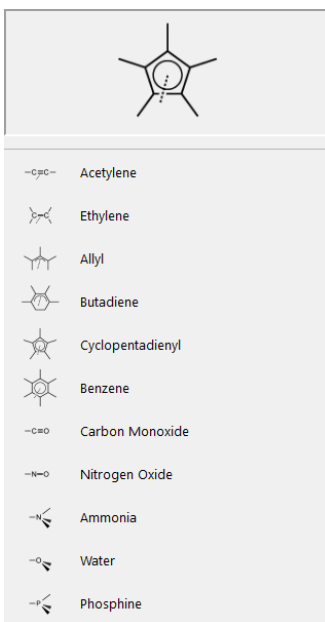
\* The existence of either an all-electron basis set or a pseudopotential is sufficient to classify a model as available for a particular element. Note that while molecular mechanics models are available for all elements, they have been carefully parameterized for only a relatively few of these. Use 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 Model Kit**).

described for the organic model kit as is **Clipboard**. One additional fragment collection is 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 model kit.



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 Point** in the **Geometry** menu; **Chapter 18**).

## Peptide Model Kit

This builder is not intended to be used for constructing proteins (although this is actually 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 15**). A model kit for construction of

polypeptides is reached by selecting **Peptide** from among the tabs at the top of the model kit.



At the middle of the peptide model kit are icons designating the amino acids (specified by their usual 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 the building blocks in the peptide model kit. Because these other building blocks are missing, modifications of peptides, aside from modifications in sequence and in overall conformation, need to be carried out using the organic or inorganic model kits.

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:

### Specify Amino Acid Sequence

This is accomplished by *clicking* in the desired order on the amino acid codes. 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 model kit. 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 model kit. (It may be changed back to l by *checking l*). d amino acids are indicated by **.d** following the code in the box.

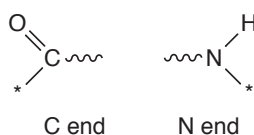
The sequence may be altered by changing the text in the box. Existing amino acid codes may be deleted or changed or new codes can be added. The entire sequence may be specified in this way if desired. Specification of a non-existent code will result in an error message. The sequence can be cleared by *clicking* on **Clear**.

### Specify Macroscopic Structure

Once sequencing is complete, macroscopic structure ( $\psi$  and  $\phi$  angles), is specified by *clicking* on one of  **$\alpha$  Helix**,  **$\beta$  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 need to be entered.

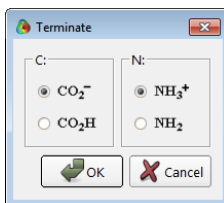
### Terminate


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.

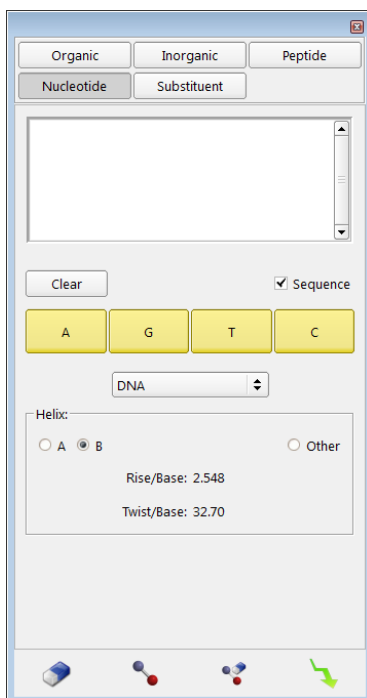


C and N terminating groups may be selected by repeated *clicking* on the C and N icons, respectively. Selection will rotate among the available terminating groups. *Clicking* on **OK** removes the dialog and terminates the polypeptide. *Clicking* on **Cancel** or  removes the dialog but does not terminate the polypeptide.

The peptide (or single amino acid) may now be used either to initiate building, by *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 also be joined onto 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 Model Kit

Finally, *Spartan* provides a model kit for construction of polynucleotides. It is reached by selecting **Nucleotide** from among the tabs at the top of the model kit.



At the middle of the model kit is a menu designating the type of polynucleotide.

DNA  
 DNA (single strand)  
 DNA-RNA  
 RNA  
 RNA (double strand)  
 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 model kit, 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 model kit. 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 model kits.

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:

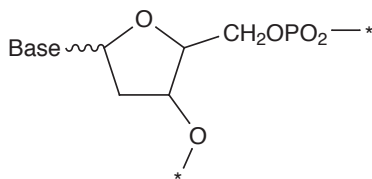
### **Specify Base Sequence**

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 **Clear**.

### **Specify 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.



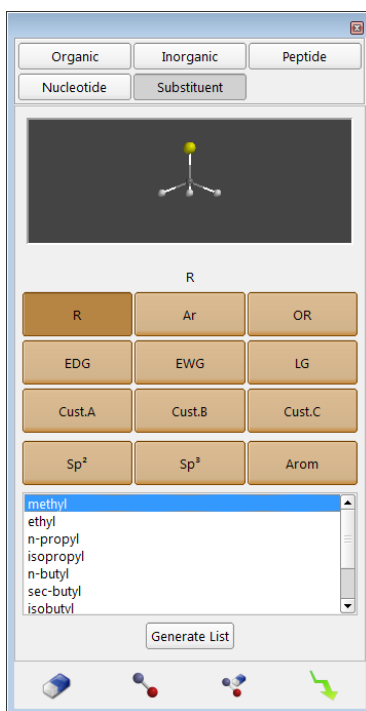
\* indicates a free valence. Hydrogens occupy all free valences (except the \*'ed positions at the two ends of the chain).

The polynucleotide (or single base pair) may now be used to either initiate building, by *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* on 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 Model Kit

The substituent model kit 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** model kit. 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 Molecular Database and Spartan Spectra and Properties Database (see **Databases** under the **Search** menu; **Chapter 22**) or as elements of reaction energy calculations (see **Reactions** under the **Display** menu; **Chapter 21**). 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 21** and **22**.

The substituent model kit is accessed by selecting **Substituent** from among the tabs at the top of the model kits.



At the center of the model kit 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 dialog. 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 21**).

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 may be displayed in the scroll box at the bottom of the model kit by *clicking* the appropriate button. A ball-and-wire model of a substituent may be seen at the top of the model kit 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 model kit. 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 may be used in the usual way.

Substituents need to be specified in terms of both their structure and their connection point. One way to do this is 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 ( from the organic model kit) 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 model kit.

The three buttons in the last row designate heteroatom equivalents of carbon-based hybrids.

**sp<sup>2</sup>**: carbon () , nitrogen ()  
**sp<sup>3</sup>**: carbon () , nitrogen () , oxygen ()  
**Arom**: carbon () , nitrogen ()

## Attachment

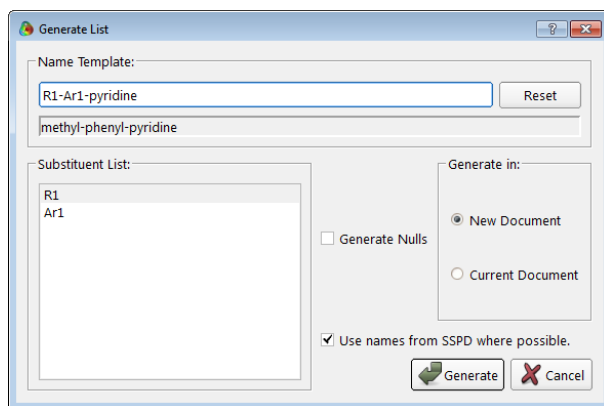
Following selection, a substituent library may be used to initiate building, to add alongside of an existing structure or to attach to an existing structure. The model is augmented by an icon containing the library name from the button, for example **EWG**, followed by a number. As many substituent libraries 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 library 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 Molecular Database (SMD) or Spartan Spectra and Properties Database (SSPD) and reaction energy calculations based on entries in SMD or 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 model kit, 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

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 Spartan Molecular Database (SMD) or in the Spartan Spectra and Properties Database (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 SMD or 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 SMD (or SSPD) where possible**

(depending on which database is “active”).

Nulls are molecules with less than the full complement of substituents, for example, monosubstituted molecules in the case where two different positions have be 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 are arbitrary and they 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 to be altered.

To access ChemDraw, select **ChemDraw** from the menu at the top of the model kit, and *click* on **Edit** at the bottom of the panel that results. ChemDraw will appear. To return to *Spartan*, close ChemDraw. The 2D drawing will appear at the center of the panel and manipulable 3D structure will appear at the top of the panel. *Clicking* on screen will move the 3D structure into *Spartan*'s main window.

---

\* ChemDraw files (.cdx) may be read with all versions of *Spartan*.

\*\* ChemDraw is not provided with *Spartan* but must be obtained from CambridgeSoft at <http://www.cambridgesoft.com>.

\*\*\* Transfer is one directional only. 3D structures that have been altered may not be transferred back to ChemDraw.

## 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 selected fragment, but may be made to apply to the entire set of fragments 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 model kits. *Double clicking* on an atom (not a free valence) while an atomic fragment in the organic model is highlighted, replaces the atom by selected fragment. Free valences are adjusted to accommodate the replacement, for example, replacement of  $\text{sp}^3$  carbon by  $\text{sp}^3$  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 model kit 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 kits.

## Invert Chirality

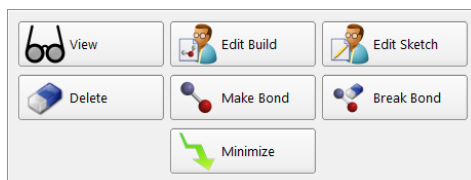
In the **Add Fragment** mode, *double clicking* on a chiral atom with the **Ctrl** key (**command** key on Mac) depressed inverts the chirality of the atom ( $\text{R} \rightarrow \text{S}$  or  $\text{S} \rightarrow \text{R}$ ). This is not available in fused ring systems. *Double clicking* on any atom 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 also found at the bottom of the model kit. They may also be included in the set of icons above the menus at the top of the screen.

### 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 15**). **Edit Build**, **Delete**, **Make Bond**, **Break Bond** and **Minimize** are for modifying existing structures.

### Edit Build (👤)

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 model kits, leads to **Edit Build**. (If a model kit is not already on screen, selection brings up the last-accessed model kit.) A fragment may be used to initiate building by *clicking* anywhere on screen, to add alongside an existing structure on screen by holding down the **Insert** key (**option** key on Mac) and by *clicking* anywhere on screen, 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. Note, however, that 2D  $\rightarrow$  3D conversion (as occurs upon exiting the sketcher) is *one way*. A 2D sketch may be brought back into the sketcher only if it has not been modified from one of the 3D model kits or has not been replaced with an entry from SMD or SSPD. In these cases, **Edit Sketch** is unavailable (greyed out). Of course, 3D structures originating from 2D sketches can be modified using the 3D model kits.

## 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. **Delete** is a one-time operation. Upon deleting, whatever was previously selected is again selected.

**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 (only) 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  $\rightarrow$  double, double  $\rightarrow$  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.

\* Free valences can be protected without altering the molecule by adding hydrogens to them ( from the organic model kit). One use of this is in conjunction with custom substituent libraries (see discussion under **Substituent Model Kit** earlier in this chapter).

## Minimize ()

This uses MMFF molecular mechanics to refine the geometry. Selection leads to a message at the bottom left of the screen.

Minimizer is active.

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  icon at the bottom right of the screen. Any geometrical constraints imposed on the structure (see **Constrain Distance**, **Constrain Angle**, **Constrain Dihedral** under the **Geometry** menu; **Chapter 18**) 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 18**) remain frozen.

Following completion, **Minimize** reverts back to **Edit Build** only if a model kit is on screen.

---

\* 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.

# Chapter 20

## The Setup Menu

---

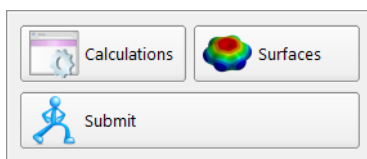
*This chapter describes functions available under the **Setup** menu. **Calculations...** is used to specify molecular mechanics calculations, semi-empirical calculations, Hartree-Fock molecular orbital calculations and correlated calculations, including density functional calculations, Møller-Plesset 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 and constructing energy profiles, although not all tasks are available for some methods. Hartree-Fock and density functional calculations may be extended to include an empirical solvation term, allowing equilibrium and transition-state structures and molecular properties to be obtained in the presence of a 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 a number of thermochemical recipes, including G3 and G3(MP2), 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. **Calculations...** also requests calculation of IR, Raman, NMR and UV/visible spectra (IR spectra may be calculated in the presence of solvent but solvated NMR, Raman and UV/visible spectra are not available) and calculation and printing of a variety of molecular properties, QSAR descriptors, thermodynamic quantities and gas-phase acidities and basicities. 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'14** take advantage of 64-bit architecture. Multi-core processors may be used to simultaneously process different jobs and, with the **Spartan'14 Parallel Suite**, may be directed at a single job (for most tasks involving Hartree-Fock, density functional and RI-MP2 models).*

***Surfaces** is used to designate graphical surfaces, including electron and spin densities, electrostatic potentials, local ionization potentials and molecular orbitals, for later 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 with the **Spartan'14 Parallel Suite**, on a remote machine running an embedded server.*

The **Setup** menu provides access to dialogs for specifying molecular mechanics and quantum chemical calculations, 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 SYBYL and MMFF 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; BP, BLYP, EDF1, EDF2 and B3LYP among other density functional calculations, including calculations based on a user-specified mix of existing exchange and correlation functionals; MP2 and RI-MP2 Møller-Plesset calculations; configuration interaction 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 MP3 and MP4 Møller-Plesset models, coupled cluster models and quadratic configuration interaction models.

A wide selection of basis sets is available, ranging from the STO-3G minimal basis set (recommended for Hartree-Fock models only), to the 3-21G split-valence basis set (recommended for Hartree-Fock models only), to 6-31G\*, 6-311G\*, and cc-pVDZ, cc-pVTZ, cc-pVQZ and cc-pV5Z polarization basis sets. 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 not presently available. 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.

Finally, a series of recipes for estimating heats of formation are available. These include the G3 and G3 (MP2) recipes as well as the T1 recipe that is based on G3(MP2) and provides nearly identical heats of formation, but is two to three orders of magnitude less costly in terms of computation.

Quantum chemical calculations also result in a variety of atomic and molecular properties, QSAR descriptors, thermodynamic quantities and gas-phase acidities and basicities as well as IR, Raman, NMR and UV/visible spectra. IR spectra are available for molecular mechanics, semi-empirical, Hartree-Fock, density functional 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 in one of two ways, both of which involve empirical formulations. The simpler and less costly in terms of computation is to add the aqueous solvation energy obtained using the semi-empirical SM5.4 model to the energy of any molecular mechanics

or quantum chemical calculation. Equilibrium and transition-state geometries as well as spectra and any properties derived from the wavefunction are unaffected. A second and computationally more costly method is based on the more recent SM8 model. Here, the correction affects the wave function in addition to the energy. This means that it can be used to obtain equilibrium and transition-state geometries as well as infrared spectra and most properties that depend on the wave function. The SM8 solvation correction is limited to Hartree-Fock and density functional models, and applies to water in addition to a variety of organic solvents. A listing of available solvents is provided in the **Properties FAQ** section accessed from the **Help** menu. Solvent-corrected Raman, NMR and UV/visible spectra are not available.

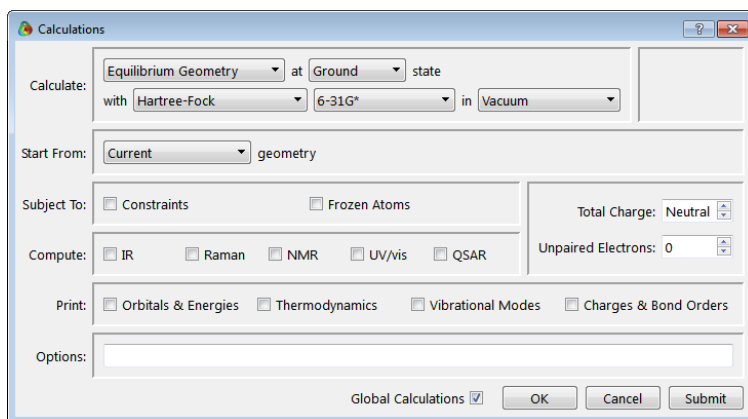
Discussion of molecular mechanics and quantum chemical methods in general, focusing on the specific methods available in *Spartan*, is provided both in **Topics** under the **Activities** menu and in *A Guide to Molecular Mechanics and Quantum Chemical Calculations* which is provided as a PDF under the **Help** menu. The guide also discusses how different methods may be combined, as well as illustrates graphical modeling techniques. A more thorough discussion of specific quantum chemical models is provided in the Q-Chem Users Manual (version 4.0).

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 either molecular mechanics or quantum chemical calculations but depends only on the geometry of the molecule (pharmacophore).

Selection of **Calculations...** results in the **Calculations** dialog\*.

---

\* The dialogs associated with similarity analysis and with conformer library generation are different from that associated with the other entries under the **Calculate** menu, and will be presented later in this chapter.



This contains a number of pull-down menus, buttons, check boxes and a space to enter additional information (“options”):

## 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.

## 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  
Conformer Library  
Energy Profile  
Similarity Analysis

**Energy** specifies calculation of energy (and in the case of quantum chemical methods, a wave function) at a single geometry.

*Spartan* reports energies from molecular mechanics calculations in kJ/mol, from semi-empirical calculations and from T1, G3(MP2) and other thermochemical recipes as heats of formation in kJ/mol, and from Hartree-Fock and correlated calculations as total energies in atomic units (hartrees).

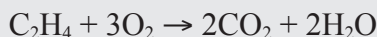
The molecular mechanics energy comprises two parts: the strain energy and the non-bonded energy. The strain energy is the difference in energy between a molecule and its strain free analog. It is nearly always positive

and less than a few hundred kJ/mols 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 to 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,

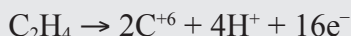


where graphite and hydrogen molecule are the carbon and hydrogen standards, respectively. In practice, the actual measurement is typically carried out for a combustion reaction, for example, for ethylene:



Heats of formation may either be positive or negative quantities and generally span a range of only 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:



Total energies, as the energies of such reactions are termed, are always negative and may be very large (tens of thousands of kJ/mol). They are most commonly given in atomic units (hartrees).

$$1 \text{ atomic unit} = 2625 \text{ kJ/mol}$$

Aside from a difference in units (see **Appendix E**), it makes no difference whatsoever which standard is employed to investigate thermochemistry.

To summarize, the heat of formation differs from the total energy both with regard to the standard reaction and with regard to units. Either provides a suitable basis for thermochemical calculations.

**Equilibrium Geometry** specifies that the nearest energy minimum will be located, and **Transition State Geometry** that the nearest

*first-order saddle point* (energy maximum in one dimension and energy minima in all other dimensions or commonly known as a *transition state*) will be located. **Equilibrium Conformer**, **Conformer Distribution** and **Conformer Library** attempt to characterize the conformers available to a molecule based on different criteria (see discussion following). **Energy Profile** steps along a user-defined coordinate or set of coordinates. **Similarity Analysis** quantifies the likeness among molecules based either on structure or chemical functionality or between molecules and a template (a pharmacophore).

Three different conformational analysis procedures are available in *Spartan*. The first two (**Equilibrium Conformer** and **Conformer Distribution**) may be employed with a variety of molecular mechanics and quantum chemical models (although practical concerns may limit application to molecular mechanics), while the third (**Conformer Library**) is restricted to the MMFF molecular mechanics model.

*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 23**). Finer control may be exercised using **Set Torsions** under the **Geometry** menu (**Chapter 18**).

**Equilibrium Conformer** replaces the initial conformer of a molecule by the lowest-energy conformer. This procedure is typically used to get a good guess at the best (lowest-energy) conformer for calculation of properties using quantum chemical methods.

**Conformer Distribution** provides a selection of low-energy conformers. This procedure is commonly used to identify a set of energetically accessible conformers and construct a *Boltzmann distribution* for calculation of average molecular properties.

Except for molecules with very few degrees of conformational freedom full *systematic searches* (looking everywhere) are not likely to be practical. 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 is 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. Both **Equilibrium Conformer** and **Conformer Distribution** may be restricted to “looking at” a user-specified maximum number of conformers (**Maximum Conformers Examined**). Selection forces a Monte-Carlo search, irrespective of the number of conformational degrees of freedom. **Conformer Distribution** also allows the user to select how many conformers are returned (**Percent Conformers Kept**). This is in terms of a cumulative Boltzmann distribution at room temperature, for example, **95** means keep all conformers that make up 95% of the room temperature Boltzmann distribution.

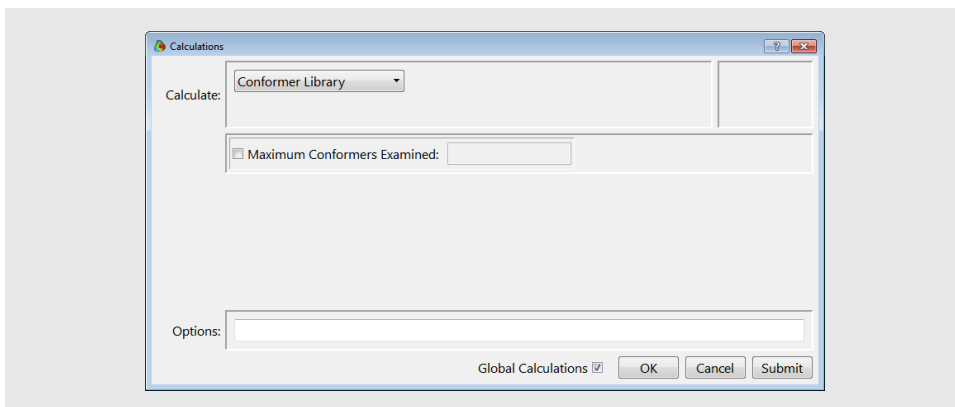
**Conformer Library** replaces the initial conformer of a molecule by the lowest-energy conformer and attaches the coordinates of a set of conformers spanning the possible shapes accessible to the molecule. Used to build libraries for similarity analysis.

The following recipe is used to generate a conformer library:

1. Step through all possible conformers in a systematic manner and calculate equilibrium structures and energies for each using the MMFF force field. Eliminate duplicates and very high-energy conformers.
2. Place spheres on each nitrogen, and oxygen atom and at the center of each hydrophobic region for each conformer.
3. Find the smallest set of conformers for which spheres do not overlap, that is, eliminate any conformer for which all of its spheres overlap with the corresponding spheres on another conformer.

Because this procedure necessarily involves a systematic search of all possible conformers, it may be desirable to place a limit on the total number of conformers actually considered. This is known as a *restricted systematic search*, a procedure that walks through all possible conformers but randomly discards some without further minimization in order to enforce a fixed number of conformers. This is triggered by setting **Maximum Conformers Examined**.

Selection of **Conformer Library** results in a different dialog.



**Energy** applies to all methods. **Equilibrium Geometry**, **Equilibrium Conformer**, **Conformer Distribution** and **Energy Profile** does not apply to advanced correlated methods and thermochemical recipes. **Transition State Geometry** does not apply to molecular mechanics methods, advanced correlated methods and thermochemical recipes. **Conformer Library** applies only to MMFF molecular mechanics. **Similarity Analysis** does not involve any molecular mechanics or quantum chemical methods.

## 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 . Different methods are available to handle ground and excited-state species.\* Hartree-Fock, Møller-Plesset as well as a variety of advanced correlated methods are available only for ground-state species, while configuration interaction methods are available only for excited-state species. Density functional methods are available for both ground and excited-state species. **State** is ignored for molecular mechanics calculations, and semi-empirical calculations are limited to the ground state.

---

\* Note that gradients for density functional methods for excited states need to be calculated numerically. This means that determining an equilibrium geometry for an excited state of a molecule is significantly slower than determining an equilibrium geometry for the ground state. Analytical gradients for CIS and CIS(D) methods for excited states are available.

## Type of Calculation

Specification of type of calculation is by way of a menu, the selections in which depend on electronic state. For ground states:

Molecular Mechanics  
Semi-Empirical  
Hartree-Fock  
Density Functional  
Møller Plesset  
Advanced Correlated  
Thermochemical Recipes

For excited states:

Density Functional  
Configuration Interaction

Some calculation types require additional information:

Molecular Mechanics

MMFF  
MMFFaq  
SYBYL

Møller-Plesset

MP2  
RI-MP2

Semi-Empirical

AM1  
RM1  
PM3  
PM6  
MNDO

Advanced  
Correlated\*

MP3  
MP4  
CCSD  
CCSD(T)

Density Functional\*\*

BP  
BLYP  
EDF1  
EDF2  
B3LYP  
M06  
wB97X-D

Thermochemical  
Recipes\*\*\*

T1  
G3(MP2)

Configuration  
Interaction

CIS  
CIS(D)

The most important application of **molecular mechanics models** is conformational searching, in particular, for molecules with several degrees of conformational freedom. In these cases, even the simplest quantum chemical models may not be practical. The MMFF force field is known to properly assign equilibrium conformation in a variety of simple molecules for which experimental data are available. It also has been shown to generally (but not always) provide a reasonable account of conformational energy differences in larger organic molecules as

\* A number of other advanced correlated methods are available, but need to be entered directly into the **Options** box: CCSD(2), OD, OD(T), QCCD, QCCD(2), QCISD and QCISD (T).

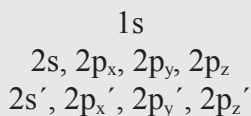
\*\* A variety of other functionals are available and may be entered directly into the **Options** box. A full listing is provided in **Appendix B**. Additionally, any combination of supported exchange and correlation functionals may be specified.

\*\*\* G2 and G3 recipes are also available but need to be entered into the **Options** box.

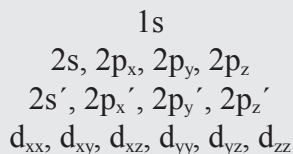
obtained from high-level quantum chemical calculations. Aside from conformational energy differences, molecular mechanics models are not suitable for thermochemical calculations.

**Semi-empirical models** are 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. The MNDO, AM1, RM1, PM3 and PM6 models generally provide geometries in good accord with experimental structures. The PM3 and PM6 models have 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. Also, none of the methods reliably accounts for conformational energy differences, as would be required to establish whether a particular conformer has a good chance at actually being present or to calculate the average value of a property based on the relative (Boltzmann) populations of different conformers. 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. This is known as 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 **Linear Combinations of Atomic Orbital approximation**, distinguishes different Hartree-Fock models. Different models use different basis sets, that is, different numbers and kinds of atomic orbitals. For example, the model with 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.



The model with 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.



Larger basis sets involving even higher splitting of valence shells, including higher-order (f and g type) functions as well as diffuse functions can be and have been defined and are available in **Spartan**. 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 and 6-31G\* and larger basis set Hartree-Fock models generally provide good accounts 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 6-31G\* model (and to a lesser extent, the 3-21G model) also properly accounts for thermochemistry, at least insofar as being able to say whether a particular reaction is weakly or strongly *exothermic* or weakly or strongly *endothermic*. It is also able to properly order the stabilities of isomeric products and to reliably identify the lowest-energy tautomer. Finally, Hartree-Fock 6-31G\* calculations provide a qualitatively correct account of the relative stabilities of different conformational arrangements. Hartree-Fock models with larger basis sets show similar behavior but 3-21G calculations do not provide reliable results.

Hartree-Fock models are 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 bond dissociation. This includes activation energies, that is, the difference in energy between reactants and transition state. This is the case even with very large basis sets, and 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. This leads 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 and the calculated dissociation energy will be too small.\*

Hartree-Fock models are available for the 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. Density functional models are to be preferred.

Models that do away with the Hartree-Fock approximation, or at least lessen its effect, are termed **correlated models**. These divide into two broad categories, **density functional models** and **wave function-based models**. Density functional models explicitly introduce an approximate (“empirical”) correlation term (a functional). They are not much more costly in terms of computation time than Hartree-Fock models, although it is not apparent how to improve on a particular choice of functional. Wave function-based models, typified by **Møller-Plesset models** and **configuration interaction models** extend the flexibility of Hartree-Fock models by mixing ground-state and excited-state wave functions. They are significantly more costly than both Hartree-Fock and density functional models, but offer the advantage over the latter of a clear path to improvement. In the limit of complete mixing, so-called full configuration interaction and infinite order Møller-Plesset models, wave function-based models lead to the exact result, although in practice this limit cannot be reached.

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 simplest and most popular models are the B3LYP/6-31G\* density functional model and the MP2/6-31G\* (second-order) Møller-Plesset model.

Density functional models and MP2 models provide an excellent account of the equilibrium geometries of organic molecules, and are generally to

---

\* Consider the bond dissociation energy of  $\text{H}_2$ . The (limiting) Hartree-Fock energy of the products (hydrogen atoms each with only one electron) is exact, while the energy of the reactant ( $\text{H}_2$  with two electrons) is too large. Therefore, the bond dissociation energy will be too small.

be preferred over Hartree-Fock models for this purpose. 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 and including activation energies.

Density functional 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.

A number of thermochemical recipes or combinations of different quantum chemical methods have been formulated to reproduce experimental heats of formation to within 4-8 kJ/mol. (Except for very simple molecules, experimental heats are seldom more accurate than this.) G3(MP2) is the simplest of the recipes that have been formulated. It 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. The problem with G3(MP2) is its large computational cost, which limits the size of systems that can be dealt with. G3(MP2) may easily be applied to molecules comprising up to ten heavy (non-hydrogen) atoms, but applications to molecules comprising fifteen or more heavy atoms will generally not be practical.

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;  $E_{\text{zero-point}}$ , and a correction to the enthalpy (energy) to account for finite temperature (298K),  $H(T)$ .

$$E_{\text{zero-point}} = \frac{1}{2} h \sum_i^{\text{vibrations}} \nu_i$$

$$\Delta H(T) = H_{\text{trans}}(T) + H_{\text{rot}}(T) + \Delta H_{\text{vib}}(T) + RT$$

$$H_{\text{trans}}(T) = \frac{3}{2} RT$$

$$H_{\text{rot}}(T) = \frac{3}{2} RT \text{ (RT for a linear molecule)}$$

$$\Delta H_{\text{vib}}(T)^* = H_{\text{vib}}(T) - H_{\text{vib}}(T) - H_{\text{vib}}(0)$$

$$= Nh \sum_i^{\text{vibrations}} \frac{\nu_i}{(e^{h\nu_i/kT} - 1)}$$

$h$  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 based on G3(MP2) and intended to closely reproduce G3(MP2) heats of formation. It eliminates the QCISD(T)/6-31G\* energy calculation ( $E_C$ ), the zero-point energy calculation ( $E_{\text{zero-point}}$ ) 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  $\approx 1,050$  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  $\sim 2,000$  experimental heats of formation in the NIST thermochemical database with a mean absolute error of 9 kJ/mol. The T1 recipe also provides a very good account of conformational energy differences (although the

---

\* Note, that the expression for  $\Delta H_{\text{vib}}(T)$  goes to  $RT$  instead of  $1/2 RT$  as  $\nu_i$  goes to zero. This requires that contributions that would exceed  $1/2 RT$  be set to  $1/2 RT$ .

experimental dataset is very limited), and an automated procedure has been developed to locate the best (lowest heat of formation) conformer, or at least a conformer that is very close to the best. This involves four steps: (i) examine up to 1,000 conformers using a restricted systematic search procedure and the MMFF molecular mechanics model; (ii) obtain equilibrium geometries for (up to) the six best conformers, making certain to exclude duplicate conformers and conformers that are nearly the same using the HF/6-31G\* model; (iii) perform (up to) two T1 calculations on the two best HF/6-31G\* structures; (iv) select the lower energy T1 conformer. The default numbers of structures examined in the second and third steps (six and two, respectively) may be altered by keywords typed into the **Options** box (**Appendix D**).

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** box (**Appendix D**). Semi-empirical methods and thermochemical recipes are limited to the default settings.

## 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:

STO-3G
3-21G
6-31G*
6-31G**
6-31+G*
6-311G*
6-311+G**

---

\* A number of other basis sets are available, and may be entered directly into the **Options** box: These include the Pople 6-31G or 6-311G basis sets followed by: (*idjf*) or (*idjf, kp/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 s functions on hydrogen atoms and *l* is the number of sets of p 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; cc-pVDZ, cc-pVQZ and cc-PV5Z.

For all other methods:

```
6-31G*  
6-31G**  
6-31+G*  
6-311G*  
6-311+G**  
6-311++G**  
6-311++G(2df,2p)  
cc-pVTZ
```

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, EDF2, 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, EDF2/6-31G\*. 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 without basis sets, for example, MMFF (molecular mechanics) and PM3 (semi-empirical) are themselves models.

## Dual Basis Sets

If *checked*, **Dual Basis** signifies that a Hartree-Fock, density functional, MP2 or RI-MP2 energy calculation (only) with the 6-311+G\*\*, 6-311++G\*\*, 6-311++G (2df, 2p), or cc-pVTZ basis sets will be approximated by a calculation in which SCF convergence is first achieved using the smaller 6-311G\* basis set and then corrected perturbatively for the effects of basis set extension. Use of dual basis has only a small effect on calculated relative energies\*, but 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). A dual basis set is employed for calculations with the T1 thermochemical recipe.

---

\* While it is possible to specify other combinations of small and large basis sets (see the **Small Basis** keyword in **Appendix D**) it is important that two be closely related.

## 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, a EDF2/6-31G\* 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 **EDF2/6-31G\***. Pseudopotentials pertain to Hartree-Fock, density functional, MP2 and advanced correlated calculations only, but not to RI-MP2 calculations or calculations using the STO-3G and 3-21G basis sets.

## Solvent

A solvent is specified for Hartree-Fock and density functional calculations by replacing *vacuum* in the menu to the right of **in** by water or one of a selection of common organic solvents.\*

Vacuum
Water
Acetone
Diethyl Ether
DMF
DMSO
Ethanol
Methylene Chloride
THF
Toluene

The effect of the solvent on the energy and wave function (and hence on any properties derived from the wave function) is approximated by way of the SM8 model. This has been parameterized to reproduce experimental energies of solvation and related quantities. It has not been parameterized to reproduce equilibrium geometries and properties other than the energy, and there is no solid calibration of its success or failure for anything other than energy. It should be used with caution.

The cost of SM8 in terms of computation time is roughly equal to that of an energy calculation. Therefore, solvated Hartree-Fock and density functional energy calculations are twice as costly as the corresponding

---

\* Other solvents are available and can be entered via the **Solvent** keyword in the **Options** box.

calculation in vacuum while geometry calculations are roughly 30% more costly.

## Intrinsic Reaction Coordinate

An additional control, **IRC** (Intrinsic Reaction Coordinate), 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.) 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.

## Start from (Energy, Equilibrium Geometry)

Energy and equilibrium geometry calculations (only) from one model may be preceded by equilibrium geometry calculations at another (simpler) model. Choices depend on the model:

**Molecular Mechanics**

not applicable

**Semi-Empirical**

Current  
MMFF  
MMFF Conformer  
MMFFaq Conformer

**Hartree-Fock**

Current  
MMFF  
AM1  
PM3  
MMFF Conformer  
MMFFaq Conformer

**Density Functional, Møller-Plesset,  
Advanced Correlated, Thermochemical  
Recipes**

Current  
MMFF  
AM1  
PM3  
3-21G  
6-31G\*  
MMFF Conformer  
MMFFaq Conformer

Selection of **MMFF conformer** or **MMFFaq conformer** signifies that geometry optimization will follow a search for the lowest-energy conformer. These lead to an additional menu specifying the level of geometry optimization.

**Molecular Mechanics, Semi-Empirical**      not applicable

**Hartree-Fock**

MMFF AM1 PM3 3-21G
-----------------------------

**Density Functional, Møller-Plesset, Advanced  
Correlated, Thermochemical Recipes**

MMFF AM1 PM3 3-21G 6-31G*
---------------------------------------

Note that calculation of equilibrium geometry is part of a thermochemical recipe.

**Maximum Conformers Examined (Equilibrium Conformer, Conformer Distribution, Conformer Library)** sets a maximum for the number of possible conformers that will be considered both for **Equilibrium Conformer** and **Conformer Distribution** tasks. Note, that if a systematic search has been requested (**SEARCHMETHOD** keyword; **Appendix D**) this invokes an algorithm whereby systematic conformers are randomly eliminated in order to enforce the conformer limit.

**Maximum Conformers Kept** sets a maximum for the number of conformers returned at the end of a **Conformer Distribution** task. It must be smaller or equal to the number of conformers examined.

### 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. Does not apply to energy calculation, conformer library generation or

similarity analysis. See **Constrain Distance**, **Constrain Angle** and **Constrain Dihedral** under the **Geometry** menu (**Chapter 18**) for information about constraining geometrical parameters.

NOE data 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. Does not apply to energy calculation, conformer library generation or similarity analysis. See **Freeze Center** under the **Geometry** menu (**Chapter 18**) for information about freezing atoms. See also discussion in **Chapter 22** regarding the freezing of heavy atoms in structures resulting from a search of the Cambridge Structural Database.

## 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.

### 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 21**) and in the output (**Output** under the **Display** menu; **Chapter 21**). 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 21**)\*. Vibrational modes may be animated and an IR spectrum displayed from either the simplified or more general form of the **Spectra** dialog (**Spectra** under the **Display**

---

\* Entropy (and Gibbs energy) calculations are problematic in part due to the harmonic approximation. Discussion is provided under **Properties and Spectra** in **Appendix A**.

menu; **Chapter 21**). Second derivatives required for frequency calculation are evaluated analytically for molecular mechanics, semi-empirical, Hartree-Fock and density functional models, but involves numerical differentiation of analytical gradients for RI-MP2 and MP2 models. This is much more costly in terms of computation. Frequency calculations for higher-order correlated models are not available.

## Raman

If *checked*, calculates vibrational frequencies and Raman intensities. These are available both in the **Spectra** dialog (**Spectra** under the **Display** menu; **Chapter 21**) and in the output (**Output** under the **Display** menu; **Chapter 21**). The Raman spectrum may be displayed from either the simplified or more general form of the **Spectra** dialog (**Spectra** under the **Display** menu; **Chapter 21**). Note that request for a Raman spectrum also results in an IR spectrum.

## NMR

If *checked*, calculates NMR chemical shifts. These are then available in the **Spectra** dialog (**Spectra** under the **Display** menu; **Chapter 21**), in the output (**Output** under the **Display** menu; **Chapter 21**), from the **Atom Properties** dialog (accessible from **Properties** under the **Display** menu; **Chapter 21**) and as atom labels (**Configure...** under the **Model** menu; **Chapter 17**).  $^{13}\text{C}$  (proton decoupled) and  $^1\text{H}$  (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 21**)\*. DEPT spectra ( $^{13}\text{C}$  spectra in which carbons with one or three attached protons point up, carbons with two attached

---

\* Chemical shifts for other nuclei are available in the **Output** dialog (**Output** under the **Display** menu; **Chapter 21**) and may also be attached as labels (**Configure...** under the **Model** menu; **Chapter 17**). All chemical shifts are given relative to standard compounds, for example, TMS in the case of both proton and carbon.

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.  $^{13}\text{C}$ ) spectra may be displayed from both simplified and more general forms of the **Spectra** dialog.

$^{13}\text{C}$  chemical shifts calculated from the B3LYP/6-31G\* have been empirically corrected for local topology (numbers and kinds of directly-bonded atoms).  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{19}\text{F}$  shifts calculated from the EDF2/6-31G\* model have been corrected using a more elaborate scheme involving both topology and bond orders to neighboring atoms. Corrections for other quantum chemical models are not available. Experience suggests that uncorrected chemical shifts (in particular,  $^{13}\text{C}$  shifts) are not sufficiently accurate to be useful for NMR spectra prediction, but that corrected shifts are generally successful in this regard. Uncorrected shifts are available in the printed output. Line intensities are assumed to be proportional to the number of equivalent carbons or hydrogens. HH and CH coupling constants are not calculated, but are estimated empirically from the molecular geometry. Splitting patterns are based on the number of neighboring hydrogens and on estimated 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 21**) and from the output (**Output** under the **Display** menu; **Chapter 21**). A UV/visible spectrum can be displayed from either the simplified or general form of the **Spectra** dialog (**Spectra** under the **Display** menu; **Chapter 21**). The default number of excited states is 6, but may be changed using the **UV States** keyword from the **Options** box (see discussion later in this chapter and **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.\*

## Total Charge

Total charge on the molecule. The vast majority of molecules are uncharged (neutral).

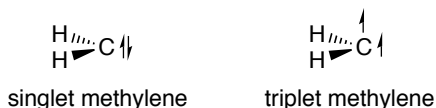
Molecules constructed from any of the 3D model kits (or from ChemDraw) are assumed to be neutral, that is **Total Charge** is **Neutral**. **Total Charge** may be changed either by *clicking* on  to the right of the box (up for **Cation**, **Dication**; down for **Anion**, **Dianion**) or by typing a 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<sub>2</sub>, may either exist as a singlet or as a triplet.



\* 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. See **Table 21-3**.

The most common odd-electron molecules are referred to as doublets. 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 on  to the right of the box.

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 previous versions of *Spartan* used the term **Multiplicity** to refer to the number of unpaired electrons: singlet = no unpaired electrons; doublet = 1 unpaired electron; triplet = 2 unpaired electrons.

## Print

If *checked*, writes the quantity to the output window. Text output may be seen by selecting **Output** from the **Display** menu (**Chapter 21**) and printed by first selecting the output dialog and then selecting **Print Output** from the **File** menu (**Chapter 15**). Additional printing may be requested from the **Options** box (see discussion later in this chapter and **Appendix D**).

## Orbitals & Energies

If *checked*, writes the orbitals and energies to the output. Orbital energies are also available in the spreadsheet (**Chapter 21**). An orbital energy diagram is available from **Orbital Energies** under the **Display** menu (**Chapter 21**).

## Thermodynamics

If *checked*, writes thermodynamic quantities to the output. Requires that vibrational frequencies (**IR** under **Compute**) have been calculated. Selected thermodynamic quantities are also available in the spreadsheet (**Spreadsheet** under the

**Display** menu; **Chapter 21**).

### **Vibrational Modes**

If *checked*, writes vibrational frequencies and modes to the output. Requires that **IR** under **Compute** have been calculated.

### **Charges & Bond Orders**

If *checked*, writes Mulliken, natural and electrostatic-fit charges and Millikan and Löwdin bond orders to the output. Note that Mulliken and natural charges from semi-empirical models are identical. Atomic charges are also available from the **Atom Properties** dialog (accessible from **Properties** under the **Display** menu; **Chapter 21**), in the spreadsheet (**Spreadsheet** under the **Display** menu; **Chapter 21**) and as atom labels (**Configure...** under the **Model** menu; **Chapter 17**). Mulliken and Löwdin bond orders are available from the **Bond Properties** dialog (accessible from **Properties** under the **Display** menu; **Chapter 21**). Atomic charges and bond orders are not available for calculations using advanced correlated models or for thermochemical recipes.

### **Options**

Program options may be specified using keywords entered into the **Options** box. 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.

### **Similarity**

*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).

<i>template</i>	<i>library</i>
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

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 all possible shapes that the molecule may take on. The latter are generated using **Conformer Library** under the **Calculate** menu (see discussion earlier in this chapter), 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.

hydrophobe (sterically crowded region)  
 aromatic hydrophobe (aromatic ring)  
 hydrogen-bond donor  
 hydrogen-bond acceptor

positive ionizable site  
negative ionizable site  
excluded volume

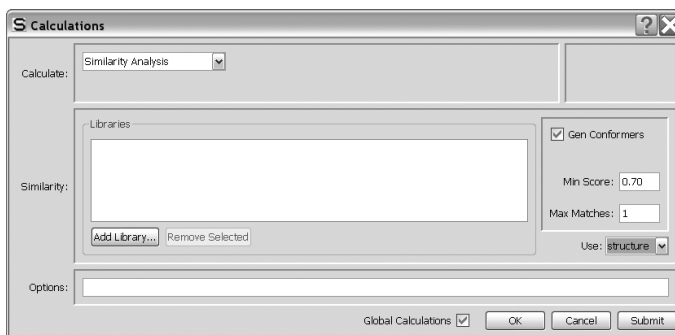
Discussion of CFD types has already been provided (see **CFD's** under the **Model** menu; **Chapter 17**).

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 18**). The **CFD Properties** dialog (accessible from **Properties** under the **Display** menu; **Chapter 21**) allows for changing or extending the definition of a CFD. For example, the nitrogen in a dimethylamino 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 any chiral molecules in a library are unresolved (they comprise an equal mixture of enantiomers). By default, similarity comparisons are carried out on both the library molecules as they are stored and (if the molecules are chiral) on their mirror images. The default procedure may be overridden (providing comparisons only to the stored molecules) by way of the **SingleEnantiomer** keyword entered in the **Options** box. 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** 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.

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**.

## Global Calculations

If *checked*, signifies that settings in the **Calculations** dialog are to be applied to all molecules in the document.

The **Calculations** dialog may be exited by clicking on **Submit**, **Cancel** or **OK** at the bottom right of the dialog, or on  at the top. (**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* allows 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.

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 immediately surrounding the atoms. We refer to the **electron density** as a surface that contains most (>99%) of a molecule's electrons and that roughly corresponds to a space-filling model, that is, a van der Waals surface. Such a surface reveals overall molecular size and shape and demarks the steric barrier seen by encroaching molecules. Another important surface, that we will refer to as the **bond density**, contains fewer (~80%) electrons in total and demarks atomic connectivity.

The **spin density** is the difference in the number of electrons of  $\alpha$  and  $\beta$  spin at a point in space. It indicates the location of the unpaired electron in a radical or unpaired electrons in a triplet or higher multiplicity state.

The **electrostatic potential** is the energy of interaction of a positive charge with a molecule. This assumes a fixed electron distribution for the molecule. It represents a balance between repulsive interactions involving the positively-charged nuclei and attractive interactions involving the 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

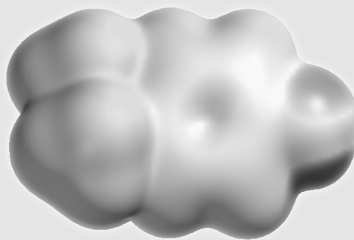
said to be electron poor (acidic) and subject to attack by nucleophiles. Electron-rich regions such as lone pairs are typically located outside the van der Waals surface. As such, they may be easily identified by constructing a surface of negative (attractive) electrostatic potential. While interesting electron-poor areas such as acidic hydrogens also lie outside the van der Waals surface, the electrostatic potential is also positive (repulsive) throughout the region inside this surface.

The **local ionization potential** indicates the ease or difficulty of electron removal (ionization). Like the negative regions of the electrostatic potential, regions of low local ionization potential are likely to be subject to attack by electrophiles.

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.

Additionally, aside from the electron density, any one of the quantities listed above may be mapped onto any surface (except a molecular orbital surface). In practice, the only maps to have received widespread attention are those based on the electron density surface (depicting overall molecular size and shape). Most common are the electrostatic potential map, the local ionization potential map and the LUMO map.

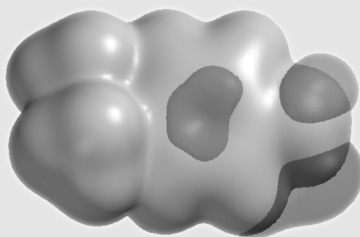
The **electrostatic potential map** paints the value of the electrostatic potential onto an electron density surface. By convention, colors toward red depict negative potential, while colors toward blue depict positive potential, and colors in between (orange, yellow, green) depict intermediate values 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  $\pi$  faces of benzene to be orange or yellow and the *tert*-butyl group to be green.



The main advantages of this presentation relative to separate electron

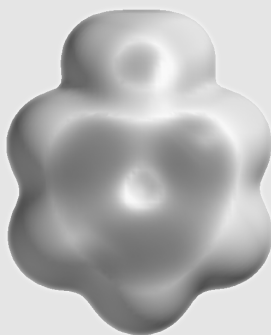
density and electrostatic potential surfaces are its clarity and its compactness. The main disadvantage is that it provides information only about the contact surface and does not reveal how far electron-rich and electron-poor areas extend beyond the surface.

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.

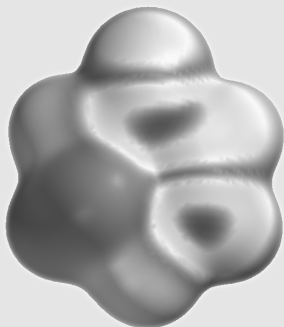


Note that the exposed electrostatic potential surface provides the same information as the electrostatic potential map. Red areas in the map correspond to regions when the negative electrostatic potential surface is likely to protrude from the electron density while blue areas correspond to regions where the positive electrostatic potential surface is likely to stick out. There is broad similarity of this graphic to CFD representations discussed previously (see **CFD's** under the **Model** menu in **Chapter 17**). 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.

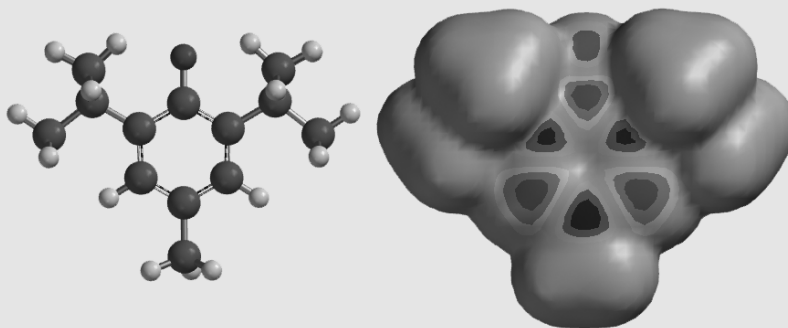
The ***local ionization potential map*** paints the value of the local ionization potential onto an electron density surface. By convention, colors toward red indicate low ionization potential, while colors toward blue indicate high ionization potential. Thus, the local ionization potential map for aniline 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. For example, a LUMO map for cyclohexenone shows concentration in two regions, one over the carbonyl carbon and the other over the  $\beta$  carbon, consistent with both carbonyl addition and Michael (conjugate) addition.



The **spin density map** paints the value of the spin density onto an electron density surface. 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 and to delineate overall molecular size and shape. Not all regions on this surface are 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 20-1**. Surfaces (and maps) may also be rendered in perspective (see **Chapter 17**) and in stereo (see **Chapter 2**).

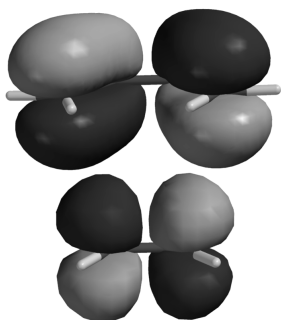
Calculated quantities may also be displayed as two dimensional contour plots (slices). Unlike surfaces and maps, these can be translated, rotated and zoomed independently of the molecular skeleton. An example of a slice display is provided in **Figure 20-1**.

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.

\* 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 23**).

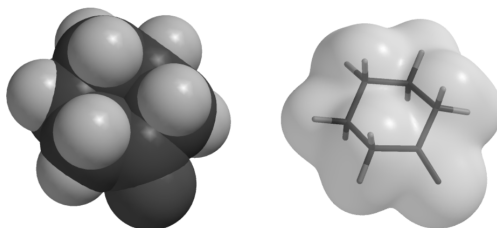
**Figure 20-1: Examples of Graphical Displays Available in *Spartan***

**Frontier orbitals for a symmetry-allowed Diels-Alder reaction,**



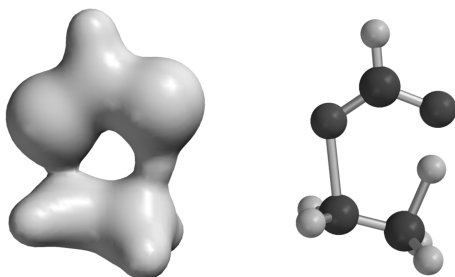
showing interaction of the HOMO of 1,3-butadiene and the LUMO of ethylene.

**Space-filling model and electron density surface of cyclohexanone,**



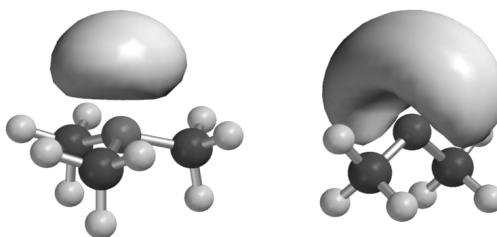
showing overall molecular size and shape.

**Electron density surface (0.08 electrons/  
au<sup>3</sup>) of transition structure for pyrolysis  
of ethyl formate,**



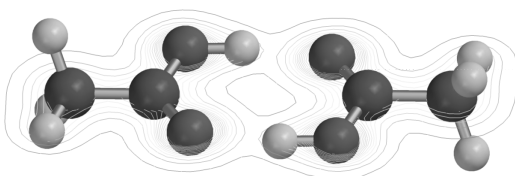
showing bonding in the transition state.

**Electrostatic potential surfaces (-40 kJ/  
mol) of trimethylamine (left) and dimethyl  
ether (right),**



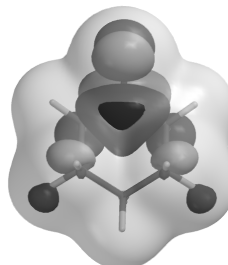
showing the lone pairs on nitrogen and oxygen, respectively.

**Electron density slice for acetic acid  
dimer,**



showing hydrogen bonding.

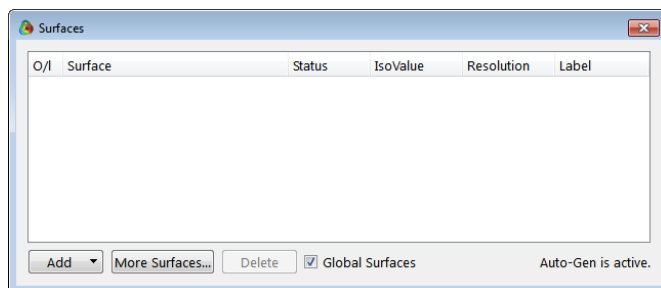
**Simultaneous display of the LUMO  
and the electron density surfaces of  
cyclohexanone,**



showing accessibility for nucleophilic  
attack.

Discussion of the utility of graphical models for describing molecular structure and chemical reactivity is provided in *Topics* available under the **Activities** menu and in *A Guide to Molecular Mechanics and Quantum Chemical Calculations*, provided as a PDF under the **Help** menu.

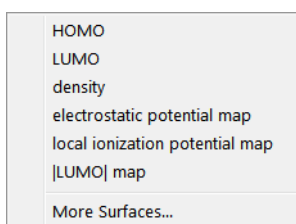
Selection of **Surfaces** from the **Setup** menu results in display of the **Surfaces** dialog.



This contains a box at the top for listing requested surfaces and property maps.

### 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.

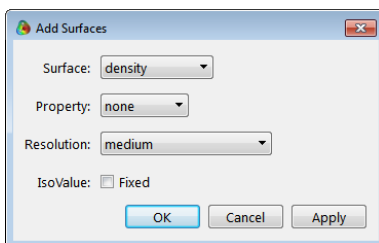


Selection of all but the last entry in the menu leads to a request for the analogous surface or map. This will be calculated at medium resolution unless a different resolution has been selected (see below). Additional surfaces and maps or the same surfaces or maps at different resolution may be requested by selecting **More Surfaces...** from the menu (or by *clicking* on **More Surfaces...** at

---

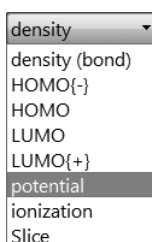
\* The number of selections will increase if the molecule has unpaired electrons.

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.

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.

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.



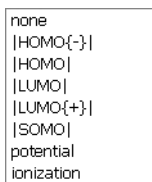
\* An “electron density” surface, an electrostatic potential surface and an electrostatic potential map may be constructed based on a molecular mechanics calculation. In this case, the density will be based on van der Waals radii and the potential will be based on charges.

\*\* These menu entries appear only for molecules with unpaired electrons.

**Slice** designates that a plane will cut through the graphic defined by **Property** (see discussion following).

## Property

Properties for maps appear in the **Property** menu.



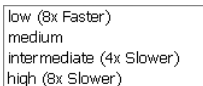
```
none
|HOMO{-}|
|HOMO|
|LUMO|
|LUMO{+}|
|SOMO|
potential
ionization
```

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). As with **Surface** above, selection of **HOMO{-}** and **LUMO{+}** leads to a decrement (increment) box.

A **Spin** button 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** box (**Calculations** dialog; see also **Appendix D**) and if **HOMO{-}**, **HOMO**, **LUMO** or **LUMO{+}** has been selected for **Surface** or for **Property**. *Clicking* on **Spin** 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  $\alpha$  spin; **Beta** designates that the molecular orbital corresponds to  $\beta$  spin.

## Resolution

Selection of surface resolution is from the **Resolution** menu.



```
low (8x Faster)
medium
intermediate (4x Slower)
high (8x Slower)
```

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).<sup>\*</sup> Fixed isosurfaces are particularly attractive for property maps (that are generally based on a density surface representing molecular size and shape) as they require 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.<sup>\*\*</sup>

*Checking* the box to the left of **Fixed** leads to display of the isovalue (0.002 electrons/au<sup>3</sup> in the case of density). 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 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

---

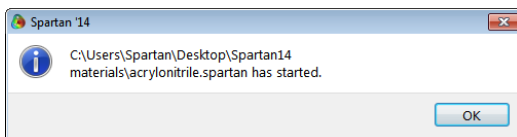
\* The “electron density” surface resulting from a molecular mechanics calculation is fixed as it is based on use of van der Waals radii.

\*\* 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 21**).

removed by *clicking* on .

## 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 Molecular Database or 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.\*



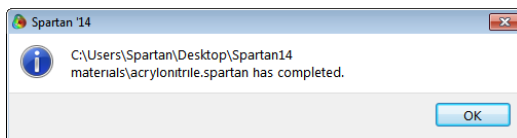
*Click* on **OK** to remove it. After a job has started, and until it has completed, all associated files will be designated read only.

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 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 23**), 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 **Local Monitor** and **Network Monitor** under the **Options** menu (**Chapter 23**) for discussion.



Click on **OK** to remove it.

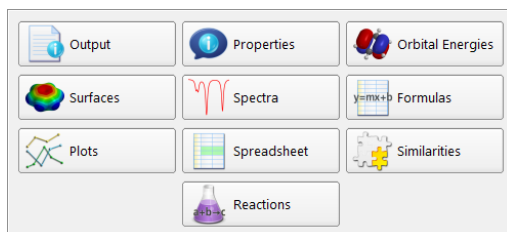
Some types of jobs, including requests for a conformer distribution and an energy profile, lead to an additional document being created for each molecule in the original document. These new documents are named *document.task.identifier* where *document* is the name given to the original document, for example, *Conf* for a conformer distribution 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, 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 15**).

# Chapter 21

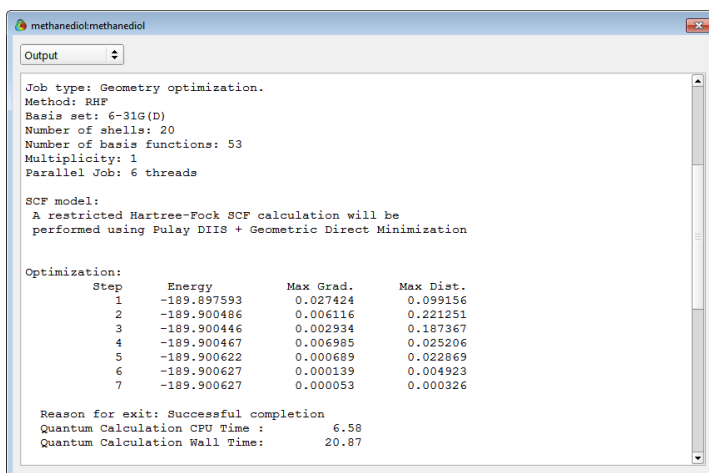
## The Display Menu

Functions available under the **Display** menu provide for text, dialog, spreadsheet and graphical displays following structure/property calculation, conformational analysis or similarity analysis. Functions are also available to query a variety of on-screen objects, display and compare IR, Raman, NMR and UV/visible spectra, animate vibrational motions, construct relationships among calculated properties or for a property between reactants and products of a chemical reaction, perform linear regression and prepare plots from spreadsheet data, display similarities between molecular structures and between molecular structures and pharmacophores and calculate reaction energies.

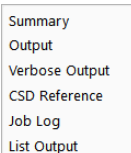



### Output ( )

Selection of **Output** opens a window.



A menu at the top left of the window selects the type of output.



**Summary** provides a brief summary of the calculated data, in particular, the energy and any spectral quantities. **Output** provides normal output. **Verbose Output** contains more detailed output, but is eliminated upon normal completion unless **Keep Verbose** is checked in the **Setting Preferences** dialog (**Preferences** under the **Options** menu; **Chapter 23**).<sup>\*</sup> **CSD Reference** provides the literature reference for data retrieved from CSD. **Job Log** and **List Output** contain 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...** and **Find Next** 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 may be open and may be closed by *clicking* on  at the top right.

Output is not accessible for jobs that are executing or are in the execution queue. Output for jobs that execute locally (only) may be examined in “real time” using the **Local Monitor** under the **Options** menu (**Chapter 23**). In addition, the current structure (in a geometry or transition-state optimization) can be examined and queried. Output from jobs executing on a remote server may not be examined until they have been completed (or terminated) and information returned.



## Properties ()

**Spartan** provides specialized dialogs for reporting (and in some cases 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

<sup>\*</sup> Irrespective of the setting, the last 100 lines of the verbose output is kept if the jobs fails for any reason.

from the spectra and plot panes), **Properties** may be used to change default plot styles, limits and fitting functions. 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 (spectra, graphical surface, substituent, etc.) of the selected molecule (**Atom Properties**, **Bond Properties**, **Surface Properties**, etc.), or to a plot (**Plot Properties**, **Curve Properties**, etc.), or fitting function (**Regression Properties**). Selection of a different molecule leads to 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 **Molecule 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 back to the **Molecule Properties** dialog.\*

Most **Properties** dialogs have an associated **Utilities** or **Style** dialog. For example, associated with the **Molecule Properties** dialog is a **Molecule Utilities** dialog. These access additional information about the molecule and its components/attributes, or provide style and color controls. This is useful for highlighting (or dehighlighting) a particular molecule, component or attribute. **Utilities/Style** dialogs 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 from the screen by *clicking* on .

### **Molecule Properties, QSAR Descriptors, Thermodynamics and Acidity/Basicity**

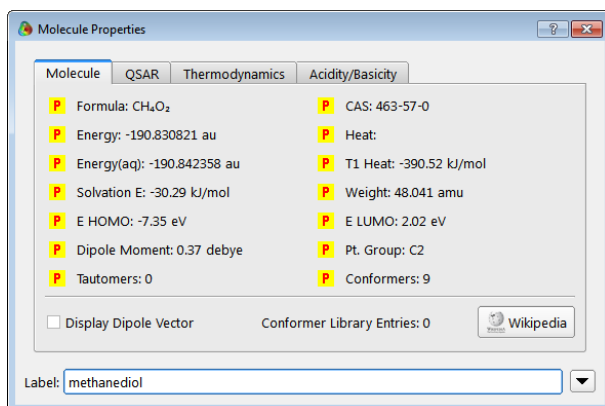
The **Molecular Properties** dialog comprises four parts: **Molecule**,

---

\* The only exception involves *clicking* on a graphical surface or property map, for example, *clicking* on a property map to obtain the value of the property at a particular surface location. *Clicking* a second time on the surface or map will report a new value of the property. *Clicking* on the background leads to the **Molecule Properties** dialog.

**QSAR, Thermodynamics and Acidity/Basicity**, controlled by tabs at the top left. Entries under the **Molecule** tab relate to common molecular properties, only some of which depend on the selected level of calculation.

## Molecule

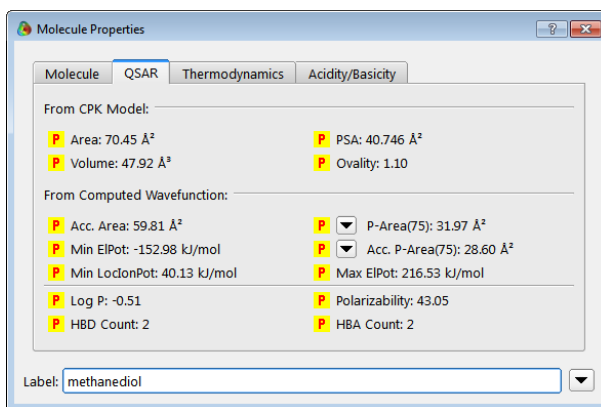


Molecule properties include the energy (in au), the heat of formation (in kJ/mol), or the sum of the strain energy and the intramolecular interaction energy (in kJ/mol) depending on the theoretical model, the energy (heat of formation, strain energy) corrected for an aqueous environment according to the SM5.4 model (in the same units as the non-corrected quantity), the HOMO and LUMO energies (in eV), the dipole moment (in Debye), the formula, the molecular weight (in amu), the point group, the predicted number of tautomers and the predicted number of conformers, and (if available) the experimental heat of formation (in kJ/mol), the heat of formation from the T1 thermochemical recipe (in kJ/mol) and the CAS number. These may be posted to the spreadsheet using the **P** buttons, or *dragged* into the spreadsheet, into the formula editor (see **Formulas** in this chapter), or onto any of the dialogs associated with the Spartan Molecular Database or Spartan Spectra and Properties Database (see **Database** under the **Search** menu; **Chapter 22**). The dipole moment vector may be added to the model by *checking* the box to the left of **Display Dipole Vector**. **Conformer Library Entries** is the number of entries in a conformer library (see discussion under **Calculations...** from the **Setup** menu; **Chapter 20**). Buttons at the bottom right of the

dialog access appropriate Wikipedia and ChemSpider pages based on InChi string. **Label** identifies the molecule in a document and appears in the first column of the spreadsheet (see **Spreadsheet** in this chapter).

## QSAR

Entries under the **QSAR** tab provide additional properties, some of which may be particularly valuable in QSAR type analyses.



These include: the area, volume, polar surface area (PSA)\* and ovality obtained from a space-filling model, and 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. Additional quantities which depend on the level of calculation and are 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

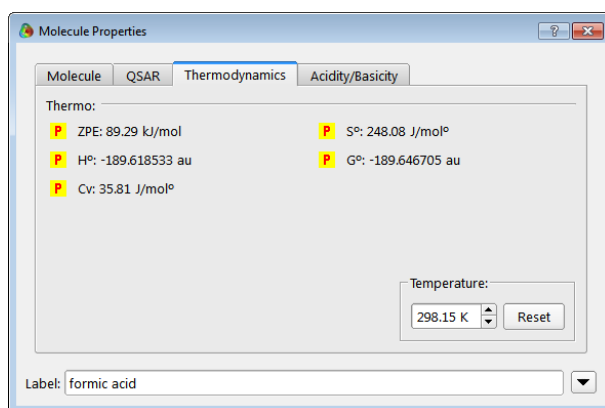
\* Polar surface area is defined as the area due to nitrogen and oxygen and any attached hydrogens. Polar surface areas corresponding to arbitrary alternative definitions are available for posting into the spreadsheet using the **PAREA** function. See **Table 21-3**.

\*\* Counts of hydrogen-bond donors and acceptors and positive and negative ionizable sites derive from CFD assignments. These in turn follow from atomic connectivity together with special case designations for common organic functional groups. CFD's assignments may be modified on a case-by-case basis using the **CFD Properties** dialog (**Properties** under the **Display** menu; this chapter).

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 20**).

## Thermodynamics

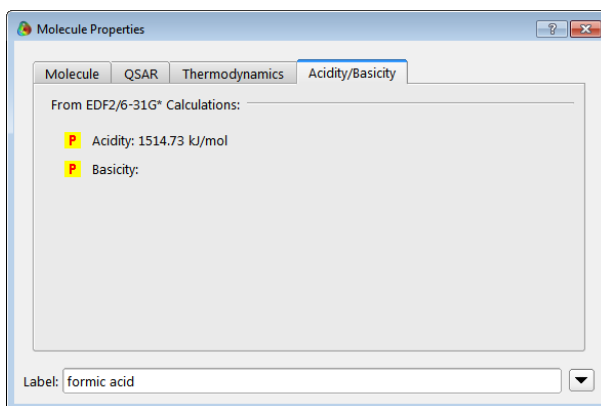
Entries under the **Thermodynamics** tab provide the zero-point energy, the enthalpy, the constant volume heat capacity, the entropy and the Gibbs energy. Except for the zero-point energy, all depend on temperature. All require vibrational frequencies.



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.



## Acidity/Basicity

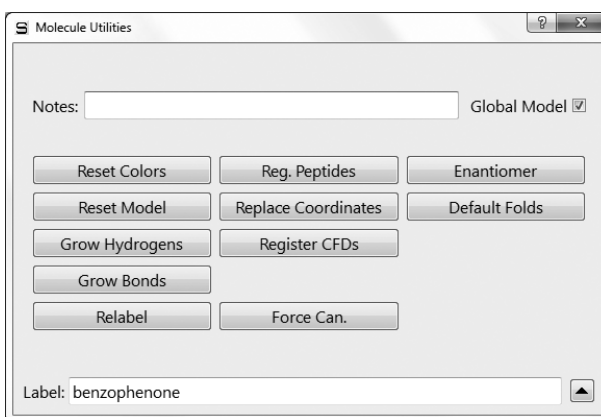
Entries under the **Acidity/Basicity** tab provide the gas-phase acidity (energy of deprotonation) and gas-phase basicity (energy of protonation) from EDF2/6-31G\* calculations. Energies for deprotonated and protonated species are extracted from the Spartan Spectra and Properties Database (SSPD).



At the present time, data for only a few classes of compounds, for example, deprotonation energies for carboxylic acids and protonation energies for pyridines, are available. These will be significantly extended in future versions of SSPD.

## Molecule Utilities

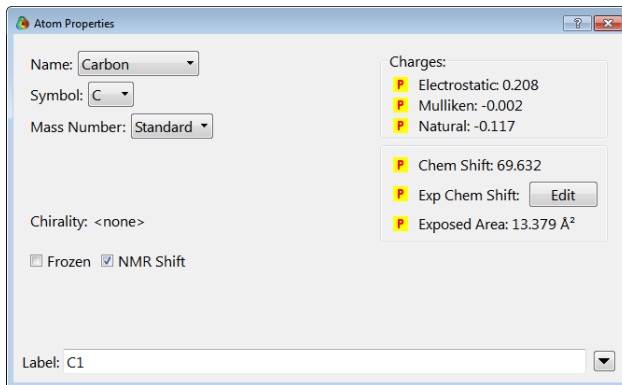
Clicking on  at the bottom right of the **Molecule Properties** dialog brings up the **Molecule Utilities** dialog (*clicking on  returns to the **Molecule Properties** dialog*).



**Notes** is a user-supplied text string that is reproduced in the output. Controls reset model color and style, add missing hydrogens and bonds, provide information about amino acids in polypeptides and about CFD's, replace coordinates by standards based only on atomic connectivities, change enantiomers and reset default conformer selections.

## Atom Properties

Selecting an atom with a **Properties** dialog on screen, or selecting **Properties** from the **Display** menu (1) following selection of an atom, leads to the **Atom Properties** dialog.





This displays the element name and symbol, mass number (isotope), R/S chirality, Mulliken, electrostatic-fit and natural charges (in electrons), calculated and experimental NMR chemical shifts (in ppm relative to an appropriate standard) and the exposed surface area of a space-filling model (in Å). Experimental chemical shifts appear if proton or  $^{13}\text{C}$  spectra have been retrieved from the public on-line databases or have been manually entered. 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.

$^{13}\text{C}$  chemical shifts from B3LYP/6-31G\* and EDF2/6-31G\* calculations only have been empirically corrected (see **Appendix A**). Corrections for the B3LYP/6-31G\* model use only topological data and show an RMS error of 2.6 ppm. Corrections for the EDF2/6-31G\* model use both topology and bond orders to neighboring atoms, and show an RMS error of 1.8 ppm. Uncorrected shifts from B3LYP/6-31G\* and EDF2/6-31G\* calculations show RMS errors that are 2-3 times larger.

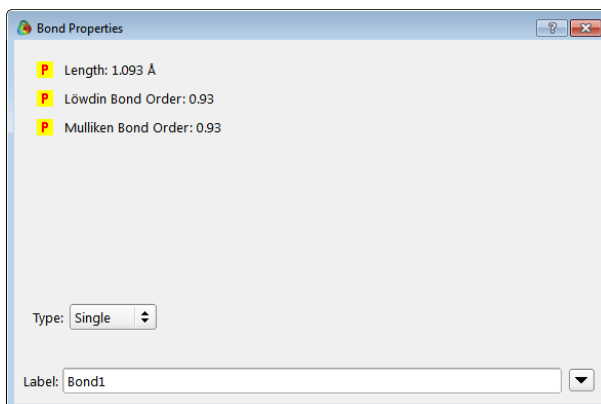
Corrections have also been made to  $^1\text{H}$  and  $^{19}\text{F}$  chemical shifts obtained from the EDF2/6-31G\* model only. These show RMS errors of 0.15 and 3.7 ppm, respectively, down by nearly a factor of two over uncorrected shifts.

Check boxes allow freezing the atom (see **Freeze Center** in **Chapter 18**), and turning off (unchecked) its line in the NMR spectrum.



*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.



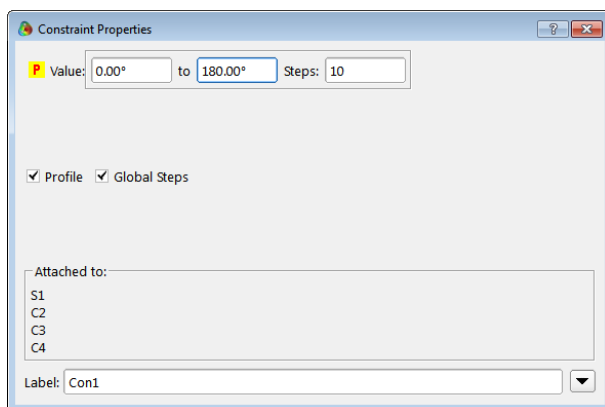
This displays the bond length, bond type (and allows changing the bond type) and Mulliken and Löwdin bond orders (the latter are used in NMR shift corrections).

*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 (not shown). 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 20**). 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 typing the desired numbers into the appropriate boxes and then *pressing* the **Enter** key (**return** key on Mac). This functionality may also be accessed from **Constrain Distance (Angle, Dihedral)** under the **Geometry** menu (**Chapter 18**). 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.


*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.

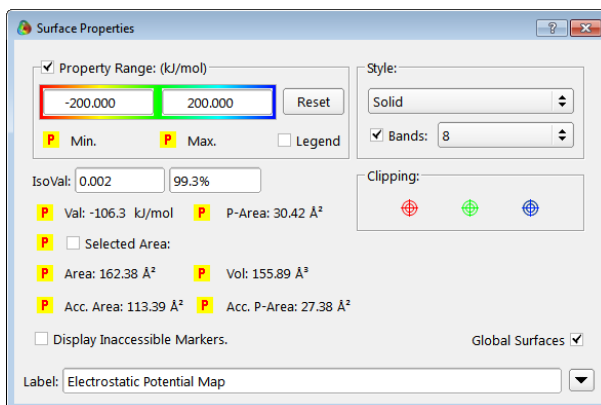
Clicking on  returns to the **Constraint Properties** dialog.

## Point and Plane Properties

Selecting a user-defined point or plane with a **Properties** dialog on screen, or selecting **Properties** from the **Display** menu () following selection of a point or plane, leads to the **Point Properties** or **Plane Properties** dialog (not shown). These allow changing point or plane labels. Clicking on  at the bottom right of the **Point (Plane) Properties** dialog leads to the **Point (Plane) Style** dialog (not shown), with controls for adjusting point (plane) colors. Clicking on  returns to the **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.

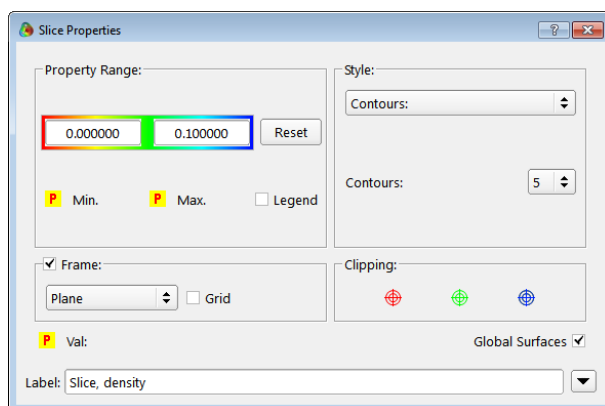


This allows changing display style (Style), selecting between continuous and banded displays and, in the case of the latter, setting the number of bands, specifying isovalue (and for electron density surfaces), percentage of the electrons contained inside the surface, turning on mapped properties, setting the property range of maps, and changing default surface labels. 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 apply to all molecules in the document.

The dialog also provides for up to three independent clipping planes which operate in toggle mode. Following selection, a plane may be rotated, translated or zoomed with the usual mouse operations.



Where the graphic 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


- 
- \* 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 23**).
  - \*\* 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 23**).
  - \*\*\* To determine property value at another position *click* on it. To bring up the **Molecule Properties** dialog, *click* on the background.

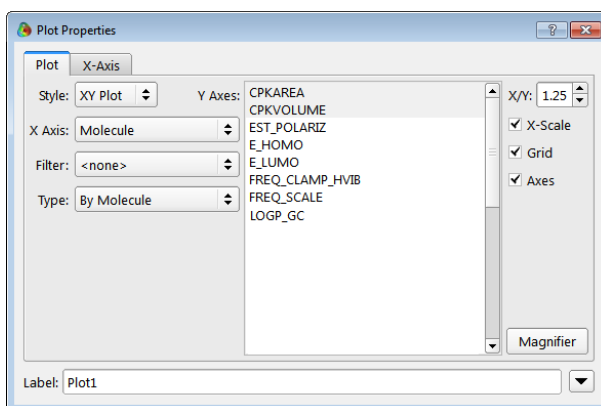
is displayed and, together with the minimum and maximum values of the property, may be posted to the spreadsheet. Check boxes allow for a legend, a frame around the slice and for a grid.

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 either be presented in separate windows (new to *Spartan'14*) or in *Spartan*'s main window (see discussion later in this chapter). While the new presentation is simpler and more intuitive, the latter is more general. 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).

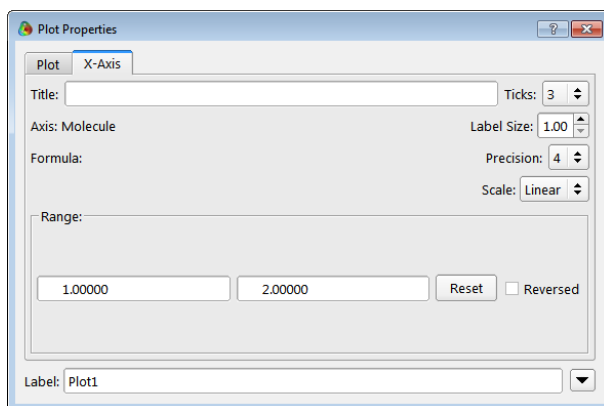
Selecting a plot frame with a **Properties** from the **Display** menu () dialog on screen, or selecting **Properties** following selection of a plot frame, leads to the **Plot Properties** dialog (**XY Plot** dialog shown).





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 contents 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 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.

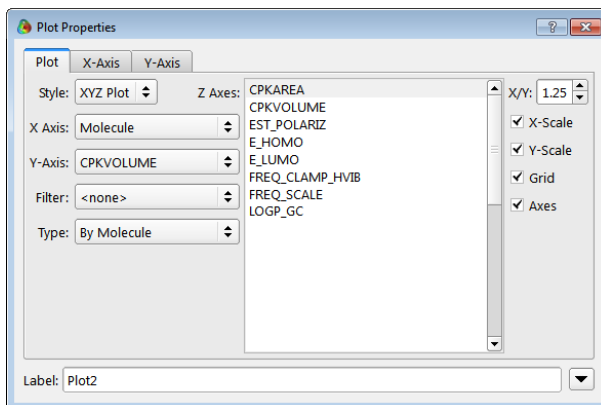


To alter the range, edit the values inside the **Range** boxes, and *press* the **Enter** key (**return** key on Mac). Controls are also available to alter the number of tic marks on the X axis, the label size and precision and change the scale from linear to logarithmic.

*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.

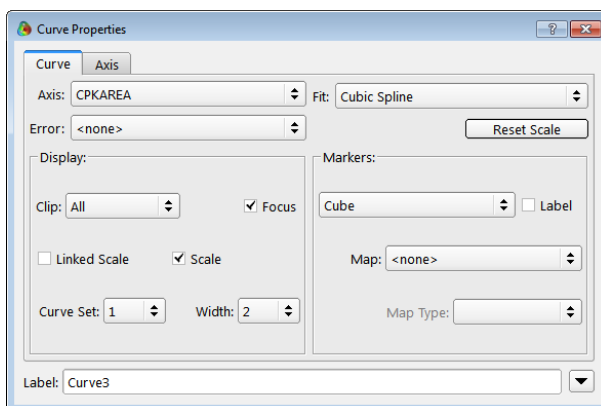
The **Style** menu in the **Plot Properties** dialog allows shifting between 2D (XY) and 3D (XYZ) plots. Selection of **XYZ Plot** leads to an extended **XYZ Plot Properties** dialog.



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**.

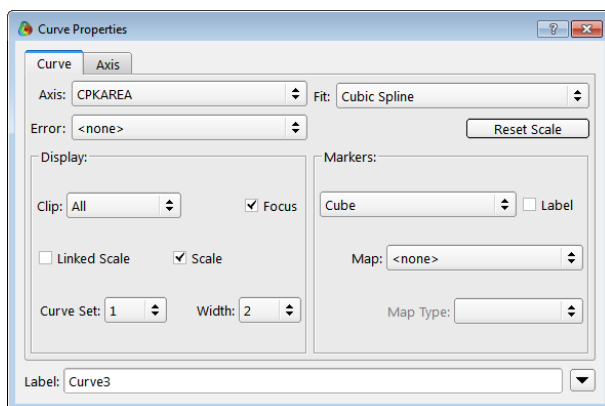
Ranges along X and Y axes may be adjusted by *clicking* on **X-Axis** and **Y-Axis** tabs to the left of the dialog. The resulting dialog (not shown) contains controls that are analogous to those previously discussed for XY plots.

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 scale, type in new values inside the boxes underneath **Range**. You need to *press* the **Enter** key (return key on Mac) following each entry. Controls are also available to alter characteristics of the Y (XY plot) or Z axis (XYZ plot) labels.

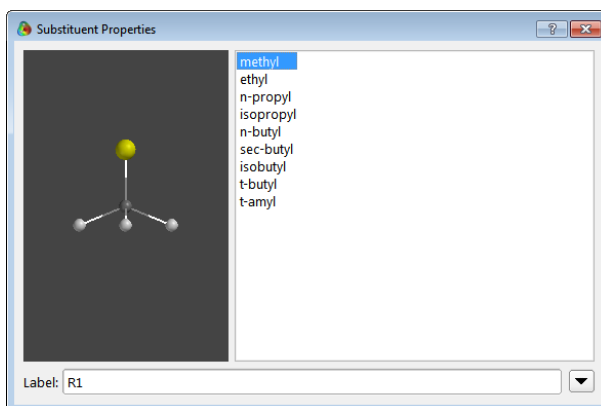
Clicking on  at the bottom right of the **Curve Properties** dialog (with either tab selected) brings up the **Curve Style** dialog (not shown). This contains controls for selecting the color of the curve (and the label associated with the Y (or Z) axis). Clicking on  at the bottom right returns to the original dialog.

---

\* Fitting functions include linear, quadratic and cubic least square, a Fourier series, a Lorentzian or a 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. The default is a linear least squares fit.



## 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.



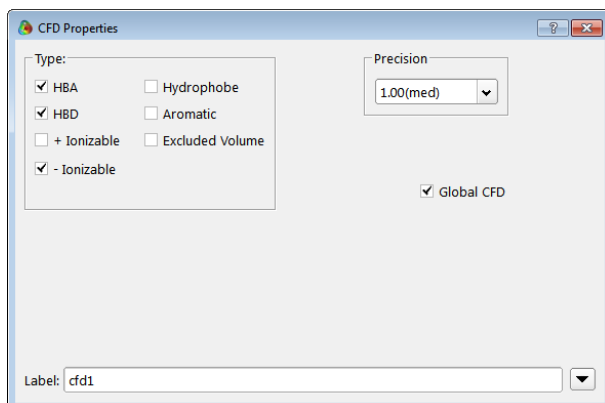
This allows substituent libraries already attached to a molecule to be edited, and the label on the substituent icon to be changed. It does not allow for attachment points to be reassigned. The dialog comprises a scroll box (listing the substituent names), a viewing window (showing a model of the substituent) 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 (1) 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.

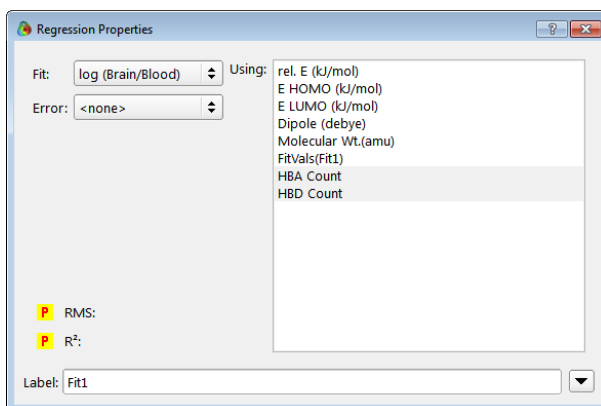
## Regression Properties

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 (1) after *clicking* on the line, leads to the **Regression Properties** dialog.

---

\* 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**, . . .



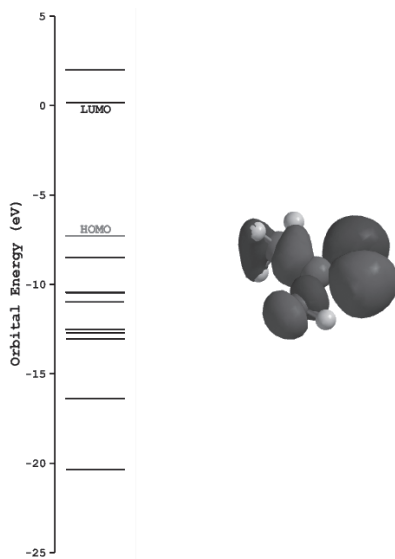
This reports RMSD and  $R^2$ , as well as allows for changing what is to be fit (**Fit**) and what it is to be fit to (**Using**). The error statistics will immediately update.

### “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). 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.



Moving a finger up or down over the diagram then lifting selects the next higher or lower energy level.

## Surfaces (🌈)

This accesses the same dialog already described in **Chapter 20**.

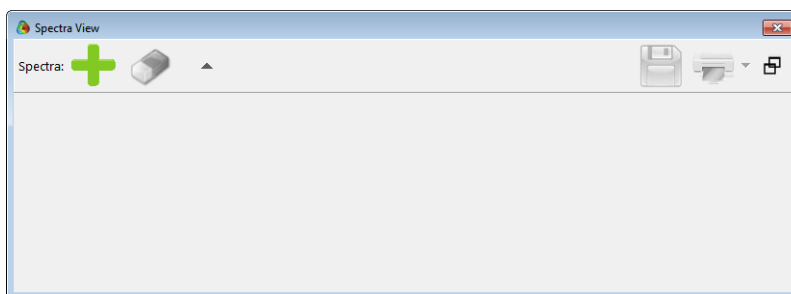
## Spectra (📈)


**Spartan** displays calculated IR, Raman, NMR and UV/visible spectra. Spectra need to have been previously requested from the **Calculations** dialog (**Calculations...** under the **Setup** menu; **Chapter 20**) or accessed from the Spartan Spectra and Properties Database. In addition, it allows on-line access and display of experimental IR, NMR and UV/visible spectra from publicly available collections,

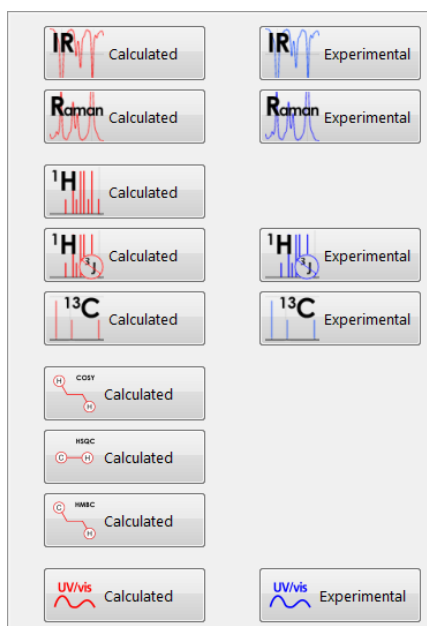
facilitating comparison with calculated spectra. *Spartan* allows two different modes for spectra display. In addition to the general presentation scheme provided in earlier versions of the program, a simpler and (we believe) more intuitive mode is available in *Spartan*'14. The choice is up to the user in the form of a preference. The new presentation mode is described first.

## Simplified Presentation of Spectra

With **Spectra Pane** in the **Settings Preferences** dialog checked (**Preferences** under the **Options** menu; **Chapter 23**), selecting **Spectra** from the **Display** menu leads to an empty display pane at the bottom of the screen.

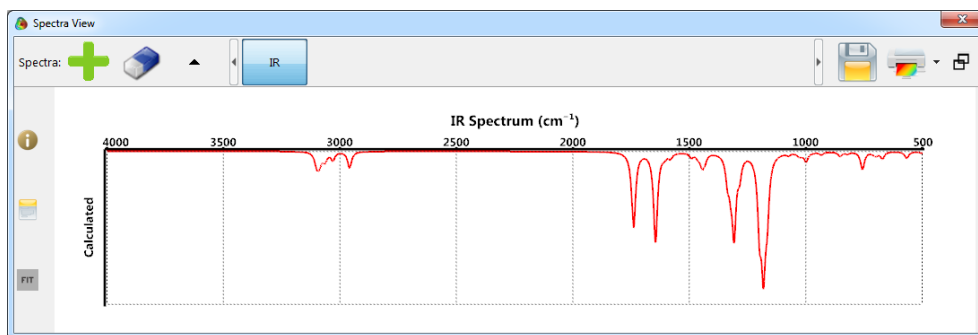


The only accessible control (in a bar at the top of the pane) is  (add a spectrum). Clicking on this leads to a palette.

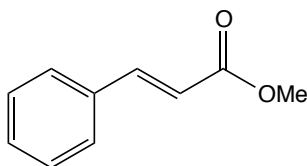



The right hand column lists the types of spectra for which calculations are available: IR, Raman, proton NMR with and without three-bond HH coupling  $^{13}\text{C}$  NMR, COSY, HSQC and HMBC 2D NMR and UV/vis. The entry is “red” if a calculation has actually been performed and the corresponding spectrum is available. The left hand column lists the types of spectra for which experimental spectra *may be available* (from on-line public databases): IR, proton NMR,  $^{13}\text{C}$  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.







Moving the mouse while holding down the left button moves the cursor (unfilled markers at the top and bottom of the spectrum) over the spectrum. When positioned directly over a special line, the markers 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  $\text{cm}^{-1}$  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 at  $1645\text{ cm}^{-1}$  corresponds to the  $\text{C}=\text{C}$  stretch while the line at  $1739\text{ cm}^{-1}$  corresponds to the  $\text{C}=\text{O}$  stretch.

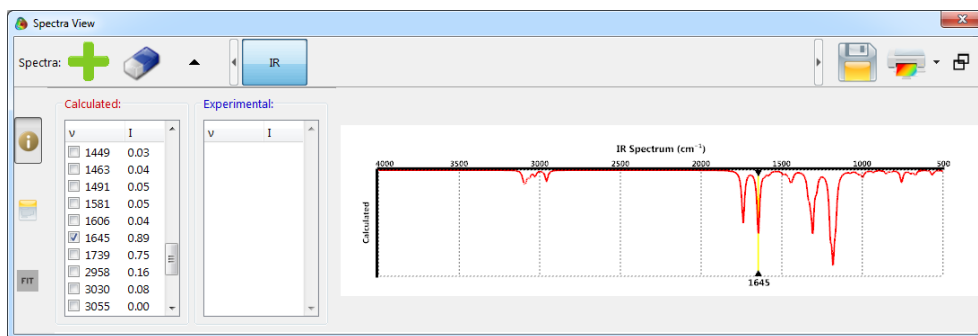


Moving the mouse while holding down the right button slides the viewable scale from  $4000\text{ cm}^{-1}$  to  $500\text{ cm}^{-1}$  but does not change the overall range (of  $3500\text{ cm}^{-1}$ ). The range is changed by using the scroll wheel. The original settings may be restored by *clicking* on  in the bar at the top of the spectra pane.



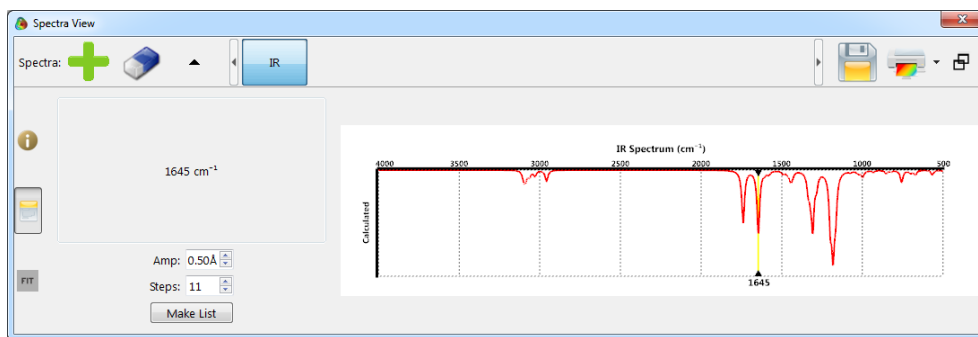
Move one finger over the spectrum to position the cursor, move two fingers to slide the viewable scale and pinch two fingers to change the range.

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.


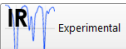


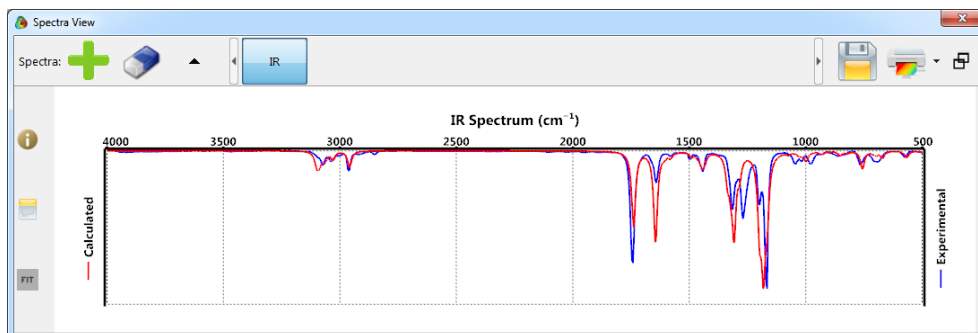
This contains a listing of calculated infrared frequencies and intensities. *Checking* the box to the left of an individual frequency moves the cursor on the spectrum over this line and animates the vibrational motion.


With a frequency selected, *clicking* on () leads to a new panel.

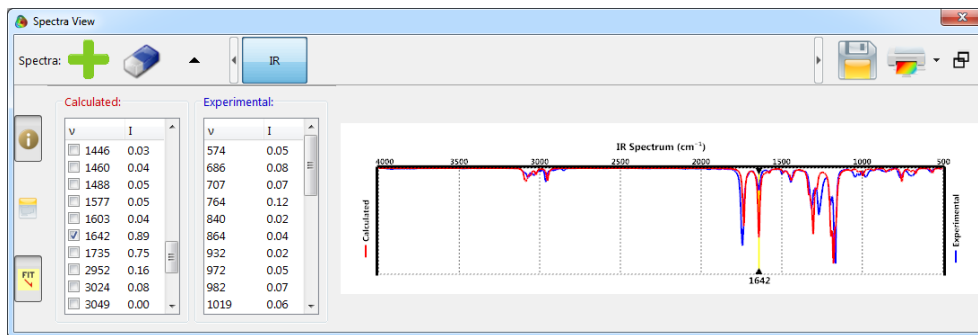


This is used to make a list of structures centering around the minimum (or maximum in the case of a transition state) with control over the amplitude of vibration (maximum displacement in Å) and number of steps. *Clicking on **Make List*** leads to a separate document.

If available, an experimental spectrum from one of the public on-line databases 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*  *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 an EDF2/6-31G\* calculation (as contained in SSPD), they are otherwise unaltered. EDF2/6-31G\* calculated infrared (and Raman) frequencies have been uniformly scaled (by 0.962) 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 **FIT** 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 spectra pane. If both calculated and experimental spectra were displayed, both will be deleted.

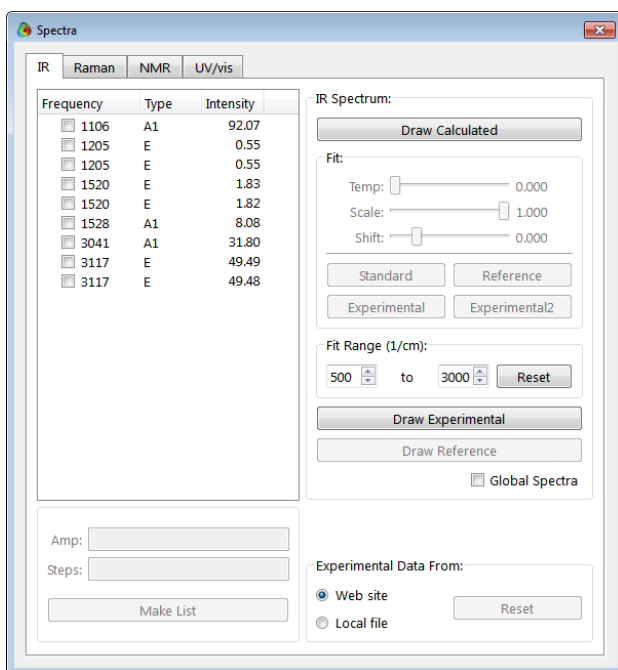
Controls at the top right of the spectra pane allow saving the spectrum as a PNG, JPEG or Bitmap image file (**📁**), printing the file (**🖨**), detaching the spectrum pane from the main window (**📄**) and closing the pane (**✖**).

## Generalized Presentation of Spectra

With **Spectra Pane** in the **Settings Preferences** dialog unchecked (**Preferences** under the **Options** menu; **Chapter 23**), selecting **Spectra** from the **Display** menu leads to a series of 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 (Raman)** tab leads to the **IR Spectra (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

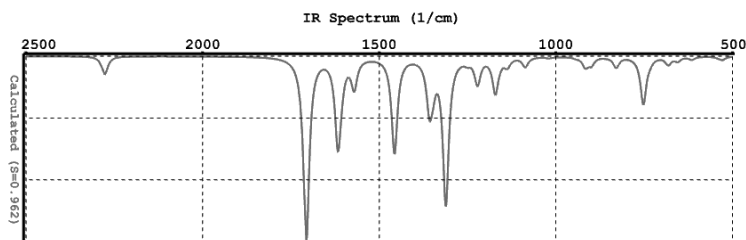
(in  $\text{cm}^{-1}$ )\* together with (infrared) intensities and symmetry labels. A frequency is selected for display (animation) by *clicking* on it. *Clicking* again deselects it (stops the animation).

The amplitude of vibrational motion (the maximum displacement away from equilibrium of any pair of atoms) may be changed from the default amplitude ( $0.5\text{\AA}$  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 right 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 right of the dialog.

*Clicking* on **Make List** creates an unnamed *Spartan* 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 not available* as part of an entry in the Spartan Molecular Database. However, it *is available* as part of the entry in the Spartan Spectra and Properties Database (see discussion in **Chapter 22**).

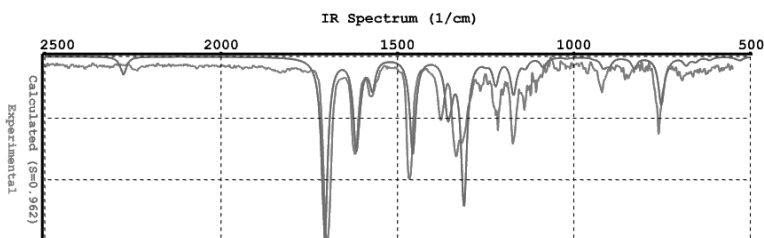
The calculated spectrum may be drawn by *clicking* on **Draw Calculated** at the top right of the dialog.



The experimental spectrum (if available) may be drawn on top

- 
- \* 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.

of the calculated spectrum (or by itself) by *clicking* on **Draw Experimental** at the center right of the dialog (IR spectrum shown).




Finally, a reference spectrum (a spectrum shared by all molecules in the document; see discussion of the Spartan Infrared Database in **Chapter 22**) 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 (<http://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. 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** (that has replaced **Draw Calculated**) in the dialog. Similarly, the experimental spectrum (infrared only) may be removed by *clicking* on **Delete Experimental** (that has replaced **Draw Experimental**) and the reference spectrum may be removed by *clicking* on **Delete Reference** (that has replaced **Draw 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 3-4% 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. This is limited in range (from 0.8 to 1.0). If it is necessary to scale outside this range, *click* on the spectra (not on an axis) to select and move the mouse up and down while simultaneously depressing the **Shift** and **Alt** keys (**option** on Mac) keys in addition to the right button. The scale factor (multiplying the frequency) will be displayed on the vertical axis. Scaling applies uniformly to all molecules in the document (not just the selected molecule).

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 20**).

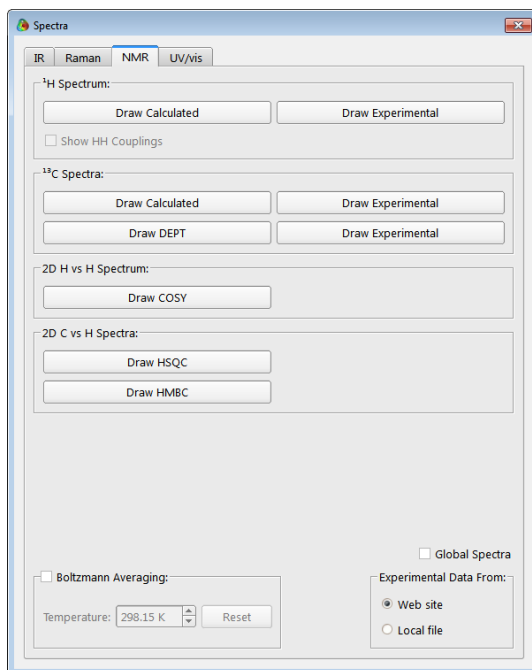
A second slider bar marked **Temp** controls peak width. This is loosely connected with the temperature at which the experimental measurement is carried out. Low “temperature” (slider to the left) will produce sharp spikes, whereas high “temperatures” (slider to the right) will produce broad bands. The 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

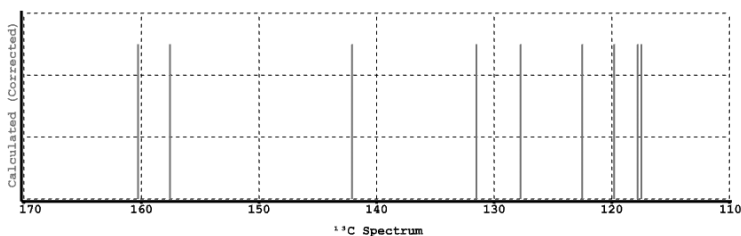
Spectra are divided into four categories\*:  $^1\text{H}$  spectra,  $^{13}\text{C}$  and DEPT spectra, COSY spectra and HSQC and HMBC spectra. Access to experimental 1D spectra is provided either on-line or from a local file. *Clicking* on the **NMR** tab leads to the **NMR Spectra** dialog.

---

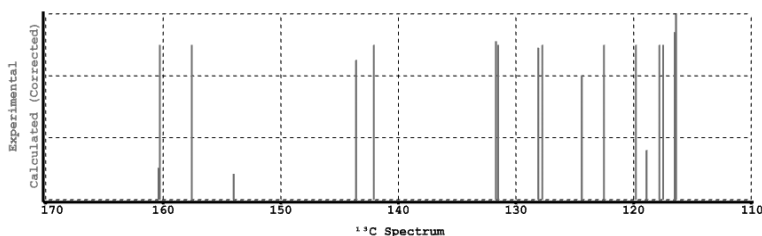
\* Chemical shifts for other nuclei are reported in the summary output (see **Output** earlier in this chapter) may be displayed as labels (see **Configure...** under the **Model** menu; **Chapter 17**).



Calculated proton,  $^{13}\text{C}$  and DEPT spectra may be drawn by *clicking* on their respective **Draw Calculated** buttons. Carbon shifts (only) for both  $^{13}\text{C}$  and DEPT spectra from the B3LYP/6-31G\* model are corrected based on topology. Proton and  $^{13}\text{C}$  and  $^{19}\text{F}$  shifts from the EDF2/6-31G\* model are corrected based on topology and bond orders. Shifts from other models are not corrected.

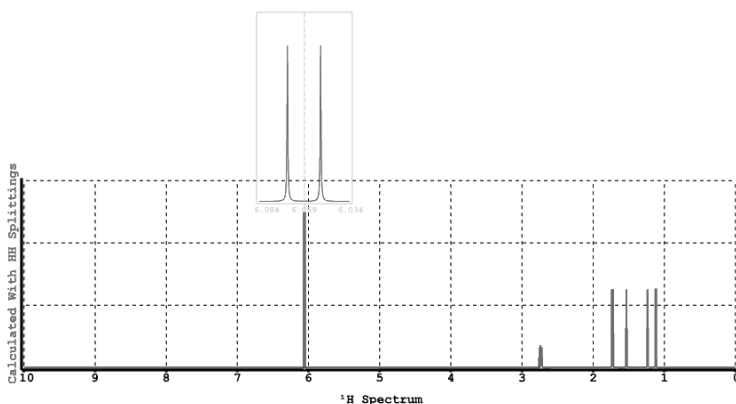


Experimental  $^{13}\text{C}$  and DEPT spectra (if available) may be drawn together with or independently of the calculated spectra by *clicking* on **Draw Experimental**. Draw operations apply to all molecules in the document.



Experimental proton and  $^{13}\text{C}$  spectra are accessed from <http://nmrshiftdb.ice.mpg.de>. This database contains over 40,000 experimental spectra, all for organic molecules. Alternatively, 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). 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.

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).

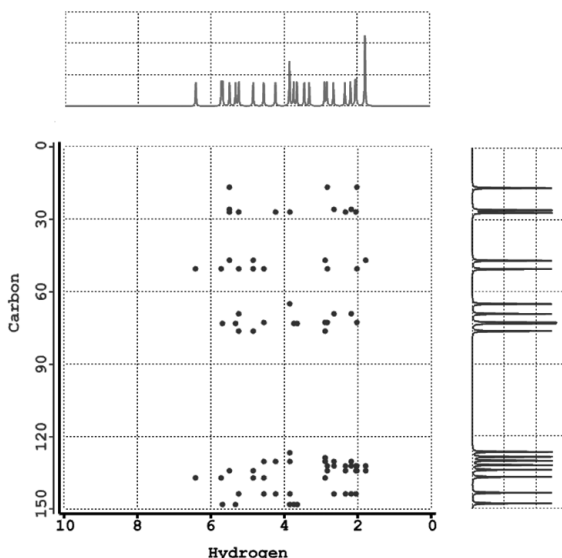


Default ranges for proton spectra (0 to 10 ppm) and  $^{13}\text{C}$  spectra (0 to 150 ppm) may be changed by selecting **Properties** from the **Display** menu, *clicking* on the horizontal axis to bring up the **Plot Properties** dialog, *clicking* on the **X Axis** tab at the left of the dialog and 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.

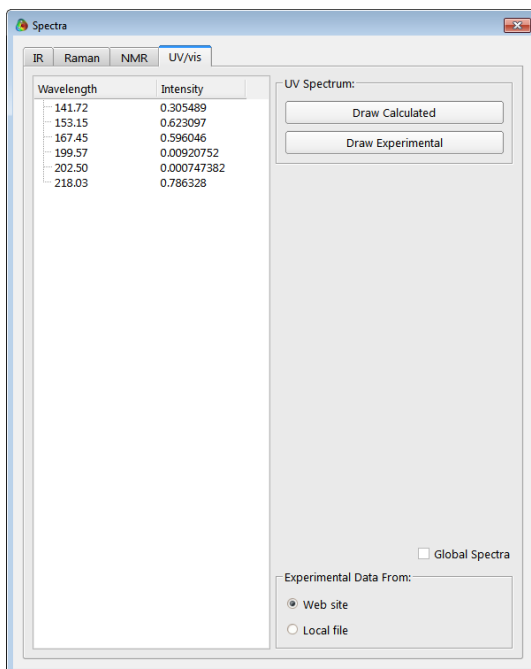
2D spectra are available from **Draw Calculated** buttons. COSY spectra identify protons three bonds removed, HSQC spectra identify protons directly attached to a carbon and HMBC spectra identify protons that are three bonds removed from a carbon. Note that proton spectra “broadened” because of the inclusion 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.



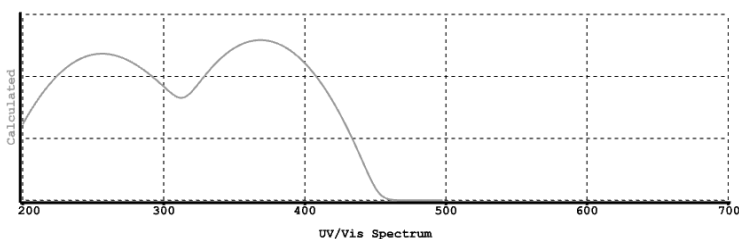
Uncorrected spectra (proton,  $^{13}\text{C}$ , DEPT and  $^{19}\text{F}$  are likely to show significant systematic errors in chemical shifts. To scale either the calculated or experimental spectra in order to provide a best visual fit, *click* on the spectra (not on an axis) to select and move the 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

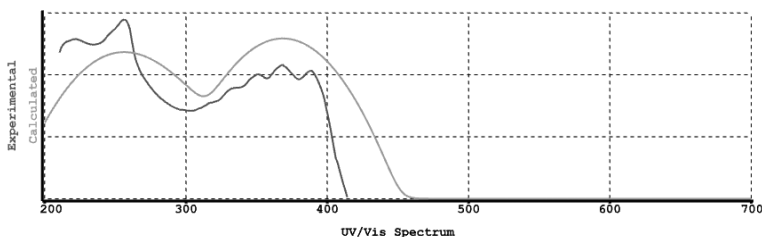
*Clicking* on the **UV/vis** tab leads to the **UV/vis Spectra** dialog that provides for display of UV/visible spectra.



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 (<http://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.

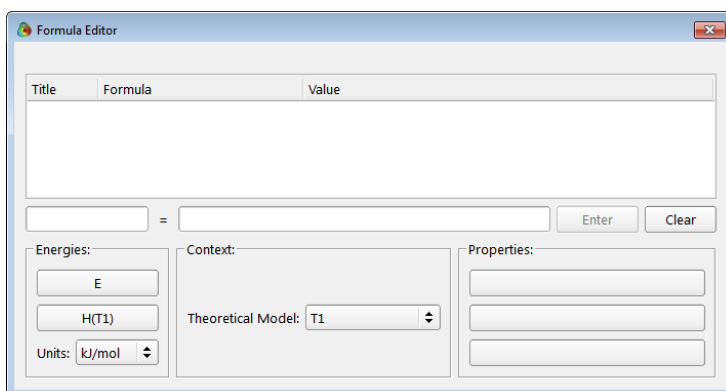
Only a single copy of one of the dialogs under **Spectra** may appear on screen; scaling will relate to the currently selected molecule. The dialog may be removed by *clicking* on .

## 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 18**), 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 is that involving 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 22**). **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 22**.

Pressing either the **E** or **H(T1)** buttons or a **Property** button onto

---

\* Heats of formation from the T1 recipe where available are included with SSPD entries.

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 (**F<sub>x</sub>**, 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 Spartan Molecular Database or Spartan Spectra and Properties Database. 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 21-1** 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-@LUMO\*@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 21-2** (preceded by **@**). Change the title to the left of the string from **F<sub>x</sub>** 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

**Table 21-1: Arithmetic and 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 21-2: 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	$\pi$

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 21-3** 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).

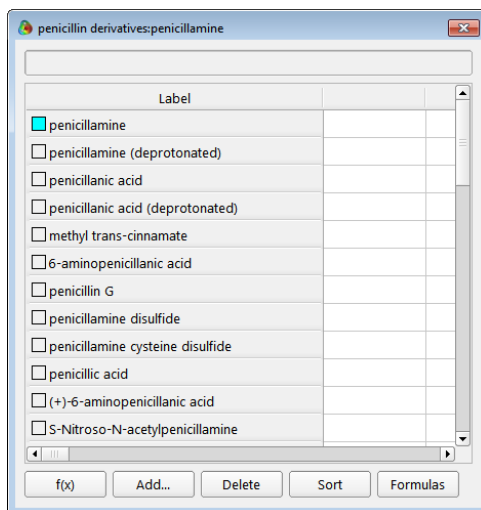
Formulas are used in conjunction with spreadsheet manipulations and plotting (both discussed later in this chapter) and with data mining (discussed in **Chapter 22**).

## 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






**Table 21-3: Calculated Quantities**

AREA	area of a user-defined plane ( $\text{\AA}^2$ )
ELECTRONEGATIVITYEV	QSAR descriptor $-(\text{HOMO}+\text{LUMO})/2$ (eV)
HARDNESSEV	QSAR descriptor $-(\text{HOMO}-\text{LUMO})/2$ (eV)
HOMO(-n)EV	energy of $n^{\text{th}}$ orbital below the HOMO (eV)
HOMOBETA(-n)EV	energy of the $n^{\text{th}}$ orbital below the $\beta$ HOMO (eV)
HYPERPOLARIZABILITY	beta hyperpolarizability parameter
INERTIA(i)	principle movements of inertia from largest (i=1) to smallest (i=3)
ISOTOPE(i)	mass number of atom i
LENGTH (i)	length of bond i ( $\text{\AA}$ )
LOGPC	LogP from Crippen model
LOGPV	LogP from Villar model
LUMO(+n)EV	energy of $n^{\text{th}}$ orbital above the LUMO (eV)
LUMOBETA(+n)EV	energy of $n^{\text{th}}$ orbital above the $\beta$ 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. ( $\text{\AA}^2$ )
POLARIZABILITY	alpha polarizability parameter

**Table 21-4: 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

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.

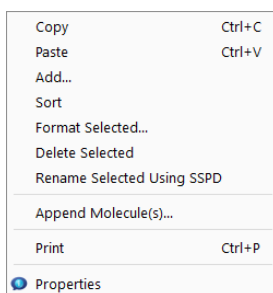
Only one molecule from one document may be selected (although several molecules may be simultaneously displayed). Molecule selection follows either by *clicking* on the spreadsheet cell containing the molecule label or identifier (leftmost column), or by using the  and  buttons or the scroll bar at the 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 23**). 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 17**). By default (**Coupled** checked) molecules move in concert. *Uncheck Coupled* to move them independently.

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 Molecular Database (SMD) by *clicking* on **Label** (the leftmost header cell), then *right clicking* and selecting **Rename Selected using SMD** from the menu that results. To replace all identifiers, *right click* inside the header cell of the leftmost column and again select **Rename Selected using SMD**. Renaming also applies to molecules in the Spartan Spectra and Properties Database (SSPD) in which case the menu entry is **Rename Selected Using SSPD**. Unless specified otherwise, molecule names are provided for data extracted from either SMD or SSPD.

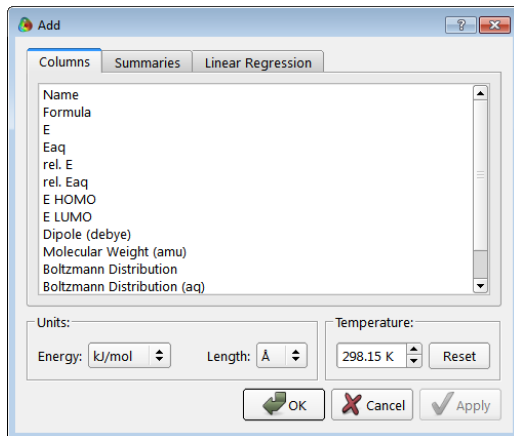
Information may be added to the spreadsheet in several different ways:

### 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 the **Add Columns** dialog.

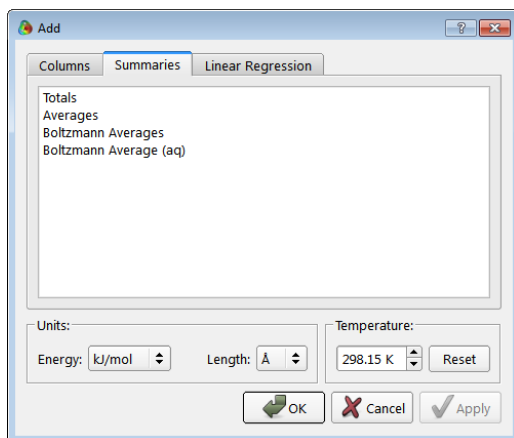


Name	molecule name as it appears in SMD or SSPD
Formula	molecular formula
E	energy (heat of formation, strain energy)
E <sub>aq</sub>	aqueous energy (heat of formation, strain energy) based on the SM5.4 model
rel. E	energy (heat of formation, strain energy) relative to selected molecule

rel. E <sub>aq</sub>	aqueous energy (heat of formation, strain energy) relative to selected molecule based on the SM5.4 model
E HOMO	energy of highest-occupied molecular orbital
E LUMO	energy of lowest-occupied molecular orbital
Dipole	dipole moment (in debyes)
Molecular Weight	molecular weight (in amu)
Boltzmann Distribution	Boltzmann distribution based on energy
Boltzmann Distribution (aq)	Boltzmann distribution based on aqueous energy as estimated from SM5.4 model
Cumulative Boltzmann Distribution	Sum of the Boltzmann weights for the selected molecule and all molecules with lower energy than the selected molecule
Cumulative Boltzmann Distribution (aq)	Sum of the Boltzmann weights for the selected molecule and all molecules with lower energy than the selected molecule based on aqueous energy as estimated by the SM 5.4 model
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

One or more properties may be added to the spreadsheet by *clicking* on their entries, then specifying appropriate units from the **Energy** and **Length** menus, varying temperature from the **Temp** menu (as applied to the Boltzmann distributions), and finally *clicking* on **OK** or on **Apply**. In the former case, the dialog is dismissed and in the latter it is left on screen. *Clicking* on **Cancel** or  removes the dialog without further affecting the contents of the spreadsheet. (Any previous additions to the spreadsheet as a result of *clicking* on **Apply** are maintained.)

Column totals, averages and Boltzmann-weighted averages may be accessed by *clicking* on the **Summaries** tab. This leads to the **Add Summaries** dialog.



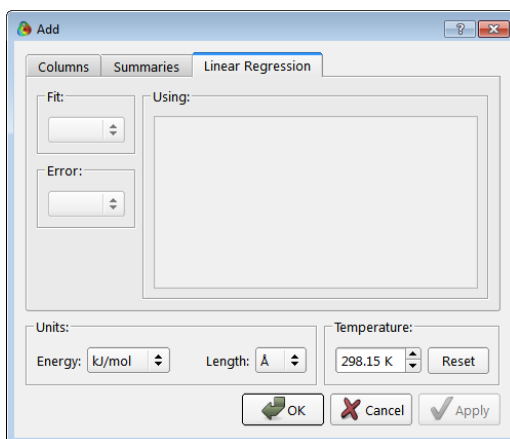
Totals	sum of column values
Averages	average of column values
Boltzmann Averages	Boltzmann weighted average of column values
Boltzmann Averages (aq)	Boltzmann weighted average of column values using energies corrected for aqueous environment based on the SM5.4 model

*Clicking* on one or more of these followed by *clicking* on **OK** or **Apply**, leads to the requested summaries as rows at the bottom of the spreadsheet, identified as **Totals**, **Averages**, etc. (one row for each summary). Note, that this dialog can be brought up directly by *clicking* on an empty row header (instead of an empty column header) inside the spreadsheet prior to *clicking* on **Add....**

Finally, a linear regression analysis may be performed on the data in the spreadsheet.

First *click* on the **Linear Regression** tab to bring up the **Add Linear Regression** dialog.

Select one entry from the **Fit** menu and one or more entries from the list under **Using**. *Clicking* on **OK** or **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 ( **P** ) Buttons

Post buttons ( **P** ) 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 18**), bond distance, angle and dihedral angle constraints (available from **Constrain Distance**, **Constrain Angle** and **Constrain Dihedral** under the **Geometry** menu; **Chapter 18**), 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.

---

\* Label reassignment is accomplished using the **Atom Properties** dialog (see discussion earlier in this chapter).

Post buttons are also available for CAS numbers and experimental heats of formation contained in SMD and SSPD and for T1 heats of formation contained in SSPD.

## From the Clipboard

Properties of one or more molecules in a document may be copied into 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 18**), bond distance, angle and dihedral angle constraints (**Constrain Distance**, **Constrain Angle** and **Constrain Dihedral** under the **Geometry** menu; **Chapter 18**), 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 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.

## 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 21-1**), specialty functions (**Table 21-4**), calculated quantities (**Table 21-3**), conversion factors and constants (**Table 21-2**) 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 21-4**). 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). Alternatively, use the ↓ key.

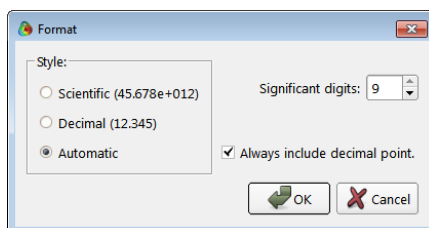
Each row in a spreadsheet corresponds to a molecule in a document, and new rows are automatically added in response to adding new molecules to the document. New molecules may be added by building (**Build New Molecule** under the **File** menu; **Chapter 15**) or sketching (**Sketch New Molecule** under the **File** menu; **Chapter 15**), by appending one or more existing documents each containing one or more molecules using either **Append Molecule(s)...** under the **File** menu (**Chapter 15**), 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 into 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. Use of the clipboard permits several molecules to 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 15**), 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.

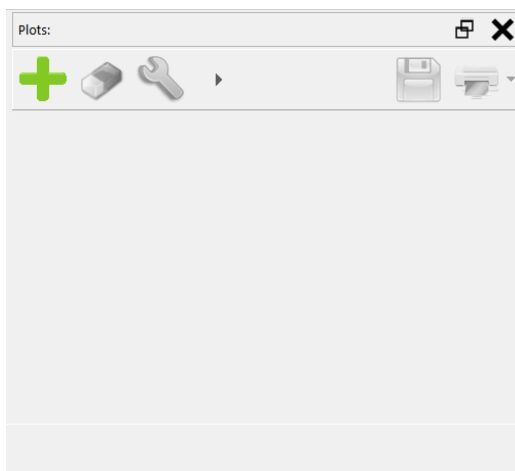
The contents of an Excel spreadsheet may be brought into *Spartan*. Copy whatever information is to be transferred to the clipboard, move into *Spartan*, then *click* on the appropriate cell and select **Paste** from the **Edit** menu (or *right click* on the appropriate cell and select **Paste** from 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...

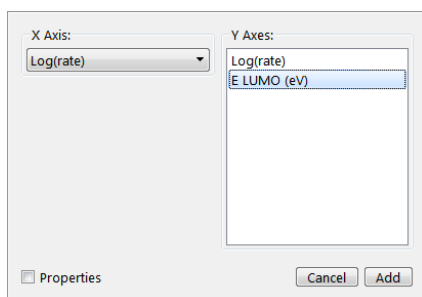
Plots may be constructed from data in a spreadsheet and a variety of simple curves fit to these data. As with presentation of spectral data (see discussion earlier this chapter), *Spartan* allows two different modes for presentation. In addition to the general presentation scheme provided in earlier versions of the program, a simpler and (we believe) more intuitive mode is available in *Spartan*'14. The choice is up to the user in the form of a preference.

### Simplified Presentation of Plots

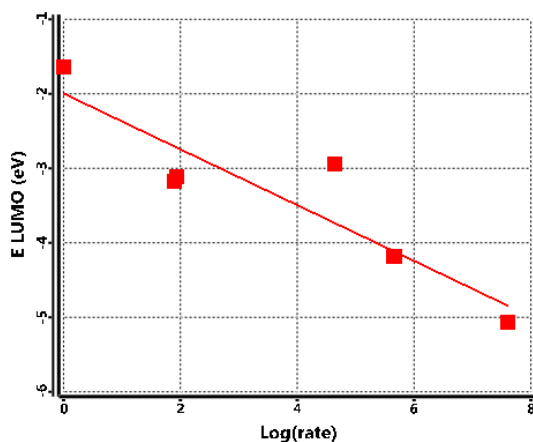
With **Plots Pane** in the **Settings Preferences** dialog checked (**Preferences** under the **Options** menu; **Chapter 23**), selecting **Plots** from the **Display** menu leads to an empty display pane at the right of the screen.




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 and the **Add Plot** dialog is dismissed.

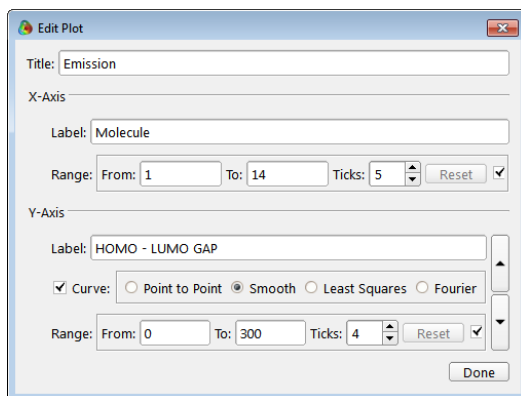


By default, the scales for both horizontal and vertical axes are set to bound the data trying to provide limits and increments that are “rounded”. Moving the mouse left and right while holding down the right button slides the horizontal scale but does not change the range. Similarly, moving the mouse 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.



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.



**Edit Plot**

Title: Emission

X-Axis

Label: Molecule

Range: From: 1 To: 14 Ticks: 5 Reset

Y-Axis

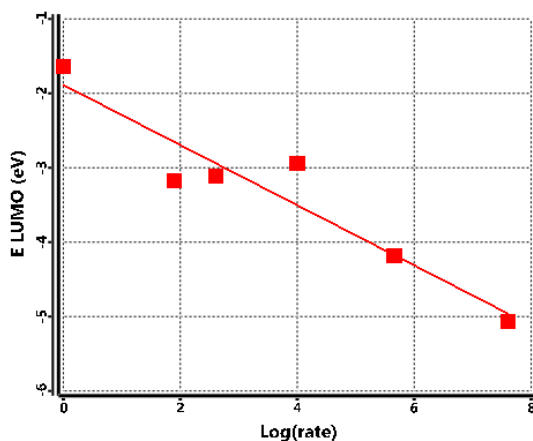
Label: HOMO - LUMO GAP



Curve: ☐ Point to Point ☒ Smooth ☐ Least Squares ☐ Fourier

Range: From: 0 To: 300 Ticks: 4 Reset

Done

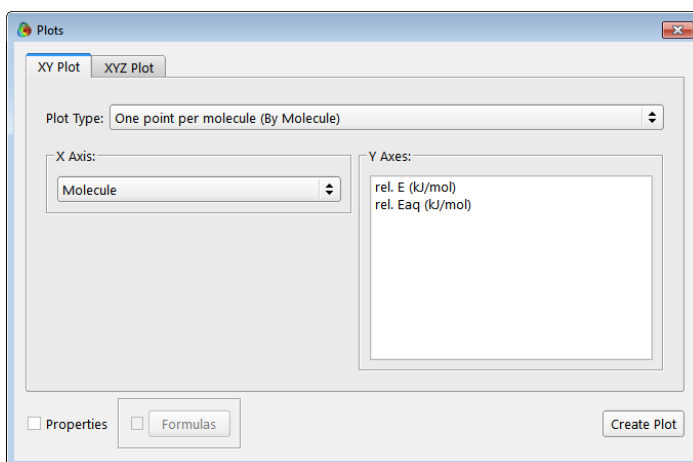
The resulting dialog also allows axis labels to be altered (from their initial values designated in the spreadsheet) the number of “tics” of horizontal and vertical axes to be changed and a plot title to be added. Finally, the default “curve” (a least-squares fit) can be changed or eliminated altogether.



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 23**), 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.

Plot Type:	One point per molecule (By Molecule)	▼
	One point per atom (By Atom)	
	Free form	

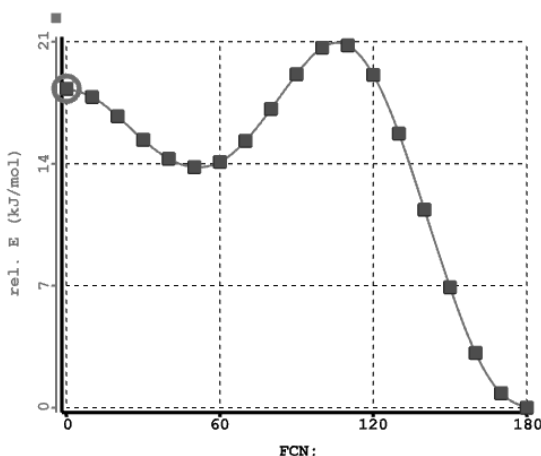
This distinguishes between plots based on the entire set of molecules in a document (**By Molecule**), to those based on the atoms in a single molecule (**By Atom**). The latter is the mode used to plot calculated IR, Raman and NMR spectra. The third choice (**Free Form**) is used to construct single 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 *Vector Spreadsheet Operations* available under **Help** from the **Help** menu (**Chapter 25**).

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 SMD or SSPD.

## XY Plots

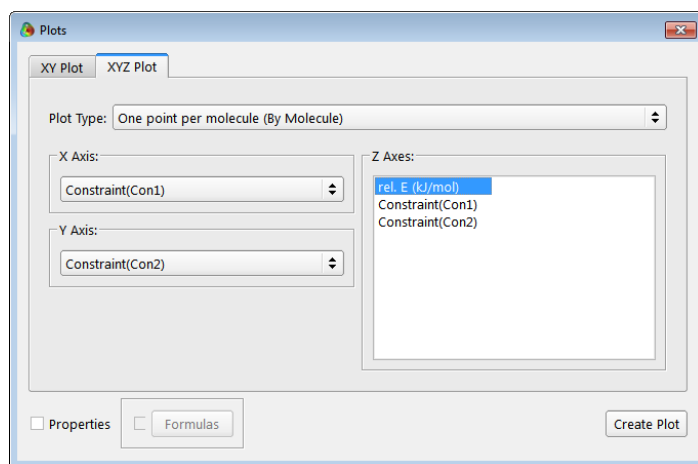
The **XY Plot** dialog incorporates an **X Axis** menu designating the molecular property to be displayed among the X (horizontal) axis, and a list of properties to be displayed along the Y (vertical) axis. These properties correspond one to one to the columns in the associated spreadsheet. To construct 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.



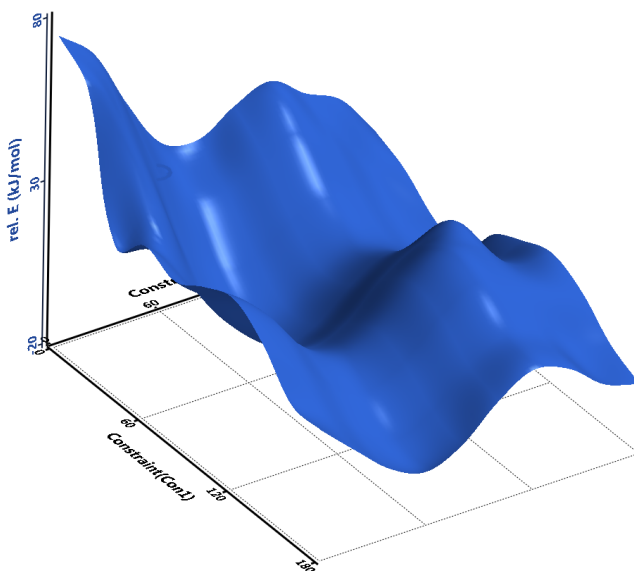
The plot may be moved about the screen. Select (*click* on) either the frame or on one of the curves (it will turn gold in response), then hold down the right mouse button and *drag* the mouse. The plot may also be scaled (expanded or shrunk) by first selecting either the frame or one of the curves and *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 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. 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.

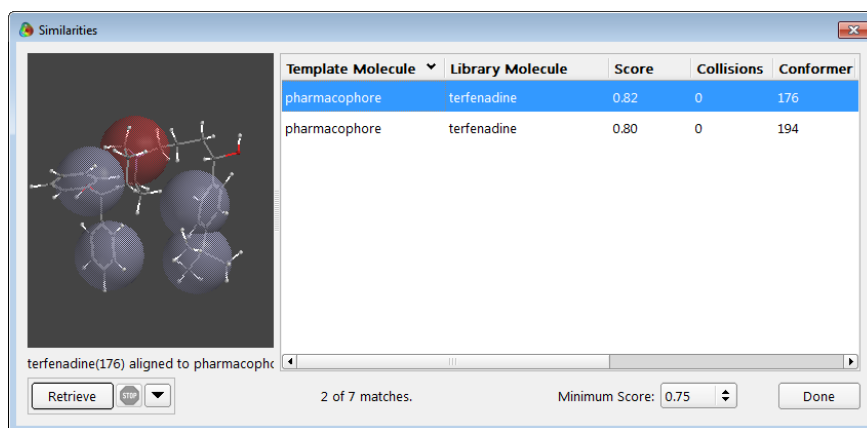
Plots initially presented connect the data points with cubic splines. A variety of curve fits (linear, least squares, Fourier, etc.) as well as a variety of different presentation formats are available under the **Plot Properties** and associated dialogs (see **Properties**

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

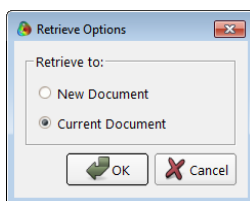
\* The score is defined as  $[(1-R^2)/N]$ -penalty, where  $R^2$  is the root mean square distance between template and library molecule centers and  $N$  is the number of similarity centers. Whereas molecule alignment as a prelude to similarity scoring is based on atomic, CFD and pharmacophore centers **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.

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.



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.

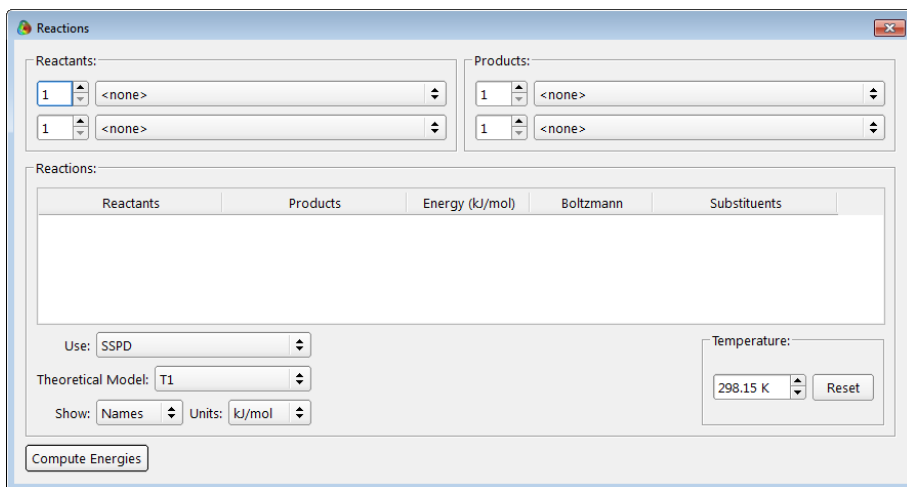
The **Similarities** dialog is closed by *clicking* on the **Done** button.

## Reactions...

Data entered in a *Spartan* document may be used to calculate reaction energies including activation energies.

$$\Delta E = E_{\text{product1}} + E_{\text{product2}} - E_{\text{reactant1}} + E_{\text{reactant2}}$$

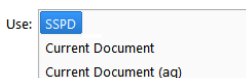
Selection of **Reactions...** from the **Display** menu leads to the **Reactions** dialog.



The screenshot shows the 'Reactions' dialog box. It has two main sections: 'Reactants' and 'Products'. Each section contains two dropdown menus, one for an identifier (showing '1') and one for a name (showing '<none>'). Below these is a table with columns: Reactants, Products, Energy (kJ/mol), Boltzmann, and Substituents. At the bottom, there are settings for 'Use' (set to 'SSPD'), 'Theoretical Model' (set to 'T1'), 'Show' (set to 'Names'), and 'Units' (set to 'kJ/mol'). A 'Temperature' section shows '298.15 K' and a 'Reset' button. A 'Compute Energies' button is at the bottom left.

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 is mass balanced.

The **Use** menu identifies the source of the energies to be used in the reaction energy calculation.

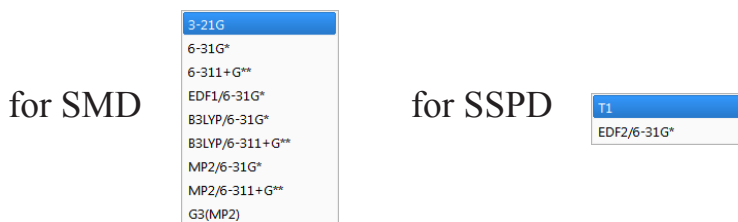


A close-up of the 'Use' menu dropdown. The options are: 'SSPD' (highlighted in blue), 'Current Document', and 'Current Document (aq)'.

**SMD** or **SSPD** refers to energies drawn from entries in the Spartan Molecular Database (SMD), or Spartan Spectra and Properties Database (SSPD), **Current Document** refers to energies in document and **Current Document (aq)** refers to energies in the document that have been empirically corrected for aqueous solvent using the SM5.4 model.

Selection of **SMD** or **SSPD** (from the **Use** menu) requires specification

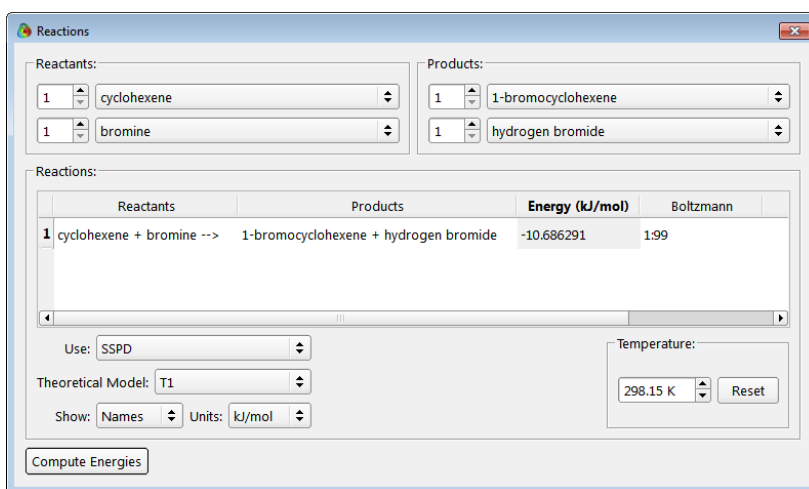
of theoretical models. This is made from the **Theoretical Model** menu.



The **Units** menu allows for selection of units.



A reaction energy is computed by *clicking* on **Compute Energies** at the bottom left of the dialog.



Substituted molecules can be used in place of real molecules as reactants and products. By accessing SMD or SSPD, this means that the energies for a series of related (by substitution) reactions can be computed at once.

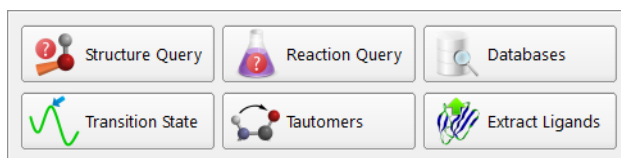
The results of a reaction energy calculation may be printed by *right clicking* inside the display area of the **Reactions** dialog and selecting **Print** from the menu that results.

The **Reactions** dialog is closed by either *clicking* on **OK** or on .

# Chapter 22

## The Search Menu

Functions available under the **Search** menu provide for defining structure and reaction queries to be used in searches of databases attached to **Spartan**. These include the Cambridge Structural Database (CSD) of experimental X-ray structures, the Spartan Molecular Database (SMD) of calculated structures, energies, and properties, 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. CSD and SRD are searchable by substructure, whereas SMD and SSPD are searchable by substructure, name, formula, molecular weight and isomer and SIRD and XIRD are searched based on matching to an unknown infrared spectrum. The **Search** menu also accesses procedures for automatically guessing transition states for reactions based on their similarity to transition states in SRD, for identifying tautomers, and for extracting ligands from PDB files.



### Structure Query (🔍)

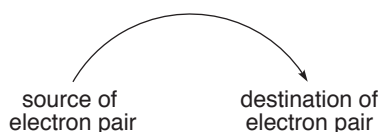
This designates one or more free valences on a structure as attachment points. *Clicking* on the free valence marks it with an orange cone and designates it as an attachment point. This means that anything directly bonded to the valence (including hydrogen) is an acceptable hit for a search on a database. 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 19**). These appear as ?'s on the sketch but as orange cones on the 3D structures.

## Reaction Query ( )

This serves either to identify the product or products of a chemical reaction from the reactants to be used for mining the Spartan Molecular Database or Spartan Spectra and Properties Database, or to provide an entry into the Spartan Reaction Database in order to access a previously calculated transition state, or to identify a similar transition state. All three functions make use of a set of curly arrows familiar to generations of organic chemists and formally corresponding to the movement of a pair of electrons.



There are two possible sources of an electron pair and three possible destinations, leading to six combinations:

lone pair $\rightarrow$ lone pair	move lone pair
lone pair $\rightarrow$ bond	use lone pair to increase bond order
lone pair $\rightarrow$ space between atoms	use lone pair to create new (single) bond
bond $\rightarrow$ lone pair	decrease bond order to make lone pair
bond $\rightarrow$ bond	decrease order of one bond to increase order of another bond
bond $\rightarrow$ 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.

The reaction query procedure presently 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 atom, bond, or two atoms as head. If two atoms hold SHIFT key.



*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 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. 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 of two detached fragments.

The process (tail selection followed by head selection) is repeated as necessary to fully define the reaction. Incorrect reaction arrows

---

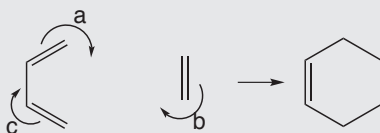
\* Note that this behavior is different from that in previous versions of *Spartan* where the **Shift** key was required to select the pair of atoms to receive the bond.

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 relate the properties of the products of chemical reactions to those of the products for the purpose of mining the Spartan Molecular Database or Spartan Spectra and Properties Database. Full discussion is provided in **Databases** later in this chapter. 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, structure queries (discussed in **Structure Query** earlier in this chapter) 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 **Guess Transition State** later in this chapter). Here, no structure queries would be defined.

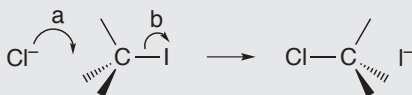
### Examples of Reaction Queries

#### *Diels-Alder reaction of 1,3-butadiene and ethylene*



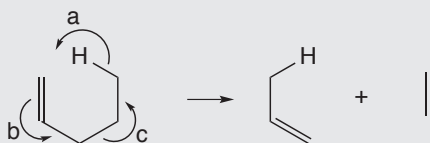
- a, b. double bond to empty space leading to single bond and single bond
- c. double bond to single bond leading to single bond and double bond

#### *S<sub>N</sub>2 reaction of chloride and methyl iodide*



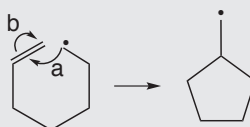
- a. lone pair on Cl<sup>-</sup> to empty space leading to ClC bond
- c. CI bond leading to lone pair on I

### *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<sub>2</sub> and C<sub>6</sub>) 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 Molecular Database (SMD) of calculated structures, energies, spectra and molecular properties, the Spartan Spectra and Properties Database (SSPD) of spectra, properties and QSAR descriptors, 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.\* CSD, SMD and SRD may be searched based on substructure. SMD and SSPD may also be searched by name, formula, molecular weight and isomer. If a selected molecule on screen exists in SMD and/or in SSPD, it may be replaced by a database entry and, even if it is not replaced, the molecule name in SMD may be used. SIRD and XIRD are searched

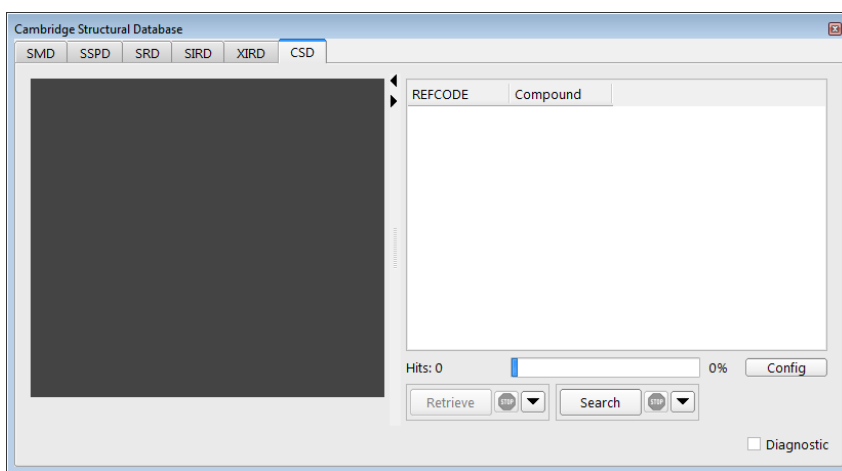
\* Databases may be located on the local machine or on a remote server. Setup details are provided in **Appendix G** for CSD and in **Databases Preferences (Preferences** under the **Options** menu; **Chapter 23)** for SMD, SSPD, SRD, SIRD and XIRD.

by comparison with an input infrared spectrum.

## Cambridge Structural Database (CSD)

The Cambridge Structural Database (CSD) is a collection of more than 650,000 experimental X-ray crystal structures for organic and organometallic molecules.\* It 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.

Access to the data in CSD is via a substructure searching procedure. 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

\* CSD is not included with *Spartan* but is available by subscription from CCDC (<http://www.ccdc.cam.ac.uk/>). For installation instructions, see **Appendix G**.

\*\* *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, that is supplied as part of the CSD subscription.

aspects of the search and subsequent transfer of CSD data into the file system. Both the preview window and the text box may be scaled independently. A substructure search is set up by specifying one or more of the following\*:

(i) **Attachment Points**


A free valence is designated as an attachment point by first selecting **Structure Query** from the **Search** menu and then *clicking* on it. It will be marked with an orange cone. *Clicking* again removes the designation and the cone. 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.

(ii) **Wild-Card Atoms**

An atom may be designated as a wild-card (meaning that element type is unimportant) by first selecting **Structure Query** from the **Search** menu and then *clicking* on it. An orange ball will surround the atom. *Clicking* again removes the wild-card designation and the ball. Use of wild-card atoms will result in structures that incorporate variants of the original substructure with different atoms at designated positions. For example, wild-card designation of an unsubstituted position of a substituted benzene will allow substituted pyridines to be located.

(iii) **Wild-Card Bonds**

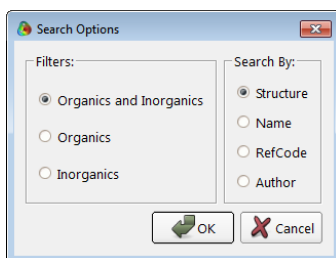
A bond may be designated as a wild-card (meaning that bond type is unimportant) by first selecting **Structure Query** from the **Search** menu and then *clicking* on it. An orange cylinder will be drawn around it. *Clicking* on this cylinder removes the wild-card designation and the cylinder. This is important in certain heterocycles where bond typing may be ambiguous.


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

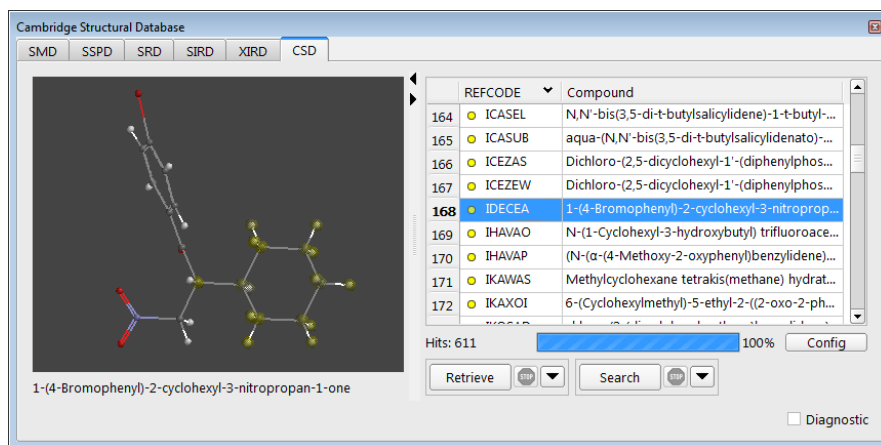
---

\* If nothing is 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.

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.



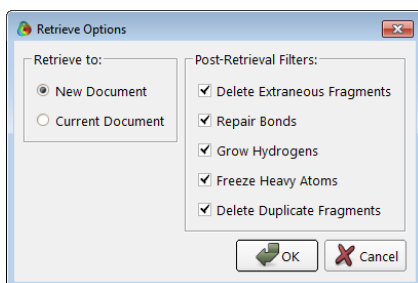
After a substructure has been specified, a search of CSD is initiated by *clicking* on the **Search** button at the bottom right of the dialog. If there are hits to be found, these will appear in the scroll box at the right of the dialog. The search can be terminated at any time by *clicking* on  to 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 (see **Output** under the **Display** menu; **Chapter 21**). A ball-and-wire 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.



\* 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) **Identify Ligands**

Identifies ligands in structures that are returned from CSD (primarily for organometallic compounds) and replace whatever bonding is supplied by a representation in which a dashed line is drawn between the geometric center of the ligand and the atom to which it connects.

(iii) **Repair Bonds**

If *checked*, attempts to fix bonding errors based on the actual geometry in CSD and on normal valence rules.

(iv) **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** from the **Display** menu; **Chapter 21**).


(v) **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.\*

(vi) **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.

## **Spartan Molecular Database (SMD) and Spartan Spectra and Properties Database (SSPD)**

*Spartan* connects to two different collections of molecules for which quantum chemical calculations have already been performed and the results of these calculations are available. The first is the Spartan Molecular Database (SMD). This comprises more than 500,000 records of calculations performed using one or more of

---

\* Hydrogen positions may be refined using molecular mechanics using **Minimize** under the **Build** menu; **Chapter 19**. Refinement using quantum chemical models is specified in the **Calculations** dialog (**Calculations...** under the **Setup** menu; **Chapter 20**) by way of **Frozen Atoms** (**Freeze Center** under the **Geometry** menu; **Chapter 18**).

the following quantum chemical models: Hartree-Fock with 3-21G, 6-31G\*, 6-311+G\*\* basis sets, B3LYP and EDF1 density functional and MP2 with 6-31G\* and 6-311+G\*\* basis sets and G3(MP2). The LACVP pseudopotential is used in place of the 6-31G\* basis set for main-group elements heavier than krypton and for second and third-row transition metals. Thus, a B3LYP/6-31G\* calculation on benzene tungsten tricarbonyl ( $\text{C}_6\text{H}_6\text{WCO}_3$ ) uses the LACVP pseudopotential for tungsten and the 6-31G\* basis set for hydrogen, carbon and oxygen. Except for the G3(MP2) entries, the geometry of each molecule has been optimized, assuming the best (lowest-energy) conformer obtained from the MMFF molecular mechanics model. G3(MP2) entries use MP2/6-31G\* equilibrium geometries. Included in the SMD entry is the geometry, the energy, the energy corrected empirically for aqueous solvation, the dipole moment and Mulliken, natural and electrostatic atomic charges. Properties related to geometry (area, volume and polar surface area) are calculated on-the-fly from the geometry. Also included, wherever they are available, is the heat of formation obtained from the T1 thermochemical model (see discussion in **Calculations...** under the **Setup** menu; **Chapter 20**) and the experimental heat of formation from the public NIST database.

The following recipe has been used to create the individual entries in SMD:

1. Identify the lowest-energy conformer using the MMFF molecular mechanics model. Molecules with fewer than 200-300 conformers are searched systematically (guaranteeing that the global minimum will be found), while Monte-Carlo sampling is used for molecules with more conformers.
2. Starting from the best conformer, perform a geometry optimization using the appropriate quantum chemical model. This implies that all quantum chemical models for a particular molecule will refer to the same conformer.

The Spartan Spectra and Properties Database (SSPD) presently comprises over 250,000 molecules. SSPD differs from SMD in several important ways. The first difference is that all calculations have been obtained from a single theoretical model, specifically the

EDF2/6-31G\* density functional model. This represents a credible attempt to get around the commonly-asked and difficult to answer question “which model is best to use?”. The EDF2/6-31G\* density functional model consistently yields reliable structures, reaction energies, spectra and properties, yet is simple enough to be routinely applied to reasonable-size molecules, that is, up to 300-500 amu.

The second difference is that, *wherever possible*, SSPD entries 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 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 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 an EDF2/6-31G\* geometry optimization on the lower energy (T1) conformer.

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**).

The third difference is that 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,  $^{13}\text{C}$  and  $^{19}\text{F}$  NMR chemical shifts have been empirically corrected and vicinal HH coupling constants have been estimated. This allows COSY, HSQC and HMBC NMR spectra (2D spectra) to be generated on-the-fly from the information in SSPD.

The fourth difference is that, in addition to the properties available from SMD, 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 both T1 and experimental heats of formation where they are available (as does each SMD entry).

The fifth difference is that each SSPD entry includes 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), SSPD has a significantly larger disk footprint than SMD (>11 Gb vs. 1.5 Gb). The in-memory footprint of SSPD (~300 Mb) is also larger than that for SMD (~175 Mb). Neither disk or in-memory footprints should seriously tax current generation computers.

SMD and SSPD are configured separately (**Databases** from **Preferences** under the **Options** menu; **Chapter 23**), and can either be used independently or in concert. SMD facilitates comparisons among different models, whereas SSPD provides simple access to a wide selection of data from a single high-quality model.

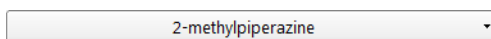
## Accessing and Processing Data From SMD and SSPD

There are two ways to access data from SMD and SSPD. The first way is to replace a structure that has been built (or read in) with a matching database entry. (Note that it is possible to replace all structures in a document for which there are matches in SMD or SSPD in a single

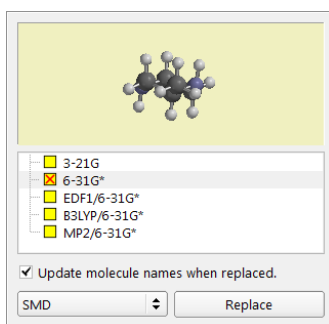
operation.) The second way is to search SMD or SSPD for entries that share particular structural features, or have part of a name in common or are isomers.

## Molecule Replacement

For the purpose of molecule replacement, SMD and SSPD are treated as independent databases. This gives the user the choice between the diversity of theoretical models represented in SMD and the simplicity brought on by the single model in SSPD. Molecule replacement is initiated by *clicking* on the molecule name displayed at the bottom of *Spartan*'s screen. The fact that a name appears means that the molecule is in the selected database (either SMD or SSPD).\*



This leads to a dialog with a menu at the bottom left that allows you to switch between the two databases.



If SMD is selected, the specific models for which data are available will be listed at the bottom of the dialog. By default, and if it is available, the 6-31G\* model (Hartree-Fock with the 6-31G\* basis set) will be selected. A different SMD entry, corresponding to a different quantum chemical model, may be selected by *checking* the box to the left of the name of the model in the list. Where SSPD is selected, only the EDF2/6-31G\* model is listed.

The model in the viewing window may be manipulated using the

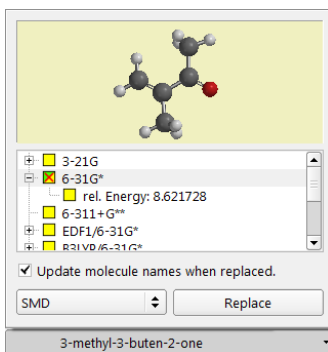
---

\* The absence of a name at the bottom of *Spartan*'s screen means that the molecule is not in the **selected** database. To see if it is in the other database, *click* inside the blank area and then change the database from SMD to SSPD or vice-versa using the menu at the bottom left of the dialog that appears.

usual mouse/keyboard commands (you need to position the cursor inside the viewing area). Model style may not be changed. Touch commands are also supported.

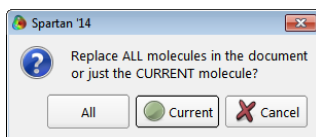
The selected model may be replaced by the selected database entry by *clicking* on **Replace** at the bottom right of the dialog. (Replacement can be undone by selecting **Undo** from the **Edit** menu; **Chapter 16**.) If the box to the left of “*Update molecule names when replaced*” is *checked*, replacement destroys all information. If it is not checked, replacement destroys all information except the molecular identifier (“name”) in the leftmost column of the spreadsheet (see **Spreadsheet** under the **Display** menu; **Chapter 21**).

Some SMD entries also include structures for higher-energy conformers (in addition to that for the lowest-energy conformer). These entries are designated by a ► to the left of the check box in the preview dialog, *clicking* on which leads to an expanded entry. This accesses a higher-energy conformer and provides their energies (in kJ/mol) relative to the best conformer. Higher-energy conformers are not available for SSPD entries.



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.

Note that the lowest-energy conformer for all other (non-selected) molecules in the document is used, even though a higher-energy conformer may be indicated for the selected molecule. (Replacement of document entries with non-lowest energy conformers needs to be done one-by-one selecting **Current** instead of **All**.) Molecules in the document for which there are no database entries at the specified theoretical model 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 SMD and SSPD

Access to data from SMD and SSPD follows from selecting **Databases**, and then *clicking* on either the **SMD** or **SSPD** tab (at the top left of the resulting window). (Aside from the title, the resulting displays are identical.) Each 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 SMD or 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.

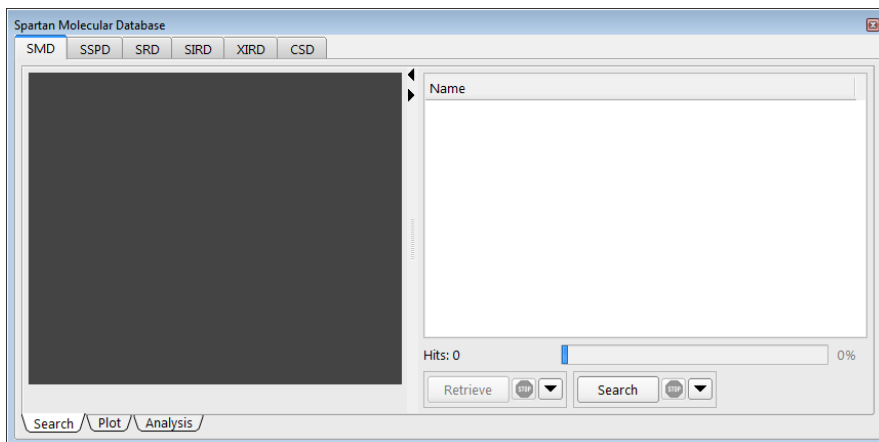
Substructure searching of SMD and SSPD is similar to that already described for substructure searching of the Cambridge Structural Database (CSD). The only required input is a three-dimensional structure. While one or more growth points (see **Structure Query** earlier in this chapter) will normally be specified, they are optional and in their absence an exact search will be carried out. Wild-card atoms and bonds used in searches of CSD are not supported. One or more substituents may be attached to the structure (see **Substituent Model Kit** under the **Build** menu; **Chapter 19**), in effect allowing multiple starting structures. (This capability is not available for CSD searches.) Name searching requires a name (or partial name), formula

---

\* For the 32 bit implementation of *Spartan*'14, the practical replacement limit is 1,500-2,000 molecules for SMD and 500-800 molecules for SSPD.

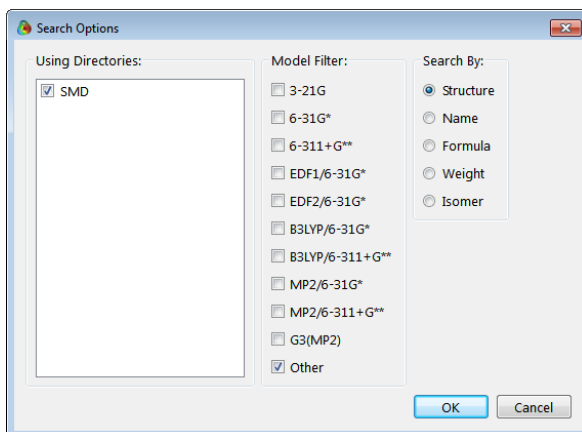
searching requires an exact formula, weight searching requires a molecular weight and isomer searching requires a three-dimensional structure.

*Clicking* on the **Search** tab leads to the **Search** dialog.



Following a search on SMD or 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 the databases to be searched, the kind of search, and (for SMD only) 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.

With the SMD tab selected, *clicking* on  to the right of the **Search** button leads to the **Search Options** dialog.



**Using Directories** lists the collections that make up SMD. A 6,000 molecule subset of SMD is supported with all copies of *Spartan'14* and the full database is provided as part of the *Spartan'14 Parallel Suite*. If the full database is installed, the subset should be removed (**Paths** from **Preferences** under the **Options** menu; **Chapter 23**). **Method Filters** specifies the theoretical models to be included in the search. **Other** accesses data with models that are not explicitly listed. **Search By** designates the type of search: **Structure**, **Name**, **Formula**, **Weight** or **Isomer**.

(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 **C<sub>6</sub>H<sub>5</sub>N<sub>1</sub>O<sub>2</sub>** will return not only nitrobenzene but also molecules such as phenyl nitrite and pyridine-4-carboxylic acid.

(iii) **Weight**

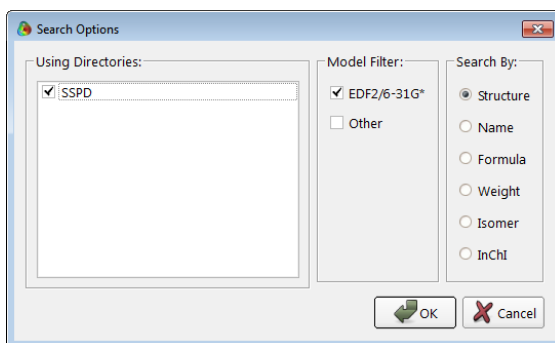
Selection followed by *clicking* on **OK** in the **Search Options** dialog leads to a box **By Weight** at the bottom right of the dialog. The search will contain all entries with the specified molecular weight  $\pm 0.05$  amu.


(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.

*Clicking* on **OK** removes the **Search Options** dialog with all settings maintained. *Clicking* on **Cancel** or  removes the dialog without maintaining the settings.

The **Search Options** dialog associated with SSPD is simpler in that **Methods Filters** is missing. Only a single theoretical model (the EDF2/6-31G\* density functional model) is available. **Using Directories** lists the collections that make up SSPD. A 6,000 molecule subset of SSPD is supplied with all copies of *Spartan'14* and the full database is provided as part of the *Spartan'14 Parallel Suite*. If the full database is installed, then the subset should be removed (**Paths** from **Preferences** under the **Options** menu; **Chapter 23**). The same search methods available for SMD are available here. In addition, a search of a (partial) InChI string may be performed.



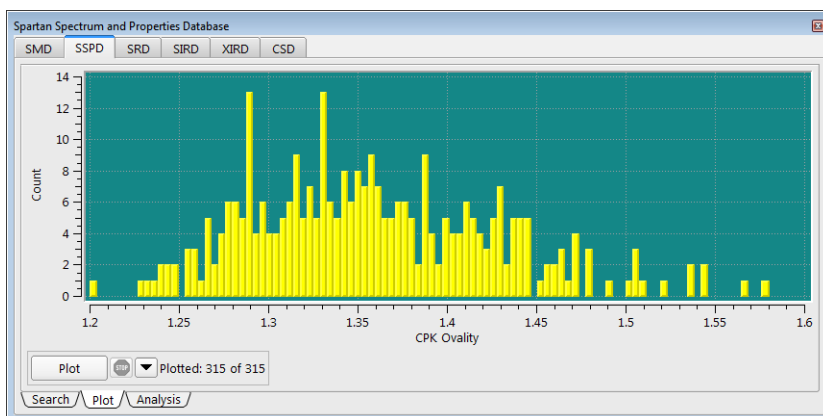
A search on either SMD or SSPD is carried out by *clicking* on the **Search** button at the bottom right of the (SMD or SSPD) search dialog, following which hits will appear at the right. The search may be terminated at any time by *clicking* on  to the right of the **Search** button. Any hits that have already been identified will be kept. A ball-and-spoke model of a hit may be examined at any time by *clicking* on its name in the box. It may be manipulated with the usual mouse/keyboard or touch-screen operations, although the model style cannot be changed.


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 21**), the formula editor (**Formulas** under the **Display** menu; **Chapter 21**) or from the main *Spartan* screen, primarily entries associated with the **Geometry** menu; **Chapter 18**. The hit list can be sorted using the property, by *right clicking* on the column header cell and selecting **Sort** from the menu.

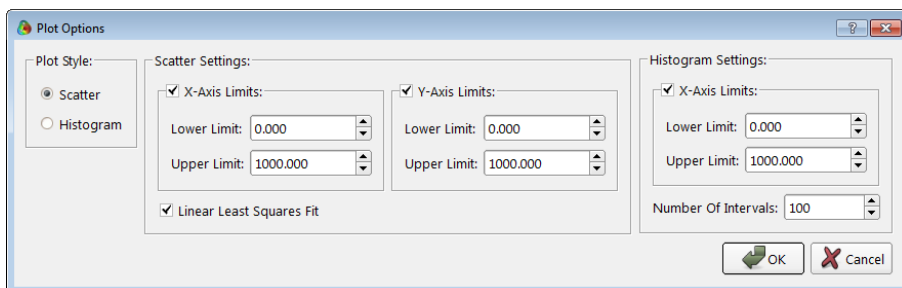
The results of a search on SMD or SSPD as well as any subsequent plotting and analyses are saved in the *Spartan* document corresponding to the original structure query or to the blank document that was used for a name or spectra query.


## Plotting Data Resulting from an SMD or SSPD Search


Data resulting from a search on SMD or 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.

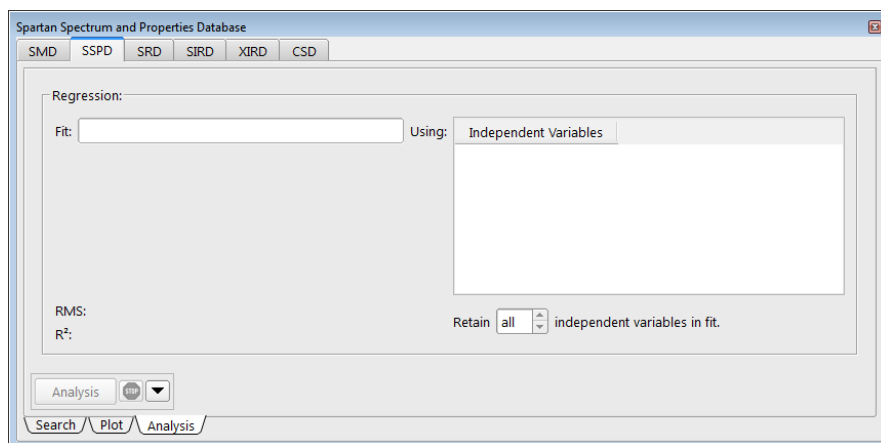



Default plot ranges are from minimum to maximum property values. *Checking Linear Least Squares Fit (Scatter Settings)* fits a straight line to the data. (A more general multi-variable regression analysis can be carried out from **Analysis**; see discussion following.) **Number of Intervals (Histogram 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 disk. 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 anytime, 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 SMD or SSPD Search

A regression analysis of data resulting from a search may be carried out by selecting the **Statistics** tab at the bottom left of the dialog.



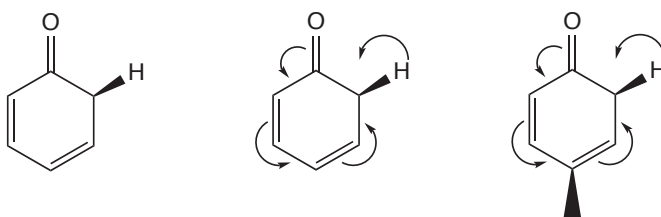
The name (not the numerical value) of the property to be fit is *dragged* from one of the properties dialogs, the formula editor or from the main **Spartan** screen, into the box to the right of **Fit**. The names of one or more properties to be used in the fit are *dragged* into the box to the right 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 disk. 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 stopping, 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 21**). The formula editor can also be employed to construct relationships between the reactants and the products of a chemical reaction of the 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 curly arrows (see **Reaction Query** earlier in this chapter). Note that 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 be favored over enols, here the reverse is certainly true as phenol benefits from aromatic stabilization. Three steps are needed to use SMD or SSPD to ask to what extent substituents alter this preference. First, a structure query (the reactant) must be built, arrows placed to designate the product of reaction and one or more growth points added to define the limits of the search. For example, use 2,4-cyclohexadieneone and limit substituents to the 4 position.



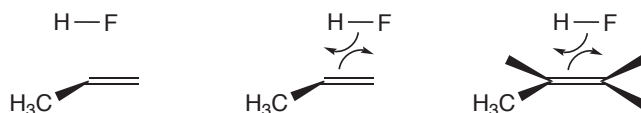
Next, a formula for the energy of reaction needs to be constructed. Inclusion of arrows extends the formula from the query itself (a 4-substituted 2,4-cyclohexadieneone), to the product of reaction (a *para*-substituted phenol) and to the difference between product and reactant. Selection of which is 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.

A search may now be carried out (*click* on the **Search** button inside the SMD or 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 SMD or 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, 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). Growth points attached to all three open valences extend the search to all available substituted propenes (for which analogous substituted 1-fluoropropanes are also available).



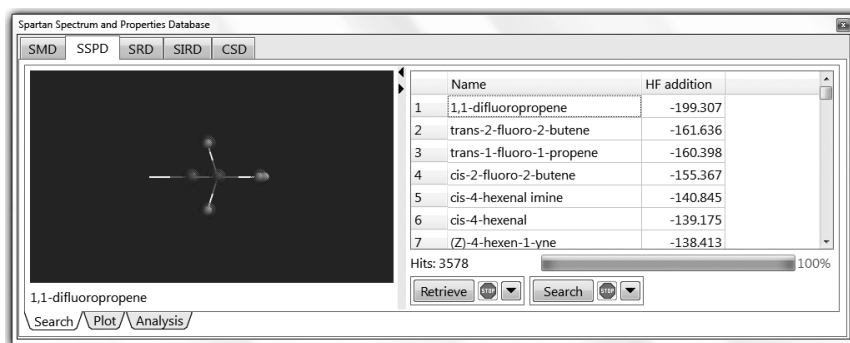
A formula for the energy of reaction may be constructed using the same sequence as in the previous example, selecting **(P-R) reaction** from the **Applied to** menu under **Context** in the formula editor, *clicking* on **E**, changing the default name and *clicking* on **Enter**.

Following a search using the query, dragging the formula for the

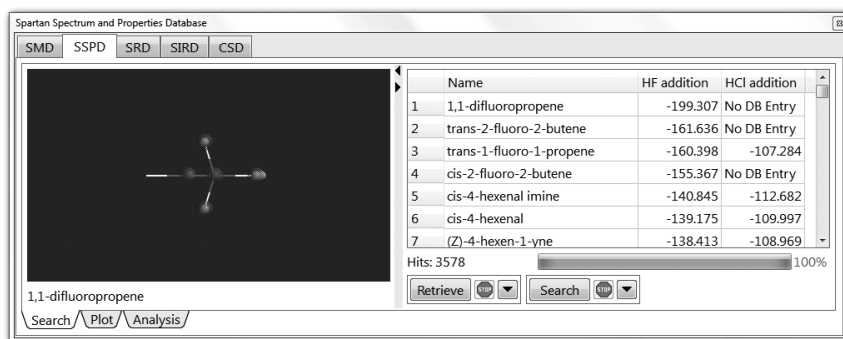
---

\* While there can be two or more reactants, only one can be provided growth points, that is, all other reactants need to be specific molecules.

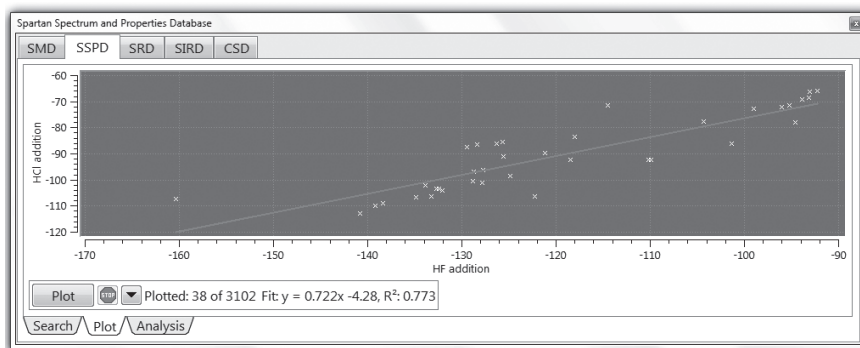
reaction energy onto an empty header cell of the hit list inside the SMD or SSPD **Search** dialog adds a column of reaction energies to the list.




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.)

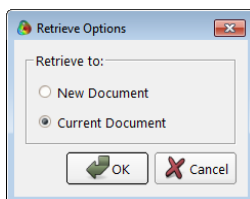



A plot of HCl vs. HF reaction energies from the SMD or SSPD **Plot** dialog shows the extent to which they are actually correlated.




## Retrieving Data from an SMD or 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 request for a file name; 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  to the right of the **Retrieve** button.

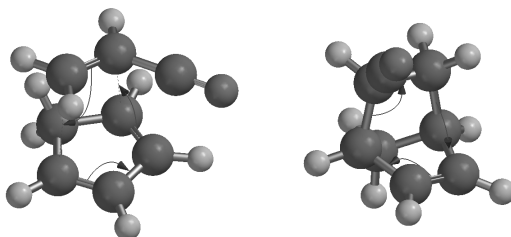
## Spartan Reaction Databases (SRD)

The **Spartan Reaction Database (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.

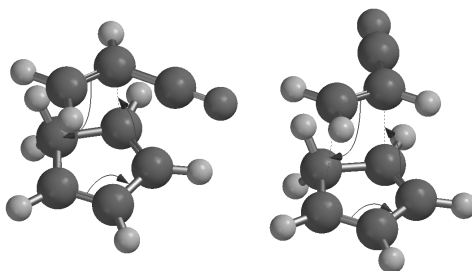
An entry in SRD is in terms of a three-dimensional structure (corresponding to the reactants or the product), together with an appropriate set of curly arrows (connecting the reactants to the product or vice versa) and designation of theoretical model.

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 on the left, while that for the corresponding retro Diels-Alder is given on the right.



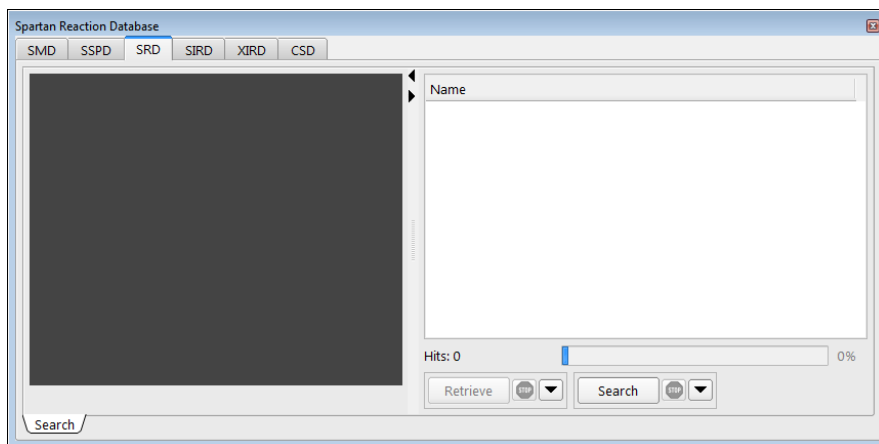
Note, that stereoisomers are distinguished by the way the reactants are positioned. For example, reactions leading to *endo* and *exo* Diels-Alder adducts are distinct.



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**).

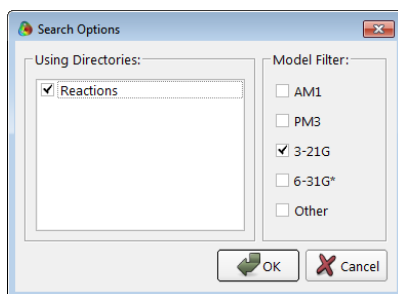
### Substructure Searching

Substructure searching from SRD is very similar to that from SMD and 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.

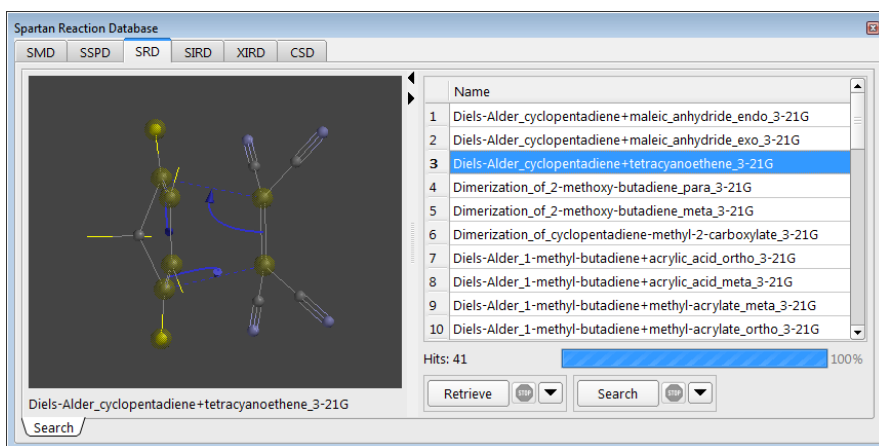
A substructure search is set up by specifying one or more attachment points on the reactant structure (see **Structure Query** earlier in this chapter) that has been augmented with reaction arrows (see **Reaction Query** earlier in this chapter). Prior to starting the actual search, it is necessary to 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 the 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 selections are not made.)

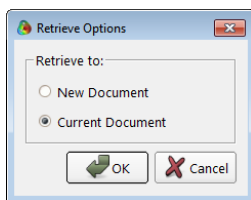
Following specification of structure/reaction arrows, attachment points and designation of theory level(s), a search on SRD is carried out by *clicking* on the **Search** button at the bottom right of the dialog. If there are hits, they will begin to appear in a few seconds in the box at the right of the dialog. The search may be terminated at any time by *clicking* on  to the right of the **Search** button at the bottom right of the dialog. Because the SRD is small, it may be difficult to stop the search before it has actually completed.





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 ball-and-wire 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 request for a file name; 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  to the right of the **Retrieve** button.

## Spartan Infrared Database (SIRD)

The Spartan Infrared Database (**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 ~250,000 molecules\* calculated from the EDF2/6-31G\* model and (where the T1 model is defined) based on the lowest-energy conformer from T1. 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.

## Spectra Searching

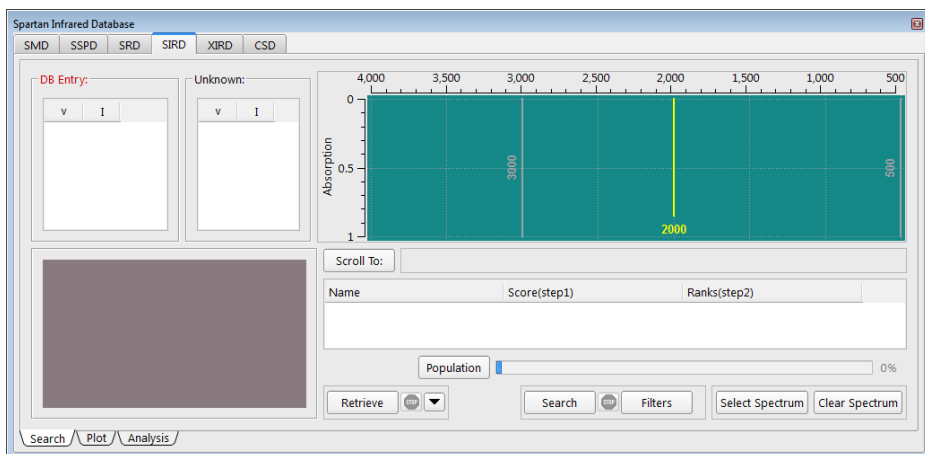
Access to SIRD is provided by *clicking* on the **SIRD** tab at the top of the **Database** dialog. A *Spartan* document must be open

---

\* As with SMD, a small subset of ~6,000 molecules from SSPD is provided with *Spartan*. The full set is provided as part of the *Spartan'14 Parallel Suite*.

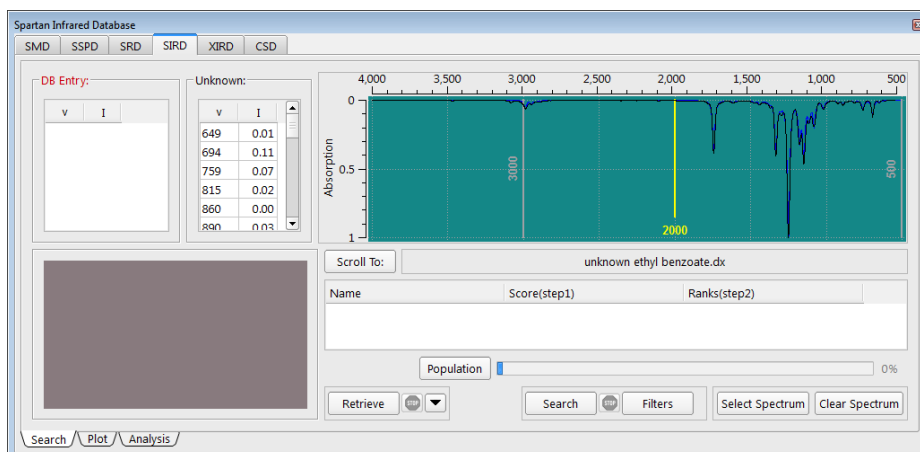
\*\* A sample .dx file is provided in **Appendix J**.

even if it is an empty document (**New** from the **File** menu). This gives rise to the **SIRD** dialog.



While this dialog is quite unlike those associated with other databases accessible from *Spartan*, it does, share two aspects with these dialogs, 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. It also contains a window at the top right for examining the match between the unknown (measured) infrared spectrum and that of a hit in the database, as well as boxes at the top left listing 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.

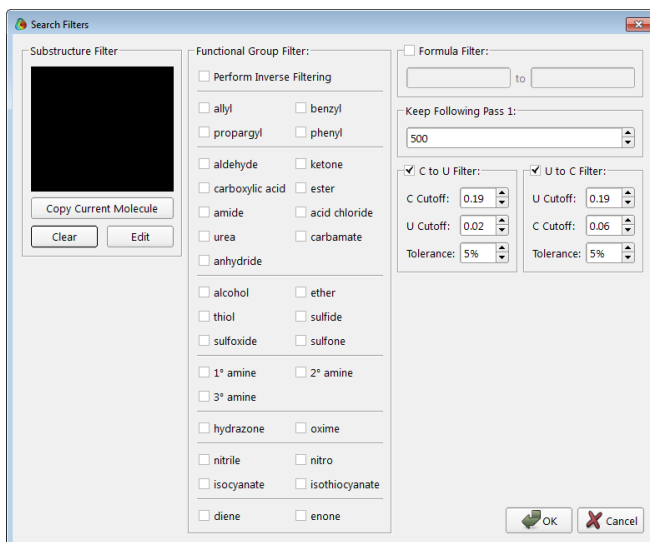
A search is initiated by *clicking* on **Select Spectrum** at the bottom right of the dialog. This leads to a file browser that has been set up to recognize only .dx files. Selection of a file in the usual manner followed by *clicking* on **Open** (or *double clicking* on the name of the .dx file) leads to the display of the unknown spectrum (in blue) in the window at the top right of the dialog.



The name of the (.dx) file is provided underneath the window. The unknown spectrum has been fit on-the-fly to a Lorentzian function. This allows a set of vibrational frequencies and intensities to be extracted, which are in turn 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 Lorentzian fit to the original spectrum is shown in black.

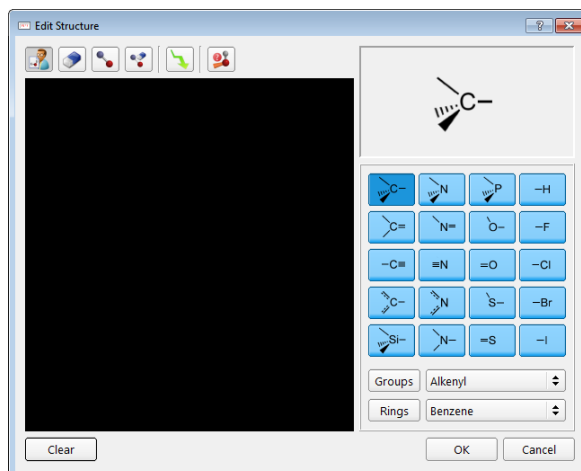
## Filters

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 to limit the molecular formula of hits on the database 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 19**) as well as the ability to define one or more attachment points (see **Structure Query** earlier in this chapter). Whatever substructure is built needs to be in the hit. Note that if no attachment points are designated, the molecule that is built 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  $C_cN_nO_o \dots$ , 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.

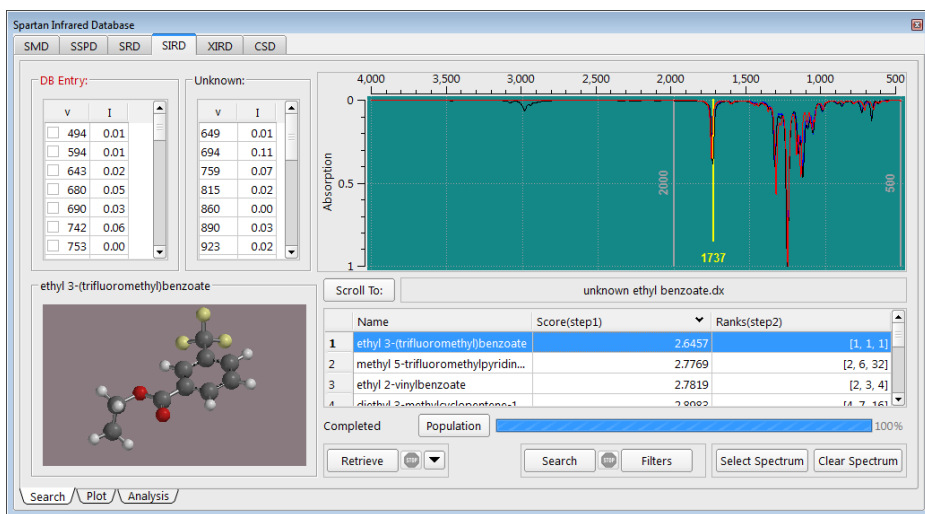
*Clicking* on **OK** exits the dialog. *Clicking* on **Cancel** or  also dismisses the dialog but any changes to the settings are lost.

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. Frequencies below  $500\text{--}600\text{ cm}^{-1}$  (which typically correspond to torsional motions and therefore may depend on conformation) require special instrumentation to measure and are seldom reported. While the experimental spectrum is usually recorded and reported to  $4500\text{ cm}^{-1}$ , the region beyond  $\sim 2800\text{ cm}^{-1}$  is dominated by CH stretching vibrations and may be too crowded to be of value in spectral searching. A reasonable range is from  $700\text{--}2200\text{ cm}^{-1}$ , but

this will depend on the details of the unknown spectrum.


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.

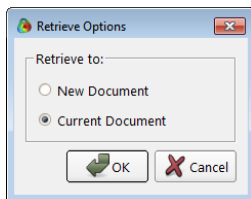
When the search has completed a listing of hits (the number of which is specified by **Keep Following Pass 2** in the **Search Options** dialog) ordered by their fit errors (lowest is best) appears in the box at the right of the **SIRD** dialog. As each hit is selected (*clicked* on in the list), its 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 an SIRD Search

Prior to retrieval on 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 destination is accomplished by  *toggling*  between **New Document** and **Current Document**. Selection of the former creates a new document which will need to be named later, while selection of the latter results in the retrieved molecules being appended to the end of the current document. The dialog is dismissed by *clicking* on **OK**. *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 the same setup and 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 (based on 2D connectivity) and not wave function and associated properties and spectra. Note, however, that the vast majority of NIST entries are in SSPD and/or SMD.

## 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 will attempt 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 similar to the linear synchronous transit method 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 curly arrows, where each arrow identifies the movement of one electron pair. The difference is that arrows are superimposed onto a three-dimensional structure rather than a two-dimensional drawing.

After all reaction arrows have been properly designated, *clicking* on  at the bottom right of the screen replaces the reactant 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.

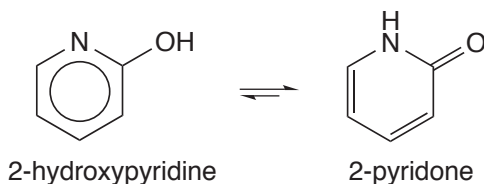
---

\* The 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.

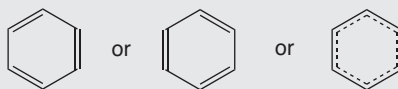
## Identify Tautomers (🧪)

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.



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.

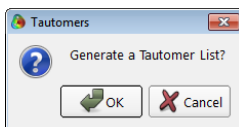



All of these translate into the same molecule.

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,  $\text{H}_3\text{N}^+\text{CH}_2\text{CO}_2^-$  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 (unnamed) 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 19**) prior to using them for quantum chemical calculations. It may also be important to perform a molecular mechanics search to obtain the lowest-energy conformer (see **Calculations...** under the **Setup** menu; **Chapter 20**).

## 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

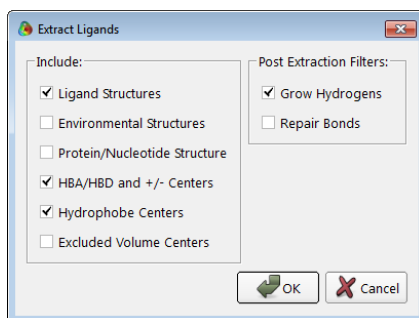
\* This capability is not restricted to files in PDB, but extends to files written in PDB format.

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 the screen.

Select Ligands.

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 whether **Ligand Structures** or 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 are to be abstracted from the PDB file, and whether amino acid and/or nucleotide residues in close hydrogen-bonding or charge-charge contact with the ligand (**Environmental Structures**) are to be identified. It also allows the **Protein/Nucleotide Structure** (minus all ligands) to be extracted.

---

\* This is the code given to small molecules that are associated with proteins/nucleotides in PDB files.

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 21**).

(ii) **Repair Bonds**

If *checked*, attempts to fix bonding errors based on the actual geometry in the PDB entry and on normal valence rules.

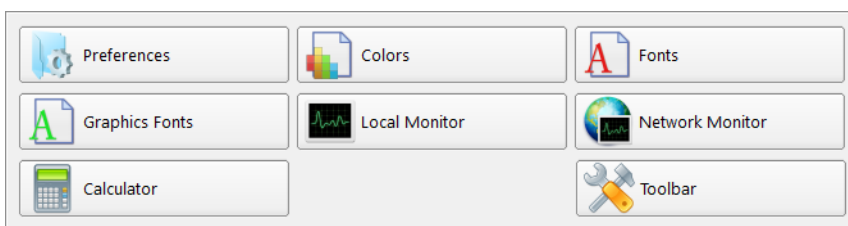
*Clicking* on **OK** at the bottom of the **Extract Ligands** dialog extracts the ligand(s) or pharmacophore(s), placing them in a single new (unnamed) document. *Clicking* on **Cancel** dismisses the dialog without ligand (pharmacophore) extraction.

# Chapter 23


## 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 program queue's, set icon displays and identify/change URL's for on-line accesses and configure **Spartan'14** to act as a remote server. They also allow for changing default colors and fonts and for monitoring executing jobs.*



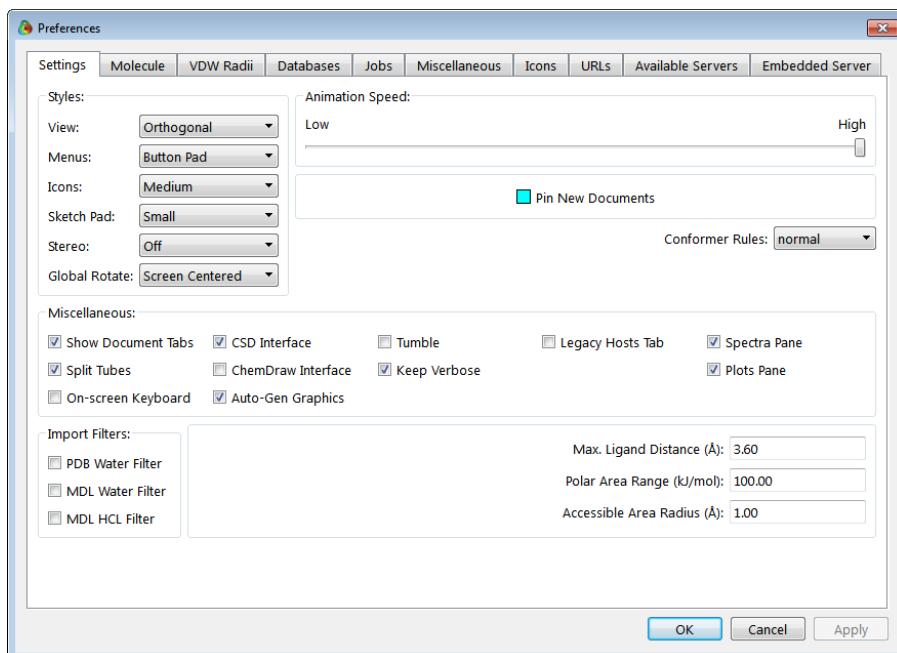
### Preferences...(⚙)

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'14** as a server for remote devices (**Embedded Server**). Selection results in one of ten dialogs, depending on which tab has been selected in the previous entry. *Clicking* on a tab brings up the associated dialog. To exit a **Preferences** dialog *click* on **OK**. *Clicking* on **Cancel** or  exits the dialog without instituting any changes.

---

\* **Preferences** is found under the **Spartan'14** menu in the Macintosh version.

# Settings



- (i) **View: Orthogonal/Perspective**  
Controls the view of structural models and graphics.
- (ii) **Menus: Classic List/Button Pad**  
Controls presentation of menus either as lists or as button pads (icon palettes). The latter is likely to be better suited to touch screen computers and tablets.
- (iii) **Icons: Small/Medium/Large/Extra Large/Jumbo**  
Controls size of the icons.
- (iv) **Sketch Pad: Small/Medium/Large**  
Controls the size of the sketch pad (palette of sketch tools) for *Spartan*'s 2D builder.
- (v) **Stereo: Off/Red-Cyan**  
Turns stereographic display on and off. This can also be done using the “3” key.
- (vi) **Global Rotate: Screen Centered/Molecule Centered**  
**Screen Centered** rotates all molecules about a common center; while **Molecule Centered** rotates each molecule about its own center.

(vii) **Animation Speed**

Controls the maximum speed for animations.

(viii) **Pin New Documents**

If *checked*, defaults to display of any new documents (from building or brought in from the **File** menu) irrespective of whether or not they are selected. Does not affect the status of existing documents. **Pin New Documents** has no meaning unless **Show Document Tabs** (see discussion following) is *checked*.

(ix) **Conformer Rules**

Chooses between rule sets for conformational searching. **Normal** should be used for **Equilibrium Conformer** and **Conformer Distribution** where Monte Carlo methods are involved. **Skeletal (Trimmed)** in previous versions of *Spartan* eliminates degrees of conformational freedom and should be used for **Conformer Library** where systematic searches are carried out. **Thorough** considers twist-boat conformers of six-member rings (in addition to chair conformers).

(x) **Show Document Tabs**

If *checked*, this displays a tab at the bottom of the screen for each open document. Only if the box to the left of the tab is *checked* will the molecule display (if it is not the selected document). If **Show Document Tabs** is not *checked*, the display corresponds to previous versions of *Spartan*, that is, all open documents are visible

(xi) **Split-Tubes**

If *checked*, the tubes in tube and ball-and-spoke models are split to designate multiple bonds.

(xii) **On-Screen Keyboard (Windows only)**

If *checked*, brings up an on-screen (virtual) keyboard for Windows 7 and 8 touch-screen computers and tablets. Should not be checked for non touch-screen devices or for touch-screen computers where the physical keyboard is to be used. Selection requires use of the **Spartan Open/Save**

**Document Set** (see **Miscellaneous** later in this chapter) which will be automatically set.

(xiii) **CSD Interface**

If *checked*, displays the **CSD** tab in the **Database** dialog (**Database** under the **Search** menu; **Chapter 22**). *Uncheck* if CSD is not installed.

(xiv) **ChemDraw Interface**

If *checked*, displays **ChemDraw** in the menu of model kits available to *Spartan*. *Uncheck* if **ChemDraw** is not installed.

(xv) **Auto-Gen Graphics**

If *checked*, graphics calculations will be performed in the interface (without having to submit the job). Does not pertain to orbital displays initiated from **Orbital Energies** under the **Display** menu (**Chapter 21**). Note, however, that graphics calculations on documents containing more than 25 molecules will not be performed in the interface irrespective of this setting.

(xvi) **Tumble**

If *checked*, allows automatic tumbling of molecule. To tumble a molecule, select it, *press* the left mouse button, move the mouse rapidly and release the button. To stop tumbling, *left click*.



To start tumbling, swipe one finger over the screen. To stop, tap anywhere on screen.

(xvii) **Keep Verbose**

If *checked*, keeps extended (verbose) output. Normally discarded upon successful completion, 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 verbose output is automatically kept for a job that has abnormally terminated.) Verbose output significantly increases the size of the *Spartan* document.

(xviii) **Legacy Host**

If *checked*, reverts to the method to access Linux remote compute hosts employed in previous versions of *Spartan*.

(xix) **Spectra Pane**

If *checked*, brings up a separate window for display of both calculated and experimental spectra. While there are some restrictions, for example, only one calculated (and its corresponding experimental) spectrum may be displayed at once, this mode is easier to use and more intuitive than that available in previous versions of *Spartan*. The previous display mode is invoked if the **Spectra Pane** is not checked.

(xx) **Plots Plane**

If *checked*, brings up a separate window for the display of 2D plots based on data entered into *Spartan's* spreadsheet. While there are some restrictions, for example, no 3D plots, this mode is easier to use and more intuitive than that available in previous versions of *Spartan*. The previous display mode is invoked if **Plots Pane** is not checked.

(xxi) **PDB Water Filter**

If *checked*, removes water molecules from imported PDB files.

(xxii) **MDL Water Filter**

If *checked*, removes water molecules from SD files.

(xxiii) **MDL HCl Filter**

If *checked*, removes HCl molecules from SD files.

(xxiv) **Max-Ligand Distance**

Set maximum distance (in Å) for an amino acid in a protein to be recognized as bonded to a ligand.

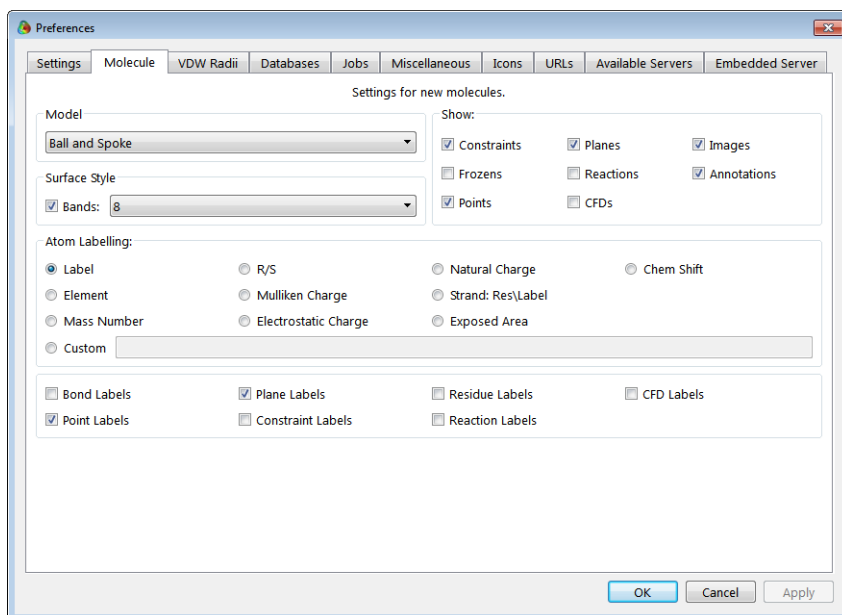
(xxv) **Polar Area Range**

Set potential (in kJ/mol) for calculating polar area from electrostatic potential map. Values > | range | contribute to polar area. Default is 100 kJ/mol.

## (xxvi) Accessible Area Radius

Set sphere radius (in Å) for determining accessible area.  
Default is 1.0Å.

## Molecule



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 21**).

### (i) **Model: Wire/Ball and Wire/Tube/Ball/Spoke/Space Filling/Line**

Controls default model style.

### (ii) **Surface Style**

If **Bonds** is *checked*, 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.

(iii) **Show: Constraints/Frozens/Points/Planes/Reactions/CFD's/Images/Annotations**

If *checked*, constraints and frozen markers, points and planes, reaction arrows, CFD's and attached images and text will always be shown as part of the model. Otherwise, they will be shown only in the appropriate mode.

(iv) **Atom Labeling: Label/Element/Mass Number/Mulliken Charge/Electrostatic Charge/Natural Charge/Strand: Residue/Label/R/S/Exposed Area/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**).

(v) **Bond Labels**

If *checked*, bond labels will be shown.

(vi) **Point Labels**

If *checked*, point labels will be shown.

(vii) **Plane Labels**

If *checked*, plane labels will be shown.

(viii) **Constraint Labels**

If *checked*, constraint labels will be shown.

(ix) **Residue Labels**

If *checked*, residue labels will be shown.

(x) **Reaction Labels**

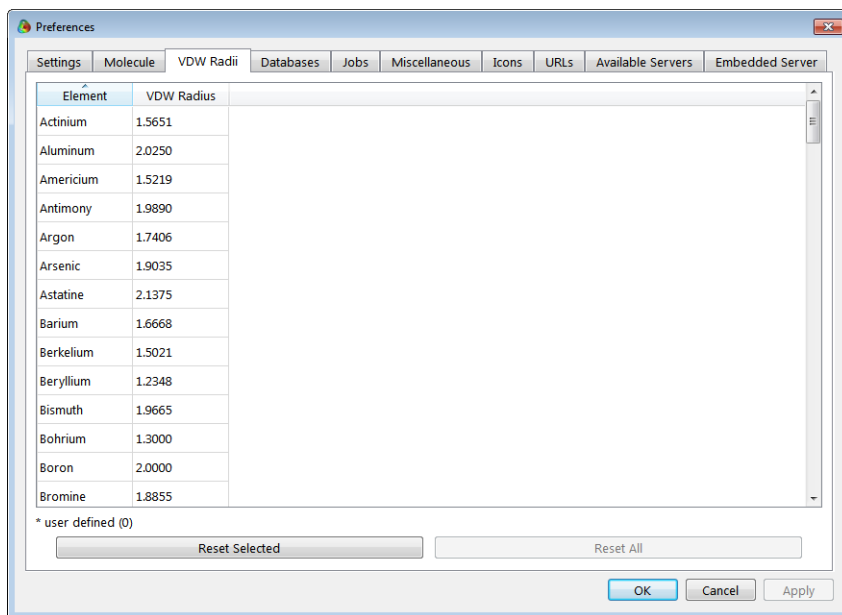
If *checked*, reaction arrow labels will be shown.

(xi) **CFD Labels**

If *checked*, CFD labels will be shown.

## VDW Radii

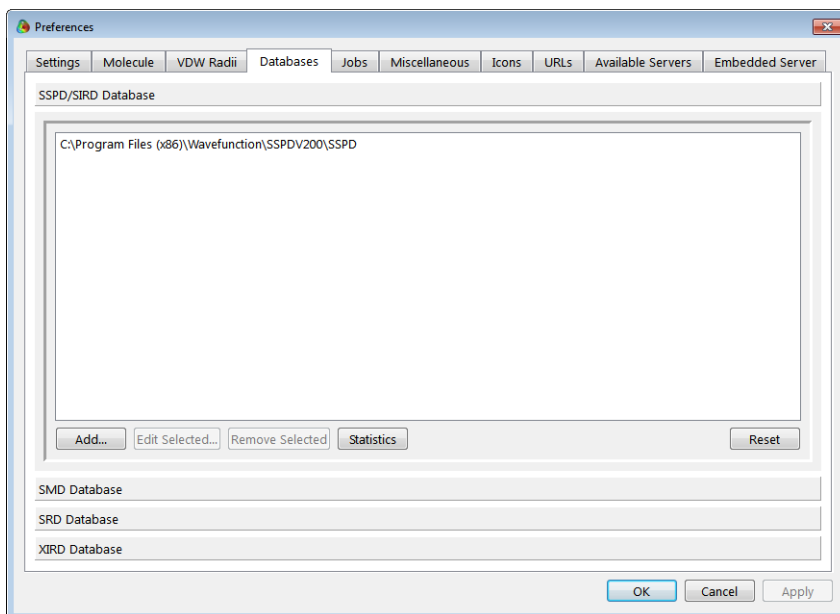
This provides a list of van der Waals radii



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

This allows setting up of paths for SMD, SSPD/SIRD, SRD and XIRD database directories.



Several directories may appear for each database. Specification of which directory (or directories) is to be searched in the **Search Options** dialogs associated with SMD, SSPD, SIRD, SRD and XIRD (**Databases** under the **Search** menu; **Chapter 22**). Note that SIRD is the same database as SSPD\*, and only the method of access (spectral matching vs. substructure matching) are different.

All copies of *Spartan* are shipped with ~6,000 molecule subsets of both SMD and SSPD databases. The full databases are provided as part of the *Spartan'14 Parallel Suite*. To avoid duplicates, the subset databases should be removed upon installation of the full SMD and/or SSPD databases.

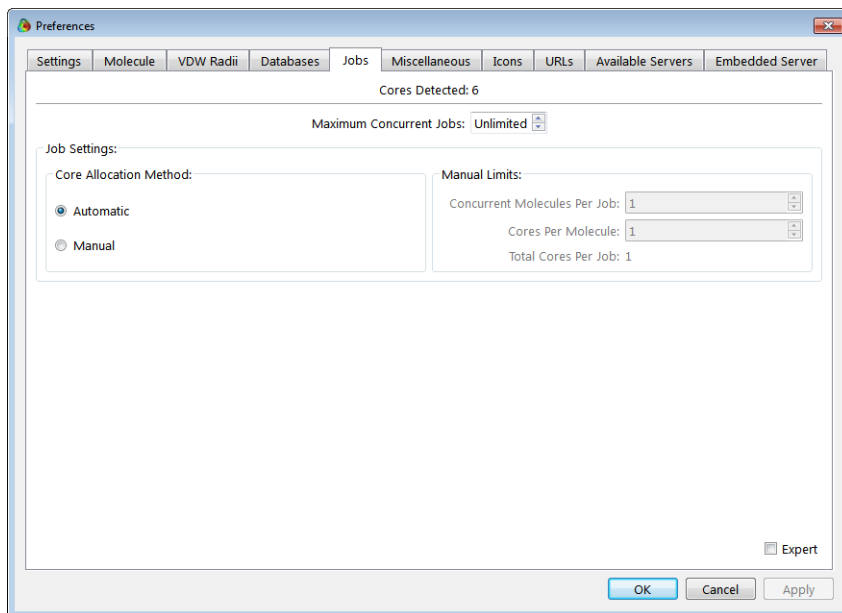
User-defined collections may be added (see **Appendix H**).

---

\* More precisely, it is the subset of entries in SSPD that contain infrared spectra.

## Jobs

This allows setting of job limits.



### (i) Maximum Concurrent Jobs

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 processors (cores) on your computer and also that you have sufficient memory to support this maximum number.

### (ii) Core Allocation Method

Toggles between **Automatic** and **Manual**. **Automatic** will assign all available cores to a single molecule (single molecule inside a *Spartan* document) if there is only one molecule in the document and will divide the cores among the molecules in the document if there is more than one. **Manual** will follow settings under **Manual Limits**.

Parallel processing, that is, allocation of more than one core to a single molecule is only available in the *Spartan'14 Parallel Suite*.

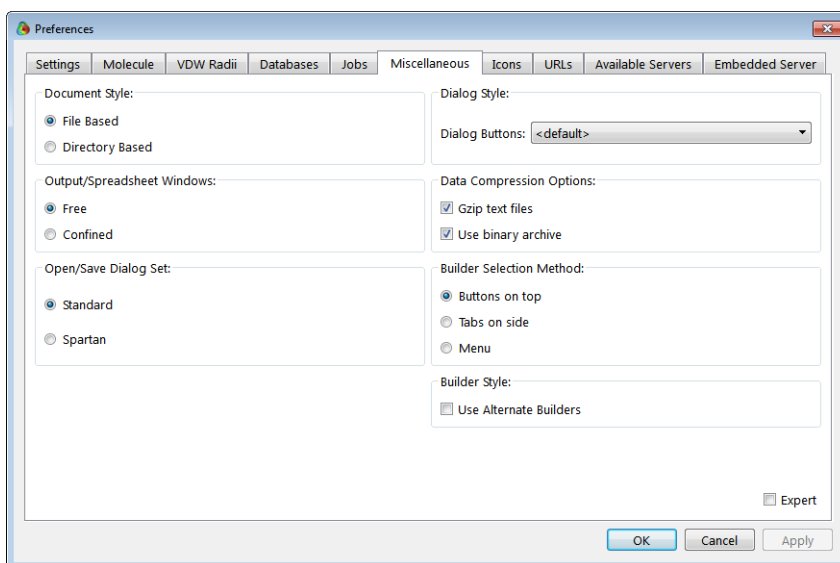
### (iii) Manual Limits

**Concurrent Molecules Per Job** sets the number of jobs (molecules inside a *Spartan* document) that will run simultaneously and **Cores Per Molecule** 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 in specifying more than one core/job. It is very easy (and, depending on available memory, potentially very inefficient) to demand more cores 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.

## Miscellaneous

Items here refer to cross-platform applications and should not be of concern to users of a single platform.



### (i) Document Style

Toggles between **File Based** and **Directory Based**. The latter is for compatibility between Windows and previous Macintosh and Linux versions of *Spartan*.

### (ii) Output/Spreadsheet Window

Toggles between allowing the output window and the

spreadsheet to be **Free** (able to move outside the main *Spartan* window) or **Confined** (restricted to move inside the main *Spartan* window). Previous versions of *Spartan* confined these windows.

(iii) **Open/Save Document Set** (Windows only)

Toggles between **Standard** and *Spartan* specific dialogs for functions under the **File** menu that involve a file browser. *Spartan* specific dialogs are required to support virtual keyboard on touch-screen devices.

(iv) **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.

(v) **Data Compression Options**

Allows compression of text files (**Gzip text files**) and switches to binary for full archive storage (**Use binary archive**). May be required for file compatibility with earlier versions of *Spartan*.

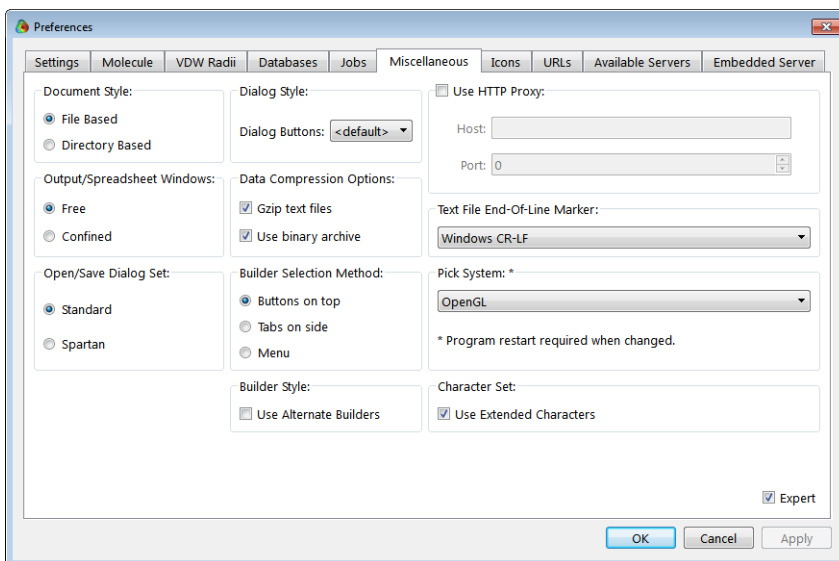
(vi) **Builder Selection Method**

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.

(vii) **Builder Style**

**Use Alternative Builder** provides a different presentation of the model kits in the 3D builder. Try both and see which one you like.

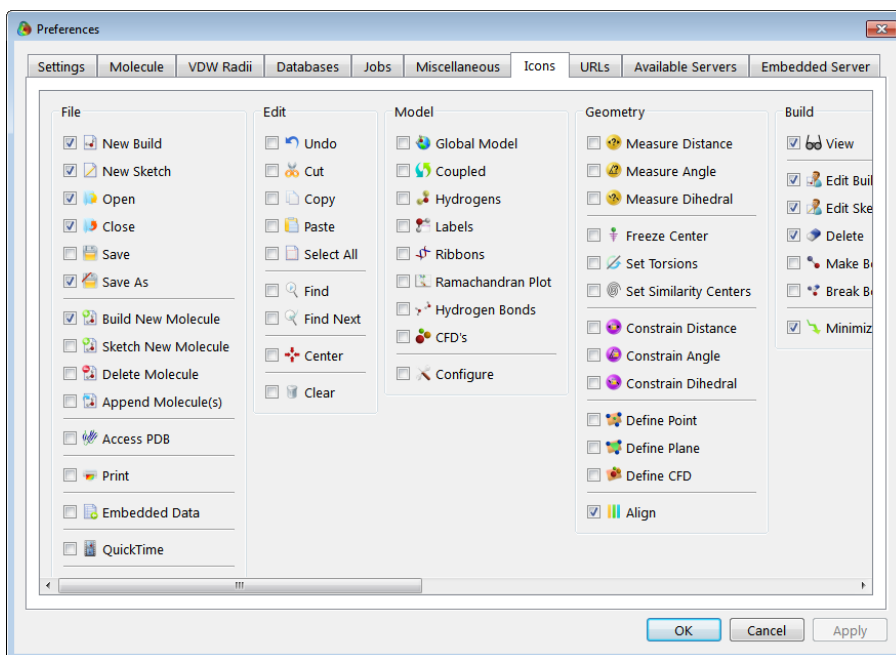
*Checking Expert* at the bottom right brings up an extended dialog, containing controls that may be needed in exceptional circumstances.



- (i) **Use HTTP Proxy**  
Allows setting up of an alternative path for access to external websites, for example, experimental spectra databases. Rarely needed.
- (ii) **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. Rarely needed.
- (iii) **Pick System**  
Toggles between **OpenGL**, **Color** and **Hybrid** picking models. **OpenGL** is the standard but causes problems for Intel HD4000 graphics (very common), where either **Color** or **Hybrid** should be used. Graphics chip is automatically detected at installation and this control should be properly set.
- (iv) **Character Set**  
**Use Extended Characters** may need to be switched off if the fonts that are chosen do not support subscripts and superscripts.

## Icons

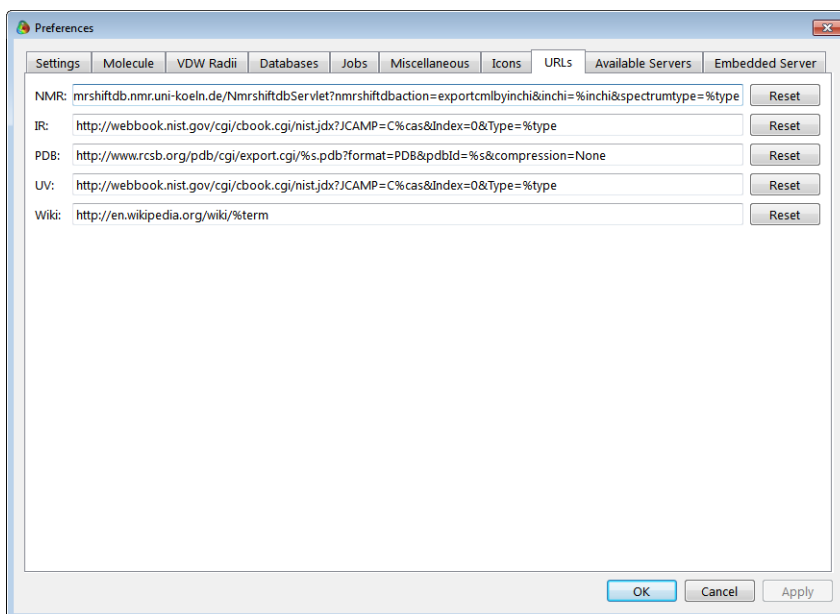
Icons for all menu entries are listed (you may need to use the horizontal slider bar to see them all). If *checked*, the icon will appear at the top of the *Spartan* screen.



Icon display is limited to one “permanent” row and approximately 20 medium size icons will fit on screen.

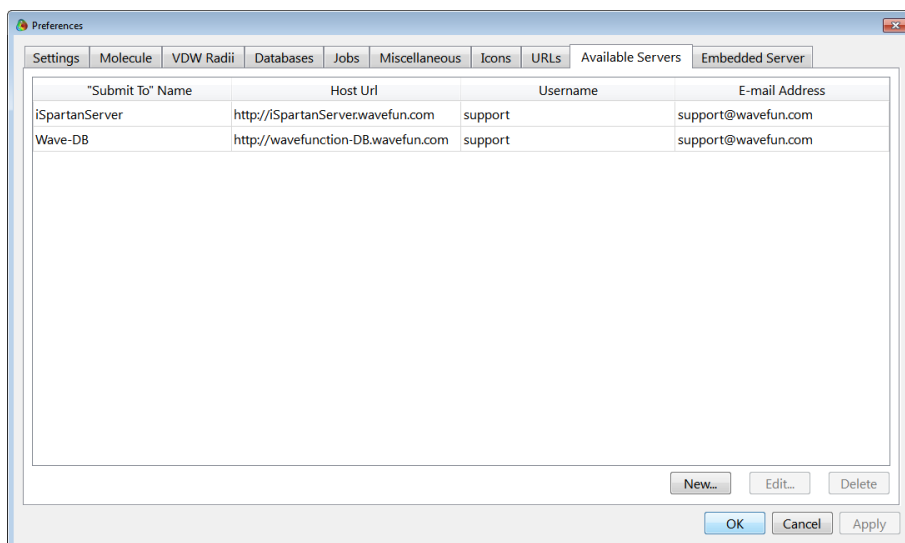
## URLs

Lists URLs for access to experimental structural and spectral databases and to Wikipedia.

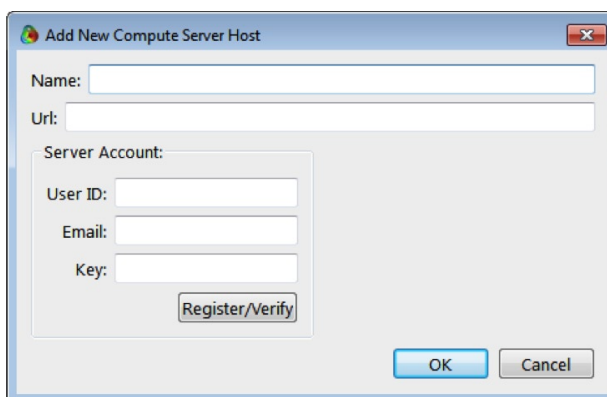


## Available Servers

In addition to processing molecular mechanics and quantum chemical calculations “locally”, *Spartan*<sup>14</sup> is able to divert calculations to a remote server on the network, in practice, a server associated with another copy of *Spartan*<sup>14</sup> located on another machine. Server capability is provided as part of the *Spartan*<sup>14</sup> **Parallel Suite**, and accounts are set up using the **Embedded Server** tab in the **Preferences** dialog (see later in this chapter). **Available Servers** 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.



- (i) **Name**  
The name of the server. This will appear in the **Submit to** menu inside the **Submit** dialog.
- (ii) **URL**  
The URL of the server. This can either be provided as a text string (DNS suffix) or as an IP address.
- (iii) **User ID**  
A text string used by the administrator of the server to identify you.

(iv) **Email**

Your e-mail address. This is needed by the administrator of the server in order to contact you.

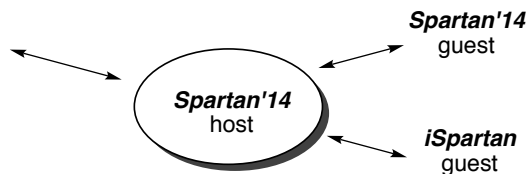
(v) **Key**

This is required only in the event that the server operates in high-security (**Disabled**) mode. This is controlled by the administrator of the server from the **Embedded Server** tab in the **Preferences** dialog (see later in this chapter).

Clicking on **Register/Verify** will setup an account (or will respond with any error conditions).

**Embedded Server** (Only Available in the *Spartan'14 Parallel Suite*)

In addition to its local capabilities, *Spartan'14* can be used as an embedded server (*host*) for licensed copies of *Spartan'14* running on other machines as well as for copies of *iSpartan* running on iPads and iPhones (*guests*).

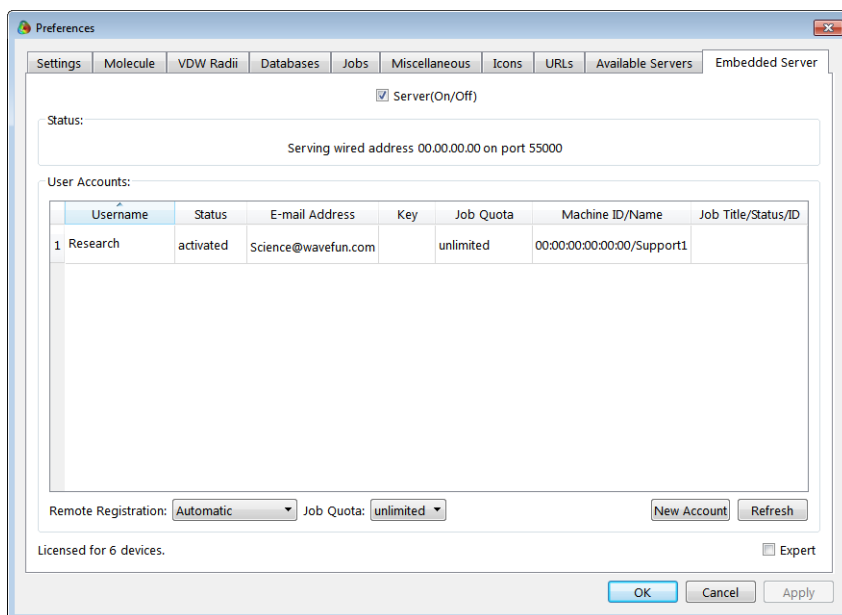


The embedded server is available only as part of the *Spartan'14 Parallel Suite* and is initially licensed for only two clients (licensing can be extended).

Operation of the server requires communication between the host and the guests. At the lowest (and default) level of security, nothing needs to be done at the host end (except turning on the server) and each of the guests simply needs to know about the host. Higher security levels require explicit action on the part of the administrator of the host, and the highest-level requirements requires explicit exchange of “keys” between the administrator and the guests.

## Turning the Server On and Off

Checking the box to the left of **Server (On/Off)** in the **Embedded Server Preferences** 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**.



## Registering Accounts

From simplest and least secure to more complicated and most secure, there are three ways to “register” a guest onto 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’14* or *iSpartan* (guest) can set up an account without the administrator (of *Spartan’14* with an embedded server which is to act as the host) 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 guest 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 **Embedded Server Preferences** dialog.

**Disabled.** In this mode, the administrator is required to set-up an account. This requires that the user supplies the administrator with the MAC address of the computer. For *Spartan'14*, this is available from the **About Spartan'14** dialog (**Help** menu). For iOS devices, this is identical to the Wi-Fi address located under Settings > General Settings > About. The MAC address 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).

### **Account Attributes**

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.

#### **(i) User ID**

This is the name of the account provided by the user at the time the account was set up.

#### **(ii) 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.

#### **(iii) Email Address**

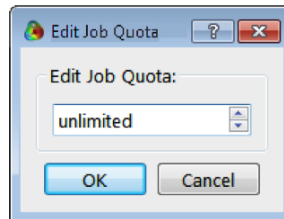
This is the e-mail address of the user provided at the time the account was set-up.

#### **(iv) 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 (**Disabled**) mode.

### (v) 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.



### (vi) Machine ID/Name

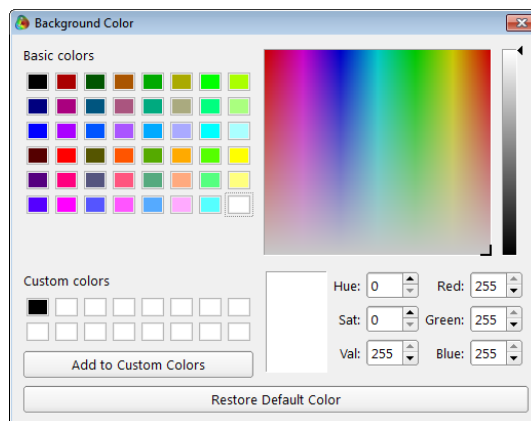
This provides the ID of the users computer running *Spartan*’14 or *iSpartan* (a “device”). The total number of devices that can access a server is limited by license (on the server). To increase the number, contact [sales@wavefun.com](mailto:sales@wavefun.com).


### (vii) Job Title/Status/ID

This identifies any queued or running jobs associated with a user account.

## Colors

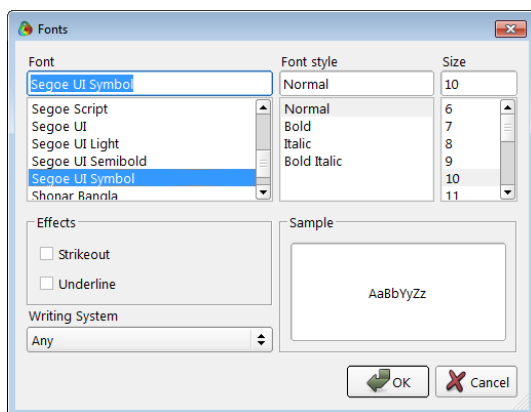
This alters default colors. Selection leads to the **Colors** dialog.




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. 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 21**). *Clicking* on  removes the dialog.

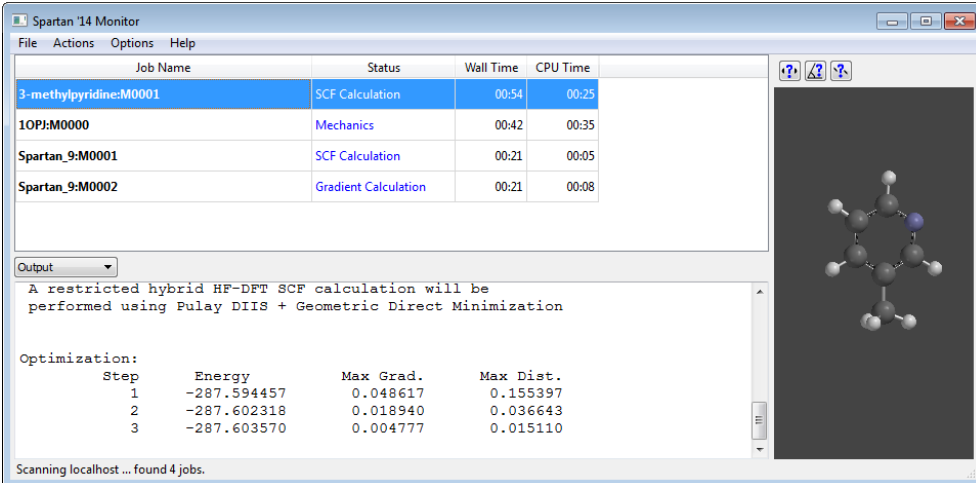
## Fonts/Graphics Fonts ( /

This selects fonts, style and size of labels attached to molecules (**Labels** and **Configure...** under the **Model** menu; **Chapter 17**), and plots (**Plots...** under the **Display** menu; **Chapter 21**). 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.

## Local Monitor



The screenshot shows the 'Spartan '14 Monitor' window. It features a menu bar (File, Actions, Options, Help) and a table of jobs. The table has columns for Job Name, Status, Wall Time, and CPU Time. The first job, '3-methylpyridine:M0001', is selected and highlighted in blue. Below the table is an 'Output' section with a dropdown menu and a text area showing calculation details and an optimization table. To the right of the output is a 3D ball-and-stick model of the 3-methylpyridine molecule. At the bottom, a status bar indicates 'Scanning localhost ... found 4 jobs.'

Job Name	Status	Wall Time	CPU Time
3-methylpyridine:M0001	SCF Calculation	00:54	00:25
1OPJ:M0000	Mechanics	00:42	00:35
Spartan_9:M0001	SCF Calculation	00:21	00:05
Spartan_9:M0002	Gradient Calculation	00:21	00:08

Output

A restricted hybrid HF-DFT SCF calculation will be performed using Pulay DIIS + Geometric Direct Minimization

Optimization:

Step	Energy	Max Grad.	Max Dist.
1	-287.594457	0.048617	0.155397
2	-287.602318	0.018940	0.036643
3	-287.603570	0.004777	0.015110

Scanning localhost ... found 4 jobs.

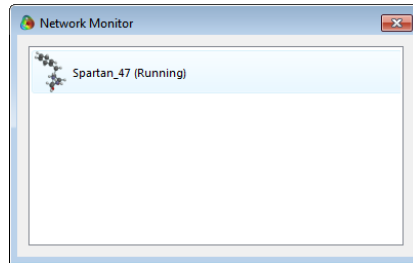
This provides a listing of all executing/queued jobs and their status. To see accumulated output for an executing job, *click* on its name. A ball-and-spoke model of the selected (executing) job will be displayed in a window to the right of the dialog. It can be manipulated using the usual mouse commands (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 kill a job, *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**), *click* on its name and select **Start** from the **Actions** menu (or *right click* on its name and select **Start** from the menu that appears).

The **Local Monitor** may be removed either by selecting **Exit** from the **File** menu or by *clicking* on  at the top of the dialog.

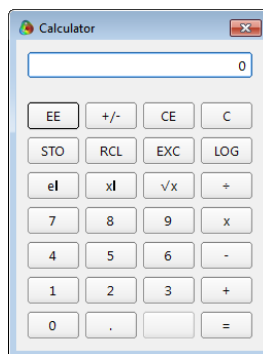
## Network Monitor


This provides the names of any jobs that are presently running on any remote servers accessed by *Spartan*. No further information is available.



## Calculator

Selection brings up a **Calculator**.



This functions the same way as a normal pocket calculator. The **Calculator** is removed by *clicking* on .

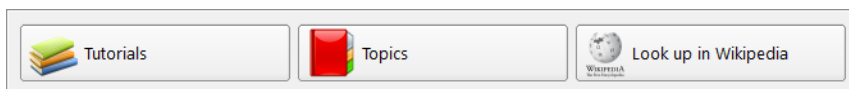
## Icons

Toggles the set of icons that appear above the menus on and off.

# Chapter 24

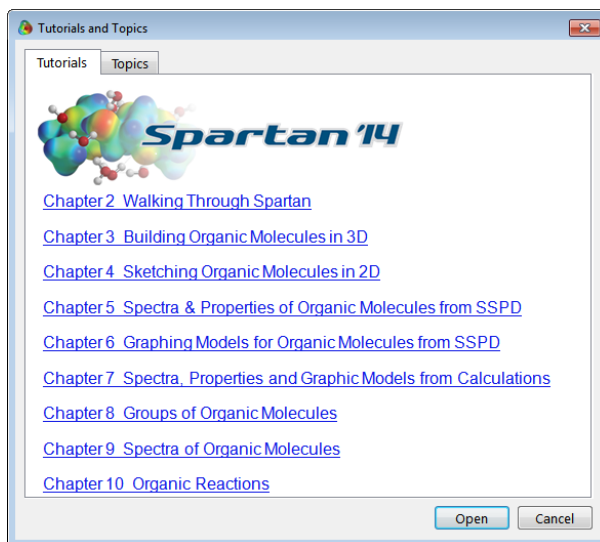
## The Activities Menu

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 ( / )

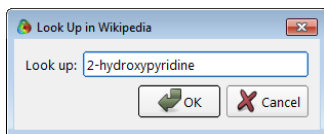
Selection of **Tutorials** or **Topics** brings up a dialog with either the **Tutorials** or **Topics** tab selected (**Tutorials** tab shown).



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.

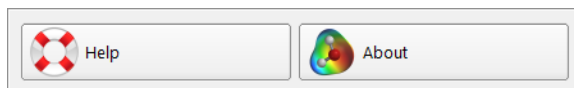


By default, the query is the name of the selected molecule. *Clicking* on **OK** or entering a different query followed by *clicking* on **OK** leads to a Wikipedia page. This occupies a window that is external to *Spartan*.

# Chapter 25

## The Help Menu

*This chapter describes help.*




### Help ( )

This provides information relating to application of computational methods available in *Spartan*, as well as technical details regarding the program's operation.

A number of topics are dealt with under **Help**; including *Mouse/Keyboard Functions*, *General Operating Features*, *Units*, *Differences between Spartan'08 and '10*, *Selecting a Theoretical Model*, *Relative Computational Times*, *Memory Usage*, *Calculation Options*, and a number of FAQ documents (*Energy Questions*, *Convergence Questions*, *Questions about Basis Sets*, *Problems with Memory*, *Questions about the Property Module*, *Questions about Conformations and Energy Profiles*, *MMFF94 Compliance*, *Units*) related to computational questions. **Help** also provides a link to Wavefunction's website. Help files are HTML documents and require that Internet Explorer is installed to access them.

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.

About...(  )\*

Provides information about the user's release of *Spartan*.



\* About is located under the **Spartan14** menu on the Macintosh version.

# Appendix A

## Topics

---

Thirteen short essays address topics that are important to molecular modeling in general and to modeling applications of *Spartan* in particular. For the most part, our discussions are at a very elementary level and lack detail.

The first two essays are the most fundamental. **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 what is actually required to reliably reproduce the experimental measure of energy.

**Dealing with Conformationally-Flexible Molecules** describes what must be done in order to identify the lowest-energy conformer of a flexible molecule. **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** shows how a more complex parameterization scheme is needed to bring proton and  $^{13}\text{C}$  spectra into close agreement with experiment.

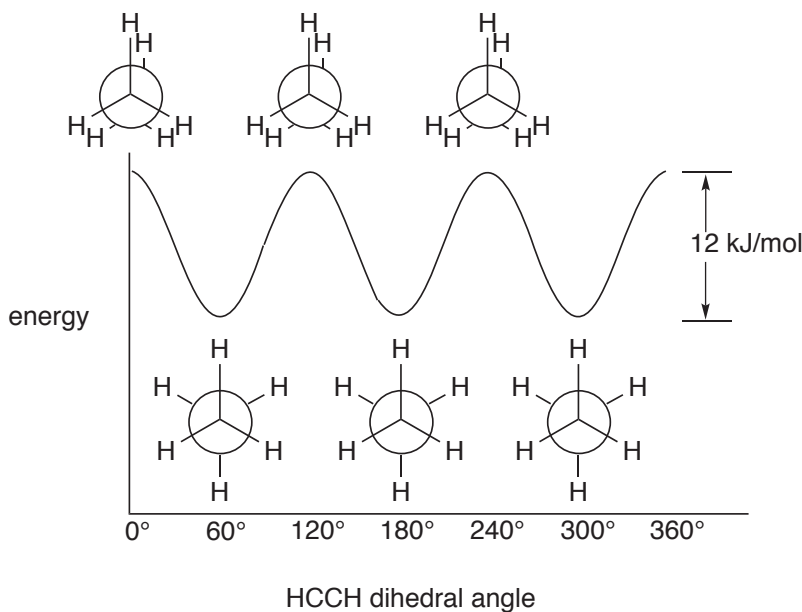
**Atomic and Molecular Orbitals** describes how molecular orbitals arise from atomic orbitals in the context of the practical quantum

chemical models introduced previously, and illustrates the information that can be drawn from them. While molecular orbitals are themselves not observable quantities, the square of the molecular orbitals (the electron density) corresponds to what is actually measured in an X-ray diffraction experiment. **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

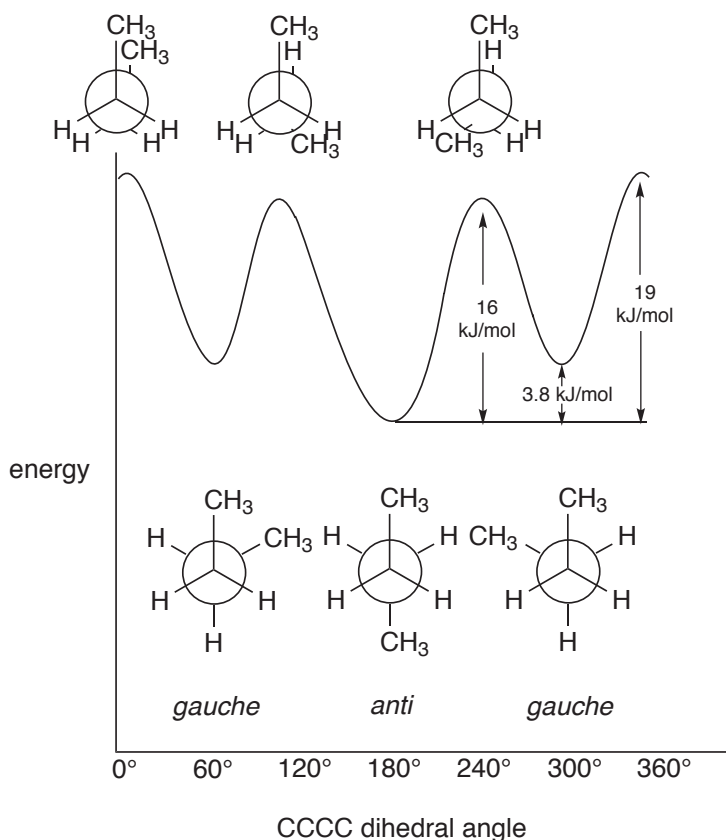
Every chemist has encountered a plot depicting the change in energy of ethane as a function of the angle of torsion (dihedral angle) around the carbon-carbon bond.






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 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.



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 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***).

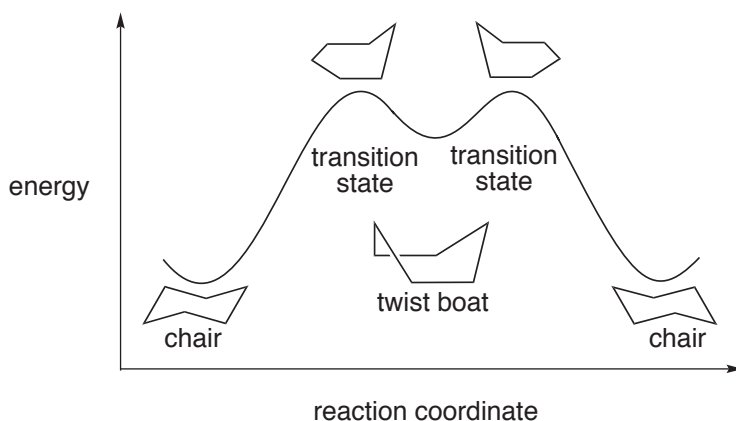
Open *n-butane rotation* in the *topics* directory. The image which appears is one frame of a sequence depicting rotation about the central carbon-carbon bond in *n*-butane. Click on the  and  keys at the bottom left of the screen to look at other frames. Verify that the staggered structures correspond 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 both of the above examples may clearly be identified as a torsion involving one particular 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

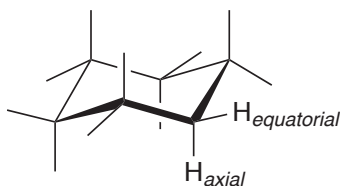
It will usually not be possible to identify a simple geometrical coordinate to designate a chemical transformation. A good example of this is provided by the potential energy surface for ring inversion in cyclohexane.



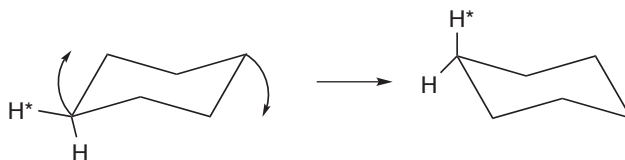
In this case, the geometrical coordinate connecting stable forms is not specified in detail (as it was in the previous two examples), but is referred to simply as the **reaction coordinate**. The two energy maxima have been designated as **transition states** to indicate that their structures may not 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 sufficiently large (around 23 kJ/mol) that 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.

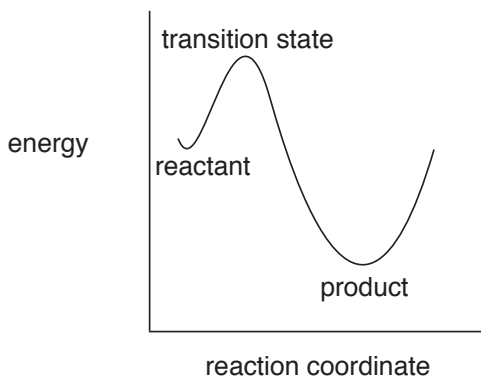


Correspondingly, the interconversion of the twist boat intermediate into the product can be viewed as rotation about the opposite ring bond. Overall, two independent bond rotations, pausing at the high-energy (but stable) twist-boat intermediate effect conversion of one chair structure into another equivalent 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.

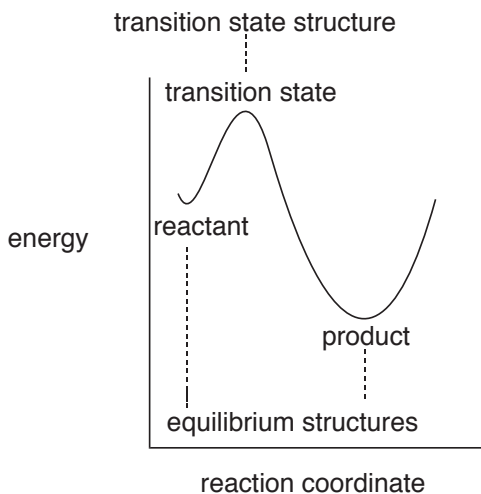
Ethane, *n*-butane and cyclohexane are all examples of the types of motions which molecules may undergo. Their potential energy

surfaces are special cases of a general type of plot in which the variation in energy is given as a function of reaction coordinate.

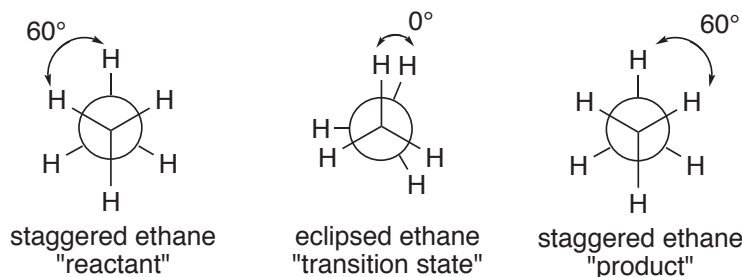


Diagrams like this provide essential connections between important chemical observables - structure, stability, reactivity and selectivity - and energy.

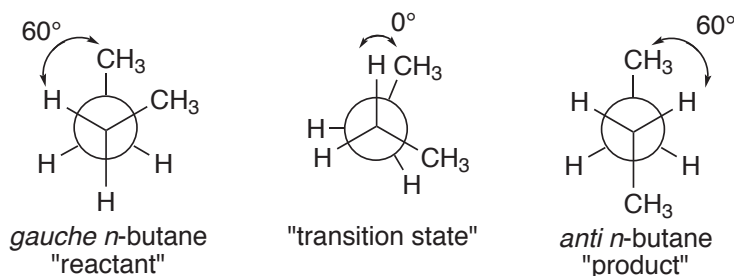
The positions of the energy minima along the reaction coordinate give the equilibrium structures of the reactant and product. Similarly, the position of the energy maximum gives the structure of the transition state. Both energy minima (which correspond to stable molecules) and the energy maximum (which 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  $60^\circ$  and eclipsed ethane is a molecule for which this angle is  $0^\circ$ .



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.



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. In practice, such a surface cannot actually be visualized for a system with more than one or at most two degrees of freedom. A better (more general) indicator is the set of frequencies associated with the vibrational motions around the structure, the same quantities measured by infrared spectroscopy. Structures for which all frequencies are real numbers correspond to stable molecules (energy minima), while structures which have one (and only one) vibrational frequency which is an imaginary number may be transition states. The coordinate (vibrational motion) associated with this imaginary frequency is the reaction coordinate. Further discussion is provided in the topic *Calculating Infrared Spectra*.

## THEORETICAL MODELS

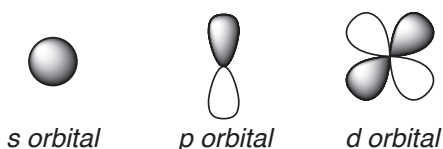
A variety of different procedures based on quantum mechanics (so-called quantum chemical models) have been developed to calculate molecular structure and properties as well as infrared, NMR and UV/visible spectra. All follow from a deceptively simple looking *Schrödinger equation* first written down in 1927.

$$\hat{H}\Psi = \epsilon\Psi$$

$\hat{H}$  (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.  $\Psi$  (the *wave function*) is a function of the Cartesian coordinates, and  $\epsilon$  (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. While the wave function itself has no physical meaning, the square of the wave function times a small volume element gives the probability of finding an electron inside the volume. This is what is measured in an X-ray diffraction experiment.

The Schrödinger equation can be 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. These correspond to the ground and excited states of the hydrogen atom.



Although the Schrödinger equation may easily be written down for many-electron atoms as well as for molecules, it cannot be solved, even for something as simple as the helium atom with only two-electrons. Approximations must be made.

## Hartree-Fock Molecular Orbital Models

*Hartree-Fock molecular orbital models* or more simply *molecular orbital models* 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 that “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  $\text{H}_2^+$  molecule, but cannot be solved for molecules with more than one-electron.
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**tomical **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.

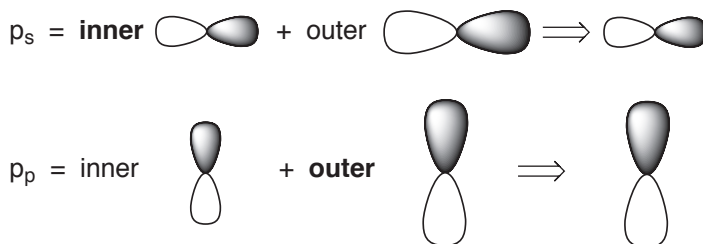
Taken together, these three approximations lead to a set of equations known as the **Roothan-Hall equations**, these can actually be solved for molecules with molecular weights in the range of 300-500 amu. In practice, the resulting models increase in computational cost as the cube of the size (number of basis functions). Note, however, that calculated atomic and molecular properties will not exactly match the corresponding experimental quantities.

## Basis Sets

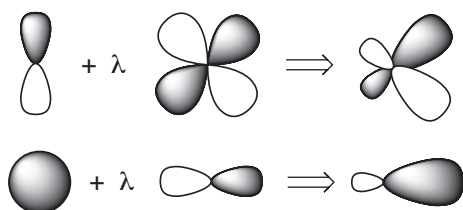
All practical molecular orbital calculations utilize Gaussian functions. These are polynomials in the Cartesian coordinates times an exponential in the **square** of the distance from the origin, and are closely related to the exact solutions of the hydrogen atom (exponential functions in the distance). Functions 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<sub>x</sub>, 2p<sub>y</sub> and 2p<sub>z</sub>) 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  $\sigma$  bond while the loose p orbital can be emphasized to construct a  $\pi$  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.



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.

In practice, Hartree-Fock models increase in computational cost as the cube of the total number of basis functions, and can easily be applied to molecules incorporating up to 100 heavy (non- hydrogen) atoms.

## Beyond Hartree-Fock Models

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 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.

There are two conceptually different approaches for calculating the correlation energy, and numerous specific models arising from each of these approaches. One approach is **Møller-Plesset theory**. This assumes that the Hartree-Fock energy  $E_0$  and wave function  $\Psi_0$  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,  $V$ .  $\lambda$  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:

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \lambda^3 E^{(3)} + \dots$$

Substituting this expansion into the Schrödinger equation and collecting terms in powers of  $\lambda$  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) limited configuration interaction and the MP2 models increase in computational cost as the fifth power of the total number of basis functions. In practice, this limits them to molecules

incorporating 20-25 heavy (non-hydrogen) atoms at most.

A completely different approach is based on the fact that the hydrogen atom is not the only problem for which the Schrödinger equation can be solved exactly (albeit numerically). Another is the so-called free electron gas (a box of non-interacting electrons), solution of which leads to a functional form for a term that accounts for electron correlation. This term, that depends on the electron density as well as the gradient of the density, is then added to the Hartree-Fock Hamiltonian.

Such a procedure is referred to as a ***density functional model***. Actual implementation requires introduction of parameters and because of this, many different ***functionals*** (combinations of functional form and parameters) have been proposed, and each can be used in conjunction with a variety of different basis set.

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). They are 50% more costly than Hartree-Fock models but can easily be applied to molecules incorporating up to 100 heavy (non-hydrogen) atoms.

### **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 in order to significantly reduce cost while still retaining the underlying quantum mechanical formalism. ***Semi-empirical molecular orbital models*** follow in a straightforward way from Hartree-Fock models:

1. Eliminate overlap between functions on different atoms (the ***NDDO approximation***). This 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. Inner-shell (core) functions are not included explicitly, and because of this, the cost of doing a calculation involving a second-row element,

e.g., silicon, is no more than that incurred for the corresponding first-row element, e.g., 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 solving the 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 large molecules, in particular, proteins and equilibrium conformations of highly-flexible molecules.

## 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 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 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. The nature and size of the system needs to be taken into account, as do the available computational resources and the experience (and patience) of the practitioner. 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 higher and higher. There is much to be done before fully reliable models will be routinely applicable to all chemical systems of interest.

The table below provides an overview of the performance of the MMFF molecular mechanics model and five of the most popular quantum chemical models selected from those available in *Spartan*, for the calculation of equilibrium and transition-state geometries and thermochemistry. The performance of an additional procedure for the calculation of thermochemistry will be discussed in the topic *Accurate Heats of Formation*. The ability of these models to account for conformational energy differences and for infrared and NMR spectra will be dealt with in later topics (*Dealing with Conformationally Flexible Molecules* and *Calculating Infrared Spectra* and *Calculating NMR Spectra*).

Three different grades have been assigned: G is good, C is good with cautious application and P is poor. NA signifies not applicable.

task	molecular mechanics	PM3 empirical	semi-Hartree-Fock 3-21G	EDF2/6-31G*	MP2/6-31G*
geometry (organic)	C	G	G	G	G
transition-state geometry	NA	C	G	G	G
thermochemistry	NA	P	C	G	G
computation time	low	-----	-----	-----	-----
		---->	high		

### In Terms of Task

- i) All models provide a good account of equilibrium geometries for organic and main-group inorganic molecules and, where they are applicable, of transition-state geometries. (Transition-state geometries cannot be judged by comparison with experimental data but only with the results of very good quantum chemical calculations.) Molecular mechanics and semi-empirical models only rarely yield very poor geometries. HF/3-21G, HF/6-31G\* and MP2/6-31G\* models do not provide a reliable account of the geometries of compounds incorporating transition metals, but the EDF2/6-31G\* model (and density functional models in general) provides a good account. The PM3 semi-empirical model generally provides a reasonable account of the geometries of transition metal organometallic compounds.
- ii) The HF/6-31G\* model generally provides an acceptable account of the energetics of reactions which do not involve bond making or breaking. The HF/3-21G model is not as satisfactory. Neither Hartree-Fock model provides an acceptable account of the energetics of reactions which involve bond breaking. Both EDF2/6-31G\* and MP2/6-31G\* models generally provide good

accounts of the energetics of all types of chemical reactions.

Neither Hartree-Fock model provides an acceptable account of absolute activation energies, but they generally provide an excellent description of relative activation energies. Both EDF2/6-31G\* and MP2/6-31G\* models properly account for both absolute and relative activation energies.

The PM3 semi-empirical model is unsatisfactory in describing the energetics of all types of processes.

### **In Terms of Model**

The MMFF molecular mechanics model is restricted to the description of equilibrium geometry, where it performs reasonably well.

The PM3 semi-empirical model is appropriate for:

- i) Equilibrium geometry determinations for large molecules.
- ii) Transition-state geometry determinations.
- iii) Equilibrium and transition-state geometry determinations involving transition metals.

The PM3 semi-empirical model is unsuitable for:

- i) Calculation of reaction energies.

The HF/3-21G and HF/6-31G\* models are appropriate for:

- i) Equilibrium and transition-state structure determinations of organic and main-group inorganic molecules (except molecules with transition metals).
- ii) Calculation of reaction energies (except reactions involving bond making or breaking).

The HF/3-21G and HF/6-31G\* models are unsuitable for:

- i) Calculation of reaction energies that involve bond making or breaking and calculation of absolute activation energies.
- ii) Equilibrium and transition-state structure determinations for transition-metal inorganic and organometallic molecules.

The EDF2/6-31G\* model is appropriate for:

- i) Equilibrium and transition-state structure determinations for molecules incorporating transition metals.
- ii) Calculation of all types of reaction energies.

The 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.

The MP2/6-31G\* model is unsuitable for:

- i) Equilibrium and transition-state structure determinations for transition-metal inorganic and organometallic molecules.
- ii) Calculation of reaction and activation energies where transition-metals are involved.

## **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. Proper choice of molecular geometry is therefore quite important in carrying out modeling studies. Experimental geometries, where available, would certainly be suitable. While experimental equilibrium geometries are available for many stable molecules\*, the problem is that many, many more have not been determined. Also, 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

---

\* The vast majority of experimental structures derive from X-ray crystallography on crystalline solids and may be different from gas-phase geometries due to requirements of crystal packing.

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.

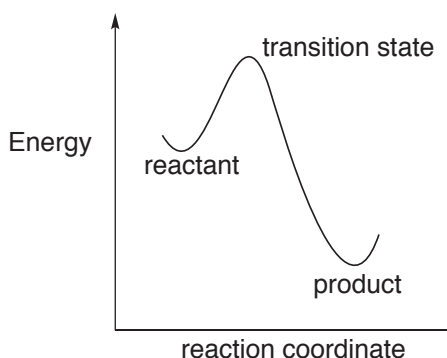
Geometry optimization does not guarantee that the final structure has a lower energy than any other structure of the same molecular formula. All that it guarantees is a **local minimum**, that is, a geometry with a lower energy than that of any similar geometry. This may not be the lowest energy geometry possible for the molecule. Finding the absolute or **global minimum** requires repeated optimization starting with different initial geometries. 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 which is not a local minimum. For example, optimization of ammonia constrained to a planar trigonal geometry ( $D_{3h}$  symmetry) will result in a structure which 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; see preceding discussion), 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*, which 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 more flat 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 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.

The same iterative procedure previously described for optimization of equilibrium geometry applies as well to transition states. However, the number of steps required for satisfactory completion is likely to be much 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:

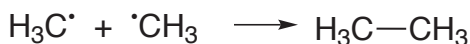
- i) Verify that the Hessian yields one and only one imaginary frequency. This requires that a normal mode analysis be carried out on the proposed transition-state geometry. The imaginary frequency will typically be in the range of  $400\text{--}2000\text{ cm}^{-1}$ , quite similar to real vibrational 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 ( $<100\text{ cm}^{-1}$ ) values. While these can usually be ignored, make certain to verify what motions these small imaginary frequencies actually correspond to (see discussion following) before doing so. 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.

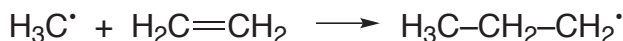
- ii) Verify that the normal coordinate corresponding to the imaginary frequency smoothly connects reactants and products. A simple way to do this is to animate the normal coordinate corresponding to the imaginary frequency, that is, to walk along this coordinate without any additional optimization. This does not 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!

## Reactions Without Transition States

It may come as a surprise that 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 combine without activation, for example, two methyl radicals to form ethane.



Radicals will often add to paired-electron species with no (or very small) activation, for example, methyl radical and ethylene forming 1-propyl radical.



*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  $\text{OH}^-$  on  $\text{CH}_3\text{Cl}$  to give  $\text{CH}_3\text{OH}$  and  $\text{Cl}^-$ .

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! Unfortunately it is very difficult to tell which is correct.

## Calculations Using Approximate Geometries

Given that semi-empirical models and even molecular mechanics models generally provide geometries 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. In fact, the answer depends on what property is being calculated and the level of confidence required. Experience suggests that, except for molecules where parameterization has not been provided, use of either semi-empirical or molecular mechanics geometries has very little effect on relative energetics. Errors in reaction energies of 4-10 kJ/mol, which may arise from the use of approximate geometries, must be balanced against the large savings in computer time. Other properties such as dipole moments, may show greater sensitivity to choice of structure, and use of appropriate geometries may lead to unacceptable errors.

Semi-empirical techniques are not as successful in reproducing the results of Hartree-Fock models for transition-state geometries as they are for equilibrium geometries. However, the energy surface is typically very flat in the vicinity of the transition state, and any errors (in energy) incurred because of the use of approximate transition-state geometries may be manageable. (The magnitude of errors for other quantities are not predictable.) Molecular mechanics techniques are not applicable to the description of transition states.

“Exact” geometries must be used for frequency (infrared 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. Frequencies evaluated at non-equilibrium (or non-transition-state) geometries are meaningless.

## TOTAL ENERGIES AND THERMODYNAMIC AND KINETIC DATA

In addition to molecular geometry, the energy is one of the most important quantities 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,  $\Delta H_f$ . This is the heat of a hypothetical chemical reaction that creates a molecule from the well defined (but arbitrary) standard states of each of its constituent elements. For example,  $\Delta H_f$  for ethylene is the energy required to create the molecule from graphite and  $H_2$ , the standard states of carbon and hydrogen, respectively. 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, in comparison with the energies of chemical bonds, are very large.

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 + ... → product 1 + product 2 + ...):

$$\Delta E(\text{reaction}) = E_{\text{product 1}} + E_{\text{product 2}} + \dots - E_{\text{reactant 1}} - E_{\text{reactant 2}} - \dots$$

Total energies will be used in the discussion that follows. A negative  $\Delta E$  indicates an *exothermic* (thermodynamically favorable) reaction, while a positive  $\Delta E$  an *endothermic* (thermodynamically unfavorable) reaction.

A special case involves differences in isomer stability. This is the energy of a chemical reaction involving only two molecules, in which the reactant and product are isomers:

$$\Delta E(\text{isomer}) = E_{\text{isomer 2}} - E_{\text{isomer 1}}$$

A negative  $\Delta E$  means that isomer 2 is more stable than isomer 1. Total energies may also be used to calculate activation energies,  $\Delta E^\ddagger$ :

$$\Delta E^\ddagger = E_{\text{transition state}} - E_{\text{reactant 1}} - E_{\text{reactant 2}} - \dots$$

Here,  $E_{\text{transition state}}$  is the total energy of the transition state. Activation energies will always be positive numbers\*\*, meaning that the transition state is less stable than reactants.

Reaction and activation energies are sufficient to know whether a reaction is *exothermic* or *endothermic* or whether it proceeds with small or large activation barrier. There are, however, other situations where energies need to be replaced by Gibbs energies in order to take proper account of the role of entropy. For example, a proper account of the equilibrium concentrations of reactants and products requires calculation of the equilibrium constant,  $K_{\text{eq}}$ , which according to

---

\* The exact energy of hydrogen atom is -0.5 atomic units.

\*\* As commented in a previous topic (*Finding and Verifying Equilibrium Transition-State Geometries*) some reactions proceed with zero activation energy, meaning that there is no transition state.

the Boltzmann equation, is related to the Gibbs energy of reaction,  $\Delta G_{\text{rxn}}$ :

$$K_{\text{eq}} = \exp(-\Delta G_{\text{rxn}}/RT)$$

Here R is the gas constant and T is the temperature (in K). At room temperature (298K) and for  $\Delta G_{\text{rxn}}$  in au, this is given by:

$$K_{\text{eq}} = \exp(-1060 \Delta G_{\text{rxn}})$$

$\Delta G_{\text{rxn}}$  has two components, the enthalpy of reaction,  $\Delta H_{\text{rxn}}$ , and the entropy of reaction,  $\Delta S_{\text{rxn}}$ . These are defined as follows:

$$\Delta G_{\text{rxn}} = \Delta H_{\text{rxn}} - T\Delta S_{\text{rxn}}$$

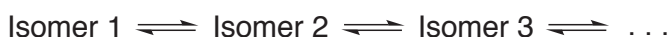
$$\Delta H_{\text{rxn}} \approx \Delta E_{\text{rxn}} = E_{\text{product 1}} + E_{\text{product 2}} + \dots - E_{\text{reactant 1}} - E_{\text{reactant 2}} - \dots$$

$$\Delta S_{\text{rxn}} = S_{\text{product 1}} + S_{\text{product 2}} + \dots - S_{\text{reactant 1}} - S_{\text{reactant 2}} - \dots$$

Although  $\Delta G_{\text{rxn}}$  depends on both enthalpy and entropy, there are many reactions for which the entropy contribution is small, and can be neglected. Further assuming that  $\Delta H_{\text{rxn}} \approx \Delta E_{\text{rxn}}$ , equilibrium constants can then be estimated according to the Boltzmann equation:

$$K_{\text{eq}} \approx \exp(-\Delta E_{\text{rxn}}/RT) \approx \exp(-1060 \Delta E_{\text{rxn}})$$

This Boltzmann equation may also be used to establish the equilibrium composition of a mixture of isomers:



$$\% \text{ Isomer } i = \frac{100 \exp(-1060 E_{\text{Isomer } i})}{\sum_k \exp(-1060 E_{\text{Isomer } k})}$$

Isomer energies,  $E_{\text{isomer}}$ , are given in atomic units relative to the energy of the lowest-energy isomer. An important special case involves an equilibrium between two isomers:



$$\frac{[\text{Isomer 1}]}{[\text{Isomer 2}]} = \exp[-1060 (E_{\text{isomer1}} - E_{\text{isomer2}})]$$

Reaction rate constants,  $k_{\text{rxn}}$ , 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:

$$k_{\text{rxn}} \approx (k_B T/h) [\exp(-\Delta E^\ddagger/RT)]$$

Here  $k_B$  and  $h$  are the Boltzmann and Planck constants, respectively. At room temperature and for  $\Delta E^\ddagger$  in au,  $k_{\text{rxn}}$  is given by:

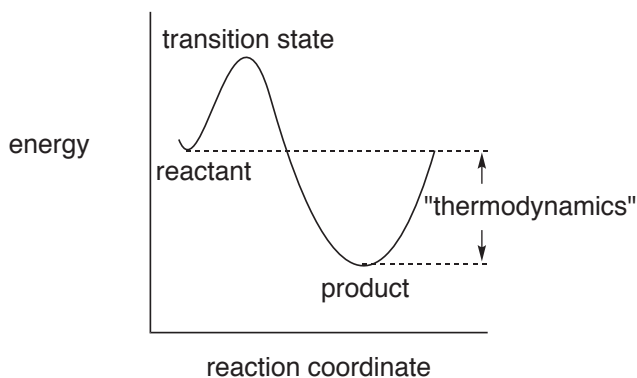
$$k_{\text{rxn}} = 6.2 \times 10^{12} \exp(-1060 \Delta E^\ddagger)$$

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,  $\text{rate} = -k_{\text{rxn}}[\text{reactant}]$ ,  $t_{1/2}$  is given by:

$$t_{1/2} = \ln 2/k_{\text{rxn}} = 0.69/k_{\text{rxn}}$$

It is useful to associate reaction energies and reaction rates with potential energy diagrams (see the topic ***Potential Energy Surfaces***). The connections are actually quite simple.

The thermodynamics of reaction is given by the relative energies of the reactant and product on the potential surface.

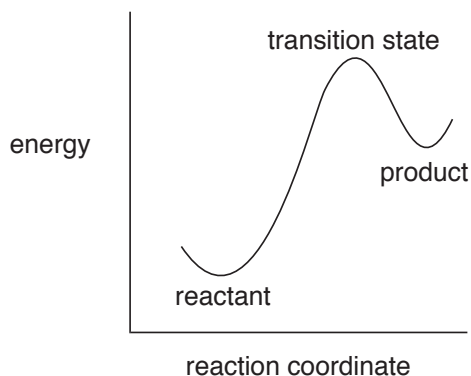


In the case of bond rotation in ethane (see the topic ***Potential Energy Surfaces***), the reactant and product are the same and the reaction is said to be ***thermoneutral***. This is also the case for the overall ring-inversion motion in cyclohexane (see the topic ***Potential Energy Surfaces***).

The most common case, as depicted in the above diagram, is where

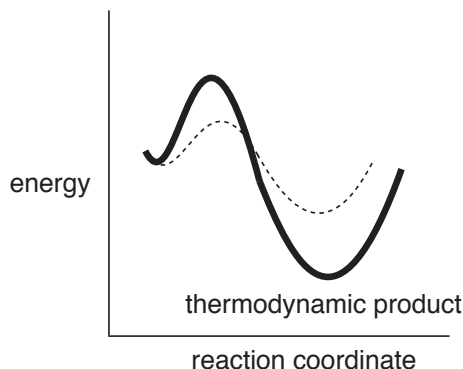
the energy of the products is lower than that of the reactants. This kind of reaction is said to be *exothermic*, and the difference in stabilities of reactant and product is simply the difference in their energies. For example, the “reaction” of *gauche n*-butane to *anti n*-butane is *exothermic* (see the topic **Potential Energy Surfaces**), 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.



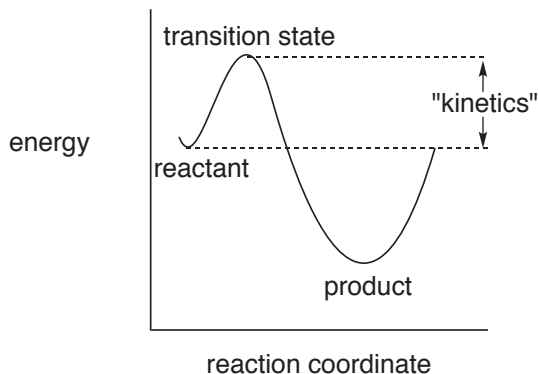
In this case, there would eventually be more reactant than product.

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.



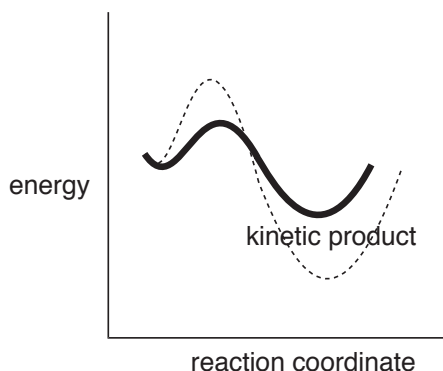
In this case, the product is referred to as the **thermodynamic product** and the reaction is said to be **thermodynamically controlled**.

The energy of the transition state above the reactants (the activation energy) provides the connection with reaction rate (kinetics).



While absolute reaction rates also depends on the concentrations of the reactants and on such factors as the likelihood that encounters between molecules will actually lead to reaction (the so-called pre-exponential or A factor), generally speaking, the lower the activation energy, the faster the reaction.

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 difference in reactant and product energies.

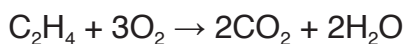
## CALCULATING ACCURATE HEATS OF FORMATION

One of the most fundamental properties of a molecule is its internal energy. This is most commonly reported as a heat of formation, 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.



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.

Heats of formation are most commonly obtained from heats of combustion. For example, the heat of formation of ethylene would likely be established from its reaction with oxygen to produce carbon dioxide and water.



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 does not correspond to the reported structure. Other more common sources of error include incomplete combustion and poorly characterized combustion products. In this regard, 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. The reason for this is obvious. While the combustion experiment is straightforward and does not require particularly expensive instrumentation, accurate measurements may require (and

destroy) significant quantities of compound. Very few chemists are willing to part with hundreds of mg of a compound that they have just spent weeks or months preparing.

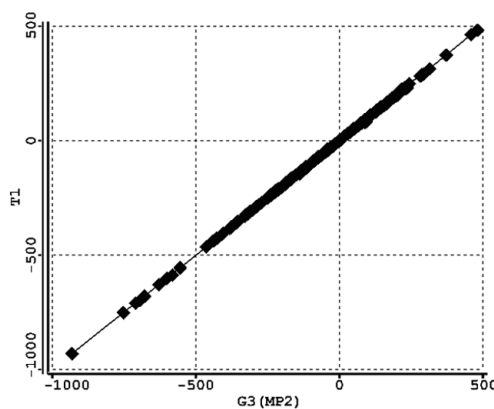
Because of the lack of experimental data and serious concerns over the accuracy of part of the data that do exist, considerable attention has been directed at the use of quantum chemical calculations to obtain heats of formation. However, this has not proven to be simple for moderate-size organic molecules (molecular weight <400-500 amu) within experimental accuracy (<4-8 kJ/mol). Among the most successful procedures in terms of overall accuracy are the GX recipes. We refer to them as “recipes” as they combine the results of several different quantum chemical models in order to get reliable estimates of what enhancements in treatment of electron correlation (beyond the MP2 model) and increases in the size of the basis set are likely to do. The simplest of the GX recipes is termed G3(MP2). It 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 seventh power of size and is applicable only to molecules with molecular weights less than 150 amu.

### **The T1 Thermochemical Recipe**

The goal behind the T1 recipe was to maintain the accuracy of the G3(MP2) recipe 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 dual basis set RI-MP2 techniques. 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.

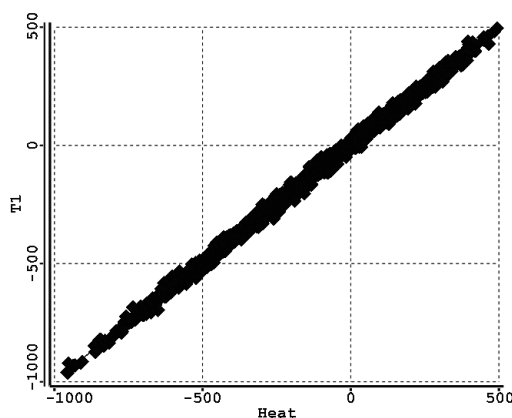
The T1 recipe, unlike G3(MP2), involves several parameters, specifically atom counts, Mulliken bond orders and HF/6-31G\*

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. A plot is shown below.



### Performance of T1 for Heat of Formation Calculations

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.



### Treatment of Conformationally-Flexible Molecules

As a consequence of their high computation cost, the GX recipes are in practice limited to rigid molecules or to molecules with

only one or two degrees of conformational freedom. This means that it is necessary to guess the conformation of most molecules. This in turn may lead to heats of formation that are too large by 10 kJ/mol or more. On the other hand, T1 can routinely be applied to molecules with several degrees of conformational freedom. This makes it possible to establish energy differences among conformers (Boltzmann weights). A more modest goal is to identify the lowest-energy conformer, and to assume that the heat of formation of the molecule is the heat of formation of this conformer. In practice, the lowest-energy conformer is identified by examining all conformers for molecules with fewer than 100 conformers and by examining a random sample of 100 conformers for molecules with more than 100 conformers.

### Database of T1 Heats of Formation

Heats of formation from the T1 recipe are included as a property in the Spartan Spectra and Properties Database (SSPD), a 6,000 molecule subset of which is included with *Spartan*. The full database presently comprises more than a quarter of a million entries and is included as part of the *Spartan Parallel Suite* and also may be separately licensed.

### 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 already 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 =  $\pm \sim 60^\circ$ ). 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.

What is required to calculate the properties of a flexible molecule? At the very least, the lowest-energy conformer needs to be identified and used for calculation of the property of interest. How well do calculations identify this conformer? It turns out the answer is not obvious. While much is known about the preferred conformation of molecules in the solid (crystalline) state, most of it from X-ray diffraction, there is actually 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.

*Spartan* allows calculation of the lowest-energy conformer using the MMFF molecular mechanics model. Of course, any of the quantum chemical models available in the program may be used to calculate energy differences between conformers that are explicitly specified.

### **Systematic vs. Monte-Carlo Conformational Searching**

There are two fundamentally different techniques available for searching conformation space. The obvious thing to do is to “look everywhere”, that is, walk through all possible single-bond rotations and ring twists. A *systematic search* guarantees that the lowest-energy conformer (the so-called *global minimum*) will be found,

and provides a proper basis for calculation of the correct Boltzmann distribution. To see why systematic searching rapidly becomes impractical, we need to know how many different conformers may actually exist for a flexible molecule. While a precise number may be hard to come by, a few rules allow a rough estimate to be made. The best-known rule is that a chain of four or more ( $N$ )  $sp^3$  centers will lead to  $3^{N-3}$  conformers. Thus  $n$ -butane ( $N=4$ ) will have three conformers,  $n$ -pentane will have nine conformers,  $n$ -hexane will have 27 conformers and so forth. Not all the conformers will be distinct, for example, only two of the three conformers of  $n$ -butane are unique and only five of the nine conformers of  $n$ -pentane are unique. Bonds between  $sp^3$  and  $sp^2$  centers will generally lead to similar conformer counts, whereas bonds between two  $sp^2$  centers will generally lead to two (*syn* and *anti*) and not three conformers. The number of conformers for rings is more difficult to estimate. Four and five-member rings can all be thought of in terms of a single conformer and six-member rings in terms of two low-energy conformers (higher-energy twist-boat conformers can usually be ignored), but seven-member and larger rings typically have a larger number of conformers. Molecules with several flexible bonds and one or more six-member or larger rings may easily have several hundred or several thousand distinct conformers. Looking everywhere quickly becomes impractical and an alternative searching technique is needed.

Many detailed alternative procedures (algorithms) have been formulated but all come down to the idea of replacing a systematic search by a so-called stochastic or random search. The procedure used in *Spartan* is referred to as the **Monte-Carlo algorithm**. 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. Without getting into the detailed mathematics, suffice it to say that the criterion used to decide whether to accept or reject a move has been chosen to guarantee that the proper Boltzmann distribution will be approached as the number of random moves is increased. In

practice, a Monte-Carlo search will nearly always find the lowest-energy conformer (or at least something very close to it) in 5-10% of the time that would be required for a systematic search. Thus, Monte-Carlo searching is likely to be an order of magnitude or more faster than systematic searching.

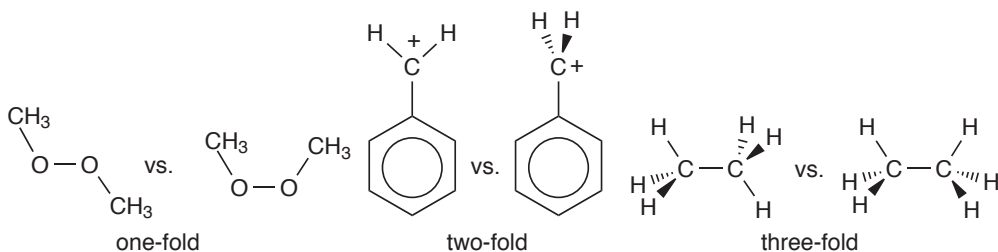
## INTERPRETING CONFORMATIONAL PREFERENCES

Rotation about single bonds is periodic, retracing itself every  $360^\circ$ . Therefore, any function that seeks to describe the energy of internal rotation must also repeat itself every  $360^\circ$ . In fact it is possible to write a general energy function,  $E^{\text{torsion}}$ , as a combination of simpler functions,  $V_n$ , each of which repeats  $n$  times in a  $360^\circ$  interval.

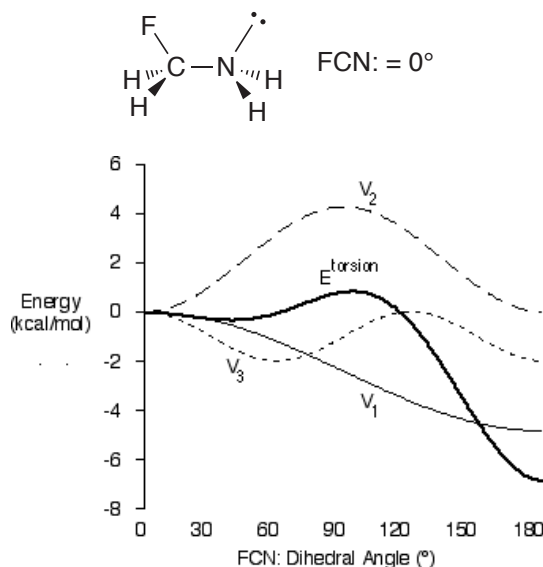
$$E^{\text{torsion}}(\omega) = V_1(\omega) + V_2(\omega) + V_3(\omega)$$

$V_1$ ,  $V_2$ , and  $V_3$  are independent functions of the torsion angle  $\omega$  that repeat every  $360^\circ$ ,  $180^\circ$ , and  $120^\circ$ , respectively. These functions are referred to as one-fold, two-fold, and three-fold potentials.

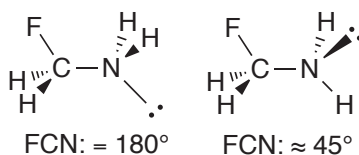
The different  $n$ -fold potentials are useful because each can be associated with a particular chemical phenomenon. For example, a one-fold potential describes the different energies of *anti* and *syn* conformers of dimethylperoxide, while a two-fold potential describes the different energies of planar 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 a symmetric molecule might be described using only one potential or a combination of two potentials, less symmetric molecules require more complex combination of potentials. This is illustrated by fluoromethylamine.

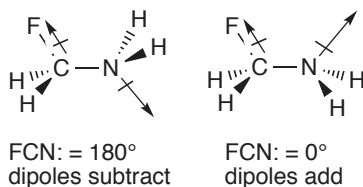


The heavy solid line in the figure describes  $E^{\text{torsion}}$  for rotation about the CN bond, while the light solid line, the dashed line and the dotted line correspond to the one-fold, two-fold and three-fold components, respectively. There are two distinct minima.



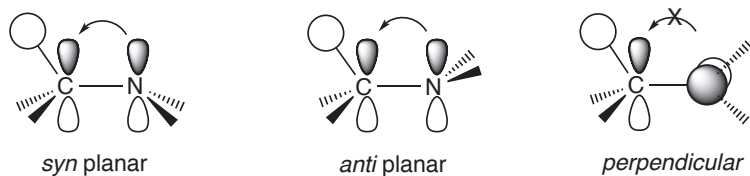
The lower (global minimum) arises when the CF bond and the nitrogen lone pair are *anti*, while the higher and much more shallow minimum is almost, but not precisely, a *gauche* structure (FCN: dihedral angle  $\sim 45^\circ$ ). Also, note that one of the two energy maxima (FCN: dihedral angle  $\sim 115^\circ$ ) does not exactly correspond to an eclipsed structure.

The unusual behavior of  $E^{\text{torsion}}$  becomes clear when it is resolved into its components (in this case, the sum of these components provides a virtually perfect fit of  $E^{\text{torsion}}$ ). 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.



The three-fold term is also easy to explain. It reflects the preference for staggered over eclipsed structures. This term contributes much less to  $E^{\text{torsion}}$  than either of the one-fold or two-fold terms, consistent with the low steric demands of the  $\text{CH}_2\text{F}$  and  $\text{NH}_2$  groups.

What is most interesting, perhaps, and what might not have been anticipated without this type of analysis, is the large contribution made by the two-fold potential. This potential reflects a strong preference for a planar arrangement of  $\text{FCN:}$ , and can be attributed to stabilization resulting from donation of the lone pair orbital on nitrogen into low-energy unfilled molecular orbital associated with the  $\text{CF}$  bond. Such an interaction requires that the molecule adopts either a *syn* planar or *anti* planar conformation and not a perpendicular conformation.



Most of the popular features of  $E^{\text{torsion}}$  can now be attributed to either  $V_1$ ,  $V_2$ .  $V_1$  accounts for the *anti* geometry being the global minimum.  $V_2$ , however, is responsible for the position of the maximum and the shift in the higher-energy minimum to smaller dihedral angles.

It is important to note that the terms that contribute to  $E^{\text{torsion}}$  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  $\text{CF}$  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), the frequency is proportional to the square root of the ratio of the second-derivative term and the reduced mass (the product of the masses of the two atoms divided by their sum).

$$\text{vibrational frequency} \propto \sqrt{[(d^2E(x)/dx^2)/\text{reduced mass}]}$$

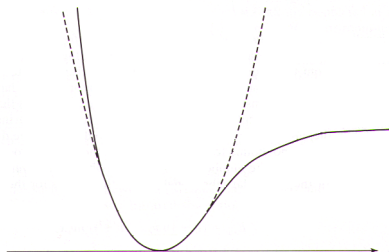
Calculation of the frequency involves expansion of the energy in a Taylor series.

$$E(x) = E(x_0) + (dE(x)/dx) x + (d^2E(x)/dx^2) x^2 + \text{higher-order terms}$$

$E(x_0)$  is a constant and  $dE(x)/dx$  (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. All practical calculations are based on the harmonic model and ignore cubic and higher-order terms (*anharmonic* terms), leaving only the second derivative term (the *force constant*). 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). This suggests that the potential curve will be too steep.



It is possible to extract the harmonic frequency 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 possible only for 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,  $\mathbf{x}$ , replaces a scalar quantity,  $x$ .

$$E(\mathbf{x}) = E(\mathbf{x}_0) + \sum_i (\partial E(\mathbf{x}) / \partial x_i) x_i + \frac{1}{2} \sum_{ij} (\partial^2 E(\mathbf{x}) / \partial x_i \partial x_j) x_i x_j + \text{higher-order terms}$$

For a molecule with  $N$  atoms, the dimension of  $\mathbf{x}$  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  $\zeta$ , such that the matrix of mass-weighted second derivatives is diagonal.  $\delta_{ij}$  is the Kronecker delta function (1 if  $i=j$ ; 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. If there is no change in dipole moment, the infrared intensity is zero.

The two major components of the earth's atmosphere,  $N_2$  and  $O_2$ , 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 second derivative calculation 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.

The EDF2/6-31G\* density functional model was specifically formulated to reproduce measured infrared frequencies. It is simple enough to be 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 perhaps more conspicuous deficiency is 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 are sharp in that they correspond to an isolated molecule at 0K. 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. These differences are more difficult to quantify and will be ignored.

It is straightforward to correct the calculated spectrum to account for the two major 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.

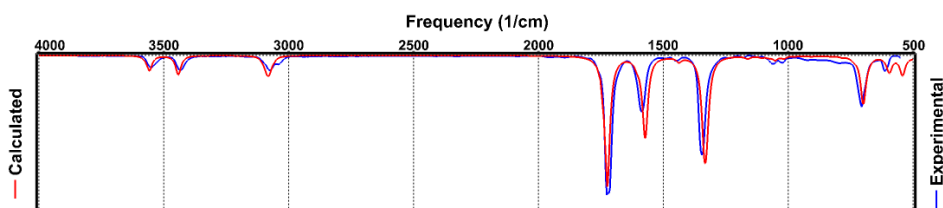
### **Quality of Calculated Infrared Spectra**

An infrared spectrometer records frequencies in the range 500-4500  $\text{cm}^{-1}$ . 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 region beyond  $\sim 2800 \text{ cm}^{-1}$  is dominated by CH stretching vibrations and may be too crowded to be of value.

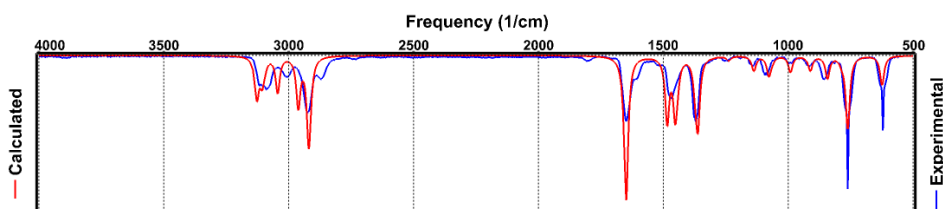
Although the focus here is on the EDF2/6-31G\* model, the comments broadly apply to other density functional models with polarization and larger basis sets. The EDF2/6-31G\* model in particular has also proven to be successful in providing accurate molecular geometries and heats of reaction, and it may fairly be described as a workhorse. That is, it is both reliable and simple enough to be routinely applied to molecules with molecular weights <500 amu. This places routine applications squarely in the range of organic molecules that can conveniently be synthesized.

Calculated infrared spectra 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, (dimethylmethyldiene)cyclopentadiene, 1,2-epoxy-*cis*-4-vinylcyclohexane and camphor are typical of the ~1000 comparisons that have been made.

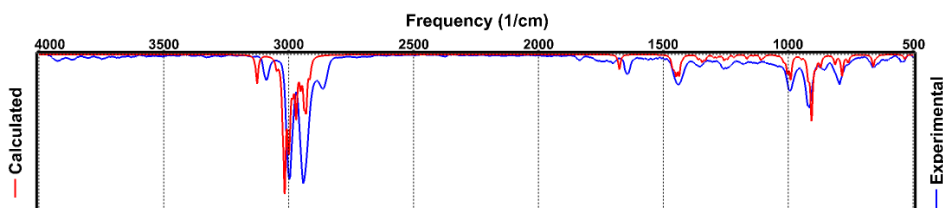
*benzamide*



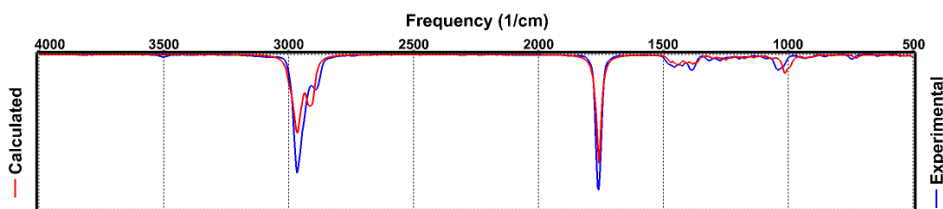
*(dimethylmethyldiene)cyclopentadiene*



*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

The Spartan Spectra and Properties Database (SSPD), a 6000 molecule subset of which is included with *Spartan*, provides includes infrared spectra as well as proton and  $^{13}\text{C}$  NMR spectra. These are based on calculations performed using the EDF2/6-31G\* density functional model starting from the best conformer assigned from the T1 recipe. The full database presently comprises more than a quarter of a million entries and is included as part of the *Spartan'14 Parallel Suite* and also may be separately licensed.

## CALCULATING NMR SPECTRA

### Introduction

There are several reasons why NMR spectroscopy, in particular proton and  $^{13}\text{C}$  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 non destructive. In the case of  $^{13}\text{C}$  NMR (the topic discussed here), the resulting (proton decoupled) spectrum is simple, comprising a single line (resonance) for each and every unique carbon. Despite its simplicity (or perhaps because of it), associating an  $^{13}\text{C}$  spectra to a particular molecular structure can be problematic and prone to error, in particular, where alternative structures might be very similar. Three-bond HH and CH coupling constants that depend on 3D geometry, as well as a variety of so-called 2D spectra that combine chemical shifts and coupling constants are more and more commonly employed to assist in assignment.

A routine and reliable method to predict  $^{13}\text{C}$  chemical shifts as a function of three-dimensional structure would clearly be of value in helping to assign experimental NMR spectra or, at the very least,

confirming or refuting a proposed assignment. One might argue that such a method already exists in the form of 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, new compounds will never give exact matches, simply because the spectrum is not in the database. Closely related are empirical relationships obtained from fitting previously assigned (and presumed correctly assigned) spectra. While these can achieve some degree of success, the fact that NMR chemical shifts (in addition to three-bond coupling constants) are sensitive to conformation means that molecules that have very similar 2D structures may give rise to entirely different spectra.

An 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, and both Hartree-Fock and density functional models have been shown to provide reasonable accounts of chemical shifts for simple molecules. However, calculations have only rarely been used to actually assist in the interpretation of spectra, and very few chemists are aware that accurate quantum chemical calculations are now possible (and practical) for real molecules. Those who are aware, are confronted and all too often stymied with what must seem to be an endless list of 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

Some but not all nuclei have a non-zero magnetic moment. Where they do, application of an external magnetic field causes the nuclear spins to align either parallel or antiparallel to the field. The difference in energy ( $\Delta E$ ) between nuclear spin states is given by.

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

$\gamma$  is the gyromagnetic ratio, a constant that depends on the nucleus,  $\hbar$  is Planck's constant/ $2\pi$  and  $B_0$  is the strength of the magnetic field **at the nucleus**. What makes nuclear magnetic resonance (NMR) spectroscopy interesting to chemists is that the magnetic field felt at the nucleus is slightly 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**.

As commented above, not all nuclei have a magnetic moment and can give rise to an NMR signal. In particular, the dominant isotope of carbon ( $^{12}\text{C}$ ) does not have a magnetic moment and is transparent to an NMR spectrometer. Fortunately  $^{13}\text{C}$  does have a magnetic moment, although it makes up only about 1% of the total. The low natural abundance of  $^{13}\text{C}$  certainly slowed the application of carbon NMR to organic chemistry by several decades (waiting for magnet and spectrometer technology to catch up). However, the low abundance of  $^{13}\text{C}$  is a blessing in disguise in that a carbon NMR spectrum is much simpler (and easier to interpret) than a proton NMR spectrum. 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  $^{13}\text{C}$  nuclei will be adjacent all but eliminates carbon-carbon coupling. Proton- $^{13}\text{C}$  coupling can occur but can be (and nearly almost always is) removed. The result is that the  $^{13}\text{C}$  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 limits, and a mere 10-20% increase in field strength (translating to

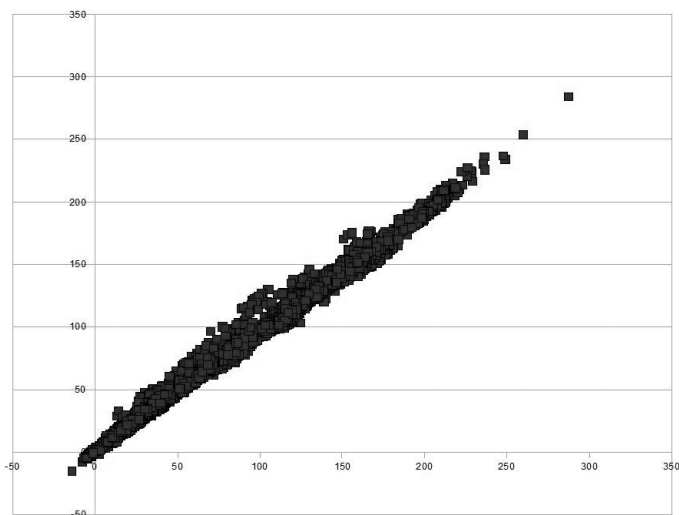
an equivalent increase in resolution) can mean more than doubling the cost of the spectrometer. Without a major breakthrough in magnet technology, the prognosis for greatly improved resolution over what is now possible (practical) is bleak. On the other hand, computer performance continues to double every few years (anticipated by Moore's law), and NMR spectra calculations on larger and ever more complex molecules continue to become more routine.

### **The EDF2/6-31G\* Density Functional Model for $^{13}\text{C}$ Chemical Shift Calculations**

*Spartan* can calculate NMR chemical shifts using two density functional models, specifically the B3LYP/6-31G\* and the EDF2/6-31G\* models. The performance of the EDF2/6-31G\* model for  $^{13}\text{C}$  shifts is documented below. This is based on a collection of ~3000 small organic molecules containing ~15,000 unique carbons for which experimental data are available. The majority of these are rigid molecules or molecules which exhibit a strong preference for a single conformer. The shift calculations are based on equilibrium geometries from the EDF2/6-31G\* model starting from the best conformation established from the T1 thermochemical recipe.

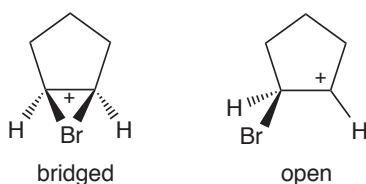
Use of the T1 recipe to establish best conformer is limited to closed-shell, uncharged molecules containing the elements H, C, N, O, F, Si, P, S and Br only. Molecules incorporating other elements use the best conformer assigned from the MMFF molecular mechanics model.

Uncorrected, the calculations yield a RMS error of 5.5 ppm. A plot of EDF2/6-31G\* vs. experimental  $^{13}\text{C}$  shifts over the range of -50 to 300 ppm reveals a number of significant outliers.



A few simple examples allow us to speculate under what conditions chemical shifts from the EDF2/6-31G\* model are likely to be of value in assigning experimental NMR spectra.

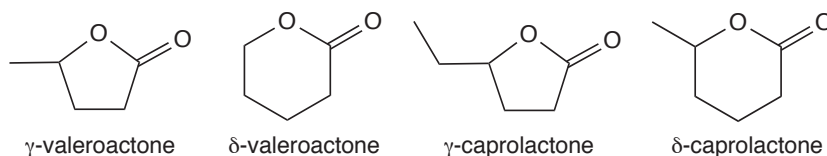
The first example typifies a situation where NMR is used to distinguish between two alternatives that are structurally very different. The observed  $^{13}\text{C}$  spectrum of cyclopentenebromonium ion contains lines at 18.8, 31.8 and 114.6 ppm. These could either correspond to a structure with bromine bonded to both  $\text{sp}^2$  carbons (bridged) or from an equilibrium between a pair of equivalent structures with bromine attached only to one carbon (open).



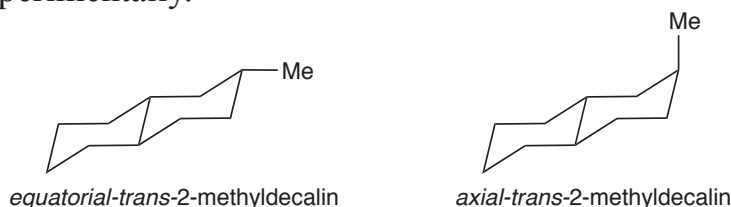
The calculated spectrum for the bridged ion shows resonances at 20, 33 and 117 ppm, while that obtained by averaging the two open ions shows lines at 28, 51 and 193 ppm. The former is a reasonable fit to the experimental spectrum whereas the latter is a poor fit. A cyclic structure for cyclopentenebromonium ion is consistent with crystal structures for related ions.

Actually, the open ion is not an energy minimum according to the EDF2/6-31G\* model, and collapses without barrier to the cyclic ion. The NMR calculations were performed using a geometry for the open ion obtained from the analogous Hartree-Fock model (where it is an energy minimum).

The second example involves the use of  $^{13}\text{C}$  NMR to distinguish structural isomers, specifically to distinguish between  $\gamma$  and  $\delta$  forms of valerolactone and caprolactone. Experimentally, the  $^{13}\text{C}$  shift for the carbonyl carbon in the  $\gamma$  isomer of valerolactone is 5.8 ppm larger than that for analogous carbon in the  $\delta$  isomer. The calculations show the proper ordering and a difference of 7.5 ppm. The same trend is found for the comparison of  $\gamma$  and  $\delta$  caprolactone. Experimentally, the  $^{13}\text{C}$  shift of the carbonyl carbon is 4.6 ppm larger for the compound with the 5-membered ring than it is for that with the 6-membered ring, compared to a calculated difference of 7.4 ppm.

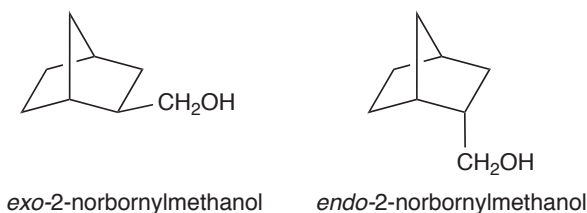


The third example examines an even more subtle situation where the alternatives are stereoisomers and therefore structurally nearly identical. Consider the choice between *equatorial* or *axial* stereochemistry in a rigid ring where equilibration cannot take place, for example, in *trans* decalin. EDF2/6-31G\* calculations show that the  $^{13}\text{C}$  shift for the methyl group in *equatorial trans*-2-methyldecalin is 3.7 ppm smaller than that for methyl in the *axial* compound, in reasonable accord with both the direction and magnitude (4.9 ppm) found experimentally.



The fourth example is similar to the previous example in that it uses NMR to distinguish between stereoisomers, in this case, between the

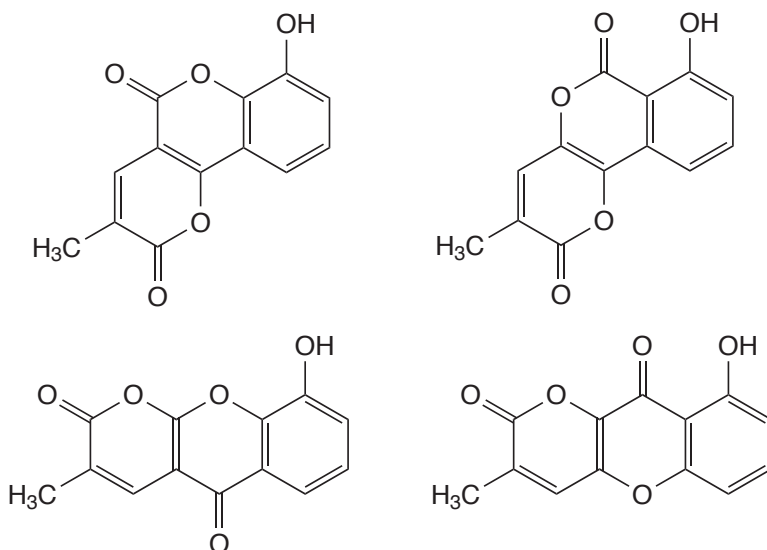
*endo* or *exo* stereoisomers of 2-norbornylmethanol. The calculated shift for C<sub>7</sub> (the methylene bridge) is larger than the experimental value for both stereoisomers, but both the direction (the shift in the *exo* isomer is smaller) and the magnitude of the difference (5.1 ppm vs. 4.6 ppm experimentally) are reproduced.



The calculations are successful for cyclopentenebromonium ion presumably because the chemical shifts for the alternative structures are sufficiently different that even sizable errors in the individual calculated <sup>13</sup>C chemical shifts can be tolerated. Success in properly assigning the structural isomers of valerolactone and caprolactone, the *equatorial* and *axial* stereoisomers of trans-2-methyldecalin and *endo* and *exo* stereoisomers of 2-norbornylmethanol is on the other hand most likely due to the fact that the alternatives are very similar and that the calculations are benefiting from cancellation of errors.

What happens for the intermediate situation where the alternatives may not be different enough to be easily distinguished and may not be similar enough to benefit from error cancellation? Consider, for example, an attempt to use calculations to decide which of the possible C<sub>13</sub>H<sub>8</sub>O<sub>5</sub> products arise from a biosynthesis related to the known pathway for lambertellol.\* This is the kind of difficult case where calculations were they reliable would be of real value. Regrettably, neither the overall spectral pattern nor the detailed values <sup>13</sup>C shifts obtained from the EDF2/6-31G\* model for any of the reasonable candidates closely matches the experimental data.

\* T. Murakami, N. Takada, W.J. Hehre and M. Hashimoto, *Bioorg. Med. Chem. Lett.*, **18**, 4547 (2008).



## Empirically Corrected <sup>13</sup>C Chemical Shifts

Successes and failures typified by the examples described above (drawn from a much more extensive list) have led to the development of a hybrid approach for calculating <sup>13</sup>C shifts. The scheme used in conjunction with the B3LYP/6-31G\* density functional model relies only on the number and kinds of directly-bonded atoms. A total of 29 parameters was involved in a linear regression to an experimental dataset made up of ~800 sp<sup>3</sup> carbons, ~1500 sp<sup>2</sup> and aromatic carbons and ~100 sp carbons. Note that this correction scheme depends only on topography and not on three-dimensional structure. This means that *changes* in chemical shifts between stereoisomers obtained from the B3LYP/6-31G\* model are unaffected.

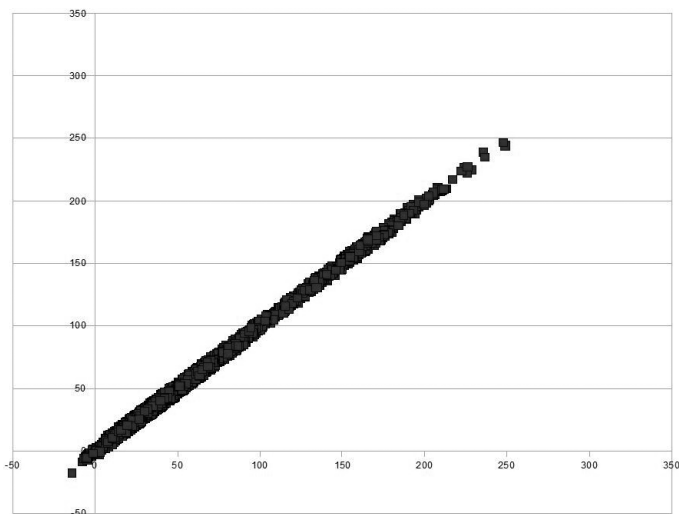
A more recent scheme used in conjunction with the EDF2/6-31G\* model also relies on bond orders, and special cases for several common (sp<sup>2</sup> and sp) centers. This scheme involves 43 independent parameters for the general case and a similar number of parameters for the special cases. The dataset used for the linear regression is much larger, comprising ~8000 sp<sup>3</sup> carbons, ~6200 sp<sup>2</sup> carbons (including aromatic carbons) and ~450 sp carbons.

The general case formula is:

$$^{13}\text{C}_i = ^{13}\text{C}_i \cdot \text{Scale} + \sum_k [\text{X}_k^{(0)} + \text{X}_k^{(1)}(\text{Q}_{i,k} - 1) + \text{X}_k^{(2)}(\text{Q}_{i,k} - 1)^2]$$

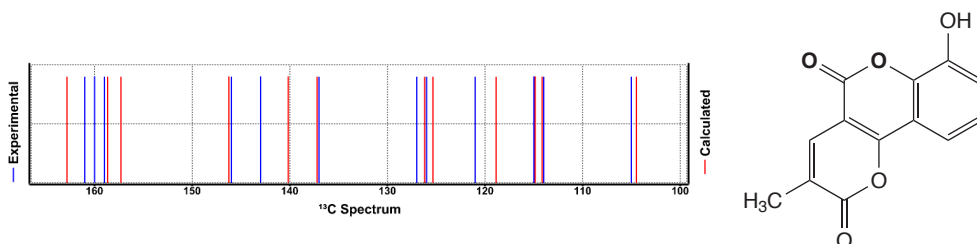
$^{13}\text{C}$  is the corrected chemical shift and  $^{13}\text{C}$  is the uncorrected chemical shift and  $Scale$  is a scaling factor. Summation is carried out over all bonds (1 or 2 for  $\text{sp}$  carbons, 3 for  $\text{sp}^2$  carbons and 4 for  $\text{sp}^3$  carbons).  $X_k^{(0)}$ ,  $X_k^{(1)}$  and  $X_k^{(2)}$  are parameters that depend on the atom bonded to carbon (H, C, N, O, F, Si, P, S, Cl or Br) and  $Q_{i,k}$  are natural bond orders. Different expressions apply to  $\text{sp}$ ,  $\text{sp}^2$  and  $\text{sp}^3$  carbons. The equation for  $\text{sp}$  carbon involves four  $X^{(0)}$  parameters, four  $X^{(1)}$  parameters and one  $X^{(2)}$  parameter (11 parameters in total), that for  $\text{sp}^2$  carbon involves all ten  $X^{(0)}$  parameters, six  $X^{(1)}$  parameters and four  $X^{(2)}$  parameters (21 parameters in total), and that for  $\text{sp}^3$  carbon involves all ten  $X^{(0)}$  parameters but no  $X^{(0)}$  or  $X^{(1)}$  parameters (11 parameters in total). As there are no special cases for  $\text{sp}^3$  carbons, this means that the new correction scheme for  $\text{sp}^3$  carbons, like the previous scheme, depends only on topography and not on three-dimensional structure. **Changes** in chemical shifts between stereoisomers involving  $\text{sp}^3$  centers obtained from the EDF2/6-31G\* model are unaffected.

A plot of  $^{13}\text{C}$  shifts obtained from the general correction formula over the range of -50 to 300 ppm is shown below provides RMS error is 1.8 ppm (vs. 5.5 ppm for the uncorrected shifts) with no significant outliers.



The corrected scheme does a much better job for the type of situation illustrated previously for the  $\text{C}_{13}\text{H}_8\text{O}_5$  isomers. The spectrum shown below provides a good match to the experimental  $^{13}\text{C}$  spectrum,

whereas the calculated spectra of the other three reasonable candidates do not. This is actually the correct structure.



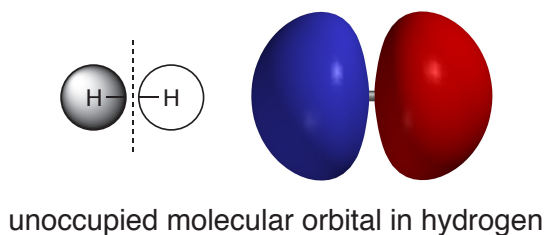
The corrected spectra also properly account for the types of problems that were previously successful (the correction did no harm).

## 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 provide important clues about chemical reactivity, but before we can use this information we first need to understand what molecular orbitals look like. The following figure shows two representations, a hand drawing and a *Spartan*-generated image of an unoccupied molecular orbital of hydrogen molecule,  $H_2$ .



\* 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.

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 *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. Equally important, the image is three-dimensional, and can be manipulated and looked at from a variety of different perspectives. Note that an orbital surface actually 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 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 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.

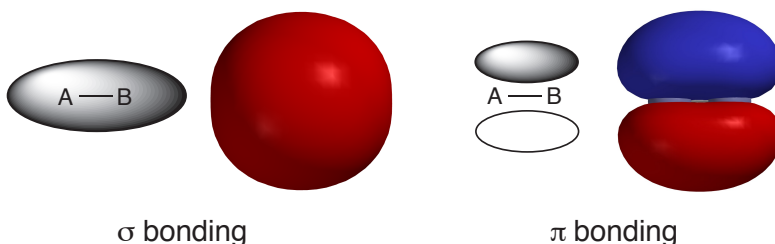
Open *fluoride and chloride* in the *topics* directory. On the top row are the four valence orbitals of fluoride anion and on the bottom row the four valence orbitals of chloride anion. You can select among them by *clicking* (left mouse button) on each in turn. First note that the three  $2p$  orbitals 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 orbital energies are useful tools for describing chemical bonding and reactivity. This section describes a few common orbital shapes and illustrates their use.

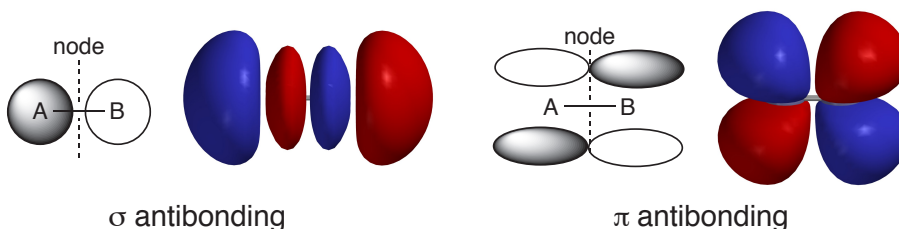
Molecular orbital surfaces can extend over varying numbers of atoms. If the orbital surface (or surfaces) is confined to a single atom or to atoms which are not close together, the orbital is regarded as non-bonding. If the orbital contains a surface that extends continuously

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  $\sigma$  bond while the drawing and surface on the right correspond to a  $\pi$  bond.



Open **nitrogen bonding** in the **topics** directory. The image on the left corresponds to the  $\sigma$  bonding orbital of  $N_2$ , while that on the right corresponds to one of two equivalent  $\pi$  bonding orbitals. Switch to a transparent or mesh model to see the underlying molecular skeleton. Note that the  $\sigma$  orbitals is drawn in a single color (insofar as NN bonding is concerned) while the  $\pi$  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  $\sigma$  and  $\pi$  type arrangements, respectively.



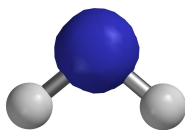
Open ***nitrogen antibonding*** in the ***topics*** directory. The image on the left corresponds to the  $\sigma$  antibonding ( $\sigma^*$ ) orbital of  $\text{N}_2$  while that on the right corresponds to one of the two equivalent  $\pi$  antibonding ( $\pi^*$ ) orbitals. Switch to a mesh or transparent surface to see the underlying molecular skeleton. Note that the  $\sigma^*$  orbital has a single node (change in color from red to blue) in the middle of the NN bond, while the  $\pi^*$  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.

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

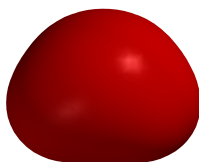
Molecular orbitals in molecules which contain many atoms are typically spread throughout the molecule (they are delocalized). Delocalized orbitals have complicated shapes and contain multiple interactions that may be bonding, non-bonding, antibonding, or any mixture of all three. Nevertheless, these shapes can still be broken down into two-atom interactions and analyzed using the principles outlined earlier. This process is illustrated for a triatomic molecule, singlet methylene,  $\text{CH}_2$ . (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***.



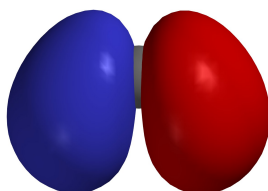
core orbital for methylene

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 ( $\sigma$ ) 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 ( $\sigma$ ) 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

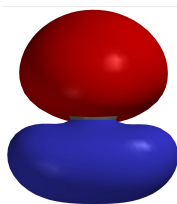
Thus, the two CH bonds in the Lewis structure for singlet methylene are replaced by two bonding molecular orbitals.

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

---

\* While the absolute signs (colors) of a molecular orbital are arbitrary, the relative signs (colors) indicate bonding and antibonding character.

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

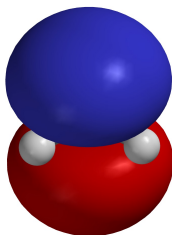
The above analysis shows that while the occupied orbitals of singlet methylene are spread over all three atoms, they are comprehensible. The orbitals divide into two groups, a single low-energy core orbital and three higher-energy valence orbitals. The latter consist of two CH bonding orbitals and a 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.

\* Because Lewis structures describe electron pair bonds and non-bonding electron pairs, they may not be related to unoccupied molecular orbitals.

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.

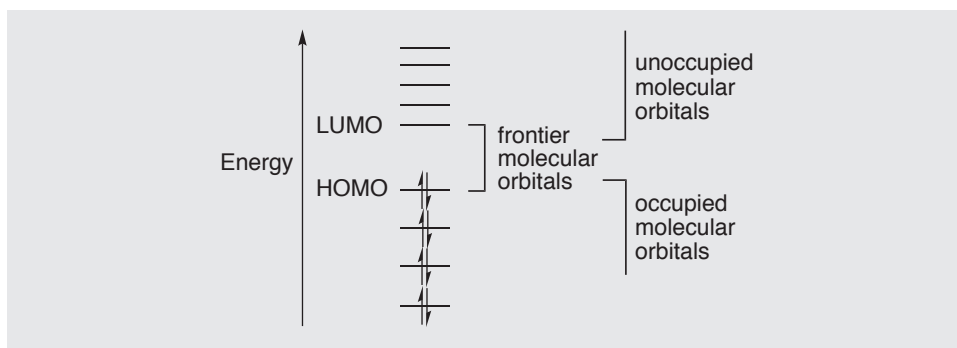
## Frontier Molecular Orbitals and Chemical Reactivity

Chemical reactions typically 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.

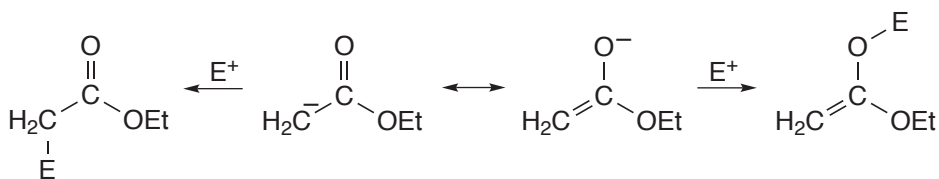
Orbital energy is usually the deciding factor. The highest-energy occupied orbital (the HOMO) is most commonly assumed to be the relevant electron-donor orbital and the lowest-energy unoccupied

orbital (the LUMO) is most commonly assumed to be the relevant electron-acceptor orbital. For example, the HOMO and LUMO of singlet methylene ( $\sigma$  and  $\pi$  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.

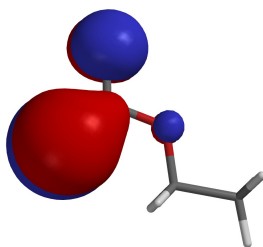


Closely related to chemical reactivity (reaction rate) is chemical selectivity. The relevant question is, where more than one combination of reagents can react, which combination will react first? The answer can often be found by examining the energies of the frontier orbitals. Consider ranking the rates 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.

A related selectivity question arises when a molecule has multiple reactive sites. In this case, the energy of the orbital is useless as a guide to site selectivity, but the shape of the relevant orbital is important. For example, the enolate of ethyl acetate shown below might react with an electrophile ( $E^+$ ) at two different sites.



Because the anion acts as an electron donor, we can find clues to its reactivity preferences by examining the shape of its HOMO. Even though the HOMO is delocalized over several sites, the largest contribution clearly comes from the terminal carbon atom. Therefore, we expect electron movement and bond formation to occur at this carbon, and lead to the product shown on the left.

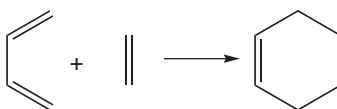


HOMO for enolate of ethyl acetate

Open *ethyl acetate enolate HOMO* in the *topics* directory, and switch to a mesh or transparent surface to see the underlying skeleton. Note that the orbital is most heavily concentrated on carbon, meaning that electrophilic attack will likely occur here. Close *ethyl acetate enolate HOMO* when you are finished.

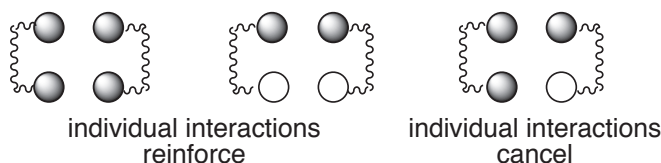
## The Fukui-Woodward-Hoffmann Rules

In certain cases, multiple frontier orbital interactions must be considered. This is particularly true of 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 (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 similar reactions 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.

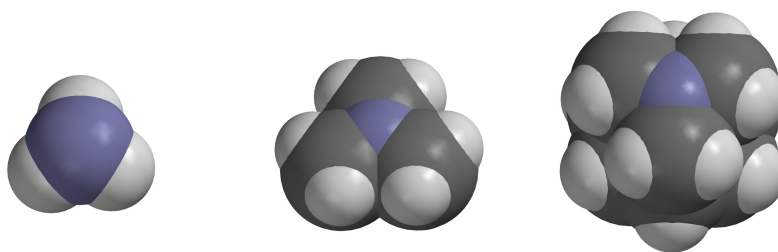
## ELECTRON DENSITIES: SIZES AND SHAPES OF MOLECULES

How big is an atom or a molecule? Atoms and molecules require a certain amount of space, but how much? A gas can be compressed into a smaller volume but only so far. Liquids and solids cannot be easily compressed. 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.

### 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 used for each element, and this radius has been chosen 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 very close together. 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.



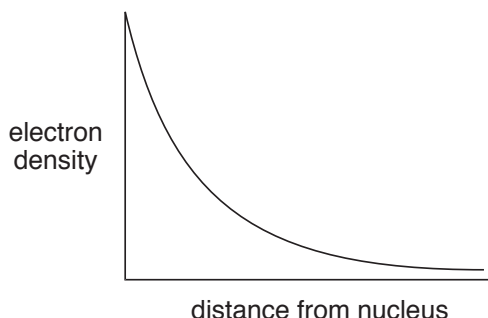
space-filling models for ammonia (left), trimethylamine (center) and 1-azaadamantane (right)

Space-filling models for ammonia, trimethylamine and 1-azaadamantane show how big these molecules are, and also show that the nitrogen in ammonia is more exposed than the corresponding nitrogen atoms in trimethylamine and 1-azaadamantane.

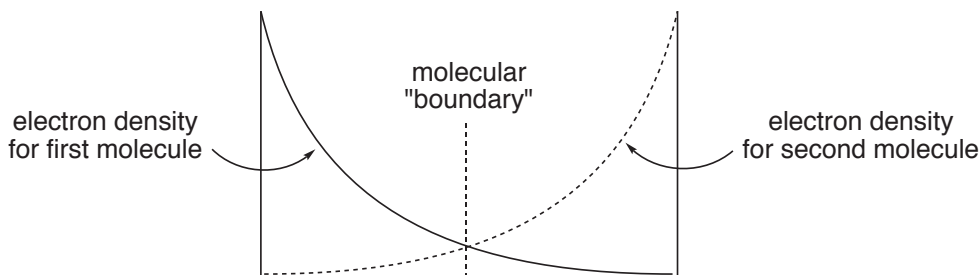
Open *amines space filling* in the *topics* directory. Space-filling models for ammonia, trimethylamine and 1-azaadamantane all appear on screen. Carbon atoms are colored dark grey, hydrogen atoms white and nitrogen blue. Note that the models clearly reveal the extent to which the nitrogen is exposed. Close *amines space filling* when you are finished.

## Electron Density Surfaces

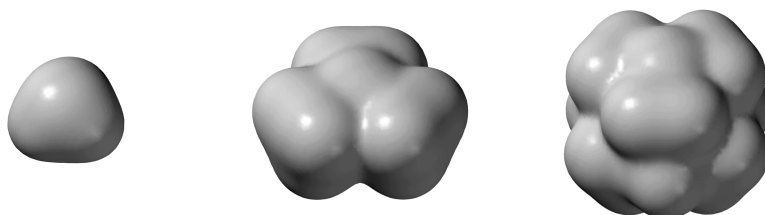
An alternative technique for portraying molecular size and shape relies on the molecule's own electron cloud. Atoms and molecules are made up of positively-charged nuclei surrounded by a negatively-charged electron cloud, and it is the size and shape of the electron cloud that defines the size and shape 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 means that it is really not possible to say how big a molecule is. 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 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. 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

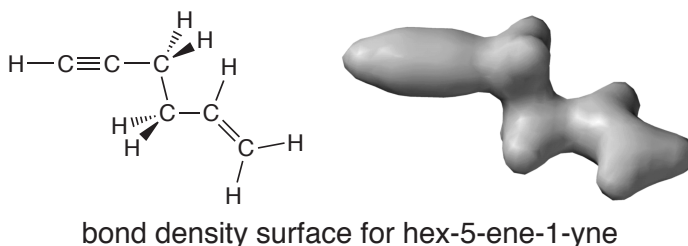
representation is to that offered by a simple space-filling model. Close *amines electron density* when you are finished.

Both space-filling and electron density models 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 atoms does not reflect reality.

### Bond Density Surfaces

Another useful surface, termed the *bond density surface*, is one that marks points corresponding to a much higher value of the electron density\*. Since points of high electron density are located much closer to the atomic nuclei, bond density surfaces enclose relatively small volumes, and do not give a correct impression of molecular size. On the other hand, bond density surfaces identify regions corresponding to bonding electron density, and the volume of these surfaces may be roughly correlated with the number of electrons that participate in bonding. In this sense, bond density surfaces are analogous to conventional line drawings and skeletal models.

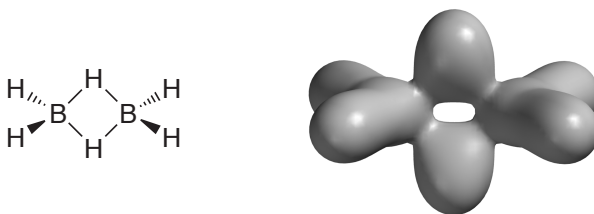
The bond density surface for hex-5-ene-1-yne clearly shows which atoms are connected, although it does not clearly distinguish single, double and triple carbon-carbon bonds.



\* 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.

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.

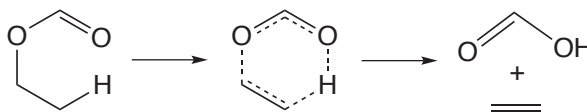
The usefulness of the bond density surface is more apparent in the following model of diborane. The surface clearly shows that there is relatively little electron density between the two borons. Apparently there is no boron-boron bond in this molecule. This is information extracted from the bond density surface model, and has been obtained without reference to any preconceived ideas about the bonding in diborane.



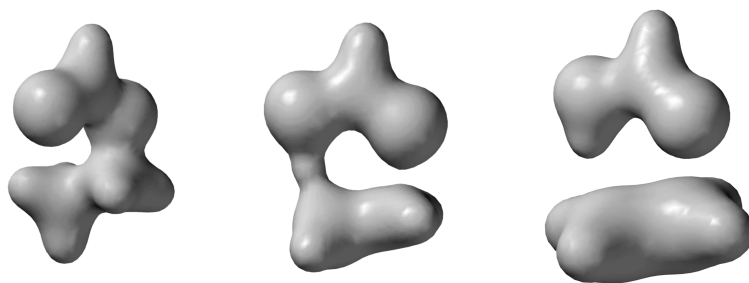
bond density surface for diborane

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.

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)

Compare the bond density surface in the pyrolysis transition state to those of the reactant and the products. The CO single bond of the reactant is clearly broken in the transition state. Also, 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.

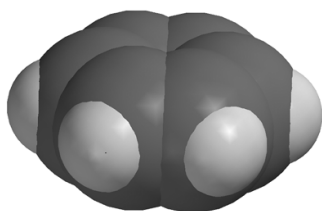
One way to describe a molecule's charge distribution is to give a numerical atomic charge for each atom. A particularly simple and familiar recipe yields so-called formal charges directly from Lewis structures. Unfortunately, formal charges are arbitrary. In fact, all methods for assigning charge are arbitrary and necessarily bias the calculated charges in one way or another. This includes methods based on quantum mechanics.

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.

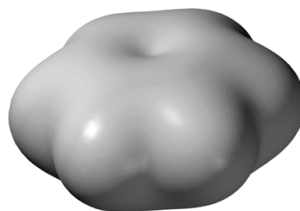
Electrostatic potentials can be depicted in various ways. For example, it is possible to make an electrostatic potential surface by finding all of the points in space where the electrostatic potential matches some particular value. A much more useful way to show molecular charge distribution is to construct an 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. It 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.



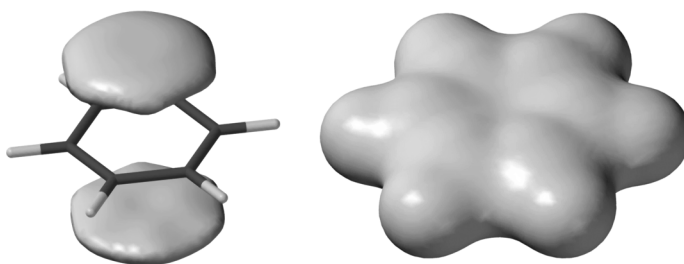
space-filling model  
for benzene



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.

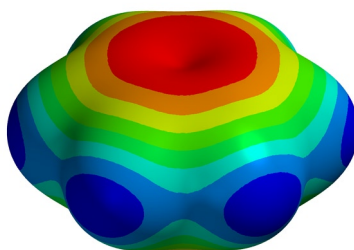
The electrostatic potential corresponding to points where the potential is negative shows two different surfaces, one above the face of the ring and the other below. Since the molecule's  $\pi$  electrons lie closest to these surfaces, we conclude that these electrons are responsible for the attraction 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.

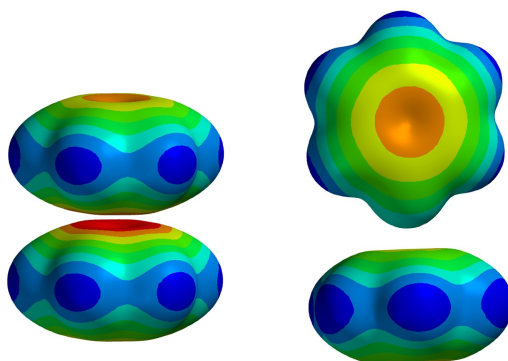
Next, combine the electron density and electrostatic potential surfaces to produce a so-called electrostatic potential map. The map for benzene conveys both the molecule's size and shape as well as its charge distribution in a compact and easily interpretable manner. The size and shape of the map are, of course, identical to that of the electron density surface, and indicate what part of the molecule is easily accessible to other molecules (the outside world). The colors reveal the overall charge distribution. The faces of the ring, the  $\pi$  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  $\pi$  faces and the blue regions are around the edges. Close *benzene electrostatic potential map* when you are finished.

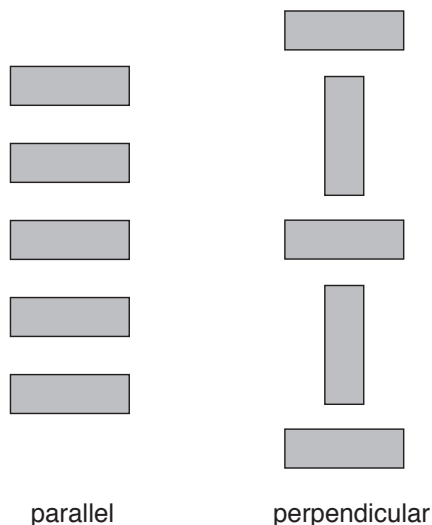
Electrostatic potential maps have made their way into mainstream general and (especially) organic chemistry textbooks as a means of displaying charge distributions in molecules. In addition, they have found application as a natural step beyond steric models for interpreting and predicting the way in which molecules fit together. A good example of this follows from the electrostatic potential map for benzene, which recall is negative on the  $\pi$  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.



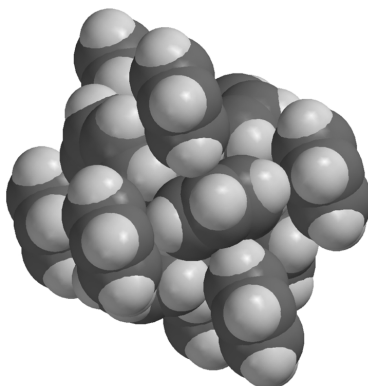
electrostatic potential maps for parallel (left) and perpendicular (right) benzene dimers

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 favors a perpendicular arrangement.



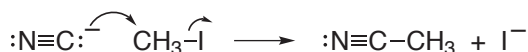
The experimental X-ray crystal structure shows a perpendicular arrangement, but in three dimensions. There are two lessons here. Intermolecular interactions go beyond steric interactions and sometimes our simple one-dimensional view of the world will lead us astray.





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.

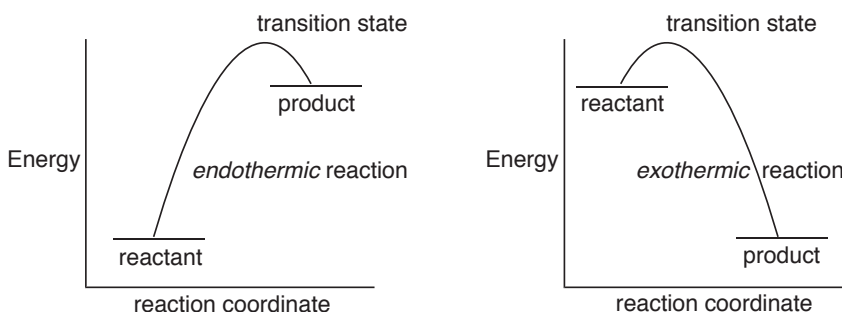
Electrostatic potential maps may also be used to describe in great detail the workings of chemical reactions. For example, a map may be used to show the transfer of negative charge during the  $S_N2$  reaction of cyanide with methyl iodide.



Open ***Sn2 cyanide+methyl iodide*** in the ***topics*** directory. One frame of a sequence of electrostatic potential maps for the S<sub>N</sub>2 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 more red 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. Conversely, 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.

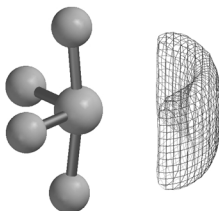


The Hammond Postulate provides a conceptual basis both for the Fukui-Woodward-Hoffmann rules (see the topic ***Atomic and Molecular***

**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 the LUMO map for nucleophilic reactions.

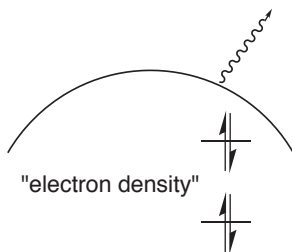
### Local Ionization Potential Maps and Electrophilic Reactivity

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.

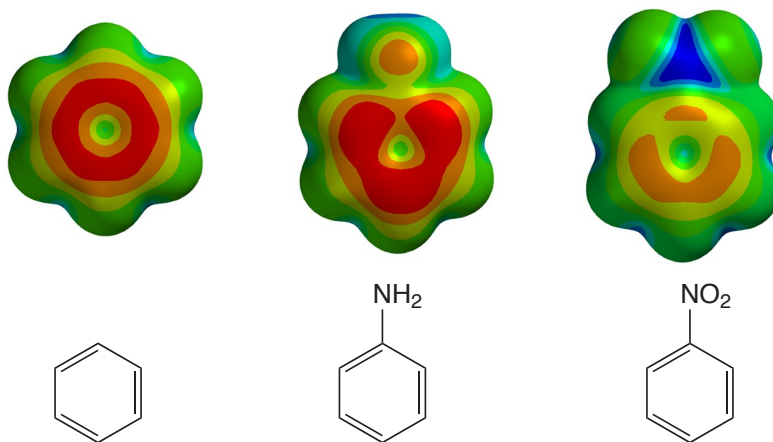


local ionization potential for sulfur tetrafluoride

The local ionization potential by itself is not generally a useful model. However, a map of the local ionization potential onto an electron density surface is a useful model, in that it reveals those regions from which electrons are most easily ionized.



Local ionization potential maps may be employed to reveal sites which are susceptible to electrophilic attack. For example, they show both the positional selectivity in electrophilic aromatic substitution ( $\text{NH}_2$  directs *ortho/para*, and  $\text{NO}_2$  directs *meta*), and the fact that  $\pi$ -donor groups ( $\text{NH}_2$ ) activate benzene while electron-withdrawing groups ( $\text{NO}_2$ ) deactivate benzene.

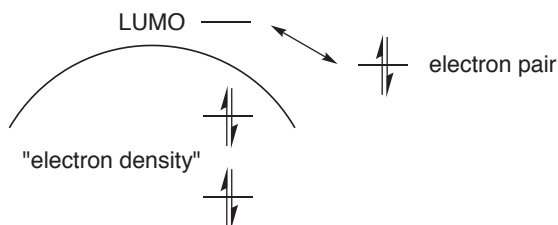


local ionization potential maps for benzene (left), aniline (center) and nitrobenzene (right)

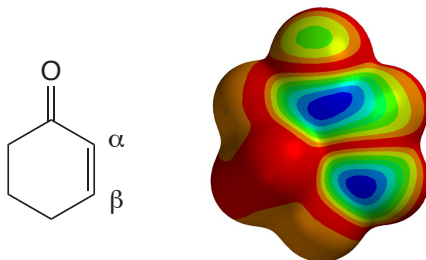
Open *benzene, aniline, nitrobenzene local ionization potential maps* in the *topics* directory. Here, the color red corresponds to regions of lowest ionization potential (most accessible to electrophiles). Note, that the  $\pi$  system in aniline is more accessible than the  $\pi$  system in benzene (the standard) and that the *ortho* and *para* positions are more accessible than *meta* positions. Note also, that the  $\pi$  system in nitrobenzene is less accessible than the  $\pi$  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.

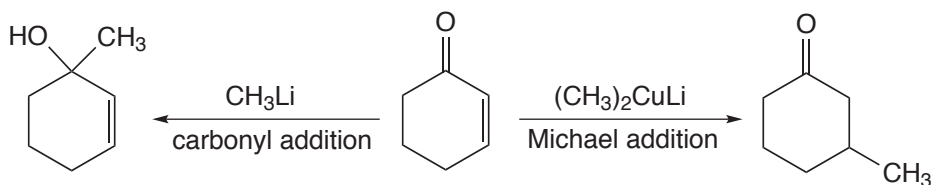


A good example is provided by the LUMO map for cyclohexenone.



LUMO map for cyclohexenone showing positions for nucleophilic attack in blue

The LUMO shows which regions of a molecule are 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  $\beta$  carbon, again consistent with the known chemistry of  $\alpha,\beta$ -unsaturated carbonyl compounds, in this case conjugate or Michael addition.



The buildup of positive charge on the  $\beta$  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.

# Appendix B

## Capabilities and Limitations

---

### Molecular Mechanics Models<sup>1</sup>

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. Energies from the latter may be corrected for the effects of aqueous solvent (MMFFaq). Structures are unaffected. There are no atom limits.

### Semi-Empirical Models<sup>1</sup>

The semi-empirical module calculates the heat of formation, wavefunction, equilibrium and transition-state geometries and vibrational frequencies, as well as provides for conformational searching. 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-Ne, Mg-Ar, 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<sup>1-3</sup>

The Hartree-Fock module calculates the energy and wave function,

---

\* MMFF is the only method available for generation of conformer libraries.

\*\* MMFF has been implemented for 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 version of PM3 has been implemented and parameterized for transition metals. Except for Zn, Cd and Hg, these include d-type orbitals in the basis set.

equilibrium and transition-state geometries and vibrational frequencies, as well as provides for conformational searching. Both closed-shell and open-shell (either ROHF or 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 (available in the *Spartan'14 Parallel Suite* only).

Preset limits for Hartree-Fock calculations are 200 atoms and 2,000 basis functions. Calculations on molecules with up to 100 atoms are practical.

### Density Functional Models<sup>1,2</sup>

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. The module supports local density calculations and a wide variety of gradient corrected and hybrid functionals. Only a few functionals are provided in the menu of functionals, but other selections can be entered into the **Options** line of the **Calculations** dialog (**Setup** menu; **Chapter 20**). The **Options** line may also be used to specify combinations of exchange and correlation components to be used in building a total functional. For example, the line **EXCHANGE=SLATER:0.9,BECKE:0.1** indicates a combination of 9 parts Slater exchange and 1 part Becke exchange. Similarly, the line **CORRELATION=VWN:0.19,LYP:0.81** indicates a combination of 19 parts VWN and 81 parts LYP correlation.

A full listing of available functionals (including exchange and correlation functions to be used as components in construction of custom functionals) is provided in **Table B-1**. Available basis sets are enumerated later in this appendix. Density functional models have been implemented for use on multi-core processors (available in the *Spartan'14 Parallel Suite* only).

Preset limits for density functional calculations are 200 atoms and 2,000 basis functions. Calculations on molecules with up to 100 atoms are practical.

**Table B-1: Available Exchange, Correlation and Combination Functionals**

<b>exchange functionals</b>	
HF	Fock exchange
Slater	Slater (Dirac 1930)
Becke	Becke (1988)
Gill96	Gill (1996)
GG99	Gilbert, Gill (1999)
B (EDF1)	Becke (uses EDF1 parameters)
PW91	Perdew, Burke, Ernzerhof (1996)
<b>correlation functionals</b>	
VWN	Vasko, Wilk, Nusair #5
LYP	Lee, Yang, Parr
PW91	GGA91, Perdew
LYP	(EDF1)
P86	Perdew (1986)
PZ81	Perdew, Zungner
PBE	Perdue, Burke, Ernzerhof (1996)
<b>combination functionals</b>	
B3PW91	
B3LYP	
B3LYP5 (Gaussian's version of B3LYP)	
EDF1	
EDF2	
BMK	
M05	
M052X	
M06	
M062X	
M06HF	
M06L	
$\omega$ B97	
$\omega$ B97X	
$\omega$ B97XD	

## Møller-Plesset Models<sup>1,2</sup>

The Møller-Plesset module calculates the energy and wavefunction and equilibrium and transition-state geometries as well as provides for conformational searching using either the MP2 or RI-MP2 models. Vibrational frequencies are also available from both models, but require numerical differentiation of analytical gradients. Available basis sets are enumerated later in this appendix. RI-MP2 models (only) have been implemented for use on multi-core processors (available in the *Spartan'14 Parallel Suite* only).

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.

MP3 and MP4 models are available for energy calculations only and are supported under **Advanced Correlated Models**.

## Advanced Correlated Models<sup>1,2</sup>

The advanced correlated module calculates only the energy using the MP3, MP4, CCSD, CCSD(T), CCSD(2), OD, OD(T), QCISD, QCISD(T), QCCD and QCCD(2) correlated models. Only a few models are provided in the menu and the others need to be entered into the **Options** line in the **Calculations** dialog (**Setup** menu; **Chapter 20**). 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 G2, G3 and G3(MP2) recipes are available for highly-accurate heat of formation calculations. Practical considerations generally limit application of the first two recipes to molecules comprising less than 10 heavy atoms and the G3(MP2) recipe to molecules comprising less than 15 heavy atoms.

The T1 recipe<sup>4</sup> 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. The result is that T1 requires 2-3 orders of magnitude less computation time than G3(MP2). Calculations on 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 that can be properly described in terms of a conventional valence structure.

### **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), while the latter supports the same functionals or combinations of functionals available for ground state density functional calculations. At the present time, only the CIS and RI-CIS(D) models are available for geometry optimization and frequency calculations, all other models being restricted to energy 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 advanced 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 polarization functions and/or with diffuse functions. Also available are the cc-pVDZ, cc-pVTZ, cc-pVQZ, cc-pV5Z and aug-cc-pCVQZ basis sets and the G3Large basis set.

STO-3G is available for H–I, 3-21G for H–Cs, 6-31G and basis sets built on it for H–Kr, 6-311G and basis sets built on it for H–Ca, Ga–Kr and I, cc-pVDZ, cc-pVTZ, cc-pVQZ, and G3 Large for H–Ca and Ga–Kr and aug-cc-pCVQZ for H–Ar.

Use of minimal and split-valence basis sets for any but Hartree-Fock calculations generally leads to poor results and is not recommended.

Dual-basis set techniques (mixing small and large basis sets) are supported for ground-state energy and equilibrium and transition-state geometry calculations for Hartree-Fock, density functional and MP2 models. These can be up to an order of magnitude faster than single-basis set calculations, depending on choice of small and large basis sets.

Pseudopotentials are available for calculations on molecules incorporating heavy atoms (for which all-electron basis sets are not available). Their use is automatic. *Spartan'14* extends pseudopotentials to lanthanides which have been specifically calibrated to provide good equilibrium geometries for density functional models using the 6-31G\* basis set.

## Solvent Models

*Spartan'14* supports two different empirical solvation models. The SM5.4 model<sup>5</sup> is available for H, C–F, S–Cl, Br and I, and 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. Because the computational cost is very low (on the order of that for an energy using a semi-empirical model), the SM5.4 solvent correction is provided automatically for all quantum chemical calculations and is reported as a molecular property.

The SM8 model<sup>6</sup> is available for H, C–F, S–Cl and Br, and is applicable only to Hartree-Fock and density functional models with the 6-31G\*, 6-31G\*\*, 6-31+G\* and 6-31+G\*\* basis sets. It affects the wave function, meaning that, in addition to energy, equilibrium or transition-state geometry, properties, infrared spectra and graphical displays may be obtained in the presence of solvent. SM8 has been parameterized for water as well as a variety of organic solvents. A few common solvents may be specified in a menu (acetone, diethyl ether, DMF, DMSO, ethanol, methylene chloride, THF and toluene) while others may be specified using the **Solvent** keyword in the **Calculations** dialog (**Setup** menu; **Chapter 20**). Available solvents are listed in **Appendix D**. The computational cost of SM8 is roughly that of an energy with the corresponding Hartree-Fock or density functional model.

## 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.

Calculation of absolute entropy (and the contribution that entropy makes to the Gibbs energy) is problematic for molecules that are not rigid, and there is no “correct” procedure that can be followed. The entropy comprises four components: translation, rotation, vibration and conformation.

$$S = S_{\text{tr}} + S_{\text{rot}} + S_{\text{vib}} + S_{\text{conf}}$$

$$S_{\text{tr}} = nR \left\{ \frac{5}{2} + \ln \left[ \left( \frac{2\pi MkT}{h^2} \right)^{3/2} \left( \frac{nRT}{P} \right) \right] \right\}$$

$$S_{\text{rot}} = nR \left\{ \frac{3}{2} + \ln \left[ \left( \frac{\pi V_A V_B V_C}{s} \right)^{1/2} \right] \right\} ; V_A = h^2/8\pi I_A kT$$

$$S_{\text{vib}} = nR \sum_i \left\{ (U_i e^{u_i} - 1)^{-1} - \ln(1 - e^{-u_i}) \right\}; \quad u_i = h\nu_i/kT$$

$$S_{\text{conf}} = \ln \left[ \sum_j g_j e^{-\Delta E_j/RT} \right] 3^{\text{Me}}$$

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,  $\nu$  are the vibrational frequencies,  $E$  are energies relative to the lowest-energy conformer,  $g$  are the number of times the conformer appears and  $\text{Me}$  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 in particular is problematic as the entropy contribution goes to  $\infty$  as the frequency goes to 0. In practice, the contribution is clamped at  $1/2 R$  (the entropy of rotation).

Note that the entropies contained in SSPD have not been corrected for the conformational contribution. This correction is zero for rigid molecules, but may range to several tens of entropy units for large flexible molecules.

Several properties based on graphical displays are also available. These 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, but only a small set of numerical values and not the surfaces and maps themselves are actually stored. These properties need to be explicitly requested by *checking* the **QSAR** box inside the **Calculations** dialog (**Calculations** under the **Setup** menu; **Chapter 20**). They are either available from the **QSAR** tab of the **Molecule Properties** dialog (**Properties** under the **Display** menu; **Chapter 21**) or from the **Formula Editor** dialog (**Formulas** under the **Display** menu; **Chapter 21**).

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 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*'14 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 wavefunction 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.

- 
1. For a general discussion and assessment of the techniques and methods available in *Spartan*, see: W.J. Hehre, *A Guide to Molecular Mechanics and Quantum Chemical Calculations*, Wavefunction, Inc., Irvine, CA 2003. This is available as a PDF under the **Help** menu.
  2. For a review of the quantum chemical methods utilized in *Spartan* (except for semi-empirical methods) with emphasis on recent developments, see: 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. Khallilulin, 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.*, **8**, 3172 (2006). A more thorough and more mathematical discussion of the Hartree-Fock, density functional and wave function based correlation methods may be found in the Q-Chem User's Manual for version 4.0 of the program (2011).
  3. For an older account see: W.J. Hehre, L. Radom, P.v.R. Schleyer and J.A. Pople, *Ab Initio Molecular Orbital Theory*, Wiley, New York, 1986.
  4. W.S. Ohlinger, P.E. Klunzinger, B.J. Deppmeier, W.J. Hehre; *J. Phys. Chem. A.*, **113**, 10, 2165 (2009).
  5. C.C. Chambers, G.D. Hawkins, C.J. Cramer and D.G. Truhlar, *J. Phys. Chem.*, **100**, 16385 (1996).
  6. A.V. Marenich, R.M. Olsen, C.P. Kelly, C.J. Cramer and D.G. Truhlar, *J. Chem. Theory Comput.*, **3**, 2011 (2007).










# Appendix C







## Menus

### *Spartan Screen*

#### **File**










---

<u>N</u> ew Build 	Brings up a model kit for 3D molecule building or substitution
New Sketch 	Brings up the sketch pad for molecule sketching in 2D
<u>O</u> pen... 	Opens (imports) a molecule or multi-molecule document
<u>C</u> lose 	Closes a molecule or multi-molecule document
<u>S</u> ave 	Saves (exports) a molecule or multi-molecule document
Save <u>A</u> s... 	Saves a molecule as a document under a user-specified name; creates files for use in the Spartan Molecular Database (SMD), Spartan Reaction Database (SRD), Spartan Infrared Database (SIRD) and Spartan Spectra and Properties Database (SSPD)
Build New M <u>o</u> lecule 	Adds a molecule to an existing document; brings up a model kit for molecule building or substitution
Sketch New M <u>o</u> lecule 	Adds a molecule to an existing document; brings up the sketch pad for molecule sketching in 2D
<u>D</u> elete Molecule 	Deletes a molecule (or molecules) from a document

Append <u>M</u> olecule(s)...		Appends molecules to an existing document
Access PDB Online...		Accesses the online Protein Data Bank (PDB)
<u>P</u> rint...		Prints on-screen display; also prints contents of output window and the <b>Spreadsheet</b>
Embedded Data		Associates external files with a document, either at molecule or document level (or both)
Start/Stop QuickTime Recording		Starts and stops QuickTime recording of contents of main <i>Spartan</i> screen (only)
<u>E</u> xit		Exits <i>Spartan</i>

## Edit

---

<u>U</u> ndo		Undoes previous operations
C <u>u</u> t		Moves the current molecule or contents of the selection box to the clipboard
<u>C</u> opy		Copies the current molecule or contents of the selection box to the clipboard
<u>P</u> aste		Pastes contents of the clipboard to the screen
S <u>e</u> lect All		Selects all atoms in the selected molecule
<u>F</u> ind...		Locates a text string in the output dialog or an on-screen molecular fragment
Find <u>N</u> ext		Locates next occurrence of a text string or molecular fragment
C <u>e</u> nter		Centers the molecule on screen; applies to all molecules in a document
<u>C</u> lear		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

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




## Geometry

---

Measure <u>D</u> istance 	Displays and/or sets bond distance
Measure <u>A</u> ngle 	Displays and/or sets bond angle
Measure <u>D</u> ihedral 	Displays and/or sets dihedral angle
<u>F</u> reeze Center 	Freezes selected atomic positions
Set <u>T</u> orsions 	Selects bond rings for conformational searching
Set Similarity Centers 	Identifies atoms /CFD's for similarity analyses
<u>C</u> onstrain Distance 	Constrains bond distance
Con <u>s</u> train Angle 	Constrains bond angle
Con <u>s</u> train Dihedral 	Constrains dihedral angle
Define <u>P</u> oint 	Defines a point as a geometric mean of a set of atoms; defines ligand point as a position that is perpendicular to the centroid of a plane made by three or more atoms
Define <u>P</u> lane 	Defines a plane made by three or more atoms
Define CFD 	Defines the location of a CFD
Align 	Aligns molecules in a document according to selected atoms, structure, or CFD's

## Build

---

<u>V</u> iew 	Removes the model kit
Edit <u>B</u> uild 	Brings up a 3D model kit (organic, inorganic, peptide, nucleotide, substituent or ChemDraw) with the presently selected molecule
Edit <u>S</u> ketch 	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 SMD or SSPD.

Delte 

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

## **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

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

Spreadsheet 

Displays spreadsheet

Plots... 

Creates 2D and 3D plots from the data in the spreadsheet. A simplified and more intuitive plot presentation mode is available

Similarities 

Displays results of similarity analyses

Reactions 

Calculates reaction (activation) energies using data either from current document or from the Spartan Molecular Database (SMD) or Spartan Spectra and Properties Database (SSPD)

## 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

Databases 

Performs substructure search on Cambridge Structural Database (CSD), Spartan Molecular Database (SMD), 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)

Guess Transition State 

Provides transition-state guess based on reaction database or, lacking a database entry, based on linear synchronous transit

Identify Tautomers 

Identifies tautomers

Extract Ligands 

Extracts ligands and pharmacophores based on protein-ligand binding from PDB files

## 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 (for *Spartan'14 Parallel Suite* only) and permits use of *Spartan'14* as a server.

Colors 

Sets screen and model colors

Fonts... 

Sets fonts for labels and plot displays

Graphics Fonts... 

Sets fonts for graphical displays

Local Monitor 

Monitors and allows for killing jobs executing on local machine and bypassing queue to start jobs

Network Monitor 

Monitors and allows killing jobs executing on a remote server

Calculator 

Pocket calculator

Icons 

Turns on and off icons displayed under the menu bar

## Activities

---

Tutorials 

Brings up *Spartan*'14 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

---

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*

About... 

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 <b>Molecular Properties</b> 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 <b>Add</b> dialog (spreadsheet) for adding calculated quantities into the spreadsheet
-----	---

Sort	Sorts the column from low to high. <i>Pressing</i> the <b>Shift</b> 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 <i>Spartan</i> document(s) to the spreadsheet (corresponding to the selected document)
--------	--

Rename Selected Using SMD/SSPD	Rename selected molecule(s) with names in the Spartan Molecular Database (SMD) or Spartan Spectra and Properties Database (SSPD); appears only when leftmost cell(s) selected
--------------------------------	---

Properties	Brings up the <b>Molecular Properties</b> 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 <i>Spartan</i> 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

---

# Appendix D

## Commonly-Used Program Options

---

This appendix describes a number of commonly-used program options. These are specified using keywords input into the **Options** box in the **Calculations** dialog (**Setup** menu). In addition to the above, any method, functional or combination of functionals and/or basis set supported in *Spartan* (and not included explicitly in the menus) may be typed into the **Options** box. For example, typing **EDF2 6-311G(2df,2pd)** will result in density functional calculation using the EDF2 functional and the 6-311G(2df,2pd) basis set. (See **Chapter 20** for further details.)

Keywords are case insensitive. An equals sign (=) separates a keyword from its integer, real or character string value. Keywords may either be single words or expressions. **Keyword=N** indicates an integer argument, **keyword=C** indicates a character argument, and **keyword=F** indicates a floating point argument. Real values can optionally include an **eNN** or **e-NN** (floating point power of ten). Default values for keywords are indicated in **bold**, and alternative values in *italics*.

### Conformational Search

keyword	value	description
<b>SearchMethod</b>	<b>C</b>	conformational searching method employed; overrides default choice which depends on complexity of system may be overridden with one of the following: <i>Systematic</i> use systematic method <i>MonteCarlo</i> use Monte-Carlo method <i>Sparse</i> use systematic method but randomly eliminate conformers

to stay within set conformer limits. Used in conformer library generation.

<b>MaxEnergy</b>	<b>F</b>	sets the maximum energy (in kJ/mol) at which a conformer will be kept to (minimum energy + F), where minimum energy is the energy of the lowest energy conformer encountered at that point in the search; default is <b>40</b> kJ/mol
<b>StartTemperature</b>	<b>F</b>	sets initial temperature (in K) for a Monte Carlo search (only) to F; high temperatures will result in the molecule further exploring global space, while low temperatures will result in the molecule further exploring local space; default is <b>5000K</b>
<b>FindBoats</b>		Extend conformation search for saturated six-member rings (“cyclohexane”) to consideration of twist-boat forms. Results in significantly increased computer time

## Similarity Analysis

keyword	value	description
<b>SingleEnantiomer</b>	<b>–</b>	Limits similarity analysis to the enantiomer actually in specified library documents. The default is to consider both enantiomers
<b>QuitSimilarity</b>	<b>N</b>	Quit similarity analysis involving a conformer list (in the graphical user interface) and/or one or more conformer libraries when N good matches have been identified; default is <b>1</b>
<b>QuitCriterion</b>	<b>F</b>	R <sup>2</sup> criterion for a good match; default is <b>0.9</b>

## Geometry Optimization

keyword	value	description
<b>GeometryCycle</b>	<b>N</b>	set maximum number of equilibrium and transition-state geometry optimization cycles to N; default is <b>number of independent variables +20</b> for equilibrium geometry optimization and <b>3x number of independent variables</b> for transition-state geometry optimization

<b>GradientTolerance</b>	<b>F</b>	set convergence criterion in equilibrium and transition-state) geometry optimization for the maximum gradient component (in hartrees/bohr) to F; default is <b>3 x 10<sup>-4</sup></b>												
<b>DistanceTolerance</b>	<b>F</b>	set convergence criterion in equilibrium and transition-state geometry optimization for the maximum change in a bond length (in Å) to F; default is <b>1.2 x 10<sup>-3</sup></b>												
<b>Hess</b>	<b>C</b>	Choose Hessian (matrix of second derivatives) for start of geometry (transition-state geometry) optimization (in the absence of a Hessian from a previous calculation, the default is <b>MMFF</b> molecular mechanics for geometry optimization and <b>PM3</b> semi-empirical for transition-state geometry optimization; may be overridden with one of the following:  <table><tr><td><i>unit</i></td><td>unit matrix</td></tr><tr><td><i>MMFF</i></td><td>MMFF molecular mechanics</td></tr><tr><td><i>AM1</i></td><td>AM1 semi-empirical</td></tr><tr><td><i>STO-3G</i></td><td>STO-3G Hartree-Fock</td></tr><tr><td><i>3-21G</i></td><td>3-21G Hartree-Fock</td></tr><tr><td><i>6-31G*</i></td><td>6-31G* Hartree-Fock</td></tr></table>	<i>unit</i>	unit matrix	<i>MMFF</i>	MMFF molecular mechanics	<i>AM1</i>	AM1 semi-empirical	<i>STO-3G</i>	STO-3G Hartree-Fock	<i>3-21G</i>	3-21G Hartree-Fock	<i>6-31G*</i>	6-31G* Hartree-Fock
<i>unit</i>	unit matrix													
<i>MMFF</i>	MMFF molecular mechanics													
<i>AM1</i>	AM1 semi-empirical													
<i>STO-3G</i>	STO-3G Hartree-Fock													
<i>3-21G</i>	3-21G Hartree-Fock													
<i>6-31G*</i>	6-31G* Hartree-Fock													

## Molecular Mechanics Calculations

keyword	description
<b>MMFF94s</b>	Use MMFF94s method instead of MMFF94. The only significant difference between the two is that MMFF94 forces nitrogens attached to aromatic rings to be planar (as in aniline). While this is widely perceived to be the case, in fact the nitrogen in aniline and related molecules are pyramidal not planar.

## Quantum Chemical Calculations

keyword	value	description
<b>Diis</b>	<b>N</b>	switch on diis all the time in SCF procedure. N is the size of the iterative subspace; it should be an integer between 2 and 10; default is <b>5</b>
<b>NoDiis</b>		turn off diis SCF convergence accelerator

<b>ScfCycle</b>	<b>N</b>	set maximum number of SCF iterations to N; default is <b>50</b>
<b>Guess</b>	<b>C</b>	choose initial wavefunction guess; in the absence of a guess from a previous calculation, the default is either <b>sad</b> or <b>PM3</b> depending on the calculation type, and may be overridden with one of the following:  <i>core</i> diagonalize the core Hamiltonian <i>sad</i> superposition of atomic densities <i>PM3</i> PM3 semi-empirical
<b>Scf</b>	<b>C</b>	SCF procedure; default is <b>restricted</b> for closed-shell systems and <b>unrestricted</b> for open-shell systems, either of which may be specified for both closed and open-shell systems
<b>SmallBasis</b>	<b>C</b>	specifies the small basis set to be used in dual basis set calculations. The large basis set is that indicated in the menu.
<b>Core</b>	<b>C</b>	use of frozen core approximation in Møller-Plesset and advanced correlated calculations; default is <b>frozen</b> which may be changed to <i>thawed</i>
<b>ScfTolerance</b>	<b>C</b>	control of all tolerances in SCF procedure and to affect the precision of the algorithm; default is <b>normal</b> which may be changed to <i>high</i> or <i>veryhigh</i>
	<b>F</b>	set SCF energy convergence criterion (in hartrees) to F; default is <b>1.0 x 10<sup>-6</sup></b>
<b>BigGrid</b>	–	uses very large grid in density functional calculations
<b>Mix</b>	–	specifies that $\alpha$ and $\beta$ HOMO's in the guess wavefunction should be constructed according to:

$$\text{HOMO}_{\alpha} = \frac{\text{HOMO} + \text{LUMO}}{\sqrt{2}}$$

$$\text{HOMO}_{\beta} = \frac{\text{HOMO} - \text{LUMO}}{\sqrt{2}}$$

useful for generating a guess wavefunction for singlet diradical

<b>Exchange</b>	<b>E</b>	specifies custom exchange functional or combination of exchange functionals for density functional calculation. See <b>Appendix A</b> for details.
<b>Correlation</b>	<b>C</b>	specifies custom correlation functional or combination of correlation functionals for density functional calculation. See <b>Appendix A</b> for details.
<b>EnergyOnly</b>		skips calculation of the density matrix for an MP2 energy calculation, reducing the time by 50%

## Semi-Empirical Calculations

keyword	value	description
<b>NoAmideFix</b>	–	turns “off” molecular mechanics amide correction in PM3 calculations
<b>hh</b>	<b>C</b>	turns “on” and “off” hydrogen-hydrogen repulsion term in PM3 calculations; default is <b>on</b> only where transition metals are involved and <b>off</b> otherwise
<b>Polarizability</b>	–	calculates and prints polarizabilities and hyperpolarizabilities

## Property and Spectra Calculations

keyword	value	description
<b>NBO</b>	<b>C</b>	specify details of NBO population analysis; default is <b>normal</b> which may be overridden for <i>ionic</i> and <i>3C</i> (three center) systems
<b>NoElCharge</b>	–	turns “off” calculation of electrostatic charges
<b>ElCharge</b>	<b>N</b>	adjusts size of grid used to calculate charge from electrostatic potential; default is <b>1</b> point/atomic unit
<b>Temperature</b>	<b>F</b>	temperature (in K) used in calculation of thermodynamic properties; default is <b>298°K</b>
<b>TEMPRANGE</b>	<b>Fi,Fj,Fk</b>	temperature range applied to thermodynamic properties when IR frequencies are calculated, where i=starting temperature, j=ending temperature, and k=increment, all values in K

<b>Pressure</b>	<b>F</b>	pressure (in atm) used in calculation of thermodynamic properties; default is <b>1</b> atm
<b>Moments</b>	–	calculates and prints principal moments of inertia
<b>NoSolvent</b>	–	turns “off” calculation of aqueous solvation energy using SM5.4 model
<b>Solvent</b>	<b>C</b>	specifies solvent for SM8 model beyond those included in menu. See <b>FAQ (Properties)</b> available under <b>Help (Help menu; Chapter 25)</b> for a list of available solvents. This will replace the menu entry.
<b>NoAreaVolume</b>	–	turns “off” calculation of surface area and volume
<b>MaxPropWeight</b>	<b>N</b>	sets limit of molecular weight for electrostatic charge, solvent and area/volume calculations; default is <b>500</b> amu
<b>Polarizability</b>	<b>F</b>	calculates and prints either static (F not specified) or frequency dependent (frequency F in cm <sup>-1</sup> ) polarizabilities and hyperpolarizabilities
<b>FreqScale</b>	<b>F</b>	scales calculated vibrational frequencies by F; default is <b>1</b>
<b>NumericalFreq</b>	–	uses numerical differentiation (of analytical gradients) for frequency calculation
<b>UVstates</b>	<b>N</b>	specifies number of states to be examined in UV/visible spectra calculation; default is <b>5</b>
<b>IRCSteps</b>	<b>N</b>	maximum number of steps allowed in constructing IRC; default is <b>40</b>
<b>IRCStepSize</b>	<b>N</b>	maximum step size (in mass-weighted atomic units x 100) for IRC; default is <b>150</b>
<b>IncludeTriplets</b>		include excited triplet states in a UV spectra calculation where the ground state is a singlet. Ignore where the ground state is a triplet

---

## Miscellaneous and Printing

keyword	value	description
<b>3DPlot</b>	–	treat dynamic constraints independently (rather than in concert). This allows data for 3D (XYZ) plot to be collected. Two dynamic constraints only.

<b>Execute</b>	<b>C</b>	either skip execution ( <b>skip</b> ) or force execution regardless of history ( <b>force</b> )
<b>PrintVibCoords</b>	–	print coordinates corresponding to normal mode vibrational frequencies
<b>PrintCoords</b>	–	prints Cartesian coordinates
<b>PrintLev</b>	<b>C</b>	print level; default is <b>normal</b> which may be changed to <i>verbose</i>
<b>PrintOrbE</b>	–	print energies of molecular orbitals
<b>PruneVirtual</b>	<b>N,C</b>	keep N virtual orbitals; default is <b>10</b> ; <b>None</b> keeps all virtual orbitals

---

# Appendix E

## Units

---

### Geometries

Cartesian coordinates are given in Ångstroms (Å), and in atomic units (au).

Bond distances are given in Å and in au. Bond angles and dihedral angles are given in degrees (°).

Surface areas, accessible surface areas and polar surface areas are available in Å<sup>2</sup> and volumes in Å<sup>3</sup>, and in au<sup>2</sup> (au<sup>3</sup>).

$$1 \text{ Å} = 0.1 \text{ nm} = 1.889762 \text{ au}$$

### 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, kJ/mol and electron volts (eV).

Experimental heats of formation as well as those from semi-empirical calculations and from thermochemical recipes are available in kJ/mol, au, kcal/mol and eV.

Strain energies from molecular mechanics calculations are available in kJ/mol, au, kcal/mol and eV.

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 kJ/mol·degree, kcal/mol·degree and au/mol·degree.

### Orbital Energies

Orbital energies are available in eV, kcal/mol, kJ/mol and au.

## 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) =  $1.593 \times 10^{-3}$

## Electron Densities, Spin Densities, Dipole Moments, Charges, Electrostatic Potentials and Local Ionization Potentials

Electron densities and spin densities are given in electrons/au<sup>3</sup>.

Dipole moments are given in Debye.

Atomic charges are given in electrons.

Electrostatic and hydride potentials are given in kJ/mol.

Local ionization potentials are given in eV.

## Vibrational Frequencies, Infrared and Raman Intensities

Vibrational frequencies are given in wavenumbers (cm<sup>-1</sup>). Intensities are given relative to the most intense peak.

## Chemical Shifts, Coupling Constants

Chemical shifts are given in parts-per-million (ppm) relative to appropriate standards: hydrogen, tetramethylsilane; carbon, tetramethylsilane; nitrogen, nitromethane; fluorine, fluorotrichloromethane; silicon, tetramethylsilane, phosphorous, phosphoric acid. These are available for Hartree-Fock, BP, BLYP, B3LYP, EDF1 and EDF2 models with the 6-31G\*, 6-31+G\*, 6-31G\*\*, 6-31+G\*\*, 6-311G, 6-311+G\*, 6-311+G\*\* and 6-311++G\*\* basis sets. Coupling constants are in ppm.

## UV/visible Spectra

$\lambda_{\text{max}}$  is given in wavenumbers (cm<sup>-1</sup>). Intensities are given relative to the most intense peak.

# Appendix F

## Citation

---

The proper citation for *Spartan'14* is as follows:

*Spartan'14*  
Wavefunction, Inc.  
Irvine, CA

Except for molecular mechanics and semi-empirical models, the calculation methods used in *Spartan* have been documented in: 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.*, **8**, 3172 (2006).

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. This is available as a PDF under the **Help** menu.

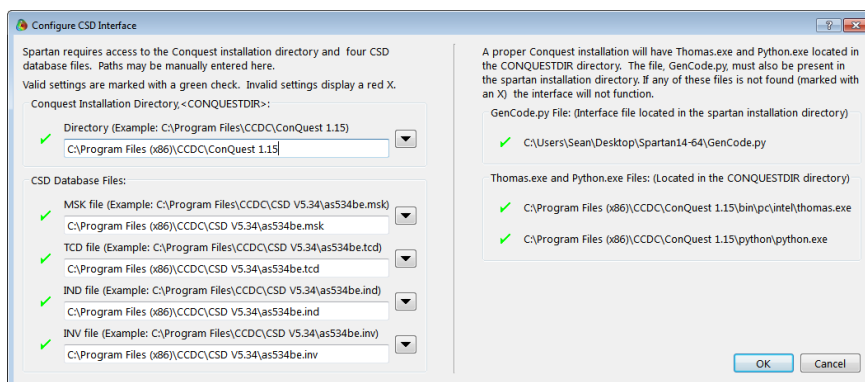
# Appendix G

## Installing the Cambridge Structural Database

Access to the Cambridge Structural Database (CSD) from *Spartan* should be automatic, following successful installation of the ConQuest program (supplied with CSD) according to the detailed instructions provided. If access is denied, the following steps need to be taken:

### Windows:

- (i) Click on **Config** near the bottom right of the dialog. This brings up the **Configure CSD Interface** dialog,



This will list any files which are missing. Click on **Setup Paths** at the bottom of the dialog to bring up the **CSD File Paths** dialog,

This dialog may show one or more red lights.

- (ii) Make certain that the path to ConQuest is correctly specified. The most common error will be that the version number is not current. (At the time of writing, the current version is 1.15.) You can get the current version number from your installation of ConQuest.

- (iii) Given the proper version number for ConQuest, update the file names in the boxes below **MSK file**, **TCD file**, **IND file** and **INV file**. You can get the current file names from your installation of ConQuest. As you enter the current file names, the red lights turn to green lights. When all lights are green, *click* on **OK** to dismiss the **CSD File Paths** dialog and then *click* on **OK** to dismiss the **Configure CSD** dialog.

**Linux:**

- (i) **CSD HOME** must be defined in order for *Spartan* to locate the CSD program.
- (ii) Please refer to the *Cambridge Structural Database System Release and Installation Notes* for instructions on setting up **CSD HOME**.

# Appendix H

## Constructing Custom Databases

---

Custom collections may be added to SMD, SIRD and SRD. 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 SMD and SIRD only) conformer libraries generated from molecular mechanics calculations. Entries for SRD (transition states) need to derive from *Spartan*'s transition state guessing facility (**Guess Transition States** under the **Search** menu; **Chapter 22**) in order that they 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 23**).

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 21**) or in the spreadsheet (see **Spreadsheet** under the **Display** menu in **Chapter 21**).

# 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  $^{13}\text{C}$  spectra is in CML (Chemical Markup Language) format. A description may be found at <http://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).