

Introduction to Molecular Modeling

One of the most important concepts in molecular modeling is the relationship between energy and structure. Mathematically this is described by a potential energy surface (Figure 1). A familiar potential energy surface is the barrier for torsion about the C2-C3 bond in butane. Another potential energy surface of interest is a typical reaction diagram.

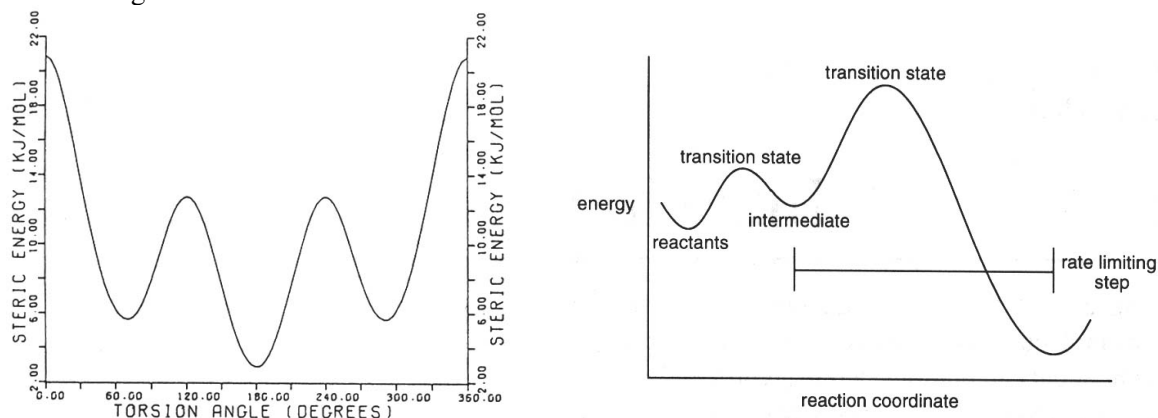


Figure 1

The global minimum on a potential energy surface is the minimum with the lowest energy. Finding this global minimum is a difficult problem in many computations. The protein folding problem is a classic example. We may know the sequence of a protein but derivation of secondary or tertiary structure is a computational challenge due to the vast numbers of local minima. On a smaller scale, consider the conformations of cyclohexane. The chair form is lowest in energy but the twist-boat is a local minimum. Computer algorithms designed to find minima tend to go downhill on a potential energy surface to reach the *nearest* minimum. The nearest minimum may not be the global minimum! Strategies exist to explore potential energy surfaces in search of a global minimum. Such strategies must be applied to any but the most rigid of systems.

Molecular mechanics are non-quantum methods based on a Hooke's Law approximation for the potential energy surface. Hooke's Law is a simple equation ($E=k(r-r_0)^2$) which describes the energetics of spring elasticity (dashed line in Figure 2). Two balls connected by a spring will have an equilibrium distance, r_0 . Distortion from this value results in higher energy. The energy increases as a function of k , the spring elasticity constant. The energetics of real diatomic behavior is best described by the solid line in Figure 2, and Hooke's Law is a useful approximation of this behavior for relatively low energy states (shaded region). In molecular mechanics, r_0 is the idealized equilibrium value for a bond length and k is the restoring force constant for distortion from the idealized value.

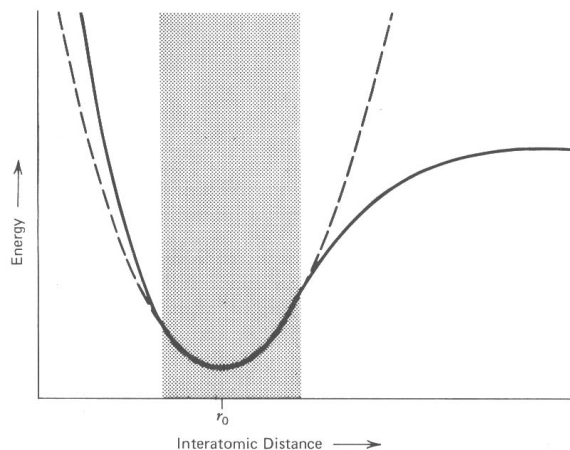


Figure 2

There is usually a set of terms for every bond, angle, and torsion in the molecule. In addition, there are usually terms for Van der Waals and electrostatic interactions.

The specific set of actual k and r_0 values developed to reproduce experimental data for a class of structures is called the force field. AMBER is a popular force field developed for proteins and nucleic acids. The k and r_0 values were derived to match experimental values for a training set of biomolecules. It is assumed that these values are transferable and useful for modeling other biomolecules which were not in the original training set. MM3 is a force field similarly derived to reproduce small organic molecules. It is advisable to use the force field most suitable for your system. Many commercial software packages provide generic force fields for molecular systems well beyond the scope of their training sets. Results derived from such generic force fields are best viewed with skepticism. Unfortunately, there are few force fields which can handle systems containing metals. Also, force fields provide little reliable information about transition-state geometries but do quite well at reproducing equilibrium geometries and relative energies of conformations. Very large molecules can be modeled with molecular mechanics.

Quantum approaches start with the Schrödinger equation. Unfortunately this equation cannot be solved exactly for systems bigger than a single H atom. Approximate solutions to the Schrödinger equation exist. Almost all methods start with the Born-Oppenheimer approximation which assumes that nuclear motion does not affect the electronic wave function. After that many methods use the Hartree-Fock approximation. This treats multiple electron wave functions as a product of one-electron wave functions. In practice, this is implemented as a linear combination of atomic orbitals (LCAO model). Mathematically, the individual atomic orbitals are actually described using a combination of Gaussian functions. The set of Gaussian functions used to make up the atomic orbitals is known as the basis set. Implementation of these approximations is called the self consistent field (SCF) or Hartree-Fock approach (HF). SCF refers to the method of iteratively solving electronic interactions between each single electron and a mean field composed of the other electrons.

An ideal solution to the Hartree-Fock approach would include an infinite basis set. Real basis sets are smaller and can have confusing terminology. Here's an example breakdown of a real basis set: 6-31G*+. This includes 6 Gaussian functions per atomic orbital for inner shells. Valence orbitals are treated by 3 functions with one value and 1 function of another value. The * is called a polarization basis set and indicates that d functions are added in for 2nd row atoms and p functions for hydrogens. Polarization basis sets are commonly used for better descriptions of H-bonded interactions. The + indicates that additional diffuse functions are added in for heavy atoms. These provide better treatment of charged systems. Overall, a basis set is chosen with an understanding of the type of molecular system under examination along with the level of accuracy required. A reasonable summary of this information is available at http://gaussian.com/g_ur/m_basis_sets.htm in the Gaussian '03 technical support material.

Electronic correlation is a correction applied to deal with the fact that SCF calculations tend to overestimate electronic repulsion. Hartree-Fock approaches assume electron motions are independent of each other but they are correlated. Electrons move in such a way as to minimize repulsions. The most common correction is based in Moller-Plesset theory and is commonly described as MP2 (for 2nd order correction). This approach involves including excited-state wave functions in the calculation. When studying bond-breaking or making, correlation corrections are important. Isodesmic equations are the exception in which the number and type of bonds of both sides of the equation are the same. These can be modeled with HF theory.

Semi-empirical calculations are low level quantum methods. The inner core of electrons is treated with empirical parameters (which implicitly include electronic correlation). Only the valence electrons are explicitly calculated using the Hartree-Fock (HF) approach. Further, the valence orbitals are described by only a very minimal basis set. Such a basis set has just enough atomic orbitals to accommodate the number of electrons in the system. These methods are useful for moderately sized systems and include parameterization for many metals.

DFT (density functional theory) begins with the relationship in which the minimal energy of a multi-electron system in a Coulombic field is seen to be a unique function of the electron density. In this approach, correlation is implicitly treated. These calculations scale less quickly than HF and MP/HF approaches and so are amenable to

larger systems. These methods are quite popular, particularly for systems containing metals. They are implemented in many modern computational software packages. Density functional models often require numerical integration techniques due to the nature of the correlation and exchange functionals. This can result in somewhat larger uncertainties in the calculated values than compared to SCF methods.

A common simple implementation of DFT is the local spin density model. This model is best for systems of uniform electron density. For more complex systems, non-local methods are preferred. More sophisticated are methods which add a correction based on the gradient of the density. Other higher level functionals add more corrections. It is also quite common to see hybrid SCF/DFT methods.

In general it is worthwhile to check the literature to see what methods have been used successfully for systems like yours. As an example, B3LYP is a commonly used functional which gives good results for a wide variety of species including metals.

Final note: *The choice of the right method or methods is the key to a useful calculation.*

References:

A. Hinchliffe, *Modeling Molecular Structures*. Wiley 1995.

W. Hehre *et al.*, *Experiments in Computational Organic Chemistry*. A publication of [Wavefunction Inc.](#), (the makers of Spartan).

J.M. Thijssen, *Computational Physics*, Cambridge University Press, 1999.

Wolfram Koch and Max C. Holthausen, *A Chemist's Guide to Density Functional Theory*. Wiley-VCH, Weinheim, 2000.

C.J. Cramer, *Essentials of Computational Chemistry*. Wiley, 2002.

F. Jensen, *Introduction to Computational Chemistry*. Wiley, 2002.