Chapter 14

"Dry Labs": Using the Spartan Spectra and Properties Database

The tutorials in this chapter comprise examples where structures, energies, spectra and other properties can be extracted from the Spartan Spectra and Properties Database doing away with the need for new quantum chemical calculations.

The Spartan Spectra and Properties Database (SSPD) is a collection of molecular structures, spectra and diverse properties initially available from a single density functional model (EDF2/6-31G*). With the release of ***Spartan'16*** the computational model has shifted to a new standard (ωB97X-D/6-31G*), which provides superior quality data (although is also more computationally costly). Both models are available. The ωB97X-D/6-31G* will be the default model, but users are able to change to EDF2/6-31G* if desired.

SSPD comprises over 300,000 molecules, each of which includes a wide range of molecular properties including a selection of QSAR descriptors (EDF2/6-31G* only), IR Spectra (EDF2/6-31G* only), NMR spectra, and the wave function, allowing "on-the-fly" generation of molecular orbitals, electron densities and other graphical surfaces as well as the electrostatic potential map and other property maps. In short, SSPD provides a rich source of diverse information for a large selection of molecules.

The emphasis of SSPD on only two theoretical models (with the ωB97X-D/6-31G* model being favored) is a deliberate attempt to turn attention away from the characteristics of a particular model (and the question "which model is best") and to focus on the chemistry at hand. While the choice of the new standard, the

ωB97X-D/6-31G* model, may be seen as a continued compromise between accuracy and practicality, in fact, ωB97X-D/6-31G* provides high quality results for a variety of properties that can be accurately measured, including molecular geometry and NMR spectra. Further, both the T1 heat of formation and the ωB97X-V/6-311+G(2df,2p) energy are included in the SSPD entry as a property. We believe that over the wide variety of molecules considered, T1 provides as reliable an indicator of ΔH 298 as the G3(MP2) recipe (which in turn is within measurable experimental error).

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

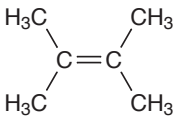
To complete the tutorials in this chapter, SSPD needs to be available. Either the sample version installed with all copies of ***Spartan*** and which comprises ≈6,000 molecules or the full version which comprises >300,000 molecules which needs to be separately downloaded.

Reactivity of Silicon-Carbon Double Bonds

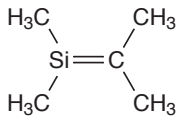
With the exception of so-called phosphorous ylides, compounds incorporating a double bond between carbon and a second-row element are quite rare. Most curious perhaps is the absence of stable compounds incorporating a carbon-silicon double bond. This can be rationalized by using local ionization potential and LUMO maps to compare the reactivities of olefins and silaolefins.

1. Build or sketch both 2,3-dimethyl-2-butene and tetramethylsilaethylene and put into the same document. Start the second molecule with **Build New Molecule** or **Sketch New**

Molecule from the **File** menu instead of **New Build** or **New Sketch**.



2,3-dimethyl-2-butene



tetramethylsilaethylene

2. Both molecules are in SSPD. *Click* on the name of whichever molecule is selected at the bottom of the screen, confirm that the ωB97X-D/6-31G* model is selected and *click* on the **Replace** button and then *click* on **All**.

3. Select **Surfaces** from the **Setup** menu (). First request a local ionization potential map. *Click* on **Add** and select **local ionization potential map** from the menu. Then request a LUMO map. *Click* on **Add** and select **|LUMO|map** from the menu. Surface generation is automatic and will take only a few seconds to complete.

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. Deselect **Coupled** from the **Model** menu () to uncouple the motions of the molecules. Position the two molecules side by side on screen. Close the spreadsheet.

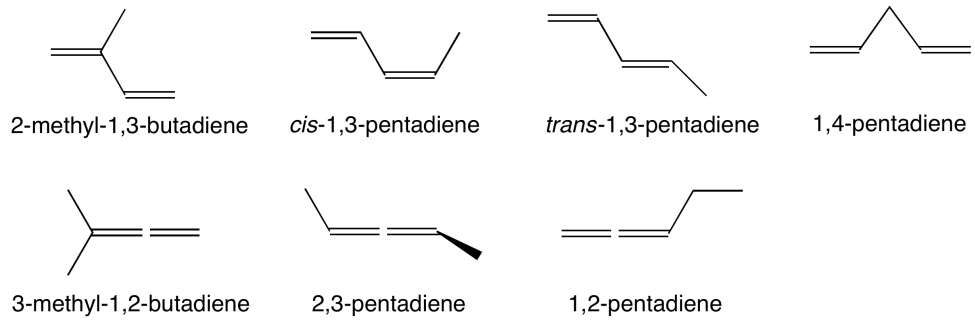
5. Inside the **Surfaces** dialog, select ***local ionization potential map***. Compare local ionization potential maps for the olefin and silaolefin, recognizing that the more red the color, the lower the ionization potential and the more susceptible toward electrophilic attack. Which molecule do you 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.

6. Close all documents and any open dialogs.

Isomeric C_5H_8 Dienes

In addition to 2-methyl-1,3-butadiene, *cis* and *trans*-1,3-pentadiene (conjugated dienes) and 1,4-pentadiene (a non-conjugated diene), there are three other dienes with the same formula: 3-methyl-1,2-butadiene, 2,3-pentadiene, and 1,2-pentadiene.



1. Build or sketch all seven C_5H_8 dienes (place in a single document). *Click* on the name of whichever diene is selected at the bottom of the screen. Make certain that ωB97X-D/6-31G* is selected, *click* on **Replace** and finally *click* on **All** from the dialog that results. This will substitute entries in SSPD for all seven C_5H_8 dienes. Select **Spreadsheet** from the **Display** menu ( ) and expand the spreadsheet to see all seven dienes. Select **Properties** from the **Display** menu ( ) and make sure that the **Molecule** tab in the resulting dialog is selected. *Click* on the buttons to the left of both **Expt. Heat** and **T1 Heat** in the **Molecule Properties** dialog. This enters the experimental heat of formation (from the NIST database) as well as the value obtained from the T1 thermochemical recipe into the spreadsheet. The T1 heat of formation will generally provide a more accurate account of relative stability than available from the ωB97X-D/6-31G* model itself.

The T1 model is presently defined only for uncharged, closed-shell molecules comprising H, C, N, O, F, Si, P, S, Cl and Br only. T1 heats of formation are not available for the other 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. Close the **Databases** window. Bring up the spreadsheet (**Spreadsheet** from the **Display** menu ( ). Note that not all the experimental heats of formation are available. In the NIST database, experimental heats of formation exist for only 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.

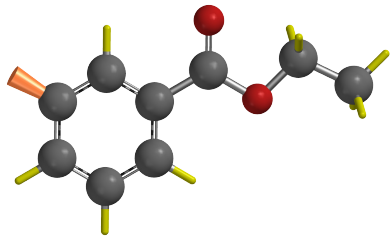
Using Infrared Spectroscopy to Identify an Unknown Ethyl Benzoate Derivative

Unlike NMR spectra (at least ^{13}C spectra) infrared (IR) spectra are rich in detail. Some chemists would view this as a negative, in that the complexity of an infrared spectrum masks its connection with the three-dimensional molecular structure. Others might see complexity as a positive, providing a detailed pattern or fingerprint with which to identify a molecule. This reflects the different ways that chemists actually use NMR and infrared spectroscopy. NMR is most commonly employed to assign structures of new molecules whereas IR is most commonly used to identify previously characterized (known) molecules.

Extensive databases of experimental infrared spectra exist, and sophisticated programs have been developed to search these databases for matches to unknown spectra. ***Spartan*** is able to search the public NIST database of experimental infrared spectra. Infrared spectra calculated from density functional models and adjusted to account for known systematic errors in frequencies and for finite temperature, generally closely match experimental spectra. This suggests that they might be used in lieu of experimental spectra to identify unknowns. This tutorial illustrates the point, attempting to identify which of a series of *meta*-substituted ethyl benzoates contained in the Spartan Infrared Database (SIRD) best matches the infrared spectrum of an unknown molecule.

> SIRD uses the infrared spectra contained in the Spartan Spectra and Properties Database (SSPD) that have been obtained using the EDF2/6-31G* model. Infrared spectra from the ωB97X-D/6-31G* model are not presently available in SSPD. The fact that there are two different tabs simply reflects the different search paradigms (spectra matching for SIRD vs. substructure matching for SSPD).

1. Build or sketch ethyl benzoate. Select **Structure Query** from the **Search** menu () and *click* on one of the free valences on the phenyl ring that is *meta* to the ester group.



2. Select **Databases** from the **Search** menu ( ) and *click* on the **SIRD** tab at the top of the dialog that results. Size the dialog to occupy the greater part of the screen. *Click* on **Select Spectrum**, move to the ***using SSPD*** subdirectory under the ***tutorials*** directory*, *click* on ***unknown ethyl benzoate*** and *click* **Open**. In a few seconds, the experimental infrared spectrum of the unknown will appear at the top right of the dialog. A list of the frequencies for all peaks may be found under **Unknown:** immediately to the left is the spectrum. This has been obtained by fitting the experimental spectrum to a Lorentzian function. You can obtain the frequency of a line in the spectrum using the yellow measurement bar. Position the cursor on top of the bar, hold the left button and move the mouse left or right until the bar is over the peak.

3. You can carry out your search with nothing other than the unknown spectrum (see the optional part of this tutorial), or you can restrict it using structural or other information that you have. Assume that you know that the unknown is a *meta*-substituted ethyl benzoate. *Click* on the **Filters** button and then *click* on **Copy Current Molecule** at the top left of the **Search Filters** dialog that results. *Click* on **OK** to exit the dialog. You have restricted the search to *meta*-substituted ethyl benzoates, that is, examine only molecules that contain this particular substructure (ethyl benzoate substituted in the *meta* position). *Click* on the **Search** button and wait for the search to complete. Hits resulting from the search appear at the bottom of the **SIRD** dialog, in

* For Windows, this directory is found in ***Program Files/Wavefunction/Spartan20***. It must be copied to another location available to the user prior to opening it in ***Spartan***. For Linux, this is found in the directory where ***Spartan*** was installed. For Macintosh, this is located at the top level of the ***Spartan'20*** disc image.

numerical order starting from the best (lowest **Score**). As you *click* on each, the corresponding calculated infrared spectrum will be superimposed onto the experimental spectrum (of the unknown) at the top left of the display. Propose a structure for the unknown.

4. Remove the substructure filter and repeat the search. *Click* on the **Filters** button, *click* on the **Clear** button at the top left of the **Search Filters** dialog and *click* on **OK**. *Click* on the **Search** button. Unlike searches of experimental databases where spectra are directly compared, each of the calculated spectra in the database are adjusted for frequency scale and peak width as it is compared to the unknown. Removing the substructure filter greatly increases the number of spectra that need to be examined. Does the *meta* substituted ethyl benzoate that came in at the top in the previous search, also come in at or very near the top in the full search?

> 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.

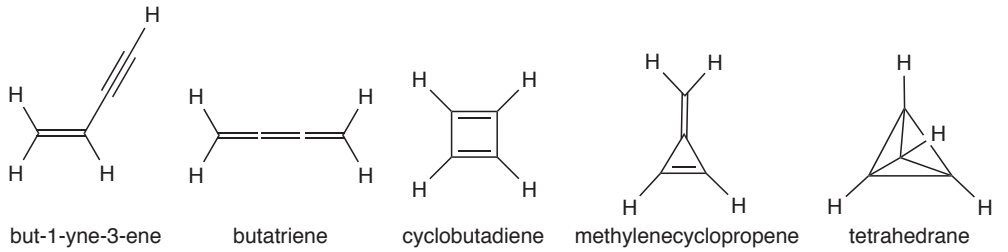
5. Close all documents and any open dialogs.

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 cm^{-1}. Possible structures include but-1-yne-3-ene, butatriene, cyclobutadiene, methylenecyclopropene and tetrahedrane.



1. Build or sketch any one of these molecules. Select **Databases** from the **Search** menu ( ) and *click* on the **SSPD** tab. *Click* on to the right of the **Search** button, ensure that **EDF2/6-31G*** (only) is selected under **Model Filter** and 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. Close the **Databases** dialog.

2. Select **Spectra** from the **Display** menu ( ) and *click* on at the top of the resulting dialog. *Click* on IR Calculated (calculated IR spectrum). Move through the list to bring up the spectra for the other isomers. Which, if any, calculated infrared spectrum of the C_4H_4 isomers best fits the unknown infrared spectrum (with peaks at 854,1608,2994 and 3080 cm^{-1})?

3. Close all documents and any open dialogs.

Using ^{13}C NMR to Distinguish Structural Isomers

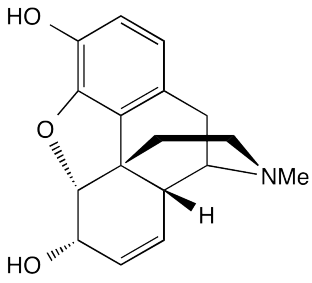
NMR spectroscopy, in particular ^{13}C NMR spectroscopy, is arguably the single most important tool available to an organic chemist to establish molecular structure. Obtaining an NMR spectrum is straightforward and the spectrum itself is very simple, comprising but a single line for each unique carbon the molecule. Because a mass spectrum is normally also available, the molecular formula will be known, and assignment of the NMR supports deciding among possible isomers. Chemical evidence (how the molecule was made) can usually be counted on to eliminate some choices and to strengthen the case for others. Still, pinning an NMR spectrum to a particular molecule can be difficult and fraught with error.

NMR assignment problems can be reduced but not altogether eliminated by requiring that both proton and ^{13}C NMR spectra are consistent with a particular structure. This is routine practice. In principal, ambiguity can be eliminated by cross-correlating the results of the ^{13}C NMR with those of proton NMR with an HMBC spectrum.

The fact that NMR spectra for reasonable size organic molecules may now be routinely calculated in a few minutes to a few hours on a personal computer costing $1000 or less raises the possibility for another tool to assist with spectral assignments. Whether this becomes common practice ultimately depends on the ability of calculations to obtain NMR spectra that are sufficiently accurate to distinguish among the different isomers.

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

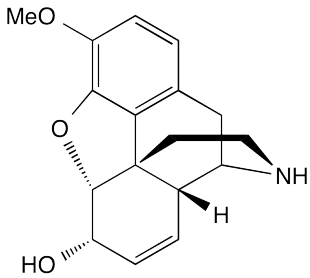
1. Sketch morphine. *Click* on its name at the bottom of the screen and *click* on **Replace**. If the name fails to appear, then you have made a mistake. In this case, select **Edit Sketch** from the **Build** menu () and correct your sketch.



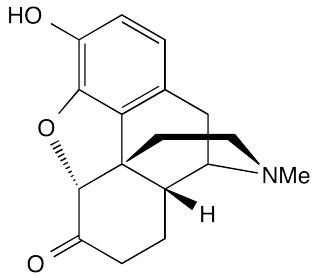
morphine

2. Select **Spectra** from the **Display** menu ( ) and *click* on in the bar at the top of the plots pane. Select from the palette. *Click* again on and this time select from the palette. Calculated and experimental ^{13}C spectra are now superimposed. Focus your attention on the four "most isolated" resonances, specifically those with calculated chemical shifts of 20.3 ppm, 59.7 ppm, 66.2 ppm and 95.2 ppm.

3. Sketch norcodeine and hydromorphone and retrieve both molecules from SSPD.



norcodeine



hydromorphone

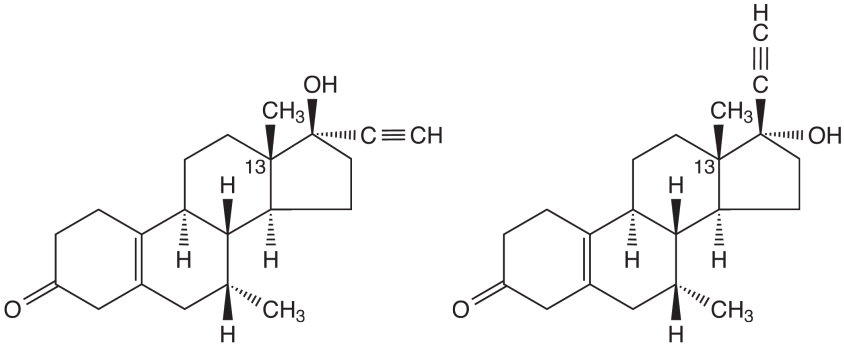
4. Select **Spectra** from the **Display** menu ( ) and *click* on and select from the palette. You only need to do this once if you put norcodeine and hydromorphine in the same document. Does either spectrum fit the experimental ^{13}C for morphine (again focus on the four isolated lines)?

5. Close all documents and any open dialogs.

Using ^{13}C NMR Spectra to Distinguish Stereoisomers

As mentioned in the previous tutorial, the combination of 1D and 2D NMR spectra is almost always able to provide unambiguous assignment of molecular structure, including assignment of stereochemistry. However, as noted, practical considerations (time on an expensive instrument) often make 2D experiments (in particular HMBC) the exception rather than the rule. Without the additional experiments, distinguishing stereoisomers based on their ^{13}C NMR spectra may be difficult experimentally, simply because the alternatives are likely to be structurally very similar. For the same reason, it is likely to be an ideal case for calculations, as comparisons between molecules with similar structures should benefit from cancellation of errors. The bottom-line question is whether or not calculated ^{13}C NMR spectra are good enough to be able to clearly distinguish between stereoisomers. This tutorial provides an illustration.

The hydroxyl CO bond in tibolone, shown below on the left, is *gauche* to the methyl group at C_{17} (the alkyne is *anti*), whereas it is *anti* (and the alkyne is *gauche*) in the stereoisomer shown on the right. You will use just three lines from the experimental ^{13}C spectrum to see if the calculations support or refute its assignment as tibolone.



1. Sketch both stereoisomers and put into the same document. *Click* on the name of whichever molecule is selected at the bottom of the screen, *click* on **Replace** in the dialog that results and *click* on **All**. NMR spectra for both molecules are now available.

2. Focus on tibolone and **Expt. Chem. Shifts** [H/C] from the **Expt. Data** menu. One after the other, enter the experimental chemical shifts for the methyl group at C_{13} (12.7 ppm) and for

the internal and external alkyne carbons (88.9 and 74.8 ppm, respectively).

2. Select **Properties** from the **Display** menu (), and *click* on the NMR tab. Which isomer provides the better agreement in terms of both RMS, mean and maximum absolute error scores as well as the DP4 score? Do the calculations support or refute the experimental assignment, or are they ambiguous?

4. Close the document and any open dialogs.