Sabbatical Report Exploration of Modern Molecular Modeling and Its Applications

> John E. Thompson Science Division Lane Community College



Sabbatical Report

Exploration of Modern Molecular Modeling and Its Applications

Introduction

Molecular modeling has moved from a supercomputer technique when I was a graduate student to a desktop computing tool for reasonably sized molecules. What had been an expensive technique is now readily available at an affordable price. My sabbatical was spent studying the theory behind molecular modeling and learning how to use Spartan – a personal computer based modeling package. Additionally explorations were also made of other modeling applications including ChemDraw and web-based applications. After my sabbatical application had been approved, the latest version of the Spartan modeling software was purchased. Consequently minor adaptations were made to my planned sabbatical to take advantage of the new features included in the software. This report will follow the sabbatical itself, beginning with a review of the theoretical underpinnings of molecular modeling and then reviewing the modeling applications which were studied.

Theoretical Basis of Molecular Modeling

Molecules can be modeled classically as point masses attached by springs. This approach gives an excellent first approximation of the character of the molecule but leaves out quantum mechanical effects. Including quantum mechanical characteristics, such as delocalized electrons and molecular orbitals, result in equations that cannot be solved analytically. Calculating molecular structure and other characteristics from first principles requires that simplifications be made to solve approximations of these equations. These include the use of symmetry point groups and orbital basis sets as a way to simplify the calculations.

Symmetry point groups are developed from symmetry operations. Symmetry operations are transformations such that the final arrangement is physically indistinguishable from the starting arrangement. The major symmetry operations are:

- \mathcal{G}_n rotation about the central axis (C_n) $2\pi/n$ radians gives a final arrangement that is physically indistinguishable from the starting arrangement.
- σ plane of symmetry where reflection gives a final arrangement that is physically indistinguishable from the starting arrangement.
- *i* center of symmetry where inversion through the center gives a final arrangement that is physically indistinguishable from the starting arrangement.
- \Im_n alternating axis of symmetry (also known as an improper axis or a rotation-reflection axis) where rotation $2\pi/n$ radians followed by reflection gives a final arrangement that is physically indistinguishable from the starting arrangement.

These symmetry operations are labeled with a ^ above the character when used as an operation.

Symmetry point groups are defined as the set of all symmetry operations on a molecule that forms a mathematical group. The major symmetry point group families are:

- 1. Groups with no central axis of symmetry
 - \mathcal{C}_1 the molecule has no symmetry element at all except the identity operation (\mathcal{E} from the German einheit meaning unity).
 - \mathcal{T}_s the molecule has the symmetry element of a plane of symmetry through the molecule, with the symmetry operations being σ and \mathcal{E} .
 - Gi the molecule only has a center of symmetry, with the only symmetry operation of *i*.
- 2. Groups with a single C_n axis (n = 1, 2, 3...)
- 3. Groups with on C_n axis and n C_2 axis (n = 1, 2, 3...)
- 4. Groups with more than on C_n axis (n > 2)
- 5. Linear molecules with rotation about the internuclear axis (n is infinite)

Symmetry point groups can be assigned using VSEPR structures as follows:

VSEPR Structure (symmetric unless noted)	Point
	Group
Linear (asymmetric)	C _∞ √
Linear (symmetric)	D _w v
Tetrahedral	T _d
Trigonal Pyramidal	C _{3v}
Trigonal Planer	C _{3h}
Bent	C _{2v}
Trigonal Bipyramidal	D _{3h}
Trigonal Pyramidal	C _{3v}
See-saw	C _{2v}
Linear	D_{mh}/C_{mv}
Octahedral	O _h
Square Pyramidal	C _{4v}
	0
Square Planer	D _{4h}

Molecular Modeling Methods

Molecular modeling methods can be broken down into three categories: molecular mechanics models, semi-empirical models, and ab *initio* or Hartree-Fock models. Molecular mechanics modeling uses Newtonian mechanics for modeling molecules. Semi-empirical modeling is based off of the Hartree-Fock method, but makes many approximations and uses some empirical data. Ab *initio* or Hartree-Fock models are based on quantum-chemical calculations where basis sets are used to approximate electron orbitals.

Molecular mechanics modeling is the least accurate of the three categories but is a significant technique as it requires the least computational power and can be used for molecules of almost any size. The potential energy is calculated using force fields assuming each atom is a particle of a given radius (usually the van der Waals radius) and bonds are treated as springs with an equilibrium length equal to the bond length. The force fields used are key to the success of the calculation and are generally calculated for a class of molecules such as proteins. This method can be applied to proteins, which are too large for other methods and can be used to predict protein folding as well as the induced fit docking of reactive molecules and potential medicinal molecules. The strengths of this method are that it uses minimal computational time, can work with large molecules, and gives reasonably accurate results for energy and van der Waals interactions. The weaknesses are that the parameters must be given for each element and the number of parameters rises rapidly for non-CHNO elements, and it does not include the majority of the cross terms in the force field.

Semi-empirical modeling uses a blend of quantum mechanical calculations and empirical data which is used for mid-sized molecules up to approximately 300 atoms in size. For the semi-empirical method the calculation is simplified by omitting the two-electron portion of the Hamiltonian. To correct for this, the method is parameterized or fitted by a set of parameters so as to produce results that agree with experimental data. Consequently the quality of the parameter set determines the quality of the calculation. These methods have been shown to be most effective for organic molecules which are composed primarily of only a few types of atoms ranging from small to moderate in size. Semi-empirical modeling techniques include MNDO, AM1, RM1, and PM3. Semi-empirical calculations require more computational time than molecular mechanics modeling but less than ab *initio* modeling techniques.

Ab initio or Hartree-Fock modeling methods are calculation techniques from first principles of quantum mechanics and are only practical for molecules smaller than 200 atoms. Since the Schrödinger equation cannot be solved exactly for anything beyond the smallest molecules, basis sets are used to make the equation solvable. These basis sets are usually built from linear combinations of atomic orbitals (LCAOs). The accuracy of the calculation is dependent upon the quality of the basis sets used. The most common basis set is the 3-21G basis which uses 3 Gaussion type orbitals (GTOs) and is a double zeta set with 2 gaussian functions that comprise the first slater type orbital (STO) and 1 gaussion function in the second STO. For more accurate calculations, especially with larger atoms, the 6-21G basis set is used. This basis set is composed of 6 GTOs and is also a double zeta set with 3 gaussian function that comprise the first STO and 1 gaussion function in the second STO. The basis set allows the Schrödinger equation to be solved as an eigenvalue equation of the electronic molecular Hamiltonion with a discrete set of solutions. These calculations are often done using a Monte Carlo technique where the calculation is made to converge to an exact solution; this solution is then perturbed and made to converge again. This process is repeated until the best solution is found. The accuracy of these calculations is dependent upon the basis set and are not corrected or limited by experimental data. While the Hartree-Fock method is the best known, other ab initio techniques are also used including configuration interaction (CI), Quadratic Configuration Interaction with Single and Double excitation (QCISD), and Møller-Plesset Models (MP2, MP3, MP4). Ab initio calculations use a great deal of computational time, memory, and disk space.

Frontier Molecular Orbitals

Frontier molecular orbitals (FMOs) are the HOMO of the nucleophile and the LUMO of the electrophile / substrate. The concept of the FMOs was developed by Kenichi Fukui and published in J. Chem. Phy. In 1952. The concept was not well received initially but has proven to be quite valuable over time. Roald Hoffman was awarded the Nobel Prize in Chemistry in 1981 for his FMO work.

FMOs are based off of three main observations:

- 1) Occupied orbitals of different molecules repel each other.
- 2) Positive charges on one molecule attract negative charges of the other.
- 3) The occupied orbitals of one molecule and the unoccupied orbitals of the other (especially the HOMO and LUMO) interact with each other causing attraction.

FMO theory simplifies reactivity to the HOMO and LUMO.

FMO theory helped to explain the Woodward-Hoffman rules for thermal pericyclic reactions which state that a ground state pericyclic reaction is symmetry allowed when the total number of (4q + 2) and (4r) components are odd. (4q+2) refers to the number of aromatic superfacial electron systems and (4r) refers to antiaromatic superfacial electron systems.

FMO is a powerful model that can effectively predict bonding behavior for many cases however its weaknesses are that the calculations can be somewhat complex and in some cases the frontier orbitals cannot be easily identified. FMO are also not useful in cases where aromaticity will be lost in the reaction.

Electron Density and Molecular Shape

The size of atoms and molecules is difficult to quantify as electrons can be found quite far from the nucleus/molecule. The classical model is the space-filled or CPK model where each atom is a sphere of a fixed radius. Different spheres are used for different applications (i.e. – atomic radius). These spaced-filled spheres are interpreted for bonding as well giving covalent or ionic radii.

Electron density surfaces are commonly set at a limit of 95% of electron density to define the atomic or molecular size. Molecular size calculations assign the electrons to the whole molecule, not to atoms. The density plot gives a working size and shape.

Bond density surfaces show regions of high electron density only showing the "bonds". This calculation shows the bonding region well, however the overall electron density is poorly represented showing the surface smaller than actual size. Also this computation does not differentiate between sigma and pi bonding, but rather shows a composite. This computation is particularly useful for studying transition state to determine if the transition structure is more like the reactant(s) or the product(s).

Electrostatic Potential Maps

Charge distribution models tend not to be accurate although they are easy to use. An example is the Lewis model of formal charges. This works well as a first approximation but is not accurate enough to

use for calculations. Even quantum mechanical models have their limitations. Electrostatic potential maps are an accurate alternative for mapping surfaces charges. These maps are represented by a rainbow of colors which represent the charge separation across the surface where hot colors indicate partial negative charge and cool colors represent partial positive charge. This map is usually laid over the space-filled surface model. This model is then used for such things as predicting polarity, reactive sights, and crystal packing.

Molecular Modeling Application

A good portion of my sabbatical was spent working with the Spartan '08 software from Wavefunction, Inc. This is professional software used both in academia and industry and is capable of substantial (powerful) computations. As such it is quite complex and requires a significant amount of practice to master. Much of my time was spent working through tutorials and practicing using the software to build molecules of interest, working through the major features of the software. The skills learned include basic skills such as building, viewing, and editing molecules through more complicated skills such as modeling strategies, working with multiple molecules simultaneously, and animations.

Several discussions were had with my advisors on this project: Dr. Alan Shustermann and Nadine Fattaleh. Conversations with both of them were quite valuable in the success of this project and have led to the possibility of future collaborations with each of them. Thank you both for your help throughout this project.