

Electronic structure modeling

Peter Pulay

Department of Chemistry and Biochemistry, University of Arkansas

2023 INBRE Conference

Workshop on Molecular Modeling

Nov. 4, 2023 10:30-11:30

Plan of the workshop

The workshop leader (Dr. Peter Pulay, Department of Chemistry and Biochemistry, University of Arkansas) will present exercises to demonstrate the power and utility of computer simulation of molecules. Participants will run the calculations on a university server which is accessed from Windows PCs via an Internet browser. Any number of participants can join and listen but only 9 groups can actively participate, that is, running the programs because of the limited capacity of the server. For the same reason, only single processor jobs can be run under the current setup (normally ~10 processing cores are used but the scheduler of the Computer Center may delay the jobs which request more than one processing core, and we do not have much time).

We plan to run the workshop in the computer lab at the computer lab in Gearhart Hall. These are only front ends; the program will run on the server. However, the same program can be downloaded **free** to the participant's PC (Windows, Mac or Linux) from Dr. Feng Wang's web site. Instructions for this are given below.

We will be using the PQS suite of programs, developed originally in Dr. Pulay's laboratory, mainly because it is free. There are a number of programs for quantum chemical simulation of molecules, and they are more or less the same, differing somewhat in efficiency, capabilities, and parallelization. The PQS suite was developed specifically for parallel computation of PC-like units. However, as explained above, we will be using the single processor ("serial") version. An important component of modeling is a Graphical User Interface (UI). Without it, it is very tedious to construct a computer model of a molecule. Equally important is a visualization of the computer results. PQS includes an efficient GUI, PQSMol, and a visualizer, PQSView. The version we use runs under the Linux operating system. Knowledge of Linux is not necessary to use the programs, although it is useful. Basic Linux can be learned in about an hour.

The address of the cloud server is <https://hpc-portal2.hpc.uark.edu>
We have created 10 user accounts: inbre0, inbre1, ... inbre9. The first one is for the workshop leader.

Their passwords will be communicated to the participants in the presentation.

The server can be accessed by a browser at the above Web address. Log in with the inbre credentials, then in the top menu go to "Interactive Apps->PQS" . Please select "comp01" and 1 hour for the job options. Those jobs should start almost immediately. It is useful to open a Linux terminal app as well to manipulate the files if needed.

These programs will also run on your Windows or Mac PC as a Virtual Machine (VM) running Linux. Instructions for this can be found on Dr. Feng Wang's Web site:

<https://wanglab.uark.edu>, click on "Modeling VM". This code can be downloaded to your computer but this requires more time than we have in this workshop.

I. Introduction: Computer modeling of molecules

Quantum mechanics can, in principle, predict the properties of all chemical systems. However, the basic Schrödinger equation, although simple enough, is very difficult to solve accurately. It is easy to get a total energy accurate to, say, 1%. However, such an accuracy is useless for chemistry. Chemistry is the science of small energy differences.

Constructing a qualitatively correct initial molecular geometry

A modeling method generally yields an approximate energy for the system in question. This depends on the electronic state and the molecular geometry, i.e., the relative positions of the nuclei. In most cases, the lowest electronic state, the *ground state*, is of interest. The lowest energy in a given state and a given qualitative structure gives the *equilibrium geometry* of the system. We consider different isomers (even conformational isomers) different systems. Modeling usually starts with a qualitatively correct molecular geometry, i.e., with the (x,y,z) coordinates of the atoms. Generating a qualitatively correct 3-dimensional structure is more tedious than it appears. Back in the dark ages (in the 70s, and 80s) people used geometry building programs, like the “Z matrix geometry” which allowed to build a geometry from bond lengths, angles and torsional (dihedral) angles. Although easier than trying to determine the Cartesian coordinates by hand, this is still a tedious and error-prone procedure.

Modern computational chemistry programs use a Graphical User Interface (GUI) to construct an approximate geometry. This is followed by optimizing the geometry using a more sophisticated method. A lot of effort was spent in the past 50 or so years to develop computer programs for this. The program we use here is called PQS, of Parallel Quantum Solutions, LLC. It has all standard methods of quantum chemistry for ground electronic states.. There are a number of such programs available; we use PQS mainly because it is free for non-commercial purposes, and has an intuitive GUI. The version we run here is the Linux version. Linux is the operating system generally used in large-scale computation. It is not absolutely necessary to understand Linux for this purpose but it can be useful if something fails and needs manual intervention.

Major levels of approximation: Force Fields, Semiempirical, wavefunction, and DFT

It is useful to distinguish three main levels of approximation in modeling. The first (after the initial graphical construction) is a variant of the Molecular Mechanics (MM), or Force Field method. This technique uses transferable molecular geometry parameters (bond lengths, angles, torsions) to construct an approximate energy expression. MM, as a rule, does not consider electrons explicitly, and therefore it is very fast. It can be very accurate for molecules it has been parametrized for (generally common organics). However, it can be widely off for unusual atoms or bonding situations, for ions, or less common elements. As it uses only local information, MM struggles to describe large-

scale conjugation. They can be used to systems consisting of many thousands of atoms, like biological macromolecules.

The next two levels aim at solving the molecular Schrödinger equation approximately. Semiempirical methods set up the Schrödinger equation formally. However, instead of calculating the quantities appearing in the equations, they are adjusted to fit known molecular properties. These were originally experimental data (relative energies, bond length and angles, dipole moments, etc.) but nowadays are mostly high-quality theoretically calculated values. (The same applies to modern MM methods as well). Semiempirical methods are more computationally demanding than Molecular Mechanics but still quite efficient computationally. However, their accuracy is limited. They appeared to vanish at one time, pushed by force field methods from below and efficient *ab initio* and DFT methods from above but they survived, mainly because they can treat long-range conjugation in graphenes, fullerenes, and biological molecules.

The most accurate but also the most expensive methods are *ab initio* and density functional (DFT) methods. *Ab initio* means “from the beginning” in Latin. It is not a single method but rather a palette of methods of increasing sophistication. At the limit, these methods can match or even surpass experimental accuracy. However, this comes at considerable computational cost. The simplest *ab initio* method is Hartree-Fock theory (1928-30). It accounts for over 99% of the molecular energy theoretically. It is computationally quite efficient because it replaces the instantaneous electron-electron repulsion by the average repulsion between charge clouds which is much easier to calculate. Unfortunately, the error of this approximation, called *correlation energy*, although less than 1% of the total energy, is quite important for chemistry. Evaluating the correlation energy in an *ab initio* fashion makes the calculations rather expensive. The simplest such method is MP2, or Second Order Moller-Plesset Perturbation Theory.

Density Functional Theory (DFT) is close to the *ab initio* Hartree-Fock theory in its techniques. However, the correlation energy is evaluated by what is essentially a semiempirical approach (although the physicists who invented it mostly deny this). As the approximation affects only a small percentage of the energy, DFT combines the efficiency of Hartree-Fock theory with the accuracy of more elaborate and expensive methods, and it is the method of choice in most accurate modeling, at least initially. It is not the method to use where absolute certainty is required but it is certainly the first method to try. DFT includes exactly the main terms in the Schrödinger equation: the kinetic energy of the electrons, the electron-nucleus attraction, and the classical part of electron-electron repulsion. However, it approximates some remaining terms by a simplified quantity called the exchange-correlation (XC) functional of the electron density. The exact XC functional is unknown and is most likely unknowable, or at least so complicated that it is not useful in practice. Hundreds of exchange-correlation functionals have been proposed, each better than all the others if you believe their authors. A large number of functionals are available in the PQS program suite but we

will use only a few well-tested ones, mainly B3LYP which is the most widely used functional.

Ab initio methods (including DFT which is sometimes called a “first principles” method but has significant empirical contribution) must describe molecular orbitals (MOs). These cannot be described (except for H-like atoms and ions) by a simple formula. Instead, we expand them in a set of *basis functions*, for example the *i*-th molecular orbital φ_i is approximated as

$$\varphi_i = C_{i,1} \chi_1 + C_{i,2} \chi_2 + C_{i,3} \chi_3 + \dots + C_{i,M} \chi_M \quad (1)$$

Here the C's are numbers forming a rectangular table, $C_{i,k}$ being in the *i*-th row and the *k*-th column, and $\chi_1, \chi_2, \dots, \chi_M$ are the basis functions. They were originally called atomic orbitals (AOs) – obviously, if we hope to describe molecular orbitals (MOs), we must be able to describe atomic orbitals (AOs). However, the χ 's don't have to be accurate atomic orbitals, the only requirement being that a number of them should approximate the MOs well. They are somewhat arbitrary, and have little physical meaning by themselves. They are defined by the user and the computer program. Equation (1) simply says that we try to represent the unknown MOs as sums of known functions (the basis functions) with unknown coefficients (the C's). Such series expansions are called *Linear Combinations*, and are widely used in physics and engineering. For instance measured functions are often approximated by polynomials, $f(x) \approx a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n$. Periodic functions are often approximated by trigonometric functions:

$$f(t) = a_0 + a_1 \cos(\omega t) + b_1 \sin(\omega t) + a_2 \cos(2\omega t) + b_2 \sin(2\omega t) + a_3 \cos(3\omega t) + b_3 \sin(3\omega t) + \dots$$

In these examples, the functions $1, x, x^2, x^3 \dots$ or $1, \cos(\omega t), \sin(\omega t), \cos(2\omega t), \sin(2\omega t)$ are the basis functions, and $a_0, a_1, a_2, \dots, b_1, b_2, \dots$ are the coefficients.

The basis functions most widely used are atomic-like. They have a *radial* part which depends on the distance from the nucleus the basis function is assigned to, and an *angular* part which is generally the same as the *s, p, d* etc. orbitals in the H atom. The radial part in the H atom can be described exactly by a few terms of the form $P(r)e^{-\zeta r} = P(r)\exp(-\zeta r)$ where $P(r)$ is a polynomial of r , the distance from the nucleus, if the exponent ζ is chosen correctly. (The notation $\exp(a) = e^a$ is used to avoid expressions with multiple levels). Therefore functions of the form $r^n e^{-\zeta r}$ are presumably the best basis functions for molecules. However, the calculation of the necessary integrals with such functions is very time-consuming. For this reason, radial functions of the form $\exp(-ar^2)$ are generally used. These correspond to the familiar Gaussian bell curve. They are less similar to atomic functions, and therefore we have to use more of them. It is not even necessary to use atomic-like basis functions. Materials scientists and physicists often use plane waves – sine and cosine functions – for describing crystals. These are definitely not atomic-like. To approximate the MOs, many more plane waves than AOs

are needed but the special properties of plane waves allow important computational simplifications.

A number of groups have developed sets of basis functions of increasing sophistication. The bigger they are, the more accurate is the calculation; results converge with increasing basis set quality. However, the basis sets are only one piece of approximation to the wavefunction, the other being the description of correlation. The two approximations should be balanced. A very large basis at the Hartree-Fock level is not accurate, nor is a high-level correlation treatment like “Coupled Cluster with single and double, and perturbational triple substitutions [CCSD(T)]” with a small basis set. CCSD(T) is the highest level of correlation treatment routinely applied. The Pople group has developed a number of popular basis sets; in order of increasing accuracy (and computational cost) these are 3-21G, 6-31G*, 6-311G(d,p), etc. The basis sets developed by R. Ahlrichs’ group in Germany: def2-SVP, def2-tzvp, etc. enjoy increased popularity. The rule about basis sets is “the bigger, the better”. However, large basis sets increase the computational expense steeply, and the larger basis sets are already quite close to convergence – further improvements may not make significant changes.

In Hartree-Fock and Density Functional theories only the occupied molecular orbitals contribute to the energy of the system, although unoccupied (“virtual”) orbitals are obtained as a by-product of the calculations. In most systems, the electrons occupy the MOs pairwise with opposite spins, and thus the number of occupied MOs in an even-electron system is $N/2$, half the number of electrons. In odd-electron systems, the MOs of alpha and beta electrons generally differ but they are almost identical for the lower energy orbitals, i.e., these are essentially doubly occupied in most cases.

In the ground state of a system, the lowest energy MOs are usually occupied (the build-up or Aufbau principle). There are, however, exceptions, when the energy difference between the highest occupied MO (HOMO), and the lowest unoccupied MO (LUMO) is small. This happens mostly in transition metal compounds..

The initial geometry, the choice of the method (say, Hartree-Fock, MP2, CCSD(T), DFT with a given functional, and the basis set) determine the calculation for the ground state of a system. The coefficients of the basis function in the molecular orbitals are determined by an iterative procedure if there is more than a single electron. This is necessary because the orbitals influence each other through the mutual repulsion of the electrons. The scheme for this was originally invented by D. R. Hartree in 1928. In a nutshell, the orbitals are optimized (i.e., the energy they yield is minimized) in the electric field generated by the nuclei and the other orbitals (which are considered fixed). This is not yet the final solution for any of the orbitals because the other orbitals change, and they influence the orbital we consider. However, the new set of orbitals is generally closer to the final solution than the original set. Iterating this method until the electric field of the system does not change yields the final solution. The iteration process is called “Self-Consistent Field”, or SCF, and converges in about a dozen or so cycles.

The original SCF procedure may diverge or converge slowly; improvements, in particular a technique called DIIS, have made it practically foolproof.

The general plan of modeling is to construct a graphical model that has qualitatively correct geometry, calculating the wavefunction, and optimize the geometry by searching for the lowest energy. We can determine molecular properties at the optimum geometry. Optimizing the geometry requires the calculation of the forces on the atoms. (In principle, the equilibrium geometry can be calculated from the energy alone, varying the geometry until the lowest energy is obtained. However, this is computationally so expensive and tedious that it is practical for only two, or at most three atoms). The ability to calculate forces was a major advance in computational chemistry. At equilibrium, the forces should be zero but at an arbitrary geometry they are not. Geometry optimization has a large literature but it is essentially similar to the optimization of the wavefunction. It is an iterative procedure: the atoms are moved in the direction of the forces acting on them until the latter become insignificant and the energy stabilized. We have thus a double iterative procedure. The outer loop is the optimization of the geometry; for each new geometry, the wavefunction is optimized. Fortunately, the computer program does this without manual intervention.

Once we have a wavefunction, various properties can be calculated. The most important property is the energy. However, a warning is in order. The only consistent energy zero point is the state in which a molecule is torn apart to a collection of nuclei and electrons, all very far from each other. Needless to say, it would require a minor nuclear explosion to accomplish this, and the energies of molecules are much lower than this hypothetical zero energy. The calculated total energies in *ab initio* and DFT methods are thus very large negative numbers, and by themselves are largely irrelevant. Only *energy differences* between various electronic states or geometries of the system are significant.

The energies of the molecular orbitals yield information (although not a very accurate one) about the energies needed to remove an electron from an occupied MO. These are the ionization energies (the smallest of this is the energy of the HOMO, and "ionization energy" alone generally refers to this first ionization energy. Similarly the energies of the virtual (unoccupied) MOs, if negative, give an idea of the energy liberated when the system binds an additional electron (electron affinity).

Another important property is the equilibrium geometry (for isomers and conformational isomers geometries). Further important properties are the rigidity of the system. This determines the natural vibrational frequencies and the infrared and Raman spectra. Interaction with external electric and magnetic fields provides a further set of properties. Interaction with external electric fields can be characterized by electric dipole, quadrupole and higher moments, and polarizabilities. Most molecules do not have permanent magnetic moments, although this is important for molecular magnets and many transition metal compounds. More important is the calculation of nuclear magnetic resonance spectra.

Interpretation of molecular geometries

It is tricky to find simple arguments to reproduce the results of sophisticated calculations. **Localized molecular orbitals** come closest to this goal. They are still not included in most physical chemistry curricula. In my opinion, the reason is that they were not known to early quantum chemists and were not included in the first textbooks on the subject (L. Pauling and E. B. Wilson, Jr., "Introduction to Quantum Mechanics with Applications to Chemistry 1935; H. Hellmann, Einführung in die Quantenchemie, 1937) from which most textbooks on the subjects were developed. It is based on a discovery by V. A. Fock that molecular orbitals with the same occupation number (most often doubly occupied) can be freely combined with each other, say

$$\varphi'_i = a\varphi_i + b\varphi_j \quad \varphi'_j = c\varphi_i + d\varphi_j$$

without changing the physical wavefunction if it is properly normalized (for computational simplifications, some conditions on the coefficients a, b, c, d are generally imposed). This shows that molecular orbitals are in general not physically existent quantities, rather a product of our effort to carve up an n -electron wavefunction to more easily understandable one-electron parts. This can be done in many different ways. The usual molecular orbitals obtained from the SCF procedure have well-defined orbital energies but are usually delocalized all over the molecule. They are called *canonical* orbitals (meaning regular or lawful). Canonical orbitals do not change qualitatively upon ionization, electron attachment, or excitation, and are therefore useful to treat these phenomena. However, they do not reflect the chemist's view of bonds and lone pairs (although they separate the core atomic orbitals from the valence ones), and are not transferable between chemical homologs (propane and butane have very different canonical MOs). Attempts to explain molecular geometries through canonical MOs, known as hybridization theories, were notably unsuccessful. (Pauling, who originated this, had such a penetrating insight and encyclopedic knowledge of chemistry that even a wrong theory worked for him. However, few of us have Pauling's insight and knowledge.)

Localized MOs are linear combination (as in the above equation) of canonical MOs which occupy as little space as possible. There are several localization criteria in use. One can minimize the average number of atoms the MOs are distributed on (Pipek-Mezey localization), or minimize the sum of the squared radii of the orbitals (Boys localization). The energies of localized MOs are not sharply defined, although one can assign an average energy. They change qualitatively upon excitation and ionization, and are not useful to treat these phenomena. However, they are transferable between chemical homologs, and conform to the chemist's view of core, bonding, and lone pair orbitals. They are useful to understand molecular geometries. Note that unpaired electrons in free radicals and excited states are generally genuinely delocalized and cannot be well localized.

Localized MOs act as ligands around a central atom. Valence Shell Electron Pair Repulsion theory (VSEPR theory) can account for subtle features in molecular geometries by using plausible assumptions about the repulsion between localized MOs.

II. Spin states and multiplicities. Is ground-state methylene singlet or triplet? What about CF₂?

1. General comments

In most stable molecules, the electrons occupy the orbitals in pairs, with opposite spins α and β . This gives rise to the misconception that spin pairing is somehow energetically advantageous. However, the exact opposite is the case. If two electrons occupy two *different* orbitals, then the parallel spin arrangement $\alpha\alpha$ or $\beta\beta$ is generally the lowest energy state, and the opposite spin $\alpha\beta$ is higher in energy (Hund's rule; it was deduced from the analysis of atomic spectra). The parallel spin state is called the *triplet* and the opposite spin state is the *singlet*.¹ If there is one singly occupied orbital, the state is a *doublet*. The *multiplicity* is $2S+1$ where S is the total spin quantum number. The name comes from spectroscopy, since an atomic state is generally split into $2S+1$ closely spaced levels if its total spin is S . This manifests itself in splitting the spectral lines. For instance, the well-known yellow D line of Na is actually a doublet, i.e. two closely spaced lines (at 589.0 and 589.6 nm).

If parallel spins are energetically more advantageous, why do most electrons occupy the orbitals with opposite spins? The answer is the Pauli principle which disallows two electrons occupying the same orbital with the same spin². If the first electron occupies the lowest unfilled orbital, the second electrons must go to a different, generally higher

¹ This is an oversimplification of the situation. In reality, while $\alpha(1)\alpha(2)$ and $\beta(1)\beta(2)$ are two components of the triplet wavefunction, differing in the direction of the total spin z component ($S_z=+1$ or -1 in \hbar units), $\alpha\beta$ is neither singlet nor triplet. Correctly, $2^{-1/2}[\alpha(1)\beta(2) - \alpha(2)\beta(1)]$ is the singlet and $2^{-1/2}[\alpha(1)\beta(2) + \alpha(2)\beta(1)]$ is the third component of the triplet wavefunction. The S_z component of the spin angular momentum is $+1 \hbar$ for $\alpha\alpha$, $-1\hbar$ for $\beta\beta$, and $S_z=0$ for the third component. See Atkins, Physical Chemistry, 8th edition, p. 347.

The terms "singlet" and "triplet" go back to the early age of spectroscopy. There is no magnetic moment in the singlet state ($S=0$) but there is a net magnetic moment in the triplet. If there is a magnetic field (usually from the orbital motion of the electrons), the projection of the spin on the magnetic field (which is taken as the z direction) can take 3 values: $S_z = -1, 0, 1 \hbar$. The 3 components of the triplet have slightly different energies, and the triplet state is split in 3 sublevels. The corresponding spectroscopic transitions are also split in 3 closely lying lines. The singlet state has no magnetic moment and is not split by the magnetic field.

² This was Pauli's first formulation of his principle but it is, again, a simplification which is valid only in the orbital approximation. The accurate statement is that the wavefunction must be antisymmetric, i.e. change its sign if two electrons are interchanged. In the orbital approximation, interchanging two electrons in the same orbital with the same spin can have, on one hand, no effect on the wavefunction, yet antisymmetry requires a sign reversal. The only number that does not change if its sign is reversed is zero. Thus the only properly antisymmetric wavefunction is identically zero (i.e., zero for any positions of the electrons) in this case. Such a wavefunction contains no electrons and is unphysical. (We considered here 2 electrons but the conclusion is valid for any number of electrons because the wavefunction is essentially a *product* of orbitals which becomes zero if one of its factors is zero.)

energy orbital if its spin is the same as the first one. The orbital energy difference is often larger than the energy gain from parallel spins. However, if the two orbitals are degenerate (their energy is the same), the ground state will be a triplet. If the two orbitals are close but not degenerate then the situation depends on the energy difference. If this is less than the spin pairing energy loss, the ground state will be triplet, otherwise singlet.

2. Methylene in the singlet and triplet state

Methylene, CH₂ is an important intermediate, and the simplest member of the *carbene* family, those few, mostly transient molecules with a lone electron pair on carbon (note however CO, CN⁻ and some fairly recently discovered stable carbenes). Methylene is too reactive to produce in macroscopic quantities. The most interesting thing about it is that its ground state is a triplet, i.e. the two highest energy electrons have parallel spin. This obviously requires that these electrons occupy different orbitals, otherwise the Pauli principle would be violated.

(a) Construct methylene, CH₂, using PQSMol. (Make sure to use the *Unrestricted* mode: the program in the *Restricted* mode insists that the C atom has 4 valences). Make sure to use the bent template to make your initial geometry bent and not linear. (Starting with a linear geometry fixes the geometry as linear but methylene is bent.) Do not pre-optimize the geometry with the built-in force field. Force fields are not appropriate for exotic molecules like methylene. Use the 6-311G(d,p) basis in PQSMol input generator, and the B3LYP DFT functional, and the default singlet wavefunction (multiplicity=1). Call the input file e.g., ch2-singlet.inp. Make sure that you specify geometry optimization (in PQSMol, use the "Optimization" option. You may want to add "GEOM GEOP" as the final line of the input. This will calculate and print the bond length, angles and dihedral (torsion) angles (if any). Of course, the same may be had by opening the output file with PQSView (usually, simply double click it). Note the final energy for singlet methylene.

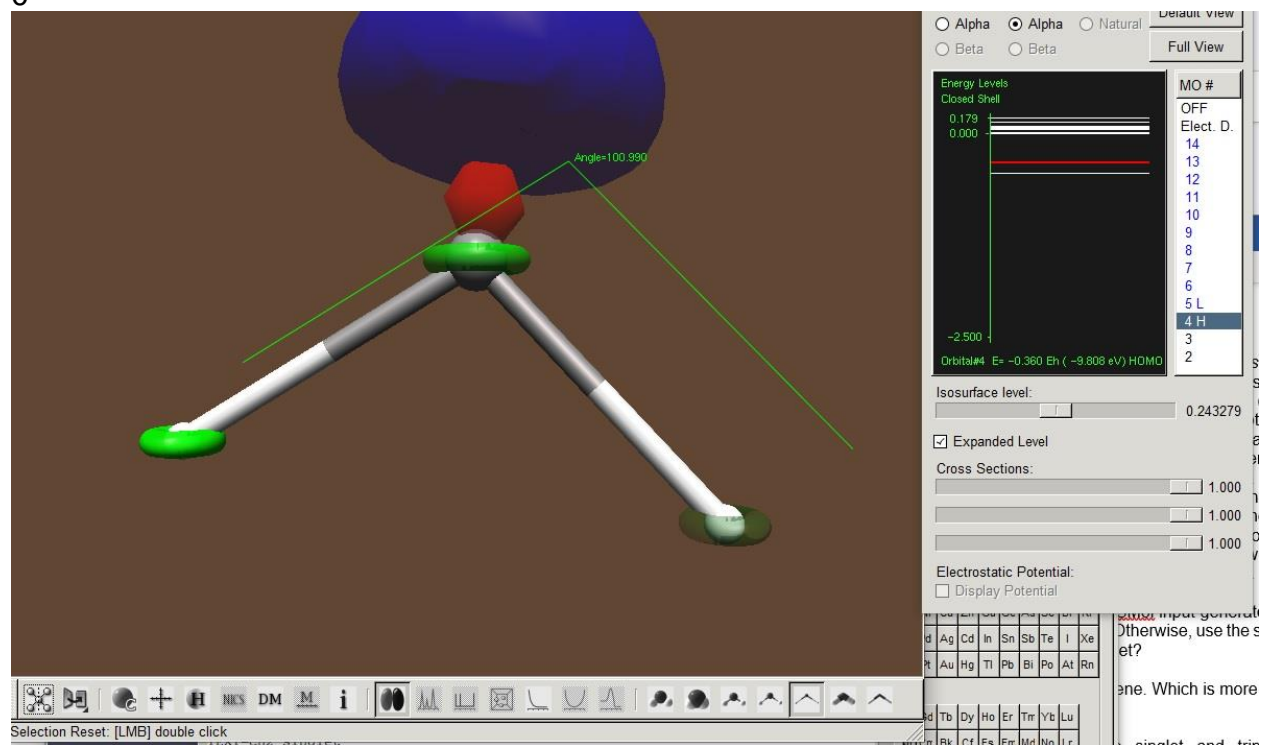
The input generator made the following input file:

```
TEXT=CH2 singlet
GEOM=PQB FILE=CH2-singlet.pqb
!!!!!!!!!!!!!! STEP 1 !!!!!!!!!!!!!!!
BASIS=6-311G(d,p)
OPTimize
SCF DFT=B3LYP LOCA=Pipek
FORCE
JUMP
GEOM GEOP
```

LOCA=Pipek and the last line were added by hand. The optimized total energy I got was -102773.5 kJ/mol or -24563.45 kcal/mol. These are huge (and largely irrelevant) numbers because the natural reference state is one in which the molecule is totally dissociated to electrons and atomic nuclei.

When the calculation finished, the visualizer popped up. I selected the bond angle (101.0°) and the highest energy localized occupied orbital which is a lone pair on the carbon.

0

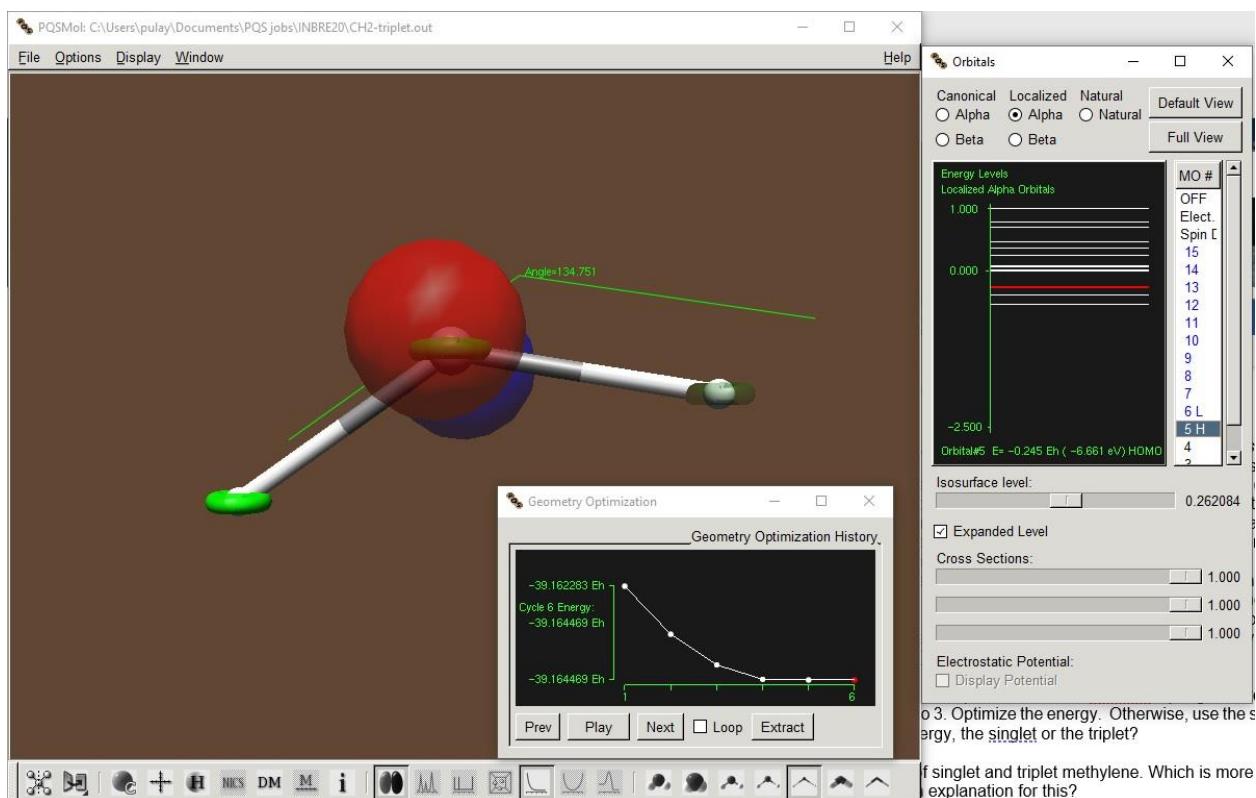


(b) Repeat the calculation for the triplet state. In the PQSMol input generator, we simply change the multiplicity from 1 to 3. Optimize the energy. Otherwise, use the same settings as in (a). Which is lower in energy, the singlet or the triplet? My input is shown below:

```
TEXT=CH2 triplet
GEOM=PQB FILE=CH2-triplet.pqb MULT=3
!!!!!!!!!!!!!! STEP 1 !!!!!!!!!!!!!!!
BASIS=6-311G(d,p)
OPTimize
SCF DFT=B3LYP LOCA=Pipek
FORCE
JUMP
GEOM GEOP
```

The visualizer output is shown below. This time I included the “Optimization history” graph, and the highest energy localized alpha orbital. The next lower α orbital is a lone pair. The two β orbitals are C-H bonds. The HCH angle in the triplet is 134.8° , and its energy at this level of theory is -102826.3 kJ/mol (-24576.08 kcal/mol). Thus the triplet is about 12.6 kcal/mol *below* the singlet. This is unusual and is caused by the fact that 00orbital energies of the HOMO and HOMO-1 in methylene are very close.

Why is the HCH angle more than 30° larger in the triplet than in the singlet? The reason, in full accordance with the Valence Shell Electron Repulsion (VSEPR) theory is that the triplet has only one electron in the lone pair localized orbital while the singlet has two. The energy of the π orbital in the triplet is largely independent of the in-plane valence angles.



(c) It is instructive to calculate the optimized energy for lowest the singlet and triplet states of difluoromethylene, CF_2 or dichloromethylene, CCl_2 . In these cases the singlet is more stable.

Interpretation of the output

Quantum chemistry calculations generally use *atomic units* in which the units of charge, mass, and action are the electron charge e , its mass m_e , and Planck's constant h . This makes the results independent of the (continuously revised) fundamental constants. They are, however, often not convenient. For instance, the atomic unit of energy, the Hartree (E_h) is very large; per mole, it is 2625.50 kJ/mol, or 627.509 kcal/mol. The program converts it to a number of units which are in use in chemistry, thermodynamics and spectroscopy, for instance kJ/mol, kcal/mol, $1 \text{ cm}^{-1} \cdot hc$, $1 \text{ Hz} \cdot h$, $1 \text{ K} \cdot k_B$. The atomic unit of distance, the Bohr (a_0) is $0.529177 \cdot 10^{-10} \text{ m} = 0.529177 \text{ \AA}$, has a reasonable magnitude for molecules. As emphasized in the introduction, calculated absolute energies are virtually useless, Relative energies (energy differences) are meaningful only if the two calculations use the same approximations, say the same basis and correlation treatment.

The experimental value of the singlet-triplet energy difference took a long time to be established with certainty but it is believed to be about 9.5 kcal/mol. This is a difficult quantity to calculate accurately.

III. Calculation of the infrared spectra and NMR chemical shifts

Introduction

Both vibrational (infrared and Raman) and NMR spectra are widely used to identify substances, particularly in organic chemistry. NMR has the advantage that the spectra can often be interpreted without sophisticated calculations because the main NMR phenomenon, the shielding of the external magnetic field by the electrons, is fairly local, depending mostly on the immediate neighborhood of the nucleus which is observed in NMR, except in molecules with extended conjugation. This results in a relatively simple correlation between structure and the NMR spectrum. Molecular vibrations, with the exception of a few characteristic vibrations (for instance X-H stretching vibrations, C=O and in some cases C=C stretchings, some vibrations of the amide group, out-of-plane bendings of the hydrogens in aromatic systems) are quite delocalized and there is no simple rule connecting the structure to the vibrational (infrared or Raman) spectrum. However, NMR is quite insensitive compared to infrared (IR) spectroscopy, and cannot be easily used to characterize, e.g., adsorbed species or other minute components. Modern electronic structure methods can be used to establish a connection between vibrational spectra and structure, facilitating the interpretation of vibrational spectra. This has resulted in a comeback of vibrational spectroscopy.

Distinguishing 2,3-dihydrofuran and 3,4-dihydrofuran by comparing experimental and calculated infrared spectra and NMR shieldings

Comments on the meaningful comparison of calculated and experimental vibrational spectra

We begin with some general comments about comparing calculated and experimental vibrational (infrared=IR and Raman) and NMR spectra. Raman spectra

are easy to measure these days but used to be difficult experimentally, and the sensitivity is still much worse than that of IR. Most spectral collections are still incomplete in the Raman domain. Thus we will concentrate on IR. The program can calculate both but the Raman spectrum is more expensive.

Compare the calculated infrared spectrum with experiment. Here are a few hints for a *meaningful* comparison. The IR spectrum consists of three separate regions. Vibrational wavenumbers below 400 cm^{-1} constitute the far infrared spectrum. These IR bands are generally rather weak (they get more intense as the frequency, or what is the same, the wavenumber, increases). Older IR spectrometers could not generally observe in this region, and many spectral collections don't show this part of the spectrum. The next region is between 400 cm^{-1} to 2000 cm^{-1} . This is the *fingerprint* region. It is highly characteristic of the compound. (The fingerprint region in older textbooks is the $700\text{-}2000\text{ cm}^{-1}$ region because the antiquated spectrometers used prisms of ordinary salt (NaCl) which transmits only above 700 cm^{-1}). For most compounds, the region between 1800 and 2700 cm^{-1} is essentially empty. Exceptions are acetylenes and allenes, and molecules containing Si-H, P-H and S-H bonds. Most X-H stretching vibrations, where X is a first-row atom (C,N,O), are in the $2800\text{-}3700\text{ cm}^{-1}$ range. The frequency is mainly determined by the bond length: the longer the bond, the lower the wavenumber. C-H stretchings are found around 3000 cm^{-1} .

It is best to compare the fingerprint region and the C-H stretch region separately. Indeed, many spectrometers use a different, more compressed scale above 2000 cm^{-1} . You may find experimental IR and NMR spectra in several online databases. One convenient database is in Japan:

https://sdb.sdb.aist.go.jp/sdb/cgi-bin/cre_index.cgi?lang=eng

The calculated spectrum does not match the real spectrum exactly because of the approximate nature of the wavefunction but mainly because we use the *harmonic approximation*. This model assumes that the molecular potential energy, as a function of the atomic displacements from the equilibrium geometry, is a quadratic function, i.e., it does not contain cubic, quartic, etc., terms. Applied to a spring, this is equivalent to Hooke's law, i.e., it postulates that the restoring force is

proportional to the displacement from equilibrium.³ This is a good approximation for small displacements but it is clear that all springs break sooner or later if stretched too much, and molecules are no exceptions. Indeed, if Hooke's law would hold all the way, the energy of a chemical bond would increase without limit as the atoms are pulled apart. We know that this does not happen: the energy increases for a while but then it goes into saturation, and does not change whether the atoms are 10 Å apart or 10 km. The neglected *anharmonic* terms are generally small but give rise to observable effects. First, they influence the frequencies. The observed frequencies are generally lower than the harmonic frequencies, particularly for bond stretchings. Second, infrared bands which have zero intensity in the harmonic model ("forbidden" in the language of spectroscopists) appear with weak but non-zero intensity. In the harmonic model, the only allowed transitions are in which one normal vibration changes its quantum number by ± 1 . In the presence of anharmonicity, transitions corresponding to jumps by 2,3,.. quanta (*overtones*) become weakly allowed. Also, two (or more) vibrations can change their quantum numbers simultaneously. For this reason, the observed IR spectrum usually contains extra weak bands, compared to the computed harmonic spectrum.

Experimental IR spectra are usually plotted with the wavenumber increasing to the left. This curious convention stems from plotting the absorption (or transmission) against the wavelength (and not the frequency or wavenumber). After the discovery of the quantization of light and the Planck formula for the energy of the photon, $E=h\nu$, people switched to plotting against the frequency ν , or, because the frequencies are huge, against wave number, ν/c where c is the speed of light. However, to preserve the approximate look of the spectra, the frequency axis increased to the left. The display in the current PQS plots the spectrum from left to right but the latest release plots it like most IR spectrometers, from right to left.

Agreement between observed and calculated spectra can be improved by *scaling* the calculated spectra. Using different scale factors for different types of motions (various stretches, bendings, torsions, etc.) essentially quantitative accuracy

³ The word "harmonic" derives from the fact that a musical instrument that has no anharmonicity will emit a pure sine wave sound. The "harmonic" designation is misleading because most people find a pure sine wave sound very unpleasant.

can be achieved. A less accurate but simpler method is to use a single scale factor. The scale factor corrects approximately for the effect of anharmonicity, and the deficiencies of the quantum chemical procedure (electron correlation and basis set defects). For B3LYP calculations, a scale factor of 0.97 often gives improved agreement with experiment. One simply multiplies the calculated frequencies by 0.97.

Calculation and comparison of the spectra

Construct one of these molecules using PQSMol, and optimize its geometry with the default force field, SYBYL, to get a reasonable first geometry. (One or more of your colleagues will construct the other one.) I started with an ethylene molecule (Fragments -> Chains -> Ethylene), connected an -O- group at one of the hydrogens, a tetrahedral C to the H on the other ethylene C the same side, and then closed the ring with another tetrahedral C. Click on Calculation -> New -> PQS Job. It will prompt you to save the build file (the default filename is the same for both isomers; it is best to save it as, say, 2,3-dihydrofuran in the Documents directory). This creates a build file 2,3-dihydrofuran.pqb and starts the construction of the input file (2,3-dihydrofuran.inp). **Do not use spaces in the file names.**

Select Geometry Optimization with DFT (the default exchange-correlation functional is B3LYP), and the 6-311G-d basis set. Select Vibrational Frequencies and NMR Chemical Shifts as properties to be calculated, and click Save. My input looked like this:

```
GEOM=PQB FILE=2,3-dihydrofuran.pqb
!!!!!!!!!!!! STEP 1 !!!!!!!!!!!!!!!
BASIS=6-311g-d
OPTimize
SCF DFT=B3LYP THREshold=6.0
FORCe
JUMP
NMR
HESS
FREQ
```

I also added, after the last line, the following line:

```
GEOM GEOP
```

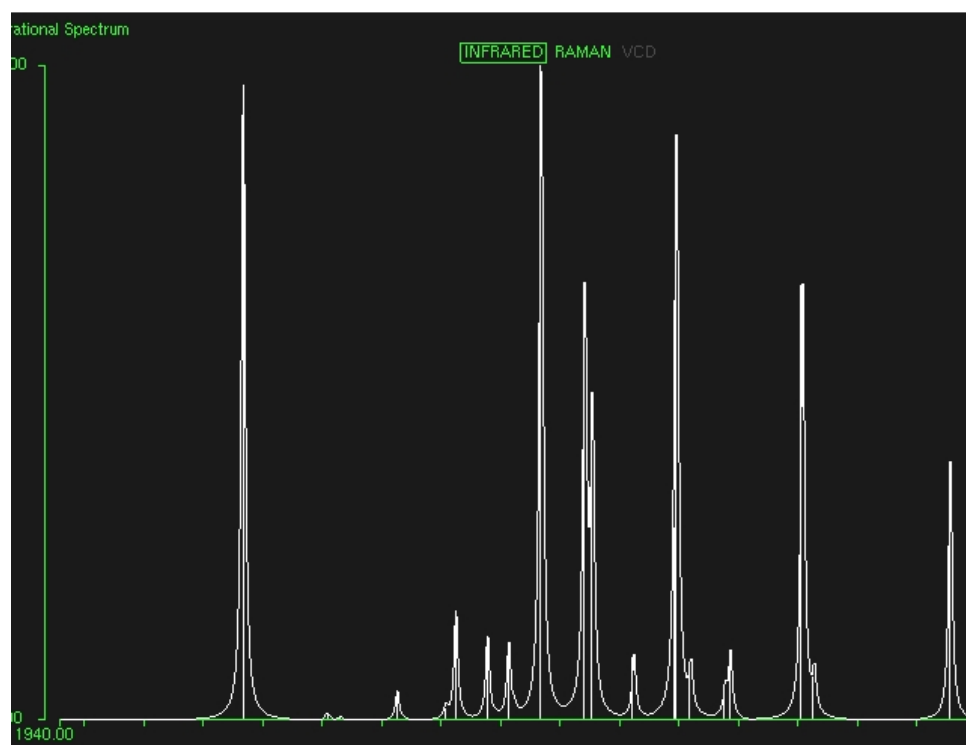
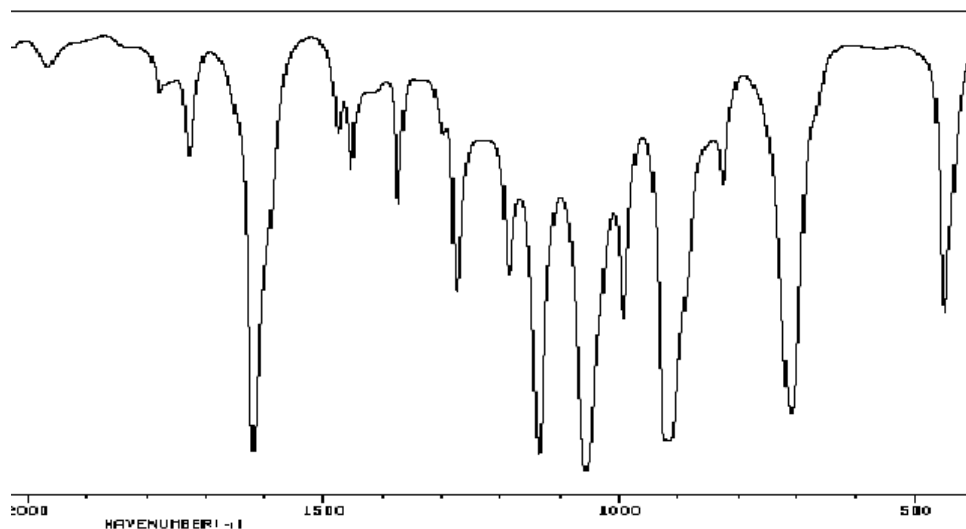
This makes the program to calculate the bond lengths, angles, and torsional angles. These can be read off from the GUI but it is often faster later to just look them up in the output.

Click Calculations. Click on 2,3-dihydrofuran.inp and then on "Submit Serial Job". The program should display the output file. When it has finished, it will ask you if you want to update the initial geometry in the build (*.pqb) file to the optimized geometry. It does not matter; I usually do not update.

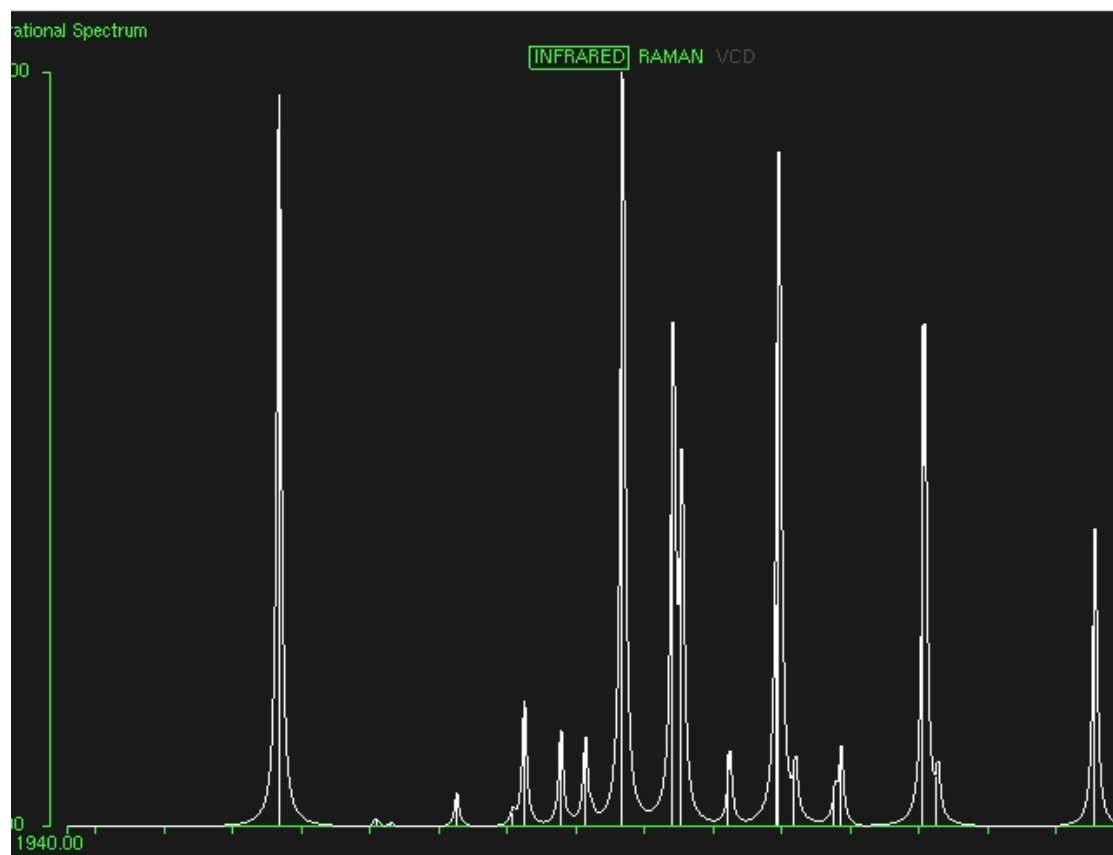
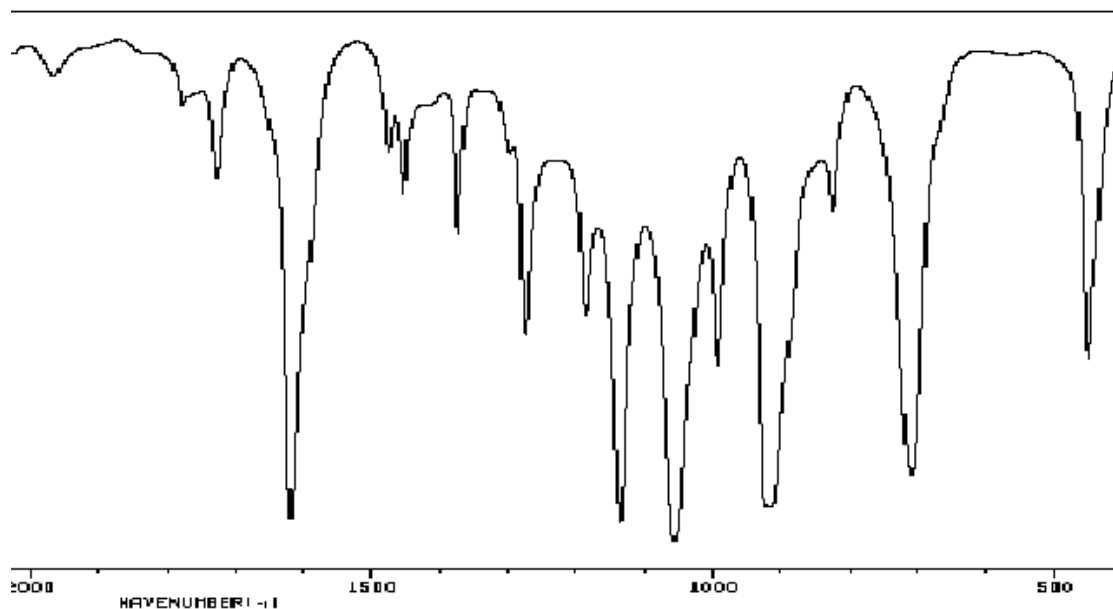
The results screen has a light brown default background. One can drag, rotate (left-right, up-down and in the plane), and scale (middle mouse button) the molecule. At this level, which is usually as good for organic molecules as most experiments), I got, for instance, I got for the =CH-O bond length 1.386 Å, for the O-CH₂- bond length 1.455 Å, and for the C-O-C angle 106.4 ° in 2,3-dihydrofuran. The molecule does not have a symmetry plane. Note that the sp² hybridized carbon is significantly more compact than the sp³ hybridized one, resulting in a shorter C-O distance. The agreement between the calculated and experimental IR spectra is quite pleasing, considering that the calculation on one CPU core took less than 8 minutes, and a normal calculation on 10 cores would be less than 1 min.

Comparison of the IR spectra

A comparison of the experimental thin-film IR spectrum from the AIST database and my calculated spectrum is shown below for both isomers. The band positions agree better than the relative intensities. For the latter, diffuse basis functions are important. The calculated spectra leave no doubt which isomer is which.



Comparison of the experimental and calculated (bottom) IR spectra of 2,3-dihydrofuran.



Comparison of the experimental and calculated (bottom) IR spectra of 2,5-dihydrofuran

Comparison of the calculated and experimental NMR shifts

Interpreting the NMR spectrum for such a small molecule does not require sophisticated calculations. The difference in symmetry alone suffices to distinguish the isomers. They are given here to indicate the accuracy of the calculations. Only the C-13 shifts are shown. The following Table compares the calculated chemical shifts, using 179.0 ppm for the shielding in TMS, and the AIST database values. For a discussion of the TMS reference value, see the next exercise (cyclohexene).

Molecule	C ₂	C ₃	C ₄	C ₅
2,3-dihydrofuran calc.	67.9	28.5	95.2	147.8
2,3-dihydrofuran exp. (AIST)	69.58	29.28	99.54	145.99
2,5-dihydrofuran calc.	75.6	127.8	127.8	75.6
2,5-dihydrofuran exp. (AIST)	75.42	126.34	126.34	75.42

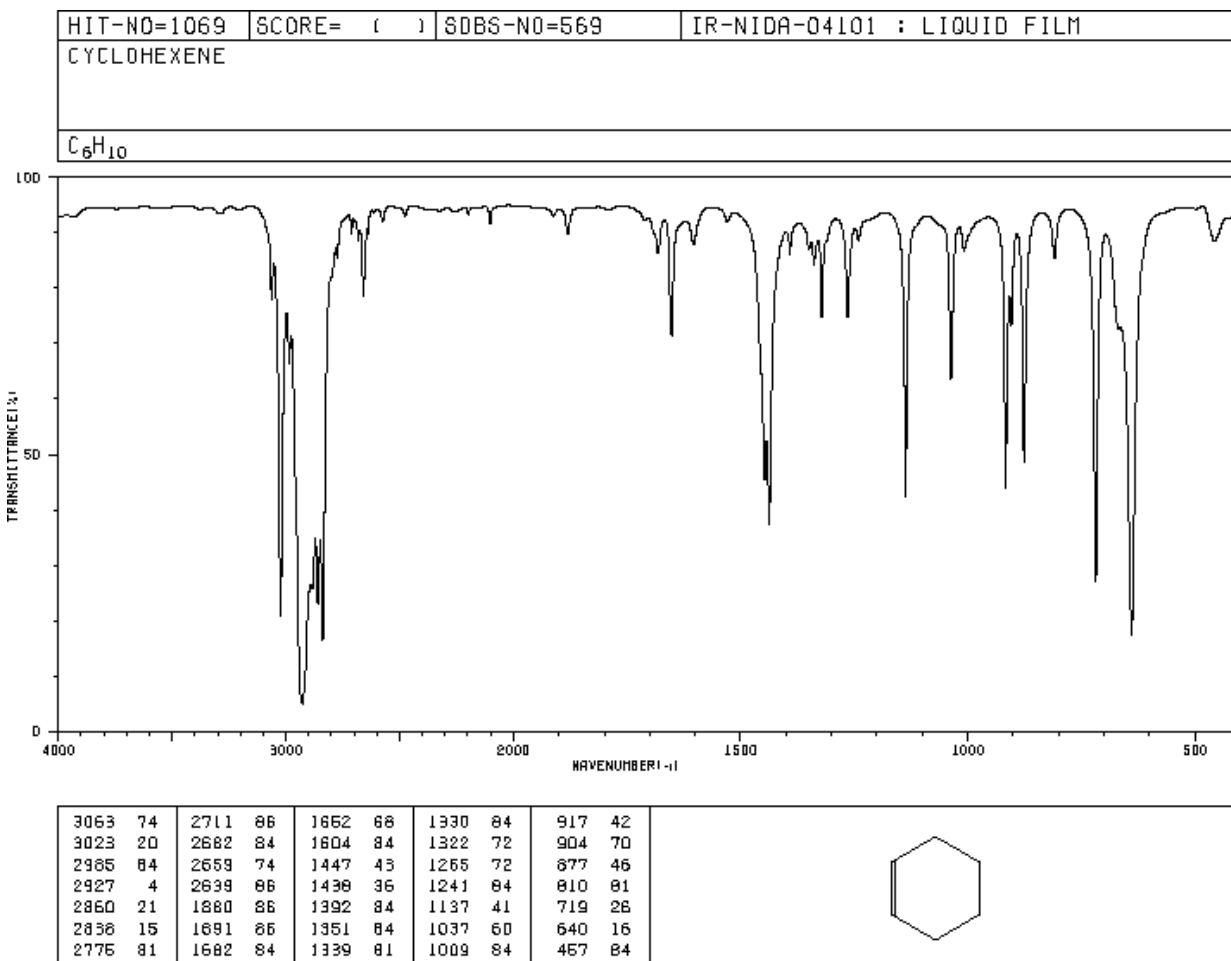
Calculation of the infrared spectrum and NMR shieldings in cyclohexene

In this exercise, we calculate both the expected infrared spectrum and the NMR shieldings of cyclohexene. A single calculation can yield both quantities. The shieldings can be converted to chemical shifts if the shieldings of the standard (most often tetramethylsilane, $\text{Si}(\text{CH}_3)_4$, for protons, ^{13}C and ^{29}Si), are known.

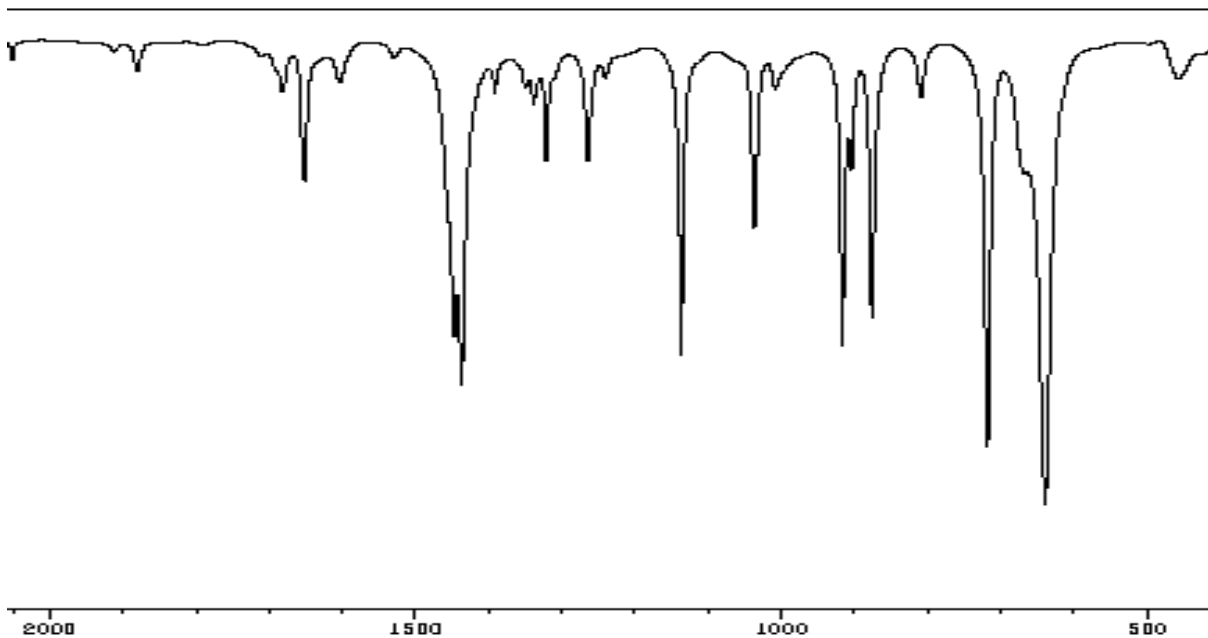
Construct cyclohexene using PQSMol. It is not planar: the ideal valence angles for the sp^2 and sp^3 hybridized carbon atoms (two 120° angles and four 109.5° angles) add up to less than the sum of the angles in a planar hexagon (720°). As a result, the molecule folds up somewhat, diminishing the angles. This can happen in two different ways, giving either C_2 or C_s symmetry. The stable form is C_2 ; make sure that your initial force field optimized structure does not have C_s symmetry. Use DFT with the B3LYP exchange-correlation potential, and the relatively modest 6-311G(d) basis set. This combination is usually abbreviated as B3LYP/6-311G(d). Both NMR shielding and vibrational frequencies are so-called second order properties, and take longer to calculate than first-order properties, and therefore we try not to use a very large basis set. Optimize the molecular geometry, and select *NMR* and *Vibrational Frequencies* as properties to be calculated. Note that vibrational frequencies are well defined only if calculated at a stationary geometry, i.e., a minimum or a transition state.

Let us identify the most intense bands of cyclohexene in the fingerprint region ($400\text{-}2000\text{ cm}^{-1}$) of the calculated infrared spectrum. You can identify these bands in the graphical display. To select the fingerprint region, click on "Standardize" in the Vibrational Frequencies display. Select the 400 (min) to 2000 (max) cm^{-1} region. As a default, the spectrum is drawn so that the most intense band (in this case a C-H of little interest) is not cut off. This makes all other bands very small. Change the scale of intensity from 0 to, say, 30 (km/mol, which is the accepted unit of IR intensity). You may also want to change the simulated width of the IR bands. A half-width near 5 cm^{-1} is often the best. Identify all fundamental vibrations with an intensity over 2.5 km/mol. For vibrations in the crowded $1490\text{-}1520\text{ cm}^{-1}$ range, add the intensities and calculate the intensity-weighted average frequency. To get the best agreement with

experiment, use a scale factor of 0.97. Compare the calculated and experimental spectra. You may want to make a copy of the reference spectrum, and draw (by hand) the strongest calculated frequencies. Also determine, if possible, what kind of a motion do these bands correspond to (click on the “Animate” or “Vectors” box in the IR display).



Cyclohexene infrared (IR) spectrum from the AIST database (400-4000 cm^{-1}). Note the change of scale at 2000 cm^{-1} .



Cyclohexene IR spectrum in the fingerprint (400-2000 cm^{-1}) region



Cyclohexene calculated IR spectrum (unscaled)

NMR

The program gives the calculated isotropic shielding; these can also be displayed in PQSView. Experimentally, chemical shift *differences* are measured. They are related to the shielding as $\delta = \sigma_0 - \sigma$, where δ is the chemical shift, σ_0 is the absolute shielding in the reference compound, and σ is the absolute shielding of the sample. The reference used today is almost exclusively tetramethylsilane for ^1H , ^{13}C and ^{29}Si , although in the early history of NMR other substances, for instance benzene, were also used. TMS is convenient experimentally but is not ideal for calculations because it is chemically very different from most organic molecules, and therefore requires high accuracy to reproduce the shifts accurately. It is best to use an intermediate standard, say benzene. For this assignment, you can take the ^{13}C shielding of TMS 179.0 ppm and the ^1H shielding of TMS 32.12 ppm. These values were determined using intermediate standards.

Let us compare the calculated ^{13}C and ^1H chemical shifts of cyclohexene with experiment. For the methylene hydrogens, there is a slight complication. The two hydrogens on the same carbon are not equivalent at the equilibrium C_2 geometry. This difference cannot be observed in the experimental NMR spectrum at room temperature because the molecule flips back and forth between the two mirror image C_2 structures. (This is the reason why cyclohexene, although chiral in principle, cannot be separated into D and L conformers). The rate of interconversion is too high for NMR to separate the peaks, and they appear at the average frequency of the two distinct methylene hydrogens (on the same C atom).

In the same AIST database, the experimental proton chemical shifts (in CDCl_3 , 400 MHz) are listed as 5.66, 1.99 and 1.61 ppm. The calculated values (after averaging the NMR-equivalent protons) are 5.73, 1.97, and 1.58 ppm, using 32.12 ppm for the absolute proton shielding in TMS, all within 0.04 ppm of experiment.

The experimental ^{13}C chemical shifts are 127.3, 25.3, and 22.8 ppm. The calculated absolute shieldings are 49.63, 157.43 and 154.54 ppm. Using 179.0 ppm for the C shielding in TMS, the chemical shifts are 129.37, 24.46, and 21.57 ppm. The largest deviation from experiment is about 2 ppm.

A strange molecule: sulfur tetrafluoride

In this exercise, we will investigate the molecular geometry of SF₄. Most people are surprised when encounter this for the first time. Its shape is not at all tetrahedral but is like a seesaw, with an approximately linear F-S-F arrangement, the two other fluorine atoms being in a plane perpendicular to the F...F axis, and making an F-S-F angle of about 90°. Yet rotational spectroscopy, electron diffraction data and, most convincingly, vibrational (infrared and Raman) spectra prove conclusively this, at first sight counterintuitive, equilibrium geometry. We will also simulate the vibrational spectra, and will try to explain the strange geometry.

To construct the geometry, change to “Unrestricted” mode (restricted mode is useful only for plain vanilla organic-type molecules). Click in the GUI “Building Blocks” and the sulfur atom. You can choose the bonding pattern. As we shall see, the strange geometry of this molecule arises from 5-coordination around the central S atom. Sulfur has 6 valence electrons (2 s and 4 p), and each fluorine contributes one bonding electron. The other 6 valence electrons of fluorine form 3 largely inactive lone pairs. This gives 10 valence electrons around the central sulfur. According to Valence Shell Electron Repulsion Theory (VSEPR), 5-coordination generally gives a trigonal bipyramidal structure, with 5 electron pairs around the S atom. Four of these positions are occupied by a fluorine, the fifth by a lone pair of electrons. The axial positions are preferred for F because they are more distant from the other ligands. Accordingly, F atoms fill the two axial positions and two of the equatorial positions; the last equatorial position is occupied by a lone pair of electrons.

Choose the 5-coordinated trigonal bipyramid pattern for S. Right-clicking in the field while the main cursor is in its normal (leftmost) position will put a sulfur in the field. Now click on F. Right-clicking on the S valences will put an F atom in the field, bound to S. Make SF₄ with two axial and two equatorial fluorines but *do not* try to optimize it using one of the force fields. They will probably go to a qualitatively incorrect geometry. Before we go to the DFT calculation, we have to remove the phantom fifth valence on S. Click on the “Select” item in the bottom row (sixth from left), and also change the “Select” and “Torsion” options to “Manual” from “Auto”. After putting the cursor to the “Select” mode, click on the excess valence on the sulfur atom, and then click “Cut” (top). Now we can click “Calculation” and save the initial modeling geometry as a *.pqb (PQS build) file. You may save it in your “PQS Jobs” folder. Make sure that the file name you give it *does not contain blanks* or other questionable characters.

Saving the PQS build file brings up an input builder screen. You must give a name for the input file, and optionally increase the memory assigned to the job. The latter is generally only necessary if vibrational frequencies or MP2 energies or forces (gradients) are to be calculated. It does not hurt, though, to increase the memory from the standard 12 MW (megawords) to 30 or so, keeping in mind that 1 MW=8MB (megabytes). You have to specify the job type (single point energy or geometry optimization). Choose “optimization”, and for the type of calculation select “DFT”. I recommend a pretty good

basis set, say def2-tzvp (“triple zeta valence polarization”). Save the input file (SF4.inp is a natural choice) in the folder “PQS Jobs”. (Note that under Linux, upper and lower case characters are different, i.e., SF4.inp is a different file than sf4.inp). We want to edit the input a little before running the job. For understanding what is the reason for the strange geometry of SF₄, we would like to calculate the *localized molecular orbitals* of SF₄. This can be done simply by saving the input file, clicking on “Calculations”, browsing for our input file, and opening it for editing. It will open in the manual edit mode. Simply add, anywhere on the SCF line, the phrase LOCA=Boys (separated by blanks from the rest).

We can run the calculation by selecting the “Calculation” (on top), and, if the input file we want to run is not showing, clicking “Browse” and selecting the file. If the input is showing than left-clicking on it brings up a menu, the main points of which are “Edit” (hand editing) of the input, “Submit Serial job” (“Submit Parallel job is not yet implemented on the cloud platform), “Job Output” and “Visualize Output”. Job Output comes up automatically when the calculation starts, and if it finishes successfully, “Visualize Output” comes up.

Click on SF4.inp and then on “Submit serial job”. My GUI-generated and slightly edited input file looked like this:

```
%MEM=30
GEOM=pqb FILE=SF4.pqb
BASIS=def2-tzvp
OPTimize
SCF DFT=B3LYP
FORCE
JUMP
GEOM GEOP
SCF DFT=B3LYP LOCA=Boys
HESS
FREQ
```

The first line increases the memory somewhat; vibrational frequency calculations require more memory. The second line specifies the source of the initial geometry as a *.pqb file. The third line specifies the basis set. The fourth line requests geometry optimization. The fifth is the iterative (SCF) energy calculation. The sixth is the calculation of the forces on the atoms. The JUMP line jumps back to optimization unless the geometry and the energy have converged. The eighth line requests the printing of bond lengths and angles; these can, of course, also read off the visualization of the results. The ninth line requests a higher accuracy SCF (Threshold is in pH form, 6 means 10⁻⁶), and also the calculation of localized orbitals. HESS calculates the force constants, that is the second derivatives (Hessian) of the energy with respect of the nuclear coordinates. FREQ calculates the vibrational frequencies, IR intensities and

thermodynamic data. Raman intensities are more expensive to calculate and have to be specifically requested.

The following little table compares the calculated and experimental (microwave) geometries. Bond distances in Å.

B3LYP: $SF_{eq}=1.565 \text{ \AA}$, $SF_{ax}=1.668$, $F_{eq}-S-F_{eq}=101.1^\circ$, $F_{ax}-S-F_{ax}=173.2^\circ$

MP2: $SF_{eq}=1.551 \text{ \AA}$, $SF_{ax}=1.649$, $F_{eq}-S-F_{eq}=101.4^\circ$, $F_{ax}-S-F_{ax}=172.8^\circ$

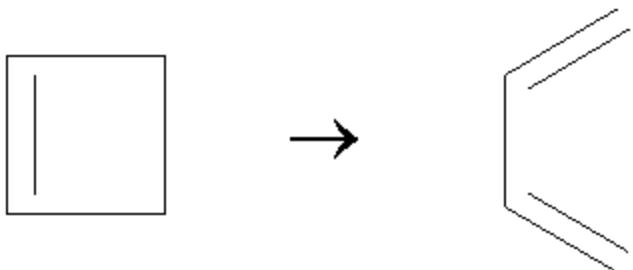
MW: $SF_{eq}=1.545 \text{ \AA}$, $SF_{ax}=1.646$, $F_{eq}-S-F_{eq}=101.5^\circ$, $F_{ax}-S-F_{ax}=173.1^\circ$ (Gwinn, 1962)

DFT gives slightly too long bonds for 3rd and higher period elements. A more expensive method, MP2, gives a better geometry, accurate to a few thousandths of an Å.

Minima and transition states. The barrier to thermal ring opening in cyclobutene

The determination of transition states and reaction barriers is much more involved than the calculation of equilibrium geometries. The simplest reactions are monomolecular, a single molecule reacting. If the reactant is stable, it is a true minimum on the potential energy surface. A chemical reaction consists of going from a given minimum to another one. During this, the molecule must pass through a transition state, or saddle point. Transition states (also called first-order transition states) are stable (i.e. minima) in every direction except in the direction of the reaction path. In that direction they are a maximum. Because of the difficulty of finding transition states, we will go through several stages, locating first the rough position of the transition state, and then its exact position. We will also use inexpensive methods first. The nature of minima and transition states can be checked by calculating the vibrational frequencies. Vibrations corresponding to motion along the reaction path are formally imaginary. A stable molecular geometry has all forces on the atoms zero and all vibrational frequencies real. An imaginary frequency signals an unstable situation like a pencil balanced on its tip. There should be one, and only one imaginary vibrational frequency at the transition state, corresponding to the motion along the reaction coordinate.

Cyclobutene can undergo thermal ring opening, isomerizing to 1,3-butadiene. In the first step *gauche* butadiene is formed (CCCC torsion about 30°); this converts largely to the slightly more stable *trans* form by rotation around the central C-C bond. The reaction path is more easily mapped in the reverse direction. The main step on the reverse path is the ring closure of *gauche* butadiene to cyclobutene. The reason for going in the reverse direction is that the carbon skeleton of cyclobutene is planar while the transition state and the initial product are not. The program keeps the initial high C_{2v} symmetry; in the actual reaction, symmetry-breaking bumps by other molecules lead to the correct (nonplanar) reaction path.



Input preparation: Construct *cis*-butadiene and rotate the groups around the central C-C bond by ~30 degrees (**rotate around bond at the bottom**). Do not reoptimize. We will use an **optimized scan** to trace the reaction path and locate the approximate position of the transition state. This consists of systematically varying (“scanning”) a coordinate (in our case the distance between the terminal C atoms) and optimizing all geometrical variables at each given value of this coordinate except the coordinate itself that we scan. Note that we scan backwards, from butadiene to cyclobutene, rather than forward, for reasons explained above. Cyclobutene has higher symmetry (C_{2v}) than the transition state (C_2). The program (and all other programs) preserves the initial symmetry but the molecule cannot react if it stays planar. The reverse reaction does not have this problem, as *gauche*-butadiene has only C_2 symmetry. The PQSMol input generator does not include the SCAN facility, and therefore the input has to be edited in by hand. See the input example below. We first use the inexpensive AM1 semiempirical method to yield a qualitatively correct input geometry.

Semiempirical methods like AM1 are less accurate than *ab initio* and DFT methods but much faster. Start by noting the atom numbers of the two terminal C atoms in *your* molecule (in my case they were atoms 1 and 7), and the distance between them (it should be a little over 3 Å. The visualizer (PQSView) keeps the initial bond orders. In our case, however, a bond forms between the terminal C atoms, and the double bonds of butadiene change to single bonds, and the single C-C bond changes to double. To have a better visual representation of the transition state, we add a bond between the terminal carbons and change its bond order to $\frac{1}{2}$, and change the remaining C-C bond orders in *cis*-butadiene to 1.5, intermediate between single and double. This editing of the bond orders is only cosmetic, it gives a better visualization of the transition state.

In the calculation, we systematically close the C...C bond distance from 3.2 Å (longer than in *gauche*-butadiene) in steps of 0.05 Å to 1.4 Å: the latter value is somewhat less than the C-C bond length opposite to the double bond in cyclobutene. The input options and examples of the SCAN feature are in the PQS manual. For help, I am enclosing part of my input file. Depending on your atom numbering, your input is likely to be slightly different from mine. Start with constructing a simple geometry optimization in PQSMol and modify it according to the example below.

Partial optimization, when some geometry parameters are kept constant, is called **constrained optimization**. We then can plot the energy as a function of the C-C bond as we go from butadiene to cyclobutene. There should be two minima on this curve, separated by a maximum.

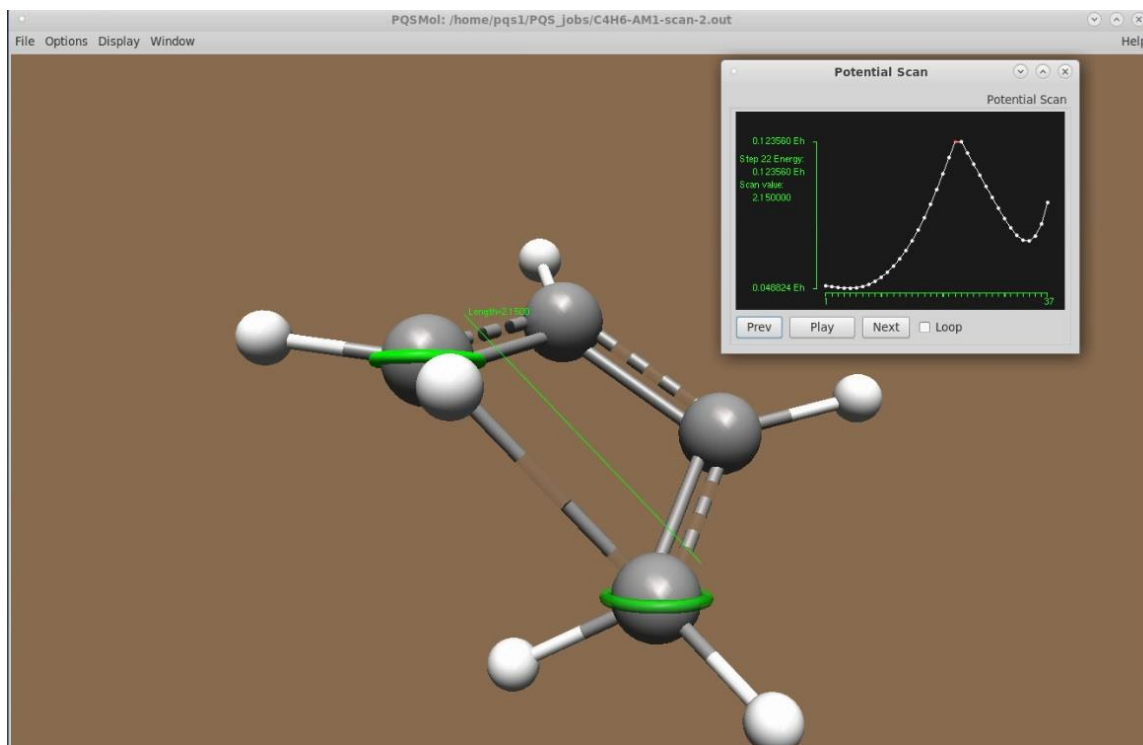
My input file looks like the one below. Remember, your input file will be similar but different because the atom numbering and the geometry will not be the same. The lines that you have to change are set in bold type. Note that anything beyond an exclamation mark ! is ignored, and the whole line is ignored if its first character is a question mark (?).

```
Text= butadiene-cyclobutene ring closure scan, AM1
Geom=PQB FILE=C4H6-AM1.pqb
SCAN stre 1 7 FROM 3.2 1.4 -0.05 !Change 1 and 7 as needed
OPTI
SEMI=AM1
FORCE
JUMP
JUMP
```

Note the two JUMP statements – this is a double loop, first changing systematically the C...C distance, and within each scan optimizing the geometry. The **SCAN** line shows that we vary the distance from 3.2 to 1.4 Å.

The visualization program, PQSView, can display the scanned energy against the value of the scan coordinate. The maximum of this curve approximates the transition state. We can determine the accurate AM1 transition state by performing a transition state optimization, starting with the approximate transition state obtained by scanning. First, determine which scan value corresponds to the transition state, i.e., the highest

point on the energy profile. In this case, it is scan point 22, corresponding to C...C distance 2.15 Å.



This will be our starting point for finding the transition state more accurately using Density Functional Theory. We can import the result of this calculation into PQSMol. We need to import scan geometry 22, the approximate transition state.

To save time, we use a small basis set, "3-21G". This yields only qualitatively correct results. We use the B3LYP exchange-correlation function in DFT. To find the transition state, the program must have some idea of the *force constants* or *Hessian matrix*. This is not generally needed for geometry optimization where every step that lowers the energy is good. However, for transition state (saddle point) calculations, the energy is not informative: it can go up along the reaction path. We will therefore calculate the Hessian at the initial geometry, optimize the geometry, and calculate the Hessian matrix and vibrational frequencies again. There is one and only one imaginary vibrational frequency (the program prints this out as negative) in a proper transition state. For higher accuracy, one should use a larger, more accurate basis set for the optimization and second Hessian and frequency calculation. My input looked like this:


```

%MEM=30                                ! Probably an overkill
TEXT=Cyclobutene TS B3LYP/3-21G
!!!!!!!!!!!!!! STEP 1  !!!!!!!!!!!!!!! Comment only
BASIS=3-21G
SCF DFT=B3LYP
HESS                                    ! Initial force constants
FREQ                                    ! Needed only for check
OPTimize type=ts
SCF DFT=B3LYP
FORCE
JUMP
HESS
FREQ

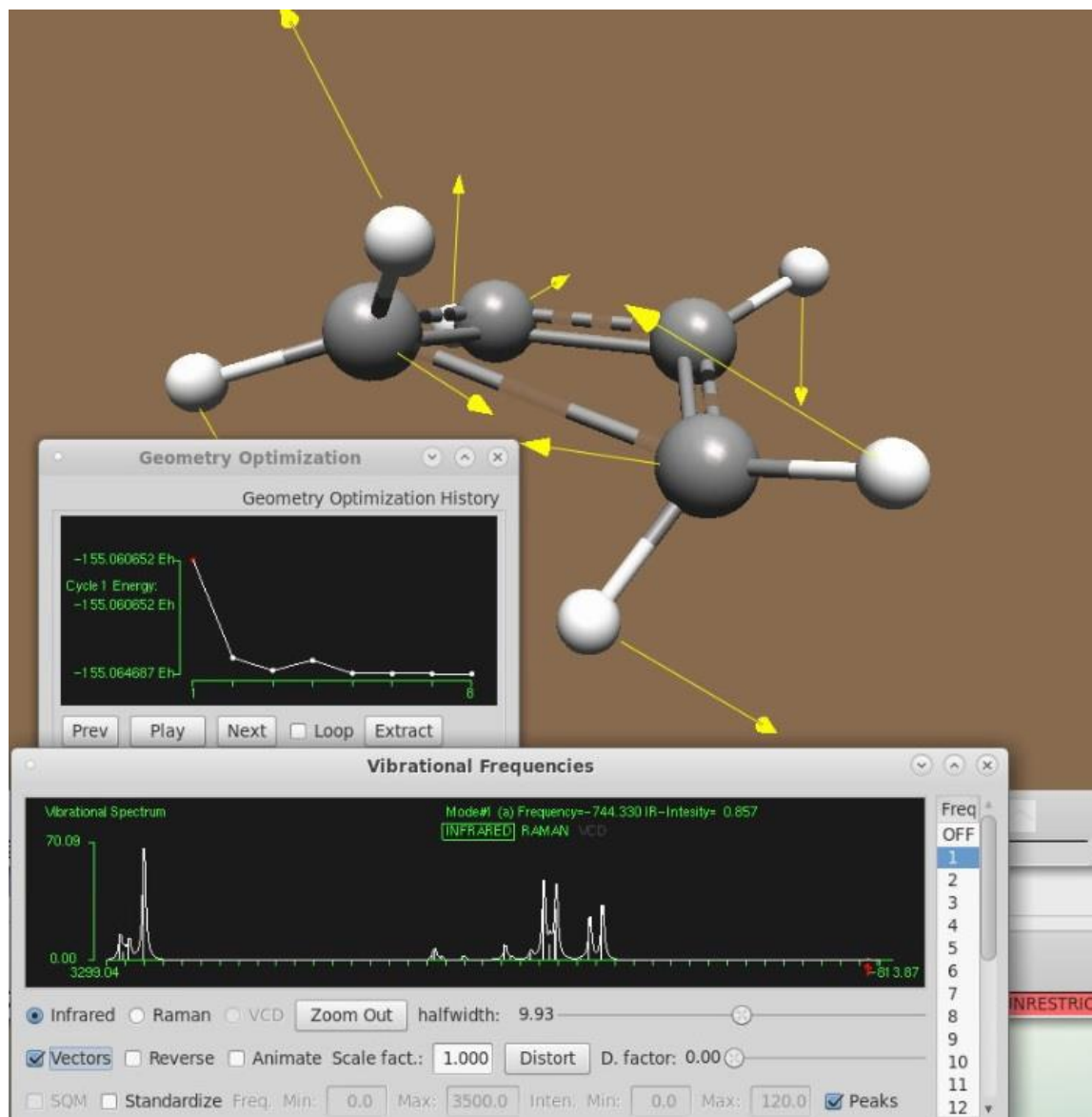
```

Visualization of the results, including Optimization History and Vibrational Frequencies gives the picture shown on the next page. The transition state is qualitatively the same as with the semiempirical AM1 method. The frequencies can be animated but in this picture the vectors corresponding to the imaginary vibrational frequency are plotted. This is basically the motion over the reaction barrier at the transition state.

To get an estimate for the *activation energy* and the *enthalpy of the reaction*, we need to determine the energies of butadiene and cyclobutene *at the same theoretical level*, in this case B3LYP/3-21G. Since the minimum energy form of butadiene is the *trans* form, this should be used.

The electronic energy difference is the dominant but not the only contribution to reaction energetics. The second leading contribution is vibrational energy. The program calculates this in the harmonic oscillator – rigid rotor approximation if a frequency calculation is specified. According to *Transition State Theory*, the vibrational energy corresponding to the imaginary vibrational frequency should be simply omitted.

More accurate results can be obtained by using a larger basis set and higher theoretical level, say MP2 or even CCSD(T).



Experimental data about this reaction can be found in the following publications:

- (a) Cooper, E.; Walters, W. D. *J. Am. Chem. Soc.* **1958**, *80*, 4220-4224.
- (b) Wiberg, K. B.; Fenoglio, R. A. *J. Am. Chem. Soc.* **1968**, *90*, 3395-3397.
- (c) Criegee, R.; Seebach, D.; Winter, R. E.; Börretzen, B.; Brune, H. *Chem. Ber.* **1965**, *98*, 2339-2352. For a review see
- (d) Dolbier, W. R., Jr.; Koroniak, H.; Houk, K. N.; Sheu, C. *Acc. Chem. Res.* **1996**, *29*, 471-477 and references therein.