Computational Chemistry Laboratory

Modeling Molecular Structures with HyperChem™

A study of molecular structure and reaction mechanism with molecular mechanics and semi-empirical methods using HyperChem™ molecular modeling software

Introduction

In this laboratory you will study some basic methods of molecular structure investigation. You will be using *HyperChem*^{TM} molecular modeling software and employ semi-empirical quantum mechanical methods to study the structure and energetic of various alkenes, intermediates and final products

Concepts you should get familiar with

1. What are molecular mechanics and semi-empirical methods of quantum mechanics?

2. Optimization of the molecular geometry

3. How to use *HyperChem* to generate molecules (the builder), perform structural optimizations, and analyze molecular orbitals and its relation to functionality.

4. General mechanisms of electrophilic addition of Bromine to Alkenes.

5. Molecular dynamic simulations, Simulated Annealing and Monte Carlo approaches

Introduction to molecular modeling

Molecular Modeling
Molecular modeling involves the development of mathematical models of molecules that can be used to predict and interpret their properties. There are two types of molecular modeling - molecular mechanics and quantum mechanics.

Molecular mechanics: a classical mechanical model that represents a molecule as a group of atoms held together by elastic bonds. Molecular mechanics methods give predictions of molecular geometries and heats of formation.

Quantum mechanics: a quantum mechanical model of the electronic structure of a molecule, which involves solving the Schrödinger equation. Quantum mechanics can be used to predict electronic properties of molecules, such as dipole moments and spectroscopy.

Molecular Mechanics

Molecular mechanics is a classical mechanical model that represents a molecule as a group of atoms held together by elastic bonds. In a nutshell, molecular mechanics looks at the bonds as springs, which can be stretched, compressed, bent at the bond angles, and twisted in torsional (dihedral) angles[1]. Interactions between nonbonded atoms also are considered. The sum of all these forces is called the **force field** of the molecule. A molecular mechanics force field is constructed and parameterized by comparison with a number of molecules, for instance a group of alkanes. This force field then can be used for other molecules similar to those for which it was parameterized. To make a molecular mechanics calculation, a force field is chosen and suitable molecular structure values (natural bond lengths, angles, etc.) are set. The structure then is optimized by changing the structure incrementally to minimize the strain energy and spread it over the entire molecule. This minimization is orders of magnitude faster than a quantum mechanical calculation on an equivalent molecule.

Molecular mechanics is a valuable tool for predicting geometries and heats of formation of molecules for which a force field is available. It is a good way to compare different conformations of the same molecule, for instance. However, molecular mechanics have two weaknesses. First, force fields are based on the properties of known, similar molecules. If one interested in the properties of a new type of molecule an appropriate force field probably will not be available for that molecule. Second, because molecular mechanics models look at molecules as sets of springs, they cannot be used to predict electronic properties of molecules, such as dipole moments and spectroscopy. To make predictions about the electronic properties of a molecule, you must use quantum mechanical models.

Molecular Dynamics

This method uses the Newtonian equations of motion, a potential energy function and associated force field to follow the displacement of atoms in a molecule over a certain period of time, at a certain temperature and a certain pressure. Calculations of motion are done at discrete and small time intervals and a velocity calculated on each atom position which in turn is used to calculate the acceleration for the next step. Starting velocities can be calculated at random (necessary when starting at 0 Kelvin where the kinetic energy is 0) or by scaling the initial forces on the atoms. Simulations can also be run with differing temperatures to obtain different families of conformers. At higher temperatures more conformers are possible and it becomes feasible to cross energy barriers.

---

[1] A torsion angle $\varphi=0$ is defined when the two planes coincide at a *cis* configuration.

When doing calculations on biological molecules it is becoming more frequent to do the calculations in the presence of solvent (usually water). However, this brings further complications due to two main problems. The first being increased CPU time due to the larger number of atoms. The second is that the water molecules surrounding the molecule tend to drift away from the molecule of interest and get "lost" from the calculation if only a certain area of space is being monitored as is usually the case. This causes nasty "edge effects". There is one method currently used to get around this problem. That is to place your molecule surrounded in water in a box of a specific size and then to surround that box with an image of itself in all directions. The solute in the box of interest only interacts with its nearest neighbor images. Since each box is an image of the other, then when a molecule leaves a box its image enters from the opposite box and replaces it so that there is conservation of the total number of molecules and atoms in the box. Those are known as periodic boundary conditions.

Simulated Annealing

Simulated annealing is a special type of dynamics. The molecule is heated and then cooled very slowly so that conformational changes taking place will lead to a local minimum being located. This process is generally repeated many times until several very closely related, low energy conformations are obtained. This assumed to be the global minimum.

Monte Carlo

Related to molecular dynamics are **Monte Carlo** methods which randomly move to a new geometry/conformation. If this conformation has a lower energy or is very close in energy it is accepted, if not, an entirely new conformation is generated. This process is continued until a set of low energy conformers has been generated a certain number of times.

Quantum Mechanics

To make a quantum mechanical model of the electronic structure of a molecule, we must solve the Schrödinger equation.

$$H\Psi = E\Psi$$

The Hamiltonian operator, **H**, depends on the kinetic and potential energies of the nuclei and electrons in the atom or molecule. In more complicated situations, e.g., the presence of an external electric and magnetic fields, in the event of significant spin-orbit coupling in heavy elements, taking account of relativistic effects, etc., other terms are required in the Hamiltonian. The wavefunction, Ψ, will give us information about the probability of finding the electrons in different places in the molecule. The energy, E, is related to the energies of individual electrons, which can be used to help interpret electronic

spectroscopy.

Solving the Schrödinger equation is a very difficult problem and cannot be done without making approximations. Two types of approximations are the Born-Oppenheimer approximation and the independent electron approximation.

The Born-Oppenheimer Approximation

In the Born-Oppenheimer approximation, the positions of the nuclei are taken to be fixed so that the internuclear distances are constant. This is a sensible approximation because the massive nuclei are essentially immobile in comparison with light electrons. We first choose geometry (with fixed internuclear distances) for a molecule and solve the Schrödinger equation for that geometry. We then change the geometry slightly and solve the equation again. This continues until we find an optimum geometry with the lowest energy. It should be noted that without the Born-Oppenheimer approximation we would lack the concept of a potential energy surface: The PES is the surface defined by E_{el} over all possible nuclear coordinates. We would further lack the concepts of equilibrium and transition state geometries, since these are defined as critical points on the PES; instead we would be reduced to discussing high-probability regions of the nuclear wavefunctions.

Molecular Orbital (MO) Theory

Another approximation (the **independent electron approximation**) commonly made is that the wavefunction, *R*, can be written as a product of one-electron functions. The one-electron wavefunctions are called **molecular orbitals** - the molecular equivalent of atomic orbitals. Each molecular orbital then is expressed as a combination of the atomic orbitals from the atoms that make up the molecule. For example, the simplest molecular orbital function for the H_2 molecule is written as $c_1 1s_1 + c_2 1s_2$, where $1s_i$ is a hydrogen 1s atomic orbital function and c_i is a parameter. This method is called **LCAO-MO** theory for Linear Combination of Atomic Orbitals - Molecular Orbital Theory.

As an example, the two lowest energy molecular orbitals of the H_2 molecule are shown here.

Bonding   Antibonding
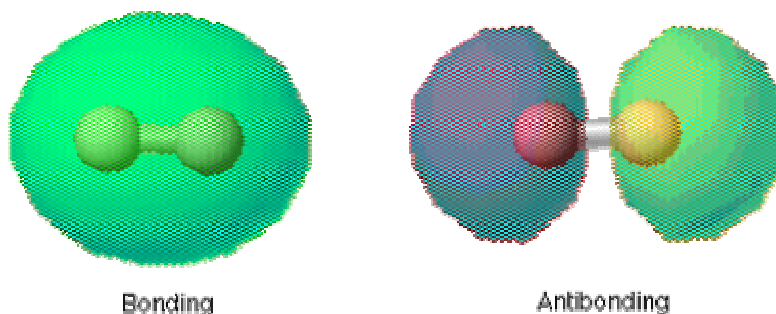
These can be thought of as combinations of 1s orbitals from the two hydrogen atoms. The molecular orbital on the left is made by two H atom 1s orbitals combining constructively. This is a **bonding** MO because it helps hold the molecule together. The molecular orbital on the right is due to the destructive interference of the two 1s orbitals and is said to be **antibonding.**

Although molecular orbitals are written as combinations of atomic orbitals from the atoms in the molecule, *molecular orbitals are not atomic orbitals.* They are analogous to atomic orbitals, but instead of being defined for atoms, molecular orbitals are characteristic of the molecule as a whole.

Ab Initio and Semiempirical Calculations

Because of the large number of particles in a molecule (benzene, for instance, has 12 nuclei and 78 electrons) computer programs are used to do the calculations necessary for the solution of the Schrödinger equation. These calculations involve an enormous number of difficult integrals for large molecules. *Ab initio* computational methods solve all of these integrals without approximation. *Ab initio* methods are the most reliable for small and medium-sized molecules, but are prohibitively time-consuming for large molecules (20 atoms or so for PCs; around 100 atoms if workstations are available).

For larger molecules, **semiempirical** methods have been developed which ignore or approximate some of the integrals used in *ab initio* methods. To compensate for neglecting the integrals, the semiempirical methods introduce parameters based on molecular data. Commercial software packages are available for both *ab initio* and semiempirical calculations.

Semiempirical Molecular Orbital Theory

Semiempirical molecular orbital theory methods have been developed which ignore or approximate some of the integrals used in the solution of the Schrödinger equation. To compensate for neglecting the integrals, the

semiempirical methods introduce parameters based on molecular data. The most commonly used semiempirical methods are included in Hyperchem computer program, offering two of the most reliable semiempirical methods (PM3 and AM1) for predicting heats of formation, ground state geometries, and ionization potentials. It also includes the ZINDO method, which does a good job predicting the visible-UV bands for molecules containing hydrogen and first or second period elements.

Chemical Accuracy

Bond lengths and bond angles
What kind of accuracy do we desire from quantum mechanical calculations? Ideally a "chemically accurate" calculation should give the results below[2].

Bond lengths: calculated values within 0. 01 - 0.02 Å of experiment.

Bond angles: calculated values within 1- 2° of experiment.

Semiempirical methods can only be used for elements for which they have been parameterized, but because these elements are common ones in organic compounds, semiempirical calculations can be quite useful. Semiempirical results do not always satisfy the criteria we set for chemical accuracy. PM3 calculations,[3] for instance, generally give bond lengths within ±0.036Å and bond angles within ±3.9°, not always "accurate" but still pretty good.

Electronic energies and heats of formation
Semi-empirical quantum mechanical calculations also can be used to predict electronic energies and heats of formation. Chemically accurate energies should be within 1 kJ/mol (0.2 kcal/mol), a challenging task.[4]

How do we obtain thermodynamical information (ΔH_f) from solution of the Schrödinger equation? The electronic energy calculated by the semi-empirical methods is, in analogy with that of the *Ab-initio* methods, the total energy relative to a situation where the nuclei (with their electrons) are infinitely separated. The electronic energy is normally converted to a *gas phase* standard heat of formation by subtracting the electronic energy of the isolated atoms and adding the *experimental* atomic heat of formation[2].

$$\Delta H_f(\text{molecule}) = N_A \cdot \{E_{\text{elec}}(\text{molecule}) - \Sigma_{\text{atom}} E_{\text{elec}}(\text{atoms}) + \Sigma_{\text{atoms}} \Delta H_f(\text{atoms})\}$$

Recall that the enthalpy is related to the energy by H=E+PV. Here the PV contribution is implicitly accounted for since the methods were optimized to

---

[2] The atomic heat of formation is the heat that realized during the formation of the stable form of the element from individual atoms at standard conditions. It should be noted that thermodynamical corrections (e.g., zero-point energies) *should not be added* to the formation energy, as these are implicitly included by the parametrization..

reproduce heat of formations at 298K, and formation enthalpies of the atoms are known experimentally. The average errors for PM3 semi-empirical calculations of heats of formation are tabulated below.[5]

Type of Compound	ΔH (kcal/mol)	Type of Compound	ΔH (kcal/mol)
All C, H, N, O	4.4	Organic cations	9.5
Hydrocarbons	3.6	Organics with F, Si, Cl, Br, I	5.7
Cyclic Hydrocarbons	2.4	Compounds with S	12.1
Hydrocarbons, double bonds	2.8	Compounds with P	11.5
Hydrocarbons, triple bonds	5.6	Closed shell anions	8.8
Aromatic hydrocarbons	4.1	Neutral radicals	7.4
Organics with N, O	5.2		

Semiempirical methods should only be used to predict the properties of molecules for which reliable parameters are available. Experience also has shown that semiempirical methods do not do well for problems involving hydrogen bonding, transition states, and poorly parameterized molecules.[6] If one interested in the highest accuracy or in problems such as these, *Ab initio* calculations should be used if possible.

When you are comparing semiempirical or *Ab initio* predictions with experiment you should remember that computational models generally are of **gas-phase molecules**. There are ways of computationally modeling molecules in solution, either at the *Ab-initio* or semiempirical level, but we will not get into this subject here.
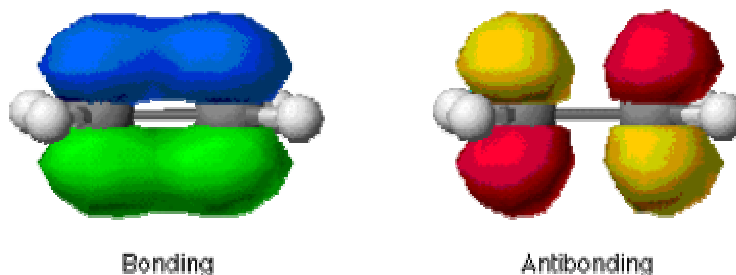

Molecular Orbital Surfaces

Molecular orbitals are not real physical quantities. Orbitals are mathematical conveniences that help us think about bonding and reactivity, but they are not physical observables. In fact, several different sets of molecular orbitals can

lead to the same energy. Nevertheless, they are quite useful. We will use ethylene as an example to illustrate MO concepts.

We often can classify molecular orbitals as sigma or pi orbitals. A sigma (σ) orbital has cylindrical symmetry about the internuclear axis. The hydrogen orbitals below are sigma orbitals. As we saw earlier a bonding orbital has a high electron probability density between the nuclei. An antibonding orbital has a node between adjacent nuclei with lobes of opposite sign (shown as different colors).

Bonding   Antibonding
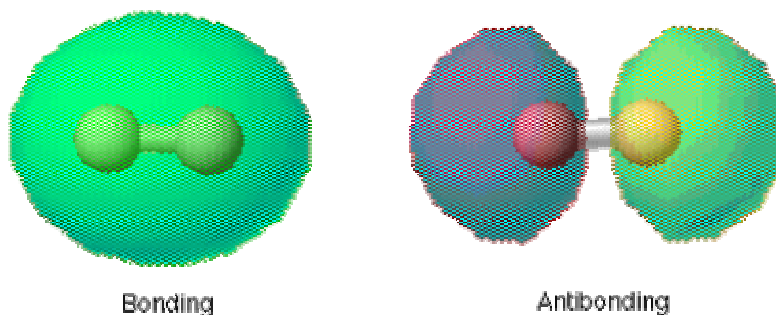
A pi (π) orbital has "up and down" properties like an atomic p-orbital. Bonding and antibonding πorbitals for ethylene are shown below.


Bonding   Antibonding

A few acetone molecular orbital surfaces displayed below, following a PM3 calculation, help illustrate these concepts. The first orbital below is a σ-**orbital** because the lobes are pointing at each other along an internuclear axis and have rotational symmetry about that axis. It is **antibonding** because lobes of a different color are adjacent to each other. The atomic orbitals that

contribute to this molecular orbital are not obvious from the picture.


LUMO +1
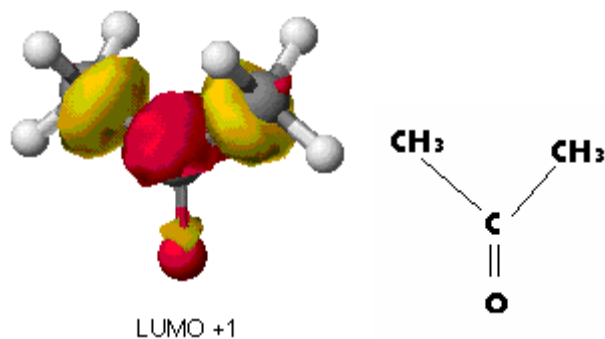
Below is the **lowest unoccupied molecular orbital (LUMO)**. This is a π-**orbital** because the lobes are perpendicular to an internuclear axis. It is **antibonding** because lobes of a different color are adjacent to each other. This molecular orbital appears to be predominantly made of a p-orbital on the central carbon atom and a p-orbital on the oxygen atom.

LUMO
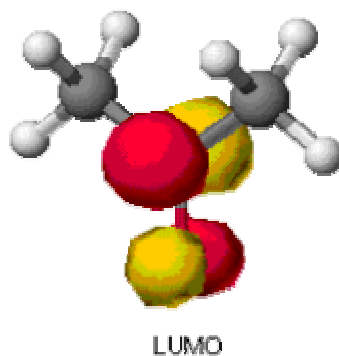
Below is the **highest occupied molecular orbital (HOMO)**. It is a **nonbonding orbital** because the molecular orbital lobes are located on atoms that are not bonded. A nonbonding orbital does not help hold the molecule together. Electrons in a nonbonding orbital are a little like lone pairs of electrons in a Lewis structure. You can see that a p-orbital on the oxygen atom is a major contributor to this molecular orbital.
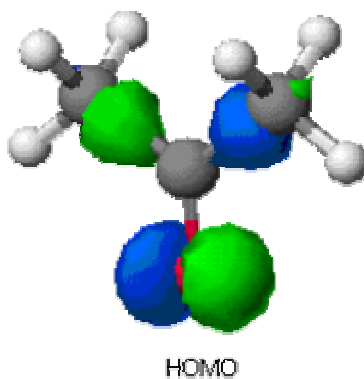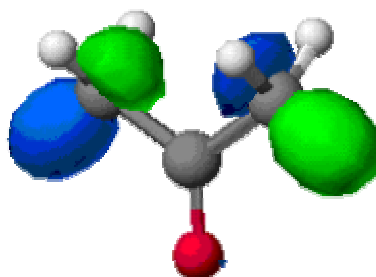

HOMO

This is a π-**orbital**. It is **bonding** because lobes of the same color are next to each other. As you can see, two p-orbitals (one on the carbon atom and one on the oxygen) overlap to form a molecular orbital that distributes electrons around the C=O bond.
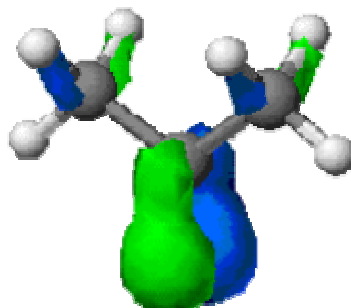
HOMO -1

This is a **bonding** σ-**orbital**. It has two portions, each of which appears to be formed by the overlap of a p-orbital on a carbon atom and an s-orbital on one of the hydrogen atoms.


HOMO -2

Frontier Orbitals and chemical reactivity

The HOMO and LUMO orbitals are commonly known as *Frontier Orbitals* and were found to be extremely useful in explaining chemical reactivity. Electrophilic attacks were shown to correlate very well with atomic sites having high density of the HOMO orbital, whereas nucleophilic attacks correlated very well with atomic sites having high density of the LUMO orbital (Kunichi Fukui was awarded the Nobel prize in chemistry in 1981 for developing this concept). The intuition behind these concepts is simple: chemical bonds are mostly the product of the valance electrons, and the spatial distribution of these electrons is determined by the HOMO orbital. Thus electrophilic attacks are prone to happen on atoms having a large density of

valence electrons, where the HOMO orbital has high values. Similarly, nuleophilic attack can be viewed (conceptually) as an electron transfer reaction from the nucleophil to a molecule. This electron is best positioned in the next empty orbital, the LUMO. But since this electron also binds the molecule to the nucleophil, there is a high probability that the bonding would be to an atom having a large value of the LUMO orbital.

The Electrostatic Potential

You may remember from physics that a distribution of electric charge creates an electric potential in the surrounding space. A positive electric potential means that a positive charge will be repelled in that region of space. A negative electric potential means that a positive charge will be attracted. A molecule is a collection of charges that will have an electric potential - commonly called the "electrostatic potential". The electrostatic potential is a physical property of a molecule related to how a molecule is first "seen" or "felt" by another approaching species[7]. A portion of a molecule that has a negative electrostatic potential will be susceptible to electrophilic attack - the more negative the better. It is not as straightforward to use electrostatic potentials to predict nucleophilic attack.

At right is an electrostatic potential surface of acetone using MM/PM3 geometry with PM3 wavefunction. The surface is color coded according to electrostatic potential (blue is negative and red is positive). What part of the acetone molecule appears to be more susceptible to electrophilic attack?
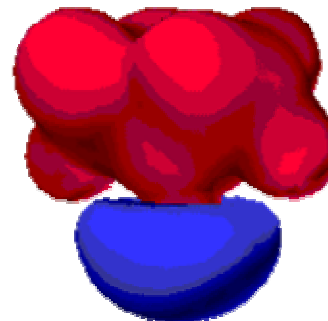
The Electron Density

The electron density surface depicts locations around the molecule where the electron probability density is equal. This gives an idea of the size of the molecule and its susceptibility to electrophilic attack.

Below is an electron density surface of acetone using MM/PM3 geometry with PM3 wavefunction. The surface color reflects the magnitude and polarity of the electrostatic potential. Gray, violet and blue colors correspond to a negative electrostatic potential - regions of the molecule susceptible to electrophilic attack.

	0.090
	0.030
	0.010
	0.000
	-0.010
	-0.030
	-0.060
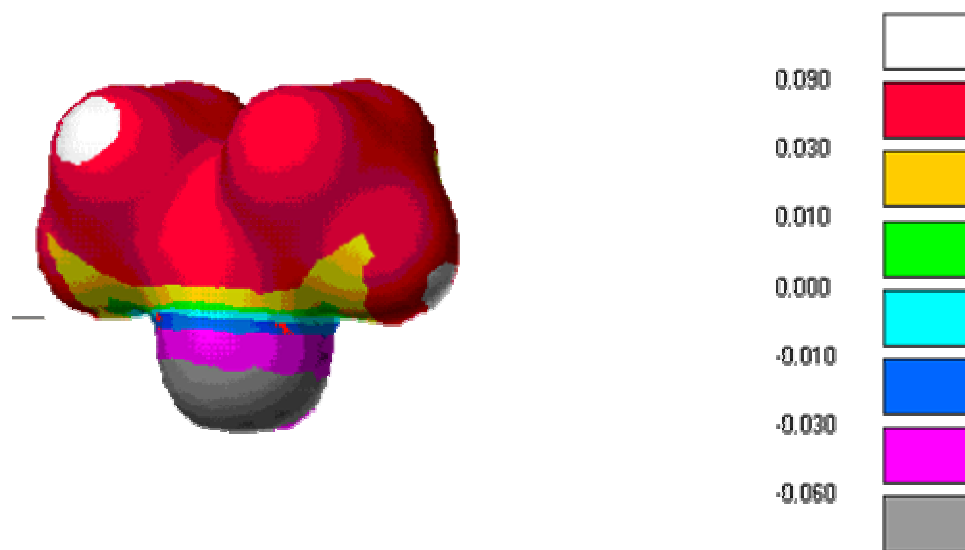
References

[1]J. H. Krieger, "Computational Chemistry Impact", *C&E News*, **1997**, May 12, p 30.; E. K. Wilson, "Computers Customize Combinatorial Libraries", *C&E News*, **1998**, April 27, p 31.

[2]J. B. Foresman and Æ. Frisch, *Exploring Chemistry with Electronic Structure Methods*, Gaussian, Pittsburgh, 1995-96, p. 118.

[3]J. Stewart, *J. Comp. Chem.*, **1989**, *10*, p. 209-220.

[4]P. W. Atkins; R. S. Friedman, *Molecular Quantum Mechanics*, 3rd Ed., Oxford, New York, 1997, p. 276.

[5]M. C. Zerner, *Rev. Comp. Chem.*, Vol. 2, VCH, New York, 1991, pp 313-365.

[6]J. B. Foresman and Æ. Frisch, *Exploring Chemistry with Electronic Structure Methods*, Gaussian, Pittsburgh, 1995-96, p. 113.

[7]P. Politzer and J. S. Murray, *Rev. Comp. Chem.*, Vol. 2, VCH, 1991, p. 273.