Colby College Molecular Mechanics Tutorial
QUANTA Version

November 2003

Thomas W. Shattuck
Department of Chemistry
Colby College
Waterville, Maine 04901

Table of Contents

Introduction to Molecular Mechanics
Section 1

Summary The goal of molecular mechanics is to predict the detailed structure and physical properties of molecules. Examples of physical properties that can be calculated include enthalpies of formation, entropies, dipole moments, and strain energies. Molecular mechanics calculates the energy of a molecule and then adjusts the energy through changes in bond lengths and angles to obtain the minimum energy structure.

Steric Energy

A molecule can possess different kinds of energy such as bond and thermal energy. Molecular mechanics calculates the steric energy of a molecule--the energy due to the geometry or conformation of a molecule. Energy is minimized in nature, and the conformation of a molecule that is favored is the lowest energy conformation. Knowledge of the conformation of a molecule is important because the structure of a molecule often has a great effect on its reactivity. The effect of structure on reactivity is important for large molecules like proteins. Studies of the conformation of proteins are difficult and therefore interesting, because their size makes many different conformations possible.

Molecular mechanics assumes the steric energy of a molecule to arise from a few, specific interactions within a molecule. These interactions include the stretching or compressing of bonds beyond their equilibrium lengths and angles, torsional effects of twisting about single bonds, the Van der Waals attractions or repulsions of atoms that come close together, and the electrostatic interactions between partial charges in a molecule due to polar bonds. To quantify the contribution of each, these interactions can be modeled by a potential function that gives the energy of the interaction as a function of distance, angle, or charge[1,2]. The total steric energy of a molecule can be written as a sum of the energies of the interactions:

$$E_{\text{steric energy}} = E_{\text{str}} + E_{\text{bend}} + E_{\text{str-bend}} + E_{\text{oop}} + E_{\text{tor}} + E_{\text{VdW}} + E_{\text{qq}} \qquad (1)$$

The bond stretching, bending, stretch-bend, out of plane, and torsion interactions are called bonded interactions because the atoms involved must be directly bonded or bonded to a common atom. The Van der Waals and electrostatic (qq) interactions are between non-bonded atoms.

Bonded Interactions

E_{str} represents the energy required to stretch or compress a bond between two atoms, Figure 1.

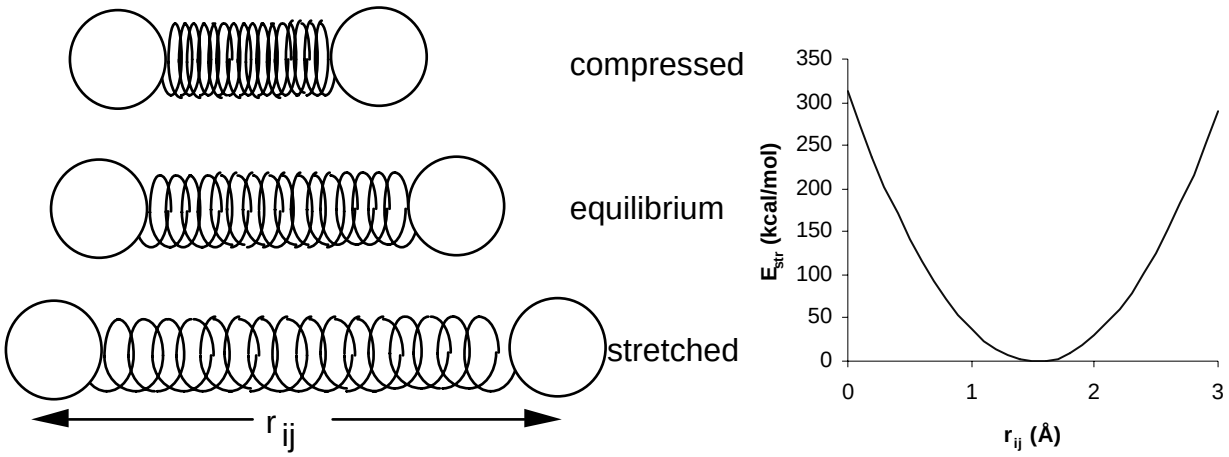


Figure 1. Bond Stretching

A bond can be thought of as a spring having its own equilibrium length, r_o, and the energy required to stretch or compress it can be approximated by the Hookian potential for an ideal spring:

$$E_{str} = 1/2\ k_{s,ij}\ (\ r_{ij} - r_o\ )^2 \qquad (2)$$

where $k_{s,ij}$ is the stretching force constant for the bond and r_{ij} is the distance between the two atoms, Figure 1.

E_{bend} is the energy required to bend a bond from its equilibrium angle, θ_o. Again this system can be modeled by a spring, and the energy is given by the Hookian potential with respect to angle:

$$E_{bend} = 1/2\ k_{b,ijk}\ (\ \theta_{ijk} - \theta_o\ )^2 \qquad (3)$$

where $k_{b,ijk}$ is the bending force constant and θ_{ijk} is the instantaneous bond angle (Figure 2).

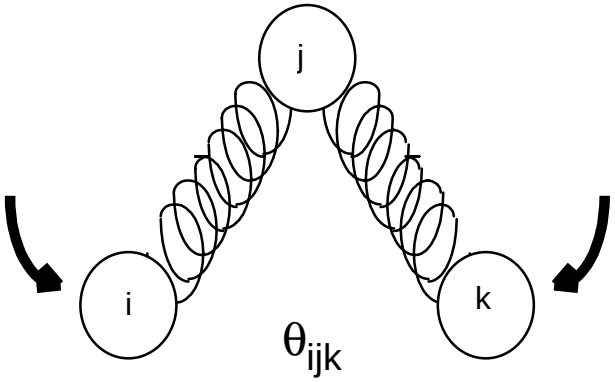


Figure 2. Bond Bending

$E_{str\text{-}bend}$ is the stretch-bend interaction energy that takes into account the observation that when a bond is bent, the two associated bond lengths increase (Figure 3). The potential function that can model this interaction is:

$$E_{str\text{-}bend} = 1/2\ k_{sb,ijk}\ (\ r_{ij} - r_o\ )\ (\theta_{ijk} - \theta_o\ ) \qquad (4)$$

where $k_{sb,ijk}$ is the stretch-bend force constant for the bond between atoms i and j with the bend between atoms i, j, and k.

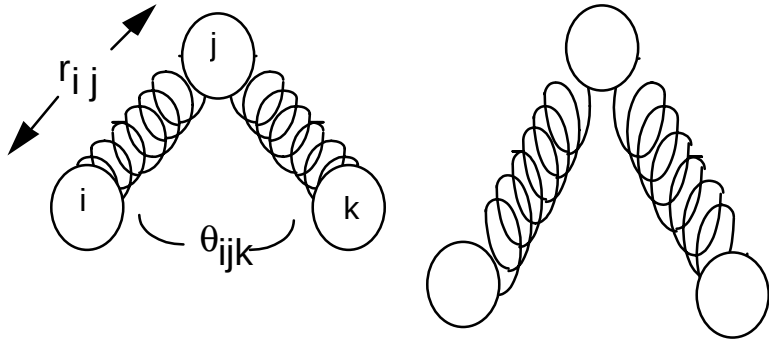


Figure 3. Stretch-Bend Interaction

E_{oop} is the energy required to deform a planar group of atoms from its equilibrium angle, ω_o, usually equal to zero.[3] This force field term is useful for sp^2 hybridized atoms such as doubly bonded carbon atoms, and some small ring systems. Again this system can be modeled by a spring, and the energy is given by the Hookian potential with respect to planar angle:

$$E_{oop} = 1/2\ k_{o,ijkl}\ (\ \omega_{ijkl} - \omega_o\ )^2 \qquad (5)$$

where $k_{o,ijkl}$ is the bending force constant and ω_{ijkl} is the instantaneous bond angle (Figure 4).

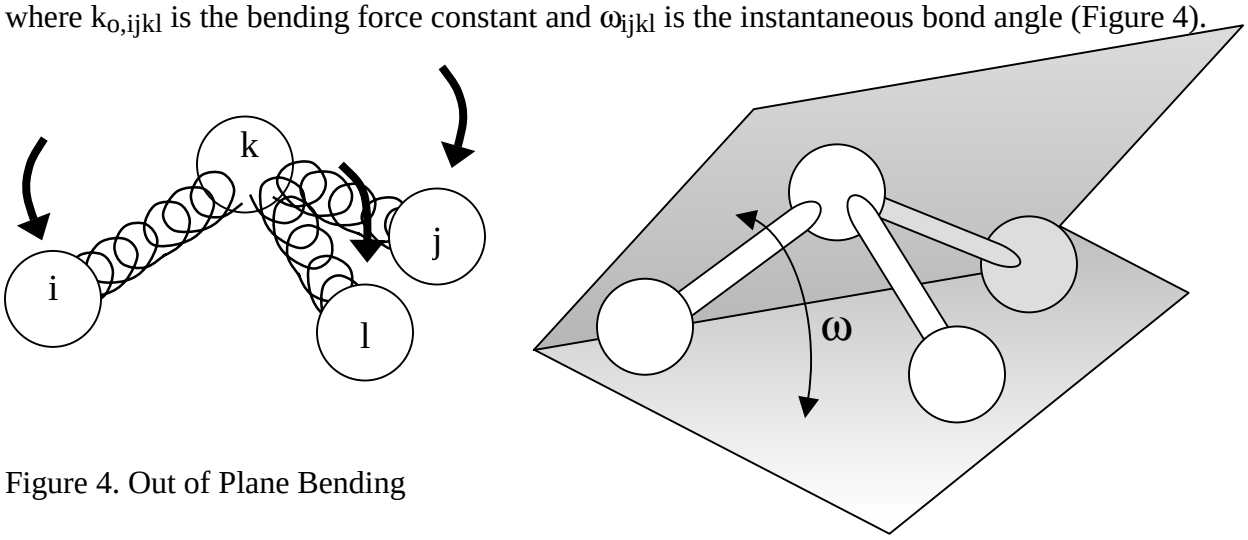


Figure 4. Out of Plane Bending

The out of plane term is also called the improper torsion in some force fields. The oop term is called the improper torsion, because like a dihedral torsion (see below) the term depends on four atoms, but the atoms are numbered in a different order. Force fields differ greatly in their use of oop terms. Most force fields use oop terms for the carbonyl carbon and the amide nitrogen in peptide bonds, which are planar (Figure 5).

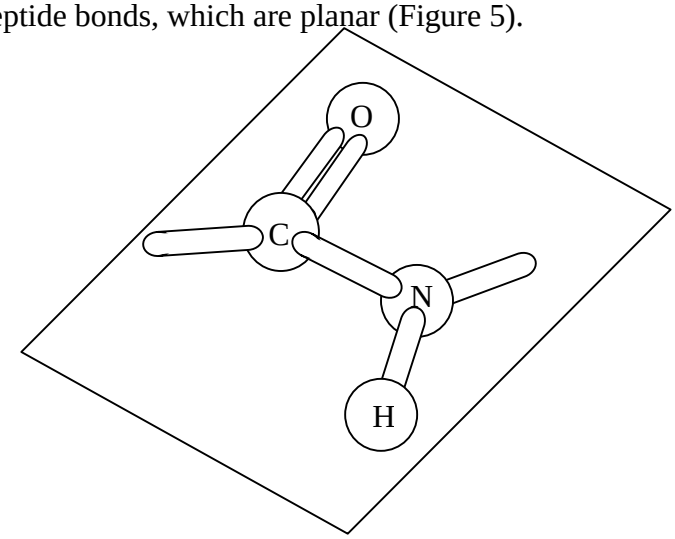


Figure 5. Peptide Bond is Planar.

Torsional Interactions: E_{tor} is the energy of torsion needed to rotate about bonds. Torsional energies are usually important only for single bonds because double and triple bonds are too rigid to permit rotation. Torsional interactions are modeled by the potential:

$$E_{tor} = 1/2\ k_{tor,1}\ (1 - \cos \phi\ ) + 1/2\ k_{tor,2}\ (1 - \cos 2\ \phi\ ) + 1/2\ k_{tor,3}\ (\ 1 - \cos 3\ \phi\ ) \qquad (6)$$

The angle ϕ is the dihedral angle about the bond. The three-fold term, that is the term in 3ϕ, is important for sp^3 hybridized systems (Figure 6a and b). The two-fold term, in 2ϕ, is needed for

halogens, for example F-C-C-F, and sp^2 hybridized systems, such as C-C-C=O and vinyl alcohols[1]. The one-fold term in just ϕ is useful for alcohols with the C-C-O-H torsion, carbonyl torsions like C-C-C(carbonyl)-C, and to a lesser extent even the central bond in molecules such as butane that have C-C-C-C frameworks(Figure 6c). The constants $k_{tor,1}$, $k_{tor,2}$ and $k_{tor,3}$ are the torsional constants for one-fold, two-fold and three-fold rotational barriers, respectively.

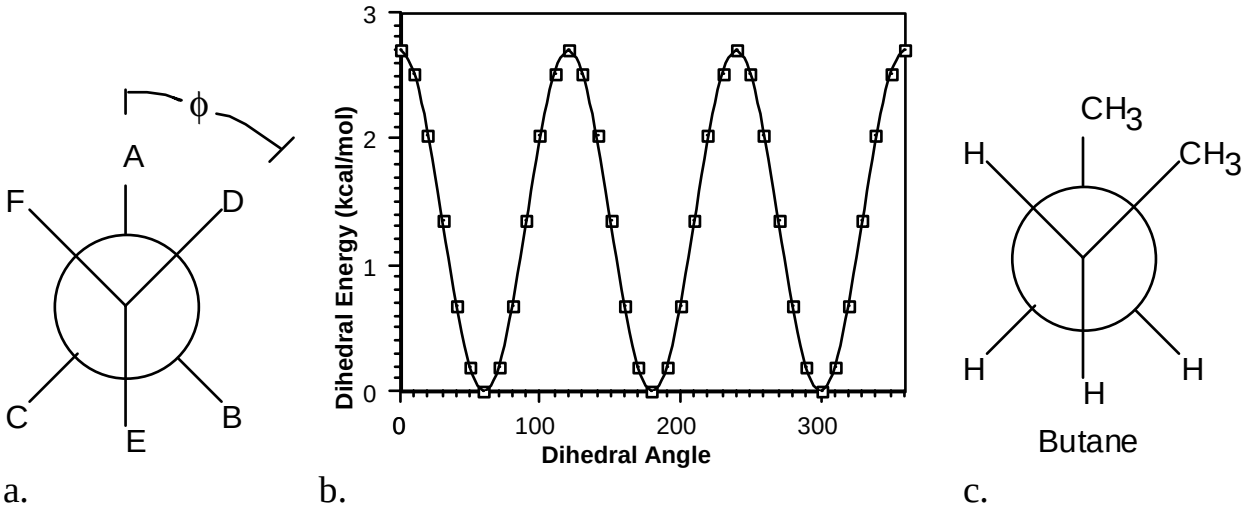


Figure 6. Torsional Interactions, (a) dihedral angle in sp^3systems. (b) three-fold, 3ϕ, rotational energy barrier in ethane. (c) butane, which also has a contribution of a one fold, ϕ, barrier.

The origin of the torsional interaction is not well understood. Torsion energies are rationalized by some authors as a repulsion between the bonds of groups attached to a central, rotating bond (i.e., C-C-C-C frameworks). Torsion terms were originally used as a fudge factor to correct for the other energy terms when they did not accurately predict steric energies for bond twisting. For example, the interactions of the methyl groups and hydrogens on the "front" and "back" carbons in butane were thought to be Van der Waals in nature (Figure 7). However, the Van der Waals function alone gives an inaccurate value for the steric energy.

Bonded Interactions Summary: Therefore, when intramolecular interactions stretch, compress, or bend a bond from its equilibrium length and angle, the bonds resist these changes with an energy given by the above equations summed over all bonds. When the bonds cannot relax back to their equilibrium positions, this energy raises the steric energy of the entire molecule.

Non-bonded Interactions

Van der Waals interactions, which are responsible for the liquefaction of non-polar gases like O_2 and N_2, also govern the energy of interaction of non-bonded atoms within a molecule. These interactions contribute to the steric interactions in molecules and are often the most important factors in determining the overall molecular conformation (shape). Such interactions are extremely important in determining the three-dimensional structure of many biomolecules, especially proteins.

A plot of the Van der Waals energy as a function of distance between two hydrogen atoms is shown in Figure 7. When two atoms are far apart, an attraction is felt. When two atoms are very close together, a strong repulsion is present. Although both attractive and repulsive forces exist,

the repulsions are often the most important for determining the shapes of molecules. A measure of the size of an atom is its Van der Waals radius. The distance that gives the lowest, most favorable energy of interaction between two atoms is the sum of their Van der Waals radii. The lowest point on the curve in Figure 7 is this point. Interactions of two nuclei separated by more than the minimum energy distance are governed by the attractive forces between the atoms. At distances smaller than the minimum energy distance, repulsions dominate the interaction. The formula for the Van der Waals energy is:

$$E_{VdW,ij} = -\frac{A}{r_{ij}6} + \frac{B}{r_{ij}12} \qquad (7)$$

where A and B are constants dependent upon the identities of the two atoms involved and r_{ij} is the distance, in Angstroms, separating the two nuclei. This equation is also called the Lennard-Jones potential. Since, by definition, lower energy is more favorable, the $-A/r^6$ part is the attractive part and the $+B/r^{12}$ part is the repulsive part of the interaction. For two hydrogen atoms in a molecule:

$$A = 70.38 \text{ kcal Å}^6 \qquad B = 6286. \text{ kcal Å}^{12}$$

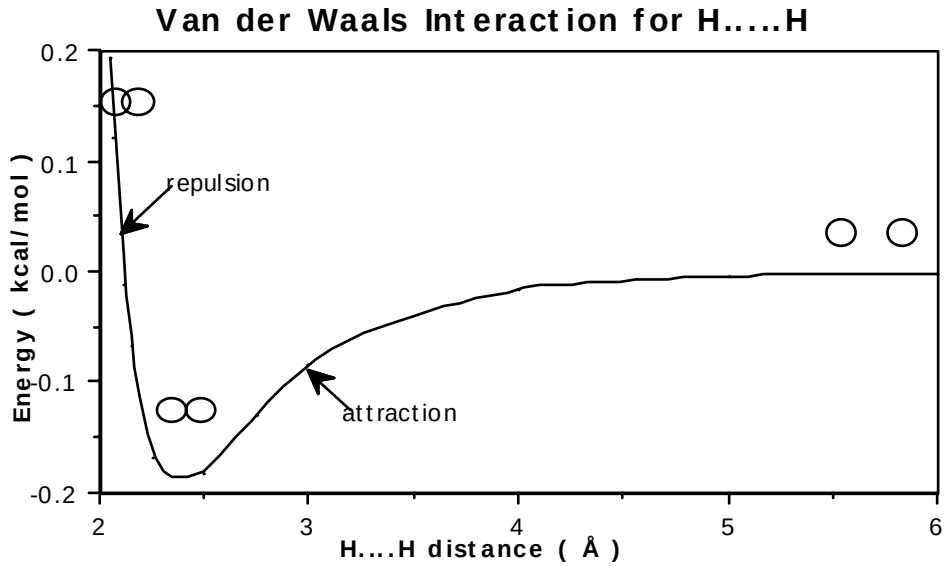


Figure 7: Van der Waals interactions between two hydrogen atoms in a molecule, such as H_2O_2 or CH_3-CH_3

An equivalent and commonly used form of the Lennard-Jones potential is

$$E_{VdW,ij} = \varepsilon\left[-\left(\frac{r_o}{r_{ij}}\right)6 + \left(\frac{r_o}{r_{ij}}\right)12\right] \qquad (8)$$

Where ε is the minimum energy and r_o is the sum of the Van der Waals radii of the two atoms, $r_i + r_j$. Comparing Eq 7 and 8 gives $A = 2\, r_o^6\, \varepsilon$ and $B = r_o^{12}\, \varepsilon$. For two hydrogens, as in Figure 7, ε = 0.195 kcal/mol and r_o = 2.376 Å. When looking for close contacts between atoms it is best to use the hard-core Van der Waals radius, σ_{HC} . This distance is the point where the Van der Waals potential is zero. When two atoms are closer than the sum of their σ_{HC} values then strong repulsions are present. For an atom $\sigma_{HC} = 2^{-1/6}\, r_i$.

Electrostatic Interactions: If bonds in the molecule are polar, partial electrostatic charges will reside on the atoms. The electrostatic interactions are represented with a Coulombic potential function:

$$E_{qq,ij} = \frac{k\ Q_i\ Q_j}{4\pi\varepsilon\ r_{ij}} \qquad (9)$$

The Q_i and Q_j are the partial atomic charges for atoms i and j separated by a distance r_{ij}. ε is the relative dielectric constant. For gas phase calculations ε is normally set to1. Larger values of ε are used to approximate the dielectric effect of intervening solute or solvent atoms in solution. k is a units conversion constant; for kcal/mol, k=2086.4. Like charges raise the steric energy, while opposite charges lower the energy. The Del Re method is often used for estimating partial charges. The Coulomb potential for a unit positive and negative charge is shown in Figure 8a and the Coulomb potential for the hydrogens in H_2O_2 is shown in Figure 8b.

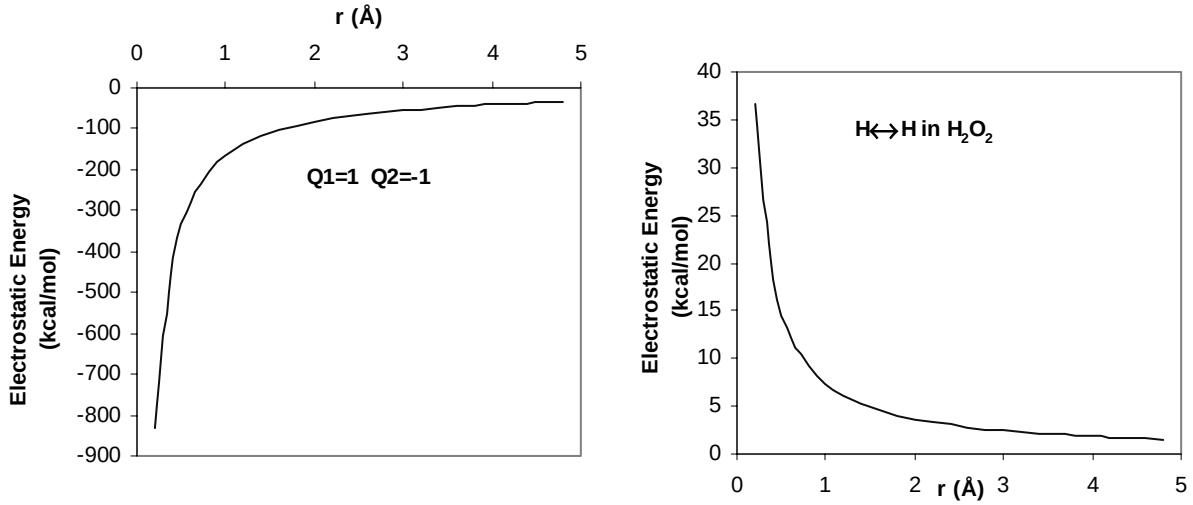


Figure 8. (a) Coulomb attraction of a positive and a negative charge. (b) Coulomb repulsion of the two hydrogens in H_2O_2, with the charge on each hydrogen as $Q_1 = Q_2 = 0.210$.

Nonbonded Summary: The Van der Waals and electrostatic potential functions represent the various non-bonded interactions that can occur between two atoms i and j. A full force field determines the steric energy by summing these potentials over all pairs of atoms in the molecule.

The bond stretching, bond bending, stretch-bend, out-of-plane, torsion, Van der Waals, and electrostatic interactions are said to make up a force field. Each interaction causes a steric force that the molecule must adjust to in finding its lowest energy conformation.

Empirical Force Fields

All the potential functions above involve some force constant or interaction constant. Theoretically, these constants should be available from quantum mechanical calculations. In practice, however, it is necessary to derive them empirically. That is, the constants are adjusted so that the detailed geometry is properly predicted for a number of well known compounds. These constants are then used to calculate the structures of new compounds. The accuracy of these constants is critical to molecular mechanics calculations. Unfortunately, no single best set of force constants is available because of the diversity of types of compounds. For example, the MM2 force field works best on hydrocarbons because most of the known compounds used in deriving the force field were hydrocarbons[1]. MM2 is less accurate for oxygen-containing

compounds and even less reliable for nitrogen and sulfur species. This is because there aren't as many hetero-atom containing compounds in the learning set for MM2 and hydrocarbons are a more homogeneous class of compounds than substances with hetero-atoms. However, the MM2 force field is one of the best available and the most widely accepted force field for use with organic compounds. MM2 was specifically parameterized to reproduce experimental enthalpies of formation.[1]

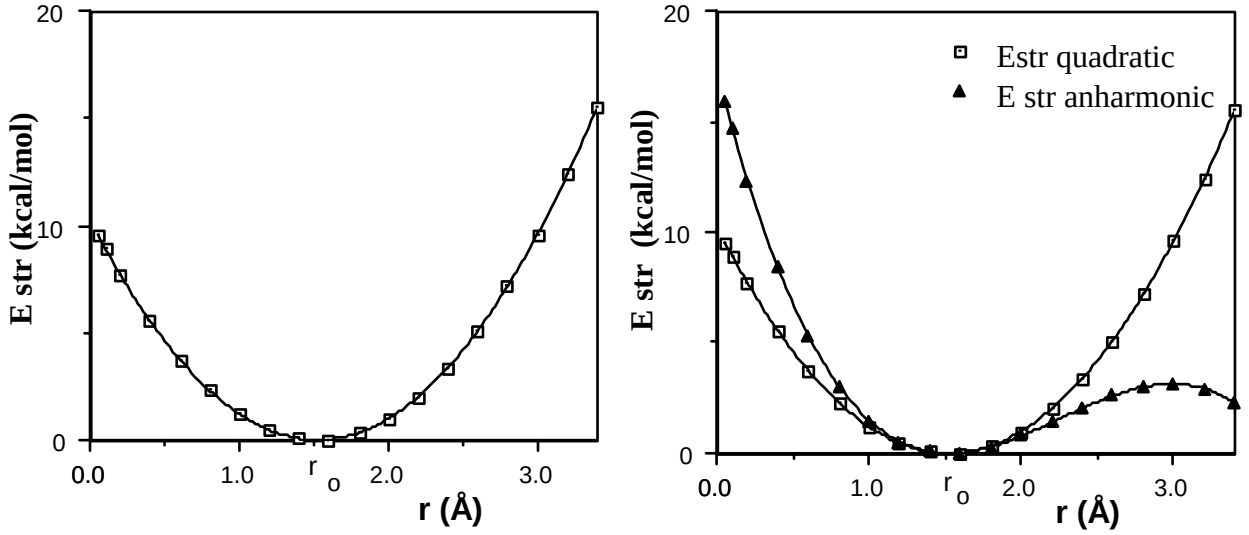
It is important to realize that the force field is not absolute, in that not all the interactions listed in Equation 1 may be necessary to accurately predict the steric energy of a molecule. On the other hand, many force fields use additional terms. For example, MM2 adds terms to the bonded interactions to better approximate the real potential function of a chemical bond. These additional terms take into account anharmonicity, which is a result of the fact that given enough vibrational energy, bonds will break. Purely quadratic potentials have steep "walls" that prevent bond dissociation (Figure 9a). Cubic terms are added to Equation 2 to adjust for this:

$$E_{str} = 1/2\ k_{s,ij}\ (\ r_{ij} - r_o\ )^2 + 1/2\ k_{s,ij}\ \chi_{ij}\ (\ r_{ij} - r_o\ )^3 \qquad (10)$$

where χ_{ij} is the anharmonicity constant. For example, for a C(sp^3)-C(sp^3) bond the anharmonicity is -2, see Figure 9b:

$$E_{str} = [4.40\ \text{mdynes/Å}]\ (\ r - 1.532\text{Å}\ )^2 + [4.40\ \text{mdynes/Å}]\ \ [-2.00]\ (\ r - 1.532\text{Å}\ )^3 \qquad (11)$$

The addition of the cubic term makes the small r portion steeper or more repulsive. This is realistic for real bonds. At larger r the curve is less steep, as desired. For r very large (r> 3Å) the energy decreases, which is unphysical; the curve should approach a constant value. Even though the large r behavior is incorrect, the bond length in compounds remains less than this value, so this region is unimportant under normal conditions. Some force fields add a quartic term, $(\ r_{ij} - r_o\ )^4$, to help improve the potential function even further.



a. b.

Figure 9. (a). Energy for the stretching of a C-C bond with only the $(r\text{-}r_o)^2$ harmonic term., Eq. 2 (b), Comparison of the harmonic term with Eq. 8, which includes the $(r\text{-}r_o)^3$ term for anharmonicity.

Force Field Atom Types and Parameters

MM2 is a good example of a molecular mechanics force field. The force constants will give a good idea of what typical force constants are like. The first step in starting a calculation is to identify the different atom types in the molecule. In some programs this must be done manually by the user. In many programs a routine does this step automatically. However, automatic atom type assignments can be incorrect, and the user should check to make sure the atom types are assigned properly. A list of some MM2 atom types is given in Table 1.

Table 1. MM2 Atom types. The typical atom symbol is listed and the radius used in the Van der Waals force field term and approximate Van der Waals radii for judging close contacts.

Atom Type	atom	Description	Type	R (Å)	σ_{HC} (Å)
1	C	C(sp^3)	C	1.969	1.75
2	C	C(sp^2) alkene	Csp2	2.097	1.87
3	C	C(sp^2) carbonyl	C=	1.992	1.77
4	C	C(sp) alkyne; C=C=O	Csp	2.077	1.85
5	H	Attached to C and Si	HC	1.485	1.32
6	O	C-O-H, C-O-C	O	1.779	1.58
7	O	=O carbonyl	O=	1.746	1.56
8	N	N(sp^3)	N	2.014	1.79
9	N	N(sp^2) amide	NC=O	1.894	1.69
10	N	N(sp)	#N	1.945	1.73
11	F	Fluoride	F	1.496	1.33
12	Cl	Chloride	CL	2.044	1.82
15	S	-S- sulfide	S	2.185	1.95
16	S+	>S+, sulfonium	>S+	2.333	2.08
17	S	>S=O, sulfoxide	>SO	2.128	1.90
18	S	>SO2, sulfone	SO2	1.998	1.78
20	LP	Lone pair	LP	1.969	1.75
21	H	-OH alcohol	HO	1.307	1.16
22	C	cyclopropane	CR3R	1.992	1.77
23	H	NH amine	HN	1.307	1.16
24	H	COOH carboxyl	HOCO	1.307	1.16
28	H	H on N(sp2); amide	HN2	1.307	1.16
36	H	ammonium	HN+	1.497	1.33
37	N	-N= ; pyridine	NPYD	1.820	1.62
39	N	N+(sp^3); ammonium	N+	2.250	2.00
40	N	N(sp^2); pyrrole	NPYL	1.900	1.69
46	N	NO_2; nitro, nitrate	NO3	1.740	1.55
47	O	carboxylate	OM	2.052	1.83

MM2 types up to type 28 are similar to MMFF types, however imines are type 9, amides are type 10, terminal S in S=C type 16, and C(sp^3) in four membered rings are type 20 in MMFF. For MM2 types: http://europa.chem.uga.edu/allinger/mm2mm3/mm2_type.html

MM2 uses the Buckingham equation instead of the Lennard-Jones equation for the Van der Waals interaction. The general form of the Buckingham equation for the Van der Waals potential energy is:

$$E_{VdW,ij} = \varepsilon \left\{ \frac{6}{\alpha-6} e^{-\alpha(r_{ij}-r_o)/r_o} - \frac{\alpha}{\alpha-6}\left(\frac{r_o}{r_{ij}}\right)^6 \right\} \qquad (12)$$

This potential uses the r^6 attractive part of the Lennard-Jones functional form, Eq. 7. The exponential part of the Buckingham potential matches the repulsive part of the Lennard-Jones 6-12 potential best with an α of 14-15. However, MM2 uses a "softer" repulsion of α=12.5:

$$E_{VdW,ij} = \varepsilon \left\{ e^{-12.5\, r_{ij}/r_o} - 2.25 \left(\frac{r_o}{r_{ij}}\right)^6 \right\} \qquad (13)$$

The MM2 force field shows that equilibrium bond lengths and angles change depending on hybridization and bonding partners. In Table 2 are listed the bond parameters that MM2 uses in its force field for a few bond types. These parameters are the starting point for energy minimizations. Any deviations from these equilibrium distance and angle values will be reflected in increases in steric energy. These parameters are derived by finding the "best fit" to experimental data for a reference set of compounds. This reference set of compounds is often called the learning set. The learning set experimental data is from electron and x-ray diffraction studies. (The k's are for the quadratic terms, there are also cubic terms included to account for anharmonicity.) The values in Table 2 are provided to show you typical values for the various force constants.

Table 2. MM2 force field parameters, bond stretch and bend.

Bond	Distance(Å)	k (kcal/Å)	Angle	Angle	k (kcal/°^{2})
C-C	1.523	317	C-C-C	109.47	32.4
C-O	1.407	386	C-C-O	107.5	50.4
Csp^2*-C	1.497	360	C-Csp^2-C	117.2	32.4
			Csp^2-C-C	109.47	32.4
C(carbonyl)-C	1.509	317	C-C(carbonyl)-C	116.60	28.8
			C(carbonyl)-C-C	107.80	32.4
C=O	1.208	777	C-C=O	122.50	67.5
H-C	1.113	331	H-C-H	109.40	23.0
			H-C-C	109.39	25.9
H-O	0.942	331	H-O-C	106.90	57.1

* sp^2 hybridized but not conjugated.

A typical stretch-bend interaction constant is the value for C-C-C of 8.6 kcal/Å/°. A typical oop force constant is the value for >C=C of 2.16 kcal/°. For torsional force constants, the expansion for the C-C-C-C torsion has one, two, and three fold terms:

$$E_{tor} = 0.051\,(1 - \cos\phi) + 0.341\,(1 - \cos 2[\phi - 180]) + 0.166\,(1 - \cos 3\phi) \qquad (14)$$

The 180° shift in the two fold term means that the trans-form is favored. When the different units of distance and angle are considered, these values show that typically the force constants have relative sizes of:

Stretch >> bend > stretch-bend ~ out-of-plane > torsion

In other words, it is difficult to stretch a bond, easier to bend a bond, and very easy to twist a bond if it is singly bonded.

The peptide bond is particularly important, since it is the linkage between amino acids in proteins. Figure 8 shows the peptide bond with the MM2 type force constants for a stretch, bend, and oop bend.

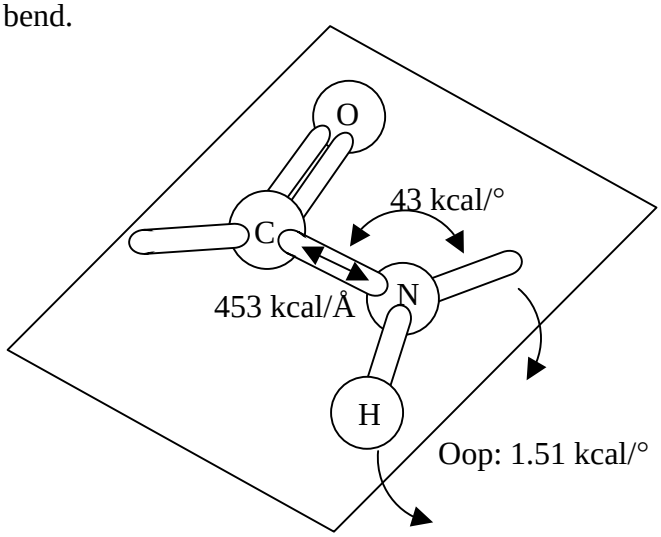


Figure 8. MM2 force field parameters for the amide nitrogen in a peptide bond.

MMFF and MM2 The Merck Molecular Force Field, MMFF, is also a very commonly used force field.[4-6] Example parameters for the MMFF force field are given in Tables 3 and 4 so that you can compare the different parameters from one force field to another. MMFF uses a 14-7 Van der Waals term instead of the more common 12-6 Lennard-Jones or Buckingham potential. Overall MMFF has more terms in the force field, including cubic and quartic terms in the bond stretch, and cubic terms in angle bending potential energy. Notice that there are large differences between MM2 and MMFF. The differences show that the specific terms in the force field make a big difference in the overall parameters. These differences also show that parameters are not transferable from one force field to another.

Table 3. Some MMFF Atom types.

Atom Type	atom	Description	Type	R (Å)
1	C	C(sp^3)	C	1.969
2	C	C(sp^2) alkene	Csp2	2.097
3	C	C(sp^2) carbonyl	C=	1.992
4	C	C(sp) alkyne; C=C=O	Csp	2.077
5	H	Attached to C and Si	HC	1.485
6	O	C-O-H, C-O-C	O	1.779
7	O	=O carbonyl	O=	1.746
8	N	N(sp^3)	N	2.014
9	N	N(sp^2) imines	N=C	1.894
10	N	N(sp^2) amides	NC=O	1.945
11	F	Fluoride	F	1.496
12	Cl	Chloride	CL	2.044
15	S	-S- sulfide	S	2.185
16	S	Terminal S=C	S=C	2.333
17	S	>S=O, sulfoxide	>SN	2.128
18	S	>SO2, sulfones and sulfates	SO2	1.998
20	C	C(sp^3) in 4-membered ring	CR4R	1.969
21	H	-OH alcohol	HO	1.307
22	C	cyclopropane	CR3R	1.992

Table 4. MMFF94 force field parameters, bond stretch and bend. The MMFF has a cubic and quartic term in the bond stretch. So even though the equilibrium distance for a C-C bond is listed as 1.508 Å, the final minimized alkane C-C bond is near 1.53 Å.

Bond	Distance(Å)	k (kcal/Å)	Angle	Angle	k (kcal/°^{2})
C-C	1.508	306	C-C-C	109.61	61.2
C-O	1.418	363	C-C-O	108.13	71.4
Csp^2*-C	1.482	339	C-Csp^2-C	118.04	54.1
			Csp^2-C-C	109.44	53.0
C(carbonyl)-C	1.492	302	C-C(carbonyl)-C	118.02	82.8
			C(carbonyl)-C-C	107.52	55.9
C=O	1.222	932	C-C=O	124.41	67.5
H-C	1.093	343	H-C-H	108.84	37.1
			H-C-C	110.55	45.8
H-O	0.972	561	H-O-C	106.50	57.1

* sp^2 hybridized but not conjugated.

References:

1. Burkert, U, Allinger, N, Molecular Mechanics (ACS Mongraph 177), ACS, Washington, DC, 1982.

2. Allinger, N. L., **"**Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms," *J. Amer. Chem. Soc*., **1977**, *99*, 8127-8134.

3. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations," *J. Comp. Chem*, **1983**, *4(2)*, 187-217.

4. Halgren, T. A., "Merck Molecular Force Field. I. Basis, form, scope, parameterization, and performance of MMFF94," *J. Comput. Chem.*, **1996**, ***17***, 490-519.

5. Halgren, T. A., "Merck Molecular Force Field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interaction". *J. Comput. Chem*., **1996**, *17*, 520-552,.

6. Halgren, T. A., "Merck Molecular Force Field. III. Molecular geometries and vibrational frequencies for MMFF94," *J. Comput. Chem.*, **1996**, *17*, 553-586.

Introduction Section 2
Enthalpy of Formation

The steric energy of a molecule can be used to calculate the enthalpy of formation. First, the steric energy is calculated from Equation 1. Then a bond energy calculation is done using standard tabular values. The bond energy, or enthalpy, is the energy needed to make all the chemical bonds in the molecule starting from the elements in their standards states. It is customary to use bond increments rather than the bond energy calculations that you did in General Chemistry for the bond energy calculation. However, the principle is the same. Thermal energy terms must then be added to account for the energy of translation and rotation of the molecule. The energy of translation (x, y, z motion of the center of mass of the molecule) is $^3/_2$RT. The rotational energy of a non-linear molecule is also $^3/_2$RT ($^1/_2$RT for each rotational axis).

The steric energy calculation in molecular mechanics corresponds to an internal energy calculation. Since $\Delta H=\Delta U+\Delta(PV)$, PV=nRT for an ideal gas, and we want the molar enthalpy of formation with n=1, we must also add RT to convert from internal energy to enthalpy.

We have not yet considered molecular vibrations, especially internal rotations. In principle, every vibration, including internal rotations, contributes to the enthalpy. However, the contribution of vibrations is difficult to calculate. In practice the contributions are often small so they can be ignored. However, the internal rotation of the methyl group is always included; in fact the effect is automatically included in the bond increment calculation. For careful work extra terms must also be added for non-methyl free internal rotations. This contribution, which is called the torsional increment, is estimated as 0.36 kcal/mol or 1.51 kJ mol^{-1} for each internal rotation[1]. For example, butane, CH_3-CH_2-CH_2-CH_3, has one additional internal rotation, other than the methyl group rotations; so the torsional increment for butane would be 0.36 kcal/mol. In summary the enthalpy of formation for non-linear molecules is then,

$$\Delta_f H° = {}^3/_2RT + {}^3/_2RT + RT + \text{bond energy} + \text{steric energy} + \text{torsional increments} \quad (1)$$

This formula also assumes that there is only one low energy conformation of the molecule. If there are several low energy conformations, each must be accounted for in Equation 1.

Bond Energy

You are familiar with bond energy calculations from General Chemistry. The energy of a molecule is assumed to be an additive function of the energy of individual bonds (Table I). The $\Delta_r H$ for a reaction is given from $\Delta H°$(bonds broken)- $\Delta H°$(bonds formed).

Table I. Bond Enthalpies, $\Delta H°$(A-B) (kJ/mol)

	H	C	O
H	436		
C	412	348 – 612 =	
O	463	360 – 743 =	146 – 497 =

C (graph) -> C (g) $\Delta H°$= 716.7 kJ/mol

For example, the enthalpy of formation of acetaldehyde is calculated as:

$$2\ C(graph) + 2\ H_2\ (g) + 1/2\ O_2\ (g) \rightarrow CH_3\text{-}CH{=}O\ (g)$$

#	Bonds Broken		-	#	Bonds Formed		
2	C (graph)	2 (716.7 kJ/mol)		1	C=O	743 kJ/mol	
2	H-H	2 (436 kJ/mol)		4	C-H4	(412 kJ/mol)	
1/2	O=O	1/2 (497 kJ/mol)		1	C-C	348 kJ/mol	
total		2553.9 kJ/mol	-	total		2739 kJ/mol	= -185.1 kJ

The experimental value is -166.19 kJ, so the value derived from Table I is not very accurate.

The bond energy calculations in molecular mechanics are done slightly differently, using bond increments. Again the bond energies are assumed to be additive. The contributions are taken not only from each bond, but increments are added for certain structures, such as tertiary carbon linkages. The bond energy calculation for acetaldehyde from the MM2 program is given below, with energies in kcal. MM2 also calculates entropies, which are also listed for your interest.

#	Bond or Structure	Each	Total	Tot S contrib.
3	C-H ALIPHATIC	-3.205	-9.615	38.700
1	C=O	-25.00	-25.00	-2.300
1	C-H ALDEHYDE	-2.500	-2.500	26.800
1	C-C SP3-SP2 C=O	-3.000	-3.000	-0.600
1	ME-CARBONYL	-2.000	-2.000	
		bond energy = -42.115 kcal		S° = 62.600 cal/K

The bond energy is -42.115 kcal or -176.2 kJ. However, caution should be used since these calculations are designed to be used in conjunction with steric energies in a molecular mechanics calculation and not as general bond energy values. Using Equation 1, with the steric energy calculated by molecular mechanics gives the final $\Delta_f H°$ = -169.33 kJ/mol, which is a significant improvement over the bond energy calculation from Table I of -185.1 kJ.

References:

1. Pitzer, Kenneth S., Quantum Chemistry, Prentice-Hall, New York, NY, 1953, pp 239-243, Appendix 18, pp 492-500.

Introduction Section 3
Comparing Steric Energies

You must be careful when comparing steric energies from molecular mechanics calculations. Strictly speaking you can only compare steric energies directly for conformational isomers or geometric isomers that have the same number and types of bonds. Some examples using MM2 will make this important point clearer.

Example 1: Different number of atoms:

Table 1 gives the MM2 results for pentane, hexane, and heptane. First note that each of the individual force field terms and the total steric energy increase on going from pentane to hexane to heptane. It would be tempting to conclude that the larger molecules have "more steric hindrance" from these numbers, but this would be incorrect. Rather, the changes are caused by the fact that you are simply adding more atoms so the number of terms in the force field are increasing causing the molecule's totals to increase. This conclusion is reinforced by the MM2 sigma strain energy results that show each molecule to have no strain energy. This example shows that you can't directly compare steric energies for molecules with different numbers of atoms.

MM2, MMX, and MM3, however, take the molecular mechanics calculation one step further. The use of bond enthalpy calculations to calculate the enthalpy of formation for the molecule adjusts for the new bonds that are formed as the molecular size increases. Enthalpies of formation can be compared directly. For example, the bond enthalpy and enthalpy of formation from MM2 are also shown in Table 1. These results show correctly that the enthalpy of formation of these molecules decreases with size, even though the total steric energy is increasing. The enthalpies of formation can, of course, be used to calculate the enthalpies for any reactions using pentane, hexane, and heptane.

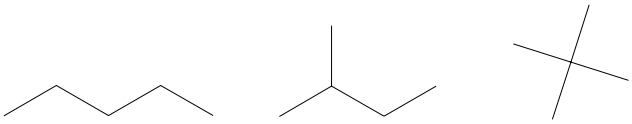
Table 1. MM2 results for linear C5, C6, and C7 hydrocarbons and branched C5 hydrocarbons.

kcal/mol	Pentane	Hexane	Heptane	2-Methylbutane	2,2-Dimethylpropane
Bond Stretch	0.2267	0.2968	0.3664	0.3180	0.4038
Bending	0.3797	0.4689	0.5553	0.6512	0.3308
Stretch-bend	0.0731	0.0938	0.1142	0.0969	0.0641
Lennard-Jones	2.1316	2.5911	3.0512	2.0967	1.4712
Dihedral	0.0116	0.0161	0.0212	0.4649	0
Total Steric	2.8226	3.4667	4.1084	3.6279	2.2699
Bond Enthalpy	-41.50	-47.91	-54.32	-42.93	-45.22
Sigma Strain	0	0	0	1.03	0
Enthalpy of Formation	-36.27	-42.04	-47.82	-36.90	-40.55

The bond enthalpy calculations in MM2 are done using tabulated values for bond increments for each specific bond and chemical environment. See the enthalpy of formation discussion earlier in this manual for more information. The sigma strain energy calculations in MM2 are done using similarly tabulated increments for each specific bond and chemical environment, but in a hypothetical "strainless environment." Differences in total enthalpy of the values based on the actual and the strainless bond enthalpies give the sigma strain energy.

Example 2: Same formula different types of bonds:

Table 1 also has the MM2 results for the branched pentanes, 2-methylbutane and 2,2-dimethylpropane, to compare with linear pentane. The corresponding structures are shown in Figure 1. Each isomer has the same number of atoms and the same number of C-H and C-C bonds. Here again, however, comparing steric energies directly is dangerous. The higher steric enegy of pentane compared to 2,2-dimethylpropane does not indicate that linear pentane has "more steric hindrance." Rather, both linear pentane and 2,2-dimethylpropane show no sigma strain. Likewise, both branched pentanes have lower enthalpies of formation than the linear isomer. Even though all three isomers have the same number of C-H and C-C bonds, the C-C bond energy increases with increased branching. That is, a tertiary C-C bond is more stable than a secondary, which is more stable than a primary.



pentane 2-methylbutane 2,2-dimethylpropane

Figure 1. Pentane geometric isomers

Once again, the final enthalpy of formation calculations adjust for these bond strength differences and are then directly comparable. Does this mean that the steric energies by themsleves are useless? No, you just need to be careful when doing comparisons.

For example, why does 2-methylbutane have a higher steric energy than linear pentane? The Lennard-Jones term is actually lower in energy for the branched isomer, because of favorable, attractive Van der Waals interactions. Looking at the other force field terms, we see that the dihedral terms increase the most. The increase in the branched isomer results from a gauche interaction. Draw a Newman projection to show that this is so. This example shows that comparing steric energies, and in particular, comparing the different force field terms can be very helpful in understanding the energetics of the molecule, especially for geometric isomers. Remember that, however, it is the enthalpy of formation of the molecule that determines its reactivity and the enthalpy of formation may or may not follow the same trends as you compare one geometric isomer to another.

An analogy might help. One person may be taller than another, but the taller person may not be the better basketball player. It is fair to compare the height of two individuals, but basketball ability depends on many more things than height alone.

Example 3: Making fair comparisons:

Most biostructure molecular mechanics programs don't use MM2 or MM3, so that the sigma strain energy and the enthalpy of formation are not calculated. In addition, MM2 and MM3 have limited parameter sets, so your compound of interest may not run with MM2 and MM3, and you must use a different force field. How can you make fair comparisons if you can't get the enthalpy of formation? Often, it is possible to build a reference structure and then look at differences with the reference structure as a fair comparison. To illustrate this point we will look at the strain energy of five, six and seven membered rings, Table 2. We will use MM2 results to check our comparisons, to make sure our reference structures provide a fair comparison. But the utility of building reference structures is really most useful when MM2 isn't available.

Table 2. MM2 Results for five, six and seven membered hydrocarbon rings.

kcal/mol	Cyclopentane	Cyclohexane	Cycloheptane
Bond Stretch	0.3264	0.3374	0.4116
Bending	2.1899	0.3652	2.8389
Stretch-bend	-0.0976	0.0826	0.2399
Lennard-Jones	2.6501	3.6100	5.3694
Dihedral	6.3279	2.1556	5.4476
Total Steric	11.4049	6.5510	14.3075
Bond Enthalpy	-32.07	-38.48	-44.90
Sigma Strain	8.12	2.61	9.71
Enthalpy of Formation	-18.27	-29.53	-28.19
(Cyclic-Linear) Steric Energy	8.58	3.08	10.20

First note that the total steric energy and the enthalpy of formation follow completely different trends. Therefore, the steric energy is a poor predictor of chemical reactivity. This example is similar to Example 1, above, in that the molecules we wish to compare have increasing numbers of atoms. However, the strain energy of rings is an important concept and has helped to guide organic chemist's intuition about chemical reactivity for over a century. Of course, MM2 calculates the strain energy, and we get the expected order cyclohexane< cyclopentane< cycloheptane. Students are often surprised at this order, thinking that the cyclopentane ring is unusually strained, but this is not so in comparison with cycloheptane.

We can make a fair comparison of the ring strain energies of these molecules by comparing each cyclic structure with a linear reference structure. The reference structure is just the cyclic molecule "opened up." We then compare this difference in energy for the cyclopentane, cyclohexane, and cycloheptane rings. In Table 2 is listed the <u>difference</u> in steric energy between the cyclic structure and the linear structure. These differences mirror the MM2 strain energies nicely. The <u>difference</u> with the linear reference structure is successful in finding the strain energy because the difference between the cyclic and linear form is the breaking of two C-H bonds and the formation of a new C-C bond for <u>each</u> of our cyclic molecules. Using the differences in energy then makes the comparison fair because we are adjusting for the fact that the rings have an increasing number of atoms. The following chart may be helpful is seeing why this difference procedure works:

<u>Incorrect comparison:</u>

	Cyclopentane	⇔	Cyclohexane	⇔	Cycloheptane
	C_5H_{10}		C_6H_{12}		C_7H_{14}
Total Change:		CH_2		CH_2CH_2	

<u>Better comparison:</u>

(Cyclopentane - pentane) ⇔ (Cyclohexane - hexane) ⇔ (Cycloheptane - heptane)

C_5H_{10} C_5H_{12} C_6H_{12} C_6H_{14} C_7H_{14} C_7H_{16}

Using differences with reference structures helps to cancel out the effects of having different numbers of atoms and bonds. In fact, the differences with the references (last row of Table 2) are each 0.47 kcal/mol larger than the corresponding MM2 sigma strain energy. So the trend in strain energy is exactly reproduced. The 0.47 kcal/mol results from the way in which MM2 tabulates the expected values of bond energy for "strainless structures."

In summary, comparisons of steric energies can be made using differences with reference structures. The reference structures should be built so that the energy term of interest is highlighted. In this example, the reference was constructed from the linear form of the cyclic molecule to highlight the strain energy. The reference structures should be as similar as possible in every other way to the compound under study.

Example 4: Different number of atoms, but ask a different question:

Steric energies, as we have seen, usually can't be compared directly when trying to predict chemical reactivity. We need enthalpies of formation for reactivity comparisons. However, we can ask a different question, for which steric energies are useful for comparisons. We can ask which terms in the force field have a big influence on the steric energy of the molecule and how that influence changes from molecule to molecule. In other words, by comparing relative contributions, we can trace through the important differences among our molecules. The relative contributions of the different force field terms to the steric energy, based on Table 2, are given in Table 3.

Table 3. Relative contributions to the total steric energy of cyclic hydrocarbons.

%	Cyclopentane	Cyclohexane	Cycloheptane
Bond Stretch	2.9	5.2	2.9
Bending	19.2	5.6	19.8
Stretch-bend	0.9	1.3	1.7
Lennard-Jones	23.2	55.1	37.5
Dihedral	55.5	32.9	38.1

The primary contributor to the steric energy for cyclopentane is the dihedral (torsional) interaction. But for cycloheptane the steric energy results more from a combination of dihedral and unfavorable Lennard-Jones (Van der Waals) contacts. For cyclohexane, angle bending is relatively unimportant, compared to the other ring systems. Comparisons such as these are invaluable for building your intuition about the energy components of molecules. These comparisons are fair because the contributions are all relative to the steric energy of the same molecule. That is, the percentages are calculated from the energies of one molecule.

However, it is important to remember what such relative contributions don't tell you. The results in Table 3, by themselves don't tell you which molecule has the highest strain, nor even the highest steric energy. These relative contributions also don't tell you which molecule has the highest enthalpy of formation. So you can't predict which molecule will be the most reactive.

Another analogy may be helpful. Jane gets a higher percentage of her points from foul shots than Susan. This statistic, however, doesn't tell you who gets more points per game. On the other hand, the statistic suggests that Susan should work on her foul shots, which is helpful information.

Comparing relative contributions is most useful when the various force field terms have comparable reference energies. For example, the various terms in Table 2 from MM2 are all very

small for linear hydrocarbons where the strain energies are quite small. However, some implementations of force fields (such as the Merck Molecular Force Field implemented in MOE but not in Cerius2 or Spartan) shift the energy zero for the torsional interaction so that even for linear hydrocarbons the torsional terms are quite large. This does not mean that linear hydrocarbons are torsionally strained! So when getting used to a new program and new force field start by minimizing trans-butane and looking at the size of the different force field terms. If the force field terms are all small for butane then comparisons of the type in this example will be easy to interpret. If one or more terms for butane are much larger than the others you will need to remember that the relative size of that interaction in your molecule will be over-emphasized when looking at relative contributions. Making comparisons in the changes in relative contributions from one molecule to another will still be useful, however.

Conclusion

The discussions in the examples above are summarized in Table 4. Comparing steric energies directly gives the most information, but you can only compare steric energies directly if the molecules have the same formula and the same number and types of bonds. We even need to consider that not all C-C single bonds are equal, when we compare steric energies. In other words the chemical environments of all the bonds must be equivalent. You can always compare enthalpies of formation.

Table 4. Molecular mechanics steric energy comparisons between molecules.

Comparison	Steric Energy Directly	Difference with Reference	Relative contributions
Conformational Isomers	yes	yes	yes
Geometric Isomers	if same environments	yes	yes
Different Formulas	never	yes	yes

Introduction Section 4
Energy Minimization

The steric energy of a molecule is the sum of the bonded and nonbonded terms (Van der Waals energy, and the electrostatic energy). The lowest energy conformation is the set of bond lengths and angles that gives the smallest steric energy. In other words, bonds find a compromise among competing forces to determine the lowest energy conformation. The goal of molecular mechanics is to determine the lowest energy conformation of a molecule. The process is called energy minimization. The computer makes small changes in the position of every atom and calculates the energy after every move. The move is kept if the energy is lowered, otherwise the atom is returned to its original position. This process is repeated many times until an overall energy minimum is reached. One full cycle, where each atom is moved once, is called a minimization step or iteration. Hundreds of steps may be necessary to find a reasonable structure for the molecule.

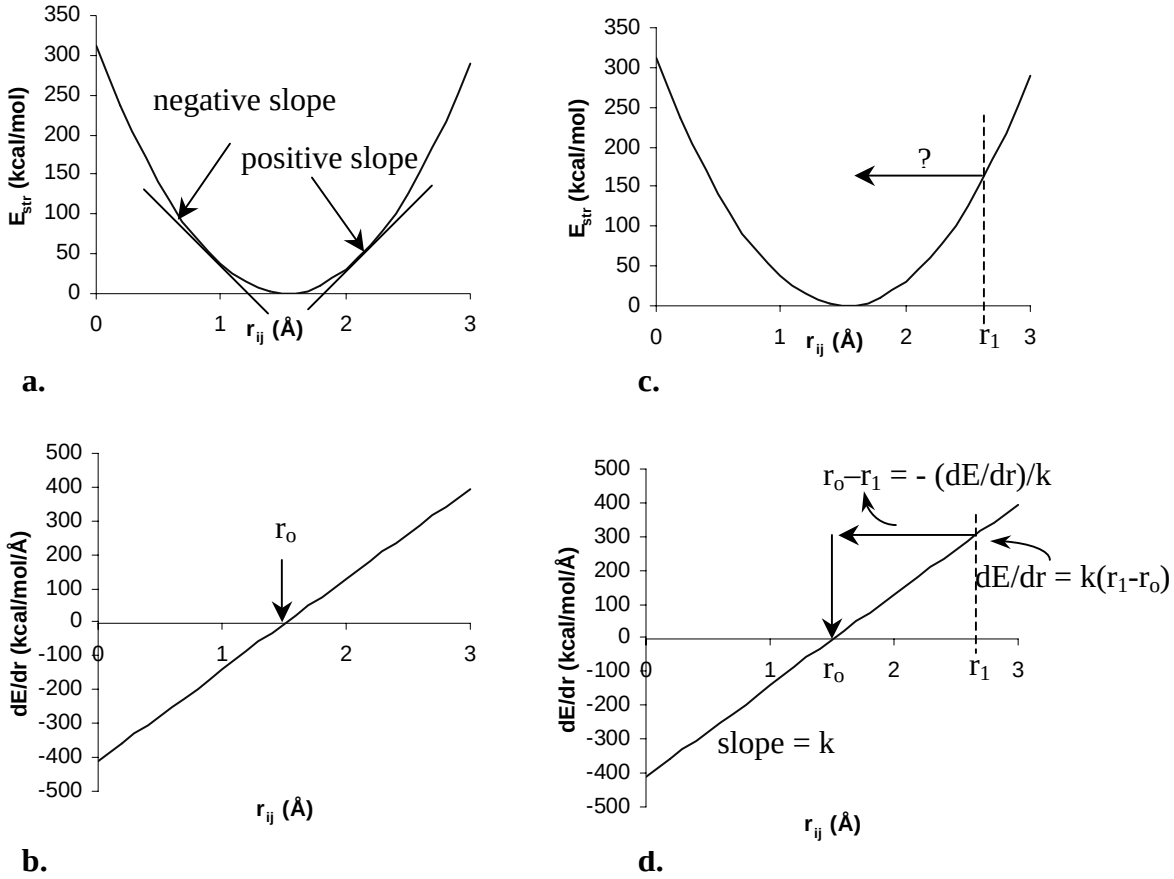


Figure 1. Finding the change in bond length to minimize the potential energy. (a.) The potential energy curve for a stretching bond. (b.) The slope of the potential energy is linear and changes sign as the molecule passes through the equilibrium bond length. (c.) The starting geometry is with bond length r_1. Now calculate the change in bond length that minimizes the potential energy. (d.) The slope of the potential energy at r_1 is $k(r_1-r_0)$, and the slope of the line in the dE/dr graph is k. To calculate the change in bond length to find the minimum potential, extrapolate down the line to the zero point.

Many methods have been developed to accelerate the minimization process. These methods use information from the derivatives of the potential energy function to calculate the change in the coordinates for each step[1]. The Newton-Raphson method is the most basic of these techniques, and we discuss this method first using a simple example. We start with a diatomic molecule. The only coordinate to minimize is the bond length, r. The potential energy function is just the bond stretching term, Figure 1a:

$$E_{str} = \frac{1}{2} k (r - r_o)^2 \qquad (1)$$

where k is the force constant for the bond and r_o is the equilibrium bond length. The derivative of E_{str} is the slope of the curve in Figure 1a:

$$\frac{dE_{str}}{dr} = k (r - r_o) \qquad (2)$$

The derivative is plotted in Figure 1b. Equation 2 shows that the slope of the potential energy is linear and changes sign as the molecule passes through the equilibrium bond length. For example, in Figure 1a, when $r>r_o$ the slope is positive, when $r<r_o$ the slope is negative, and the slope is zero at r_o. The slope of the line in Figure 1b is the second derivative of the potential energy:

$$\frac{d^2E_{str}}{dr^2} = k \qquad (3)$$

Lets say that the starting guess for the bond length before minimization is r_1, Figure 1c. Now we wish to calculate the change in bond length that minimizes the potential energy. In other words, we wish to calculate the distance we need to move to find r_o, or r_o-r_1. The change in bond length is easiest to calculate using the derivative of the potential, Figure 1d, because the derivative is a linear function. All we need do is extrapolate down the line to the zero point. In reference to Figure 1d, the derivative of the potential at r_1 is:

$$\frac{dE_{str}}{dr} = k (r_1 - r_o) \qquad \text{at } r_1 \qquad (4)$$

Solving this linear equation for the change in bond length just requires dividing by –k:

$$(r_o - r_1) = - \frac{1}{k} \frac{dE_{str}}{dr} \qquad (5)$$

This change in bond length is also shown in Figure 1d. For harmonic potentials, like Equation 1, the calculated change is exact, so only one iteration step is needed. When there are many force field terms or non-harmonic potentials (eg. torsions, Van der Waals, Coulomb) the derivative of the potential is not linear, and equation 5 is just an approximation. Therefore, in the general case many steps are necessary to find the minimum, but the derivative of the potential still gives a good guess.

Newton-Raphson: Equation 5 is specific to a harmonic potential. We can obtain a more general solution by substituting for k using Equation 3:

$$(r_o - r_1) = - \frac{1}{\frac{d^2E_{str}}{dr^2}} \frac{dE_{str}}{dr} \qquad (6)$$

Equation 6 is the basis of the Newton-Raphson method[1]. The first derivative of the potential is called the gradient. The second derivative is called the Hessian, especially when more than one dimension is involved. The Newton-Raphson method is also used for molecular orbital calculations. You will see the Hessian mentioned in Spartan and other molecular orbital software packages. When many atoms are present, the Hessian can be time consuming to calculate and to invert. The many different methods for minimization differ in the way they approximate the Hessian. The Newton-Raphson method requires the fewest steps, but each step is time

consuming. The number of steps required to minimize strychnine, Figure 2, for several methods is given in Table 1. The Newton-Raphson method was almost the fastest in this case because strychnine is a very small molecule, for larger molecules Newton-Raphson is very slow.

Table 1. Iterations necessary to minimize strychnine from a crude starting geometry.

Method	Seconds	Steps
Steepest Decents	32.4	3042
Conjugate Gradient	14.1	237
Newton-Raphson	13.0	15
Adopted Basis Newton-Raphson	3.7	279

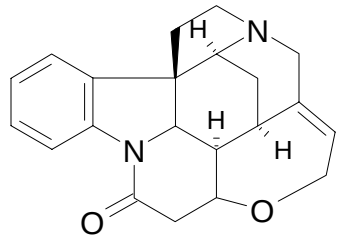


Figure 2. Strychnine

Steepest Descents: In the steepest descents method, the Hessian is just approximated as a constant, γ:

$$(r_o - r_1) = - \frac{1}{\gamma} \frac{dE_{str}}{dr} \qquad (6)$$

You can think of γ as an effective force constant as in Equation 5. γ is calculated at the beginning of the first step to give a specified step size. The dialog for the minimization parameters for CHARMm and MOE are shown in Figure 3. The Initial Step Size entry is used to fix γ.

CHARMm		MOE		
Number of Minimization Steps	50		Iteration Limit	RMS Gradient Test
Coordinate Update Frequency	5			
Energy Gradient Tolerance	0.0001	Steepest Descents	100	1000
Energy Value Tolerance	0	Conjugate Gradient	100	100
Initial Step Size	0.02	Truncated Newton	200	0.001
Step Value Tolerance	0			

Figure 3. CHARMm and MOE parameters for energy minimization.

Too small a step size can slow the minimization process. Too large a step size can prevent convergence. Table 2 lists the effect of the step size on the number of steps to give a minimized structure.

Table 2. Steps necessary to minimize strychnine for different step sizes.

	Step Size		
Method	0.01	0.02	0.04
Steepest Descents	3998	3042	no converge
Conjugate Gradient	237	237	237
Newton-Raphson	15	15	15
Adopted Basis Newton-Raphson, ABNR	311	279	331

The conjugate gradient and Newton-Raphson methods only use the step size in determining the initial gradient, so they are not strongly effected by the choice of the step size.

Table 1 shows that steepest descents has very poor convergence properties. So why is steepest descents used at all? Conjugate gradients can often fail with a poor initial structure, such as a Protein Database file for a protein. Steepest descents is less sensitive to the starting conditions.

Therefore, a few steps of steepest descents is usually used to refine a poor starting structure before switching to a better method, such as conjugate gradient.

Conjugate Gradient: Conjugate gradient is a variation of the steepest descents method. The calculation of the gradient is improved by using information from previous steps. Also after a steepest descents initial stage, a second steepest descents stage is taken in a direction perpendicular to the first direction. This perpendicular direction is called the conjugate direction. For example, for the minimization of the structure of water the OH bond length and bond angle must be adjusted to minimize the energy. A schematic representation of the potential energy surface for the two variables is shown in Figure 4. Lets say that the initial steepest descents finds the minimum along the initial direction. The next best direction to look for the overall minimum is perpendicular to the initial path, Figure 4.

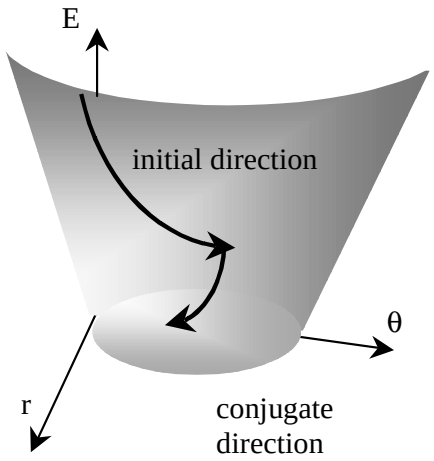


Figure 4. One iteration of conjugate gradients minimization.

Table 1 shows that the conjugate gradients method is vastly better than steepest descents while remaining nearly as fast per step. Conjugate gradients is a good general purpose technique.

Adopted Basis Newton-Raphson, ABNR: For very large systems like proteins and nucleic acids, energy minimization can require hours. The search for very efficient minimization methods for such large biological macromolecules has led to a modified version of the Newton-Raphson method that maintains excellent convergence properties but in a much shorter time. Each step of the ABNR method begins with a steepest descents stage. Then the bond lengths and angles that change the most are noted, and only these coordinates are used in a second stage of Newton-Raphson minimization. For strychnine, Table 1, and for biological macromolecules in general, ABNR is clearly the best method.

Truncated Newton-Raphson: The use of second derivatives in Newton-Raphson minimization is responsible for the excellent convergence properties. However, the inversion of the Hessian is time consuming. An approach has been developed that uses conjugate gradients to determine the directions for the minimization and then the Hessian to determine the minimum in that direction[2]. The "direction" of the minimization determines the particular bond lengths and angles that will be changed. The minimum in that "direction" determines how much to change those bond lengths and angles. Truncated Newton-Raphson has similar and often better convergence characteristics to ABNR without a significant difference in time. The Hessian is calculated directly from the second derivatives, which are evaluated numerically, but some small second derivatives between distant atoms are neglected (or truncated)[3].

The general approach to energy minimization is to use a "cascade" of techniques. First 50-200 steps of steepest descents is used to remove close contacts (atoms closer than the sum of their Van der Waals radii). Then 50-200 steps of conjugate gradients is applied, followed by final minimization using ABNR or truncated Newton-Raphson. For small molecules only 10-20 initial steepest descent steps are needed, and the intermediate conjugate gradient steps can be skipped.

Minimization Criteria: How do you determine when the molecule is minimized? Molecular modeling programs provide a number of alternate methods for deciding when to stop, Figure 3. The Number of Minimization Steps (Iteration Limit) can be used to stop the calculation. This option is dangerous; you need to realize that if the minimization stops for this reason that the molecule is not minimized and you need to continue to submit the molecule for minimization until the energy no longer changes on successive steps. A better option is to set the number of minimizations steps to a very large number and then use an energy based criterion, like the energy gradient tolerance (test).

Remember that the gradient is the derivative of the energy. The energy gradient approaches zero at the energy minimum. The criterion then is to stop if the gradient is less than a selected value. This method is illustrated in Figure 5. In program listings you will often see the term rms gradient. Rms stands for root mean squared. For some coordinates (e.g. a bond stretch) the gradient might be positive, while for other coordinates (e.g. an angle) the gradient might be negative. So that the positive and negative gradients don't cancel out, the gradients are squared to give positive numbers before adding them together to make the comparison. (The standard deviation is likewise an rms statistic).

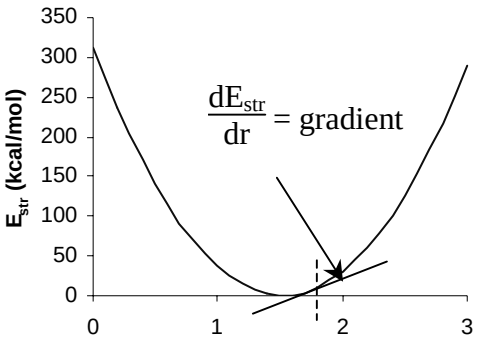


Figure 5. Stop if the energy gradient is below the tolerance.

An alternate method to stop the minimization is to compare the change in energy between the current step and the previous step, ΔE in Figure 6. If the change in energy is below the set tolerance, then the minimization is halted.

Finally, the last available criterion is the step size. The size of the change in the coordinates is monitored and when this change is smaller than the set tolerance, the minimization is halted. This criterion, where Δr is the step size, is also illustrated in Figure 6. The step size criterion is useful for shallow potentials, where the energy doesn't change much for large changes in conformation or distances between molecules. For example for complexes, the energy gradient can be small and still give large changes in the distance between the two molecules. In some programs, the step size is called the rms displacement.

When several criteria are specified, the first criterion to be met stops the minimization. For example, as mentioned above if you set the number of minimization steps to a small number, the calculation will probably stop before a minimum is achieved. To determine if the step count has stopped the calculation, look at the output and determine if the last minimization step is the same as the number of minimization steps that you specified as a control parameter (i.e. Figure 3). To avoid this problem, set the number of minimization steps to a large number. On the other hand, if you have a large molecule, there is a danger in specifying too large a number of minimization steps. The calculation my take too long to run and then the computer is tied up so that you can't do other things. Or, you may have made a mistake, and a long minimization keeps you from quickly making changes.

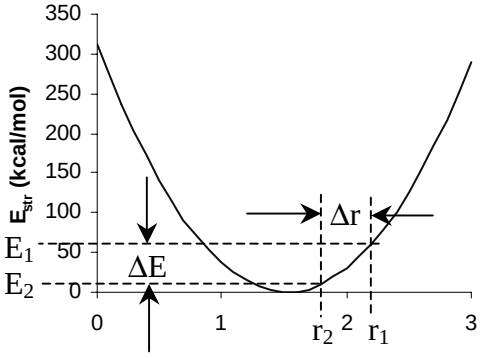


Figure 6. Stop if the energy change or step size is below the tolerance.

In CHARMm, as a default, we normally choose 500 steps for small molecules and 50 steps for large molecules. Then we make sure to resubmit the minimization if the last step matches the 500 or 50 that we set. Other than minimization steps, the criterion that you use is a matter of your choice. The very best approach is to enter a value for each and see which is statisfied first. Entering all this data is tedious, so as a default we usually use just the energy gradient tolerance. A value of 0.0001 is useful for very small molecules, however, you will find it necessary to use 0.001 or 0.01 for biological macromolecules or for solvated systems to save time. The units are in kcal/mol/Å in most programs.

References:

1. Leach, Andrew, *Molecular Modelling, Principles and Applications*, Longman, Harlow, Essex, England, 1996.

2. "Forcefield-Based Simulations," Accelerys, Corp, San Diego, CA. Chapter 4 Minimization.

3. Jensen, Frank, *Introduction to Computational Chemistry*, John Wiley, Chichester England, 1999, p322.

Introduction Section 5
Molecular Dynamics

Introduction

One of the most important developments in macromolecular chemistry is molecular dynamics. Molecular dynamics is the study of the motions of molecules. The time dependence of the motion of a molecule is called its trajectory. The trajectory is determined by integrating Newton's equations of motion for the bond stretching, angle bending, and dihedral torsions of the molecule. Molecules are always in motion. The motion of molecules is important in essentially all chemical interactions and are of particular interest in biochemistry. For example, the binding of substrates to enzymes, the binding of antigens to antibodies, the binding of regulatory proteins to DNA, and the mechanisms of enzyme catalysis are enhanced and sometimes completely determined by the conformational flexibility of the molecules. Different domains of an enzyme can have very different motional freedom. The problem of protein folding is the determination of the trajectory of the macromolecule as it assumes its active conformation after or during protein synthesis.

Most chemistry is done in solution. Molecular dynamics has proved to be an invaluable tool in studies of solvation energetics. Solute-solvent interactions are governed by the relative motions of the solute and solvent molecules and the motional-response of the solute to the presence of the solvent. Some of the earliest dynamics studies were to determine solvation Gibb's Free energies. In biochemistry, solute-solvent interactions play a particularly important role in determining the secondary and tertiary structure of biomolecules.

Another important use of dynamics is in the search for the global energy minimum in conformationally flexible molecules. Molecular mechanics find the energy minimum which is closet to the starting conformation of the molecule. This "local" energy minimum is rarely the lowest energy, or "global", minimum for the molecule. Finding the "global" minimum can be a very difficult task. In molecular mechanics a common procedure is to start with many different initial conformations and minimize them all looking for the lowest energy result. This kind of search can be very time consuming. Molecular dynamics, on the other hand, can help a molecule "explore" its conformation space more efficiently. The trajectory of the molecule is run at a high temperature, so that the atoms will move very far from their equilibrium positions. Such high temperature trajectories can overcome energy barriers that lead to more stable conformations. The trajectory often starts in one conformation and then ends up in another more stabile conformation.

Molecular dynamics is an active area of research in biochemistry, molecular biology, and polymer chemistry. Current work is directed towards making molecular dynamics a reliable tool for the estimation of Gibb's free energies of solvation, conformational equilibria, and equilibrium constants for binding interactions. These thermodynamic parameters are determined by doing free energy perturbation studies using molecular dynamics trajectories; see Section 7 for more on free energy perturbation.

The difference between molecular mechanics and dynamics can be illustrated with a simple example. Lets direct our attention to a single bond in a molecule, a C-H bond for example. Assume that we start with the bond length too large, say 2 angstroms. If we were to run molecular mechanics, the bond length would decrease until the minimum in the potential energy was reached, Figure 10.1a. Further minimization would not change the bond length. If we were to run molecular dynamics on our stretched bond, the trajectory would decrease the bond length, but the bond length would continue decreasing past the equilibrium length until it was too short. Being too short, the bond length would then begin to increase. Over time the bond length will

oscillate about its equilibrium value, never coming to rest, Figure 10.1b. In other words, in mechanics the potential energy is minimized, while the kinetic energy of the molecule is ignored. In a dynamics trajectory, both potential and kinetic energy are studied and the total energy is conserved by the motion.

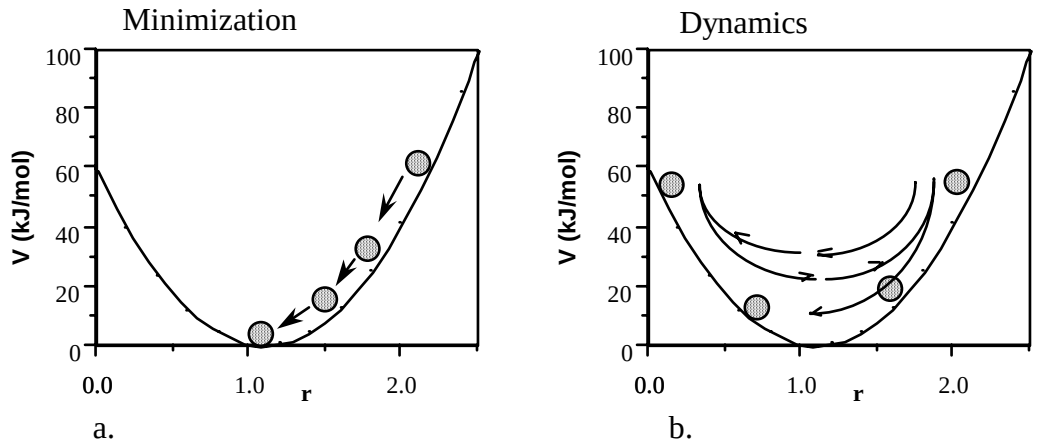


Figure 10.1. The potential energy function for a bond. The initial bond length at 2 angstroms is too long. (a) Molecular mechanics finds the lowest energy state of the molecule. b. Molecular dynamics find the time dependent motion of the molecule. The vibration continues forever.

As chemists we often have too static a picture of molecules. Our mental images of molecular structure are derived from the printed page. Rather, molecules are always in motion. The results of molecular dynamics are very instructive, because dynamics trajectories show us how important motion is in chemical interactions. We should remember that chemical reactions, by their very nature, involve the motion of atoms as bonds are broken and made.

Dynamics Trajectories: Integrating Newton's Laws

Integrating Newton's Laws of motion is actually very straight forward. First, we use the molecular mechanics force field as the potential energy for our molecule. Therefore, the potential energy of our molecule involves bond stretching, angle bending, dihedral torsions, Van der Waals interactions, and electrostatic interactions. We then solve for the motion of each atom in the molecule as a function of time using this potential energy. However, as we begin to learn about dynamics, lets simplify our system to make things less complicated. Lets start with a diatomic molecule. The results of our work on a diatomic molecule will involve everything we need to know about more complicated systems. The molecular mechanics potential energy of a diatomic system has only one term, the potential energy for bond stretching:

$$V = \frac{1}{2} k (r - r_o)^2 \qquad 1$$

where r is the current bond length, r_o is the equilibrium bond length, and k is the force constant for the bond. We can simplify Eq. 1 even further if we let $x = r - r_o$, then

$$V = \frac{1}{2} k x^2 \qquad 2$$

The force that acts on the system is the derivative of the potential:

$$F = -\frac{dV}{dx} \qquad 3$$

Taking the derivative of Eq. 2 gives:

$$F = -k\,x \qquad 4$$

which is just the familiar Hooke's Law for a mass on a spring. Here the bond is the spring. Newton's Law tells us that F= m a, where a is the acceleration. The acceleration is the rate of change of the velocity:

$$F = -k\,x = m\frac{dv}{dt} \qquad 5$$

The position of the system, x, is determined by integrating the equation:

$$\frac{dx}{dt} = v \qquad 6$$

Integrating Eq. 5 gives the velocity as a function of time, starting from an initial velocity of v_1:

$$\int_{v_1}^{v_2} dv = \int_{t_1}^{t_2} \frac{F}{m}\,dt \qquad 7$$

giving $$v_2 = v_1 + \frac{F}{m}\,(t_2 - t_1) \qquad 8$$

where m is the reduced mass for the vibrating bond. Integrating Eq. 6 gives the position as a function of time, starting from an initial position of x_1:

$$\int_{x_1}^{x_2} dx = \int_{t_1}^{t_2} v_2\,dt \qquad 9$$

giving $$x_2 = x_1 + v_2\,(t_2 - t_1) \qquad 10$$

Since the velocity and position are both changing with time, Eqs 8 and 10 are solved repeatedly over short time steps, first updating the velocity and then updating the position. The value of x for each of these successive time intervals is then the trajectory of the system. In dynamics simulations the time step is very short, usually $dt = t_2 - t_1 = 1\text{x}10^{-15}$ sec or 1 femtosec.

All that remains is to determine the initial conditions. A common choice for the position is to choose $x_1 = 0$ at t = 0. But what about the velocity? The average velocity of a system is related to the temperature; the higher the temperature the larger amplitude the motions. At x = 0 all of the energy of an oscillating molecule is in kinetic energy. The kinetic energy is given as

$$KE = \frac{1}{2} \; m \, v^2 \qquad 11$$

The Equipartition Principle of thermodynamics gives an estimate of the kinetic energy in a bond vibration as 1/2 RT, where R is the gas constant; R= 8.314 J mol^{-1} K^{-1}. Setting KE= 1/2 RT and solving for the velocity gives:

$$v = \sqrt{RT / m} \qquad 12$$

We therefore set $v_1 = \sqrt{RT / m}$ at t = 0.

Eqs 8 and 10 are all that is meant by "integrating" Newton's Laws of motion. However, our example is a "one dimensional" system: there is only one motional variable. In more complicated molecules, equations 8 and 10 would be solved for the x, y , and z motion of each atom. However, no new theory is needed; the problem just becomes more tedious. Computers are very good at solving simple, repetitive problems. In fact the advancement of molecular dynamics is very closely tied to the advancement of computer technology. The availability of fast computers means that molecular dynamics can now become one of the standard tools in computational chemistry.

Problem 10.1: Dynamics trajectories

Write a short EXCEL spreadsheet or BASIC program to determine the trajectory for a diatomic molecule. To make the problem more realistic, assume the bond is anharmonic, with potential energy function:

$$V = \frac{1}{2} \; k \, x^2 + \frac{1}{2} \; k \, anharm \, x^3 \qquad 13$$

Please see the Section 1 for more information on anharmonic potentials for bond stretching. With your dynamics trajectory you will be able to see the time dependance of the vibration. You will also be able to determine the conditions for breaking a bond. For example, you can increase the anharmonicity to determine how anharmonic the bond must be to be broken at room temperature. Conversely, you can keep the anharmonicity constant and increase the temperature until the bond breaks, which is just what synthetic chemists do when they heat a reaction mixture.
Differentiation of Eq. 13 gives:

$$F = - k \, x \; + \; \frac{3}{2} \; k \, anharm \, x^2 \qquad 14$$

Display the results graphically as two asterisks separated by the distance x. To make the graphics a little easier, you can use the program fragments below. Start with:

```
R=8.314
T=298.2
k=200
m=10
anharm=0.05
dt=0.1
x=0
```

With these constants, increasing anharm to 0.1075 will cause the molecule to dissociate at 298.2K. Solve Eq. 12 for the initial velocity, v. Because of the way that computer languages handle the "=" sign, you can drop the subscripts on v and x, for example write:

v=v+F/m dt 15

and x=x+v dt. 16

After you get your spreadsheet or program to work, change the force constant k, the anharmonicity, and the temperature to note the effect.

The Spreadsheet Version: Set up columns using the integrated Newton's equations 15 and 16 to calculate x. Then to do the graphics, set up a column with values = x+10. The 10 is an arbitrary offset to make the graphics look good. In the next column, put in statements similar to

```
=REPT(" ",15-D17/2)&"*"&REPT(" ",D17)&"*"
```

but, instead of "D17" use the cell address of the adjacent column with the x+10 values. The result should look something like:

v	x	F	x+10	plot
15.74559	0	0	10	* *
12.96835	1.574559	-277.72	11.5746	* *
8.462302	2.871394	-450.61	12.8714	* *
3.100163	3.717624	-536.21	13.7176	* *
-2.52184	4.02764	-562.2	14.0276	* *
-7.93464	3.775457	-541.28	13.7755	* *
-12.5648	2.981993	-463.01	12.982	* *
-15.5692	1.725515	-300.44	11.7255	* *
-15.9021	0.168595	-33.293	10.1686	* *
-12.7557	-1.42162	314.639	8.57838	* *
-6.27013	-2.69719	648.561	7.30281	* *
2.03583	-3.32421	830.596	6.67579	* *
9.737817	-3.12062	770.199	6.87938	* *
14.72284	-2.14684	498.502	7.85316	* *
16.1402	-0.67456	141.737	9.32544	* *

The BASIC program : The program listed below will then take care of the plotting. Just slip in your constants and initial conditions before the loop. Then put the integrated Newton's equations 15 and 16 inside the loop. The IF statement is put in to signal the dissociation of the bond. When the molecule dissociates the program will print out "rrrip." With these constants, increasing anharm to 0.1075 will cause the molecule to dissociate at 298.2K.

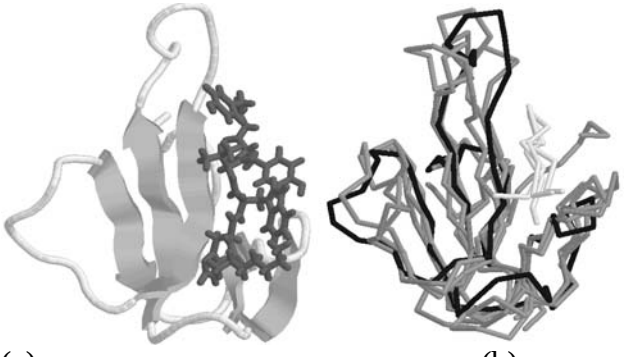
```
REM program to solve Hooke's Law dynamics
.
.
put constants and initial conditions in here
.
.
FOR i=1 TO 100
.
.
put Eq. 14, 15, and 16 in here
.
.
p=x+6
IF p>50 THEN LOCATE 1,1:PRINT"<<rrrip>>":GOTO qt
LOCATE 1,1
PRINT SPC(15-p);"*";SPC(INT(p+.5)+p);"*"
LOCATE 1,1
PRINT SPC(15-p);"  ";SPC(INT(p+.5)+p);"  ";
NEXT i
:
qt:
LOCATE 2,1
INPUT"type return to finish";a$
```

Introduction Section 6
Distance Geometry and 2D to 3D Model Conversion

Distance geometry is a general technique for generating 3D-models for chemical substances. Distance geometry is used in consort with energy minimization techniques to find low energy conformations for small molecules and large biomolecules.

Biomolecules: It is very difficult to find the global energy minimum for complex molecules. Proteins, for example, have many tens of thousands of local minima. Determining the lowest of the local minima can be a daunting task. Consider for example the ϕ and ψ angles along the protein backbone. For both angles there are roughly three low energy conformations, two gauche and one trans. Therefore each amino acid has roughly 3x3=9 possible conformations. If a protein has 20 amino acids the total possible backbone conformations is $9^{20} = 1.2x10^{19}$ conformations. However, 20 amino acids is a very small protein. The addition of side chain torsion angles greatly compounds the calculations. This problem is summarized by stating that proteins and nucleic acids have a very rough energy landscape. The valleys are the local minima and we need to visit each valley to find the lowest energy structure. We need help. Experimental information must be used to simplify the search for the tertiary structure of proteins and nucleic acids. Distance geometry is the mathematical technique that allows the construction of three-dimensional structures subject to the constraints provided by experimental information.[1-3] NMR is a particularly rich source of experimental constraints. The Protein Data Bank, PDB, has 22,000 protein structures, 15% of which were determined by NMR.[4,5]

The NMR solution structure for a model of the nicotinic acetylcholine receptor complexed with a potent natural antagonist is shown in Figure 1a. The larger structure is the antagonist, α-bungarotoxin, and the smaller is a portion of the antagonist binding site of the α-subunit of the nicotinic receptor showing amino acids 185 – 190. The distance geometry calculation used 325 distance constraints and 64 dihedral angle constraints.[6] Even so, the conformation of the peptides is still not completely specified. As a result, distance geometry was repeated producing a set of possible structures, Figure 1b. The NMR based structures in the Protein Data Bank are routinely sets of closely related structures that all satisfy the available experimental constraints. The ability of distance geometry to generate multiple structures is an important advantage for conformational searches in large and small molecules.



(a) (b)

Figure 1. (a). α-bungarotoxin (ribbon), and the antagonist binding site model of the α-subunit of the nicotinic acetylcholine receptor (stick); PDB entry 1ABT. (b) Four alternate structures for the complex derived by distance geometry, superimposed(backbone traces, with alternate solutions for bungarotoxin in different shades of gray and black and the protein receptor models in white).

2D to 3D Model conversion: Another closely related problem is the construction of the initial coordinates for molecular mechanics or molecular orbital calculations. The input for such

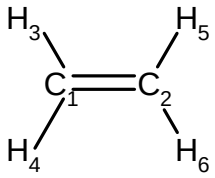
programs typically starts with the output of 2D "sketchers" or just connection tables. Molecular mechanics is a wonderful technique for predicting accurate 3D structures, however molecular mechanics programs routinely fail if the input structure is grossly distorted. Therefore, to get an accurate molecular mechanics calculation, you need to start with a structure that is not too far from a reasonable conformation. Molecular orbital programs also require that the input have a structure that is somewhat close to the geometry that you are seeking. Otherwise, the wrong atoms may end up being bonded to each other in the final structure. Therefore, it is very common to use molecular mechanics to produce the input file for molecular orbital calculations. When you use Spartan, the default is to build the input structure using the Merck Molecular Force Field when you minimize the structure in the Builder.[7] Most molecular orbital programs also use molecular mechanics to produce an initial guess for the Hessian. So even for molecular orbital calculations we have the same problem; we need a reasonable input structure even if the user isn't very adept at drawing the desired molecule on the screen.

The list of atom connections for a molecule is called the connection table. All molecular mechanics programs require a connection table for the input for each molecule, in addition to approximate atom positions. 2D-sketchers in their simplest form produce the connection table and 2D coordinates as drawn by the user. The third, z dimension needs to be added before a molecular mechanics calculation can proceed. The two common ways of building the approximate 3D-structure are functional group templates and distance geometry.

Functional group templates are simply the bond distances and angles specific to a given functional group taken from standard tables. For example, the bond angles around sp^2 hybridized carbons in alkenes and ketones are about 120°. The typical C=O bond length is 1.22Å. In other words, the ideal bond lengths and angles from standard force fields are used to guess the 3D-structure. The torsion angles present a problem since several torsion angles are possible, e.g. two gauche and trans angles for sp^3 systems. Most builders start with all trans structures unless the trans structure produces a close contact, at which point the gauche conformations are used. From the user's perspective there are two types of sketchers. The sketchers or builders in Spartan and MOE, for example, require the use of pre-built fragments to assemble the molecule.[8] These pre-built fragments already have the appropriate bond lengths and angles for the chosen functional group. As the molecule is built, the result is automatically constructed in 3D. The second approach for sketchers is to draw the molecule free-hand in 2D. Sketchers in chemical drawing programs and the Java Molecular Editor (JME)[9] are examples of this style. Free-hand sketchers present a real challenge since users can input structures that are wildly distorted. The coordinates presented by the user must be carefully adjusted to approximate real molecules. Rings systems present a particular problem with free-hand sketchers and require somewhat complex algorithms to set up using the template approach. Some sketchers maintain a database of the torsion angles for a wide variety of ring systems. Other sketchers use very approximate force fields and simplified minimization algorithms to guess the torsion angles around rings. Distance geometry is often an easier approach for ring systems. The Concord[10] and Corina programs are template based builders and provide amazingly accurate structures, even when compared to X-ray crystal structures.[11-13]

2D-sketchers work well for hands-on operation. However, completely automated procedures are also necessary. The rapid acceleration of the drug discovery process through combinatorial chemistry and high throughput screening has added an additional dimension to the 2D-3D conversion problem. Drug companies currently maintain storerooms filled with hundreds of thousands of compounds and develop combinatorial libraries (groups of compounds) of hundreds of thousands more. It is often necessary to store and retrieve information on all these compounds from exceedingly large computer databases. The efficient computer generation of 3D-models for

all these compounds is a daunting task. Often the structural information for these compounds is stored only as a connection table, so even 2D-information is not available. The connection table just specifies which atom is connected to which and the corresponding bond order. For example the connection table for ethylene, $H_2C=CH_2$, in the common format used by ".mol" files is given in Figure 2.



Atom 1	Atom 2	Bond order
1	2	2
1	3	1
1	4	1
2	5	1
2	6	1

Figure 2. Connection Table for Ethylene. The atom numbering is arbitrary.

The lack of any coordinates makes conversion of connection tables to molecular mechanics input files even harder. Corina, Concord, and distance geometry are designed to work from connection tables. One popular form of connection table is the "Smiles" string. Smiles strings are very efficient for storing large amounts of structural information.[14,15] The Smiles string for ethylene is just C=C. Some example Smiles strings are given in table I. Single bonds are assumed unless otherwise indicated. Hydrogens are omitted. Branching is shown by parentheses, i.e. *tert*-butanol is CC(C)(C)O. Ring closing connections are shown with numbers. Aromatic atoms are given in lower case. JME, drawing programs, and MOE can all be used to generate Smiles strings from sketches, so you don't really need to know the rules for generating Smiles.

Table I. Smiles strings for some molecules.

Butane CCCC	Ethanol CCO	Acetaldehyde CC=O
2-methylpropane CC(C)C	Acetone C(=O)C	Acetic acid CC(=O)O
Cyclohexane C1CCCCC1	benzene c1ccccc1	Toluene c1cccccc1C
Nitrobenzene c1ccccc1[N+](=O)[O-]	Phenylalanine NC(C(O)=O)Cc1ccccc1	

In summary, efficient calculation methods are needed for the construction of the 3D-coordinates of complicated molecules. Template based methods are very useful especially for small molecules. In many cases, however, some experimental information is known for a few distances or dihedral angles and the final structure must be built to include these structural parameters. Distance geometry can be applied to small and very large molecules and can easily incorporate experimental structural information in the form of distance constraints.

Distance Constraints

Distance constraints are ranges of allowable distances between pairs of atoms. An example is that you can specify that two atoms are to be within a normal hydrogen bond distance of each other, 1.8-2.1 Å. NMR spectra are very useful for experimentally determining distance constraints using nuclear Overhauser effects, nOe's. nOe based two-dimensional NMR spectra are called NOESY

spectra. Distance constraints from NOESY are particularly useful for studies of the tertiary structure of proteins. The combination of NOESY, distance geometry, and X-ray diffraction has spawned a new field in the molecular life sciences called Structural Biology. Using NOESY spectra it is possible to determine that pairs of atoms are within the range of about 3-4 Å of each other.[3] Just a few nOe based distance constraints can greatly simplify the search for the low energy structures of biomolecules.

Distance constraints and distance geometry can also be very useful for small molecule work. Hydrogen bond constraints and through-space nOe distances can be also useful for determining the conformation of small molecules as well as proteins. Distance constraints can also be values that you make up to help guide the conformation of the final molecule. For example, you may want a conformation that puts two parts of a long molecule close to each other rather than the default all-trans structures that most 2D-3D conversion programs generate.

Metric Matrix Distance Geometry

The input data for distance geometry are the distances between all the atoms in the molecule. The goal of distance geometry is to find the atom positions, x_i, y_i, z_i for each atom i. The metric matrix is used to calculate 3D atom coordinates using a process called embedding.[1-3] A triatomic molecule, Figure 3, will be used an example as we discuss the steps in embedding. The atom coordinates are

For atom 1: x_1, y_1, z_1 For atom 2: x_2, y_2, z_2 For atom 3: x_3, y_3, z_3 (1)

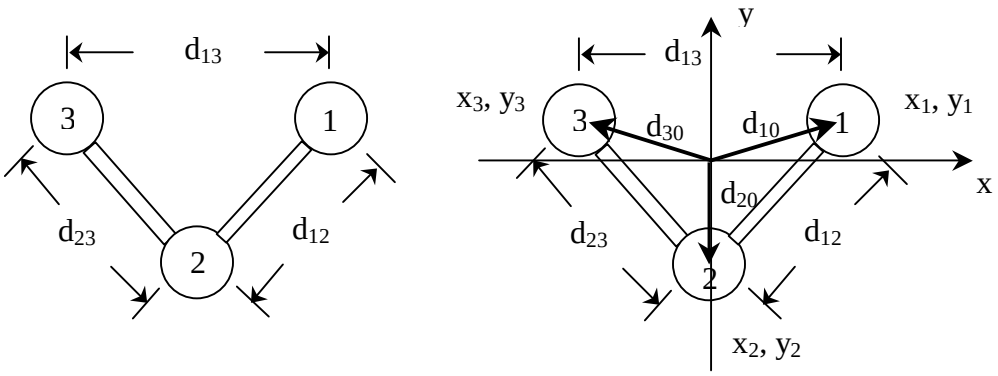


Figure 3. A triatomic molecule with (a) the input atom-atom distances and (b) the coordinate system for building the metric matrix from distances. The molecule is in the x-y plane, so z = 0 for all atoms.

The metric matrix is constructed from the dot products of the coordinate vectors. For example the dot product for atoms 1 and 2 is $(x_1x_2+y_1y_2+z_1z_2)$. The elements of metric matrix for atom pair i,j is then in general given by:

$$g_{ij} = x_ix_j+y_iy_j+z_iz_j \quad (2)$$

However, we don't know the atom coordinates at the beginning of the calculation; these coordinates are the final goal. Surprisingly, the metric matrix can also be constructed from atom

distances. The coordinate system for the important distances is shown in Figure 3b. The origin is the geometric center or centroid of the molecule. The centroid is constructed so that

$$\sum x_i = 0 \qquad \sum y_i = 0 \qquad \sum z_i = 0. \tag{3}$$

The elements of the metric matrix can then be calculated from (see appendix A).

$$g_{ij} = \frac{1}{2}\left(d_{i0}^2 + d_{j0}^2 - d_{ij}^2 \right) \tag{4}$$

However, we still have a problem. The distances to the origin, d_{i0} and d_{j0}, can't be calculated until we know the atom coordinates. However, a little bit of geometric reasoning allows the calculation of these distances (see appendix B). For N atoms:

$$d_{i0}^2 = \frac{1}{N}\sum_{j\neq i}^{N} d_{ij}^2 - \frac{1}{N^2}\sum_{j=1}^{N}\sum_{k>j} d_{jk}^2 \tag{5}$$

For our triatomic example:

$$d_{10}^2 = \frac{1}{3}\left(d_{12}^2 + d_{13}^2 \right) - \frac{1}{3^2}\left(d_{12}^2 + d_{13}^2 + d_{23}^2 \right) \tag{6}$$

A simple numerical example may help at this point. Let the bond distances, d_{12} and d_{23}, be 5 and the non-bonded distance, d_{13}, be 6. Then

$$d_{10}^2 = \frac{1}{3}\left(5^2 + 6^2 \right) - \frac{1}{3^2}\left(5^2 + 6^2 + 5^2 \right) = 10.777 \text{ or } d_{10} = 3.283 \tag{7}$$

$$d_{20}^2 = \frac{1}{3}\left(5^2 + 5^2 \right) - \frac{1}{3^2}\left(5^2 + 6^2 + 5^2 \right) = 7.111 \text{ or } d_{20} = 2.667 \tag{8}$$

Now the metric matrix entries can be calculated. For example g_{11} is easy since d_{11} in the second term of Eq 4 is 0. Substituting Eq 7 and 8 into Eq 4 gives:

$$g_{11} = \frac{1}{2}\left(d_{10}^2 + d_{10}^2 \right) = 10.778 \qquad g_{12} = \frac{1}{2}\left(d_{10}^2 + d_{20}^2 - d_{12}^2 \right) = -3.556 \tag{9}$$

Similar calculations give the final metric matrix:

$$G = \begin{pmatrix} 10.778 & -3.556 & -7.222 \\ -3.556 & 7.111 & -3.556 \\ -7.222 & -3.556 & 10.778 \end{pmatrix} \tag{10}$$

Given the metric matrix, as calculated from the atom-atom distances, we now need a way to work back to the original coordinates. The atom coordinates can be calculated from the eigenvalues and eigenvectors of the metric matrix. We find the eigenvalues λ_i, and the eigenvectors w_i by solving the equation:

$$G\, w_i = \lambda_i\, w_i \tag{11}$$

Where i corresponds to the x, y, or z axes. Eigen means "the same" in German, and Eq 11 shows that starting with w_i on the left gives back w_i on the right, multiplied by a constant, λ_i. In other

words, the same thing, w_i, appears on both sides of the equation. The eigenvalues are a measure of the size of the molecule in the x, y, and z directions (principle moments of inertia, but with unit mass for each atom). The atomic coordinates can then be calculated for each atom i:

$$x_i = \lambda_1 w_{i1} \qquad y_i = \lambda_2 w_{i2} \qquad z_i = \lambda_3 w_{i3} \qquad (12)$$

You can calculate the eigenvalues and eigenvectors using an on-line Web applet.[16] For our triatomic example, the eigenvalues and eigenvectors for the x and y directions are:

$$\lambda_1 = 18 \quad w_1 = \begin{pmatrix} 0.707 \\ 0 \\ -0.707 \end{pmatrix} \qquad \lambda_2 = 10.67 \quad w_2 = \begin{pmatrix} 0.408 \\ -0.816 \\ 0.408 \end{pmatrix} \qquad (13)$$

Giving the final coordinates:

$$\begin{aligned} x_1 &= 18^{1/2}\ 0.707 = 3 & y_1 &= 10.67^{1/2}\ 0.408 = 1.333 \\ x_2 &= 18^{1/2}\ 0 = 0 & y_2 &= 10.67^{1/2}\ -0.816 = -2.667 \\ x_3 &= 18^{1/2}\ -0.707 = -3 & y_3 &= 10.67^{1/2}\ 0.408 = 1.333 \end{aligned} \qquad (14)$$

These final coordinates are shown in Figure 4. Notice also that as Eq 3 requires, the sum of the x coordinates is zero and the sum of the y coordinates is also zero.

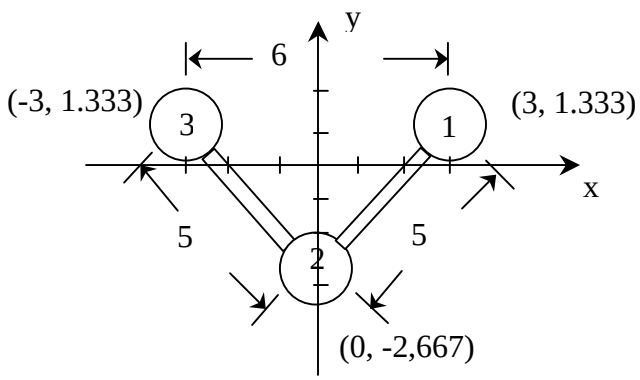


Figure 4. Final coordinates after embedding.

The remarkable thing about distance geometry is that it works just as well for thousands of atoms as it does for triatomic molecules. However, the calculation of the eigenvalues and vectors for large systems like proteins requires considerable computer time.

General Procedure[1-3]

The input for distance geometry programs is just the connection table. The complete embedding process requires four steps.

Step 1: The first step is to specify the distance range between every 1-2, 1-3, and 1-4 atom pair using standard bond lengths and angles from a table. For 1-4 distances the minimum distance is set for a 1-2-3-4 dihedral angle of 0° and the maximum distance for 180°. For non-bonded atoms the minimum is set to the sum of the Van der Waals radii. Any distance constraints that you input are also included in the list of distances. All the distance ranges are then checked for consistency using the triangle inequality, Figure 5. For example, the maximum distance between a 1-3 atom pair is the sum of the 1-2 and 2-3 bond distances, $d_{13} \leq d_{12} + d_{23}$:

If the atoms are C-C-C, the standard bond length is 1.53Å giving a maximum 1-3 distance of 2x1.53Å.

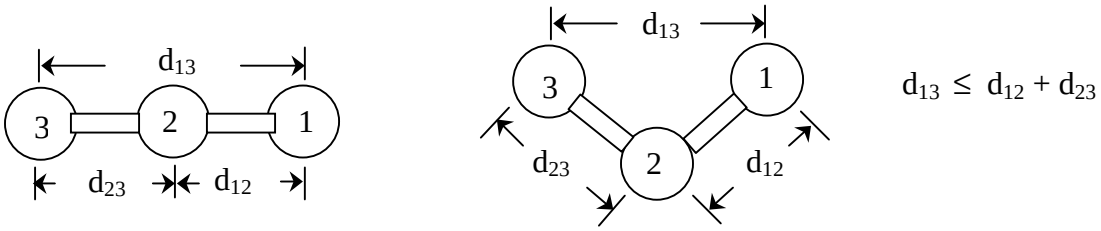


Figure 5. Checking distance maximums for consistency.

If an initially chosen maximum distance is larger than allowed by the triangle inequality, the value is lowered. This "smoothing" process helps to tighten the distance constraints. The minimum distances are also smoothed in the same way. The result is a set of consistent upper and lower bounds for all the pair-wise distances between the atoms.

Step 2: Next a distance between each atom pair is chosen at random between the upper and lower bounds set in step 1. This step is an important feature of distance geometry. The assignment of a random distance means that you will obtain a different result each time you run the algorithm. This element of randomness is one of the advantages (and disadvantages) of the distance geometry approach that can be exploited for conformational searches.

Step 3: The metric matrix is calculated from the chosen random distances. The eigenvalues and vectors are then calculated and used to find the final coordinates using Eq 12.

Step 4: The coordinates generated by distance geometry are very rough. The atom positions must be optimized using molecular mechanics with a simplified force field. This adjustment process is done in two steps. First any chiral constraints are enforced. Working with chiral constraints first is necessary because molecular mechanics minimization can switch chirality inadvertently, and also enforcing chirality first makes subsequent minimization faster. After chiral constraints are satisfied, the coordinates are adjusted with a force field that greatly penalizes atom positions that violate the distance bounds that were established in Step 1.

The force field first checks to see if the distance between atom i and atom j, d_{ij}, is outside of the distance bounds; if outside an error term is calculated:

$$e = \frac{(d_{ij}^2 - B_{ij}^2)^2}{B_{ij}^2} \qquad (15)$$

where B_{ij} is the violated upper or lower bound. This error term is summed over all atom pairs that violate the distance bounds. Violations of the chiral constraints are also added to the distance errors to complete the force field. This force field is minimized using standard conjugate gradient techniques. The force field does not include bond stretch, angle bending, out-of-plane, and torsional constraints directly. Therefore, the optimized structure is only as good as the original distance bounds. In other words, the final structure is still quite crude and must be further minimized using a traditional force field or molecular orbital calculations.

For example, Figure 6a shows the results for toluene. The ring atoms are not flat as expected for sp^2 hybridized atoms. The conformation around sp^2 atoms can be improved by specifying the atoms as chiral atoms, even though they are not, alternating (+) and (-) around the ring. The distance geometry program will enforce a flat geometry, Figure 6b.

a. b.

Figure 6. (a) Distance geometry results for toluene. (b) Distance geometry with sp^2 C atoms specified as chiral.

Even so, the results are still quite distorted from the expected planar geometry. Submitting the distance geometry results to a conventional molecular mechanics of molecular orbital calculation quickly clears up any remaining problems. The distance geometry results for complicated ring systems can often be quite good, however.

Extensions to Distance Geometry

Many extensions of the basic distance geometry procedure have been implemented provide final structures with less strain.[17,18] Distance geometry is also used in conjunction with molecular dynamics for energy minimization.[19,20] The technique is often used for conformation searches,[2, 19,20] in aspects of drug discovery,[21,23] and protein folding studies.[24,25]

Appendix A:

Given atom 1 with coordinates x_1, y_1, z_1 and atom 2 with coordinates x_2, y_2, z_2 the distance between the two atoms is:

$$d_{12}^2 = (x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2 = x_1^2+2x_1x_2+ x_2^2+ y_1^2+2y_1y_2+ y_2^2+ z_1^2+2z_1z_2+ z_2^2 \quad (17)$$

The dot product between the two atoms coordinates is $x_1x_2+y_1y_2+z_1z_2$. Rearranging Eq 17 to isolate the dot product gives:

$$d_{12}^2 = (x_1^2+ y_1^2+ z_1^2)+ (x_2^2+ y_2^2 + z_2^2) + 2(x_1x_2 +y_1y_2 +z_1z_2) \quad (18)$$

The first term in parenthesis is the squared distance of atom 1 from the origin, d_{10}^2 . The second term is the distance of atom 2 from the origin, d_{20}^2 . Rearranging Eq 18 gives:

$$(x_1x_2 +y_1y_2 +z_1z_2) = \frac{1}{2}(d_{10}^2 + d_{20}^2 - d_{12}^2) \quad (19)$$

(Many authors on distance geometry describe Eq 19 as the Law of Cosines, which is a standard geometrical construction. Given two vectors $\vec{v}$ and $\vec{w}$, the dot product is:

$$\vec{v} \bullet \vec{w} = |\vec{v}||\vec{w}| \cos\theta = d_{10}\, d_{20} \cos\theta$$

which explains the connection with the cosine of the angle and the name “Law of Cosines.”)

Appendix B:

The fact that the distance of an atom from the origin can be calculated completely from the atom-atom distances using Eq 5 is surprising. A derivation for three atoms, Eq. 6, is given in this appendix. A more general derivation is given by Havel, et. al.[26] However, the general formula is easily obtained from the three atom result. The distance of atom 1 from the origin is

$$d_{10}^2 = x_1^2 + y_1^2 + z_1^2 \qquad (20)$$

The origin is the centroid of the atoms, Eq. 3:

$$x_1 + x_2 + x_3 = 0 \qquad y_1 + y_2 + y_3 = 0 \qquad z_1 + z_2 + z_3 = 0 \qquad (21)$$

Solving for the coordinates of atom 1 gives:

$$x_1 = - x_2 - x_3 \qquad y_1 = - y_2 - y_3 \qquad z_1 = - z_2 - z_3 \qquad (22)$$

Substituting Eq 22 into Eq 20 for one factor of x_1, y_1, and z_1 gives:

$$d_{10}^2 = -x_1x_2 - x_1x_3 - y_1y_2 - y_1y_3 - z_1z_2 - z_1z_3 \qquad (23)$$

Rearranging gives two dot products:

$$d_{10}^2 = -(x_1x_2 + y_1y_2 + z_1z_2) - (x_1x_3 + y_1y_3 + z_1z_3) \qquad (24)$$

Using Eq 19 for the dot products gives:

$$d_{10}^2 = -\frac{1}{2}[(d_{10}^2 + d_{20}^2 - d_{12}^2) + (d_{10}^2 + d_{30}^2 - d_{13}^2)] \qquad (25)$$

Rearranging gives:

$$4\, d_{10}^2 = (d_{12}^2 - d_{20}^2) + (d_{13}^2 - d_{30}^2) \qquad (26)$$

Subtracting a term in d_{10}^2 from both sides gives:

$$3\, d_{10}^2 = d_{12}^2 + d_{13}^2 - (d_{10}^2 + d_{20}^2 + d_{30}^2) \qquad (27)$$

The corresponding results for the other two atoms are

$$3\, d_{20}^2 = d_{12}^2 + d_{23}^2 - (d_{10}^2 + d_{20}^2 + d_{30}^2) \qquad (28)$$

$$3\, d_{30}^2 = d_{13}^2 + d_{23}^2 - (d_{10}^2 + d_{20}^2 + d_{30}^2) \qquad (29)$$

Adding Eqs 27-29 gives:

$$3(d_{10}^2 + d_{20}^2 + d_{30}^2) = 2(d_{12}^2 + d_{13}^2 + d_{23}^2) - 3(d_{10}^2 + d_{20}^2 + d_{30}^2) \qquad (30)$$

Solving for the sum squared distances to the origin gives a result entirely in terms of atom-atom distances:

$$(d_{10}^2 + d_{20}^2 + d_{30}^2) = \frac{1}{3}(d_{12}^2 + d_{13}^2 + d_{23}^2) \qquad (31)$$

Finally substituting Eq 31 for the last term in Eq 27 gives:

$$3\, d_{10}^2 = d_{12}^2 + d_{13}^2 - \frac{1}{3}(d_{12}^2 + d_{13}^2 + d_{23}^2) \qquad (32)$$

Finally division by 3 gives Eq. 6.

Literature Cited:

1. Kuntz, I. D.; Thomason, J. F.; Oshiro, C. M., "Distance Geometry," in *Methods in Enzymology, Volume 177 Nuclear Magnetic Resonance Part B*; Oppenheimer, N. J., James, T. L., Eds.; Academic Press: San Diego, CA, 1989; pp 159-204.

2. Crippen, G. M.; Havel, T. F., *Distance Geometry and Molecular Conformation*; Bawden, D., Ed.; Chemometrics Research Studies Series 139; Research Studies Press (Wiley): New York, 1988.

3. Sutcliffe, M. J., "Structure Determination from NMR Data, II, Computational Approaches," Section 4.3.1 "Metric matrix distance geometry," in *NMR of macromolecules : a practical approach***;** Roberts, G. C. K., Ed.; Oxford: New York, 1993; pp374-377.

4. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., "The Protein Data Bank," *Nucleic Acids Res.*, **2000**, *28*, 235-242.

5. Berman, H. M.; Goodsell, D. S.; Bourne, P. E., "Protein Structures: From Famine to Feast," *Amer. Scientist*, **2002**, *90(4)*, 350 – 359.

6. Basus, V. J.; Song, G.; Hawrot, E., "NMR Solution Structure of an α-Bungarotoxin/Nicotinic Receptor Peptide Complex," *Biochemistry*, **1993**, *32*, 12290-12298.

7. Wavefunction, Inc., *Spartan*, Version 4.0; Irvine, CA, 2001.

8. Chemical Computing Group, *Molecular Operating Environment*; Montreal, Quebec, CA, 2002.

9. Ertl, P., *JME Molecular Editor*; Novartis, A.G.: Basil, Switzerland, 2003; http://www.molinspiration.com/jme/ (accessed Jan 2003).

10. Pearlman, R. S., "Rapid Generation of High Quality Approximate 3D Molecular Structures," *Chem. Des. Auto. News*, **1987**, *2*, 1-7.

11. Sadowski, J.; Gasteiger, J., "From Atoms and Bonds to Three-Dimensional Atomic Coordinates: Automatic Model Builders," *Chem. Rev.*, **1993**, *93*, 2567-2581.

12. GasteigerResearch, *Corina*, http://www2.chemie.uni-erlangen.de/software/corina/free_struct.html (accessed Jan 2003).

13. Sadowski, J.; Gasteiger, J.; Klebe, G., "Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures," *J. Chem. Inf. Comput. Sci.*, **1994**, *34*, 1000-1008.

14. Daylight Chemical Information Systems, Inc; Smiles Home Page, http://www.daylight.com/dayhtml/smiles/ (accessed Jan 2003).

15. Weininger, D., "SMILES 1. Introduction and Encoding Rules," *J. Chem. Inf. Comput. Sci.*, **1988**, *28*, 31-36.

16. Colby College Physical Chemistry Home Page, *eigen*, http://www.colby.edu/chemistry/PChem/eigen.html (accessed Jan 2003).

17. Brunger, A. T., *X-PLOR, version 3.1. A system for X-ray crystallography and NMR*; Yale University Press: New Haven, CT; 1992.

18. Crippen, G. M.; Havel, T. F., "Global Energy Minimization by Rotational Energy Embedding," *J. Chem. Inf. Sci.*, **1990**, *30*, 222-227.

19. Scheek, R. M.; van Gunsteren, W. F.; Kaptein, R., “Molecular Dynamics Simulation Techniques for Determination of Molecular Structures from Nuclear Magnetic Resonance Data,” in *Methods in Enzymology, Volume 177 Nuclear Magnetic Resonance Part B*; Oppenheimer, N. J., James, T. L., Eds.; Academic Press: San Diego, CA, 1989; pp 204-218.

20. de Vlieg, J.; van Gunsteren, W. F., ”Combined Procedures of Distance Geometry and Molecular Dynamics for Determining Protein Structure from Nuclear Magnetic Resonance,” in *Methods in Enzymology, Volume 202 Molecular Design and Modeling: Concepts and Applications Part A*; Langone, J. J., Ed.; Academic Press: San Diego, CA, 1991; pp 268-301.

21. Crippen, G. M., “Distance Geometry Approach to Rationalizing Binding Data,” *J. Med. Chem*., **1979**, *22(8)*, 988-997.

22. Ghose, A. K.; Crippen, G. M., “Use of Physicochemical Parameters in Distance Geometry and Related Three-Dimensional Quantitative Strucutre-Activity Relationships: A Demonstration Using *Escherichia coli* Dihydrofolate Reductase Inhibitors,” *J. Med. Chem*., **1985**, *28*, 333-346.

23. Sheridan, R. P.; Ramaswamy, N.; Dixon, J. S.; Venkataraghavan, R., “The Ensemble Approach to Distance Geometry,” *J. Med. Chem*., **1986**, *29*, 899-906.

24. Huang, E. S.; Samudrala, R.; Ponder, J. W., “Distance Geometry generates native-like folds for small helical proteins using the concensus distances of predicted protein structures,” *Protein Sci*., **1998**, *7*, 1998-2003.

25. Bonneau, R.; Tsai, J.; Ruczinski, I.; Chivian, D.; Rohl, C.; Strauss, C.; Baker, D, “Rosetta in CASP4: progress in *ab initio* protein structure prediction,” *Proteins: Structure, Function and Genetics*, **2001**, *45:S5*, 119-126.

26. Havel, T. F.; Kuntz, I. D.; Crippen, G. M., “The Theory and Practice of Distance Geometry,” *Bull. Math. Biol*., **1983**, *45(5)*, 665-720.

Introduction Section 7
Free Energy Perturbation Theory, FEP

The greatest value in molecular dynamics is the ability to model the internal motions of a molecule. Internal energy, enthalpy, entropy, and Gibb's Free Energy all include contributions from the motion of a molecule. Therefore, molecular dynamics provides a way to estimate these important thermodynamic parameters. The current best method for practical calculations of Gibb's Free Energies is free energy perturbation theory, based on molecular dynamics. Free energy perturbation (FEP) theory is now in use in calculating ΔG for a wide variety of processes. For example, the Gibb's Free Energy of solution of hydrophobic molecules[1], of binding of crown ethers to polar organics[2], and the binding of NADP and NADPH to dihydrofolate reductase[3] have been studied. In fact, the combined insights of x-ray crystal structure determination, NMR solution structure determination, and FEP studies have led to the consensus that the motions of proteins and nucleic acids play a major role in binding interactions. W. L. Jorgensen, in his article "Rusting of the Lock and Key Model for Protein-Ligand Binding," states simply that:

> *"These examples confirm the reasonable expectation that flexible molecules distort to form optimal interactions with binding partners."*[4]

A dynamic view of binding interactions is necessary to understand biochemical phenomena.

Molecular mechanics calculates the steric energy of a molecule at absolute zero in temperature. What is the connection of the molecular mechanics steric energy to the thermodynamic internal energy and Gibb's Free energy of a substance? The hypothesis that makes the most sense is that the internal energy, ΔU, is the time average of the total energy of the molecule. The total energy of the molecule is the kinetic plus potential energy:

$$E = \text{kinetic energy} + \text{potential energy} \qquad 1$$

The potential energy is just the molecular mechanics steric energy. Molecular dynamics provides us with the time dependent energy of the molecule; all we need do to get ΔU is average the total energy during the trajectory calculation.

Now we turn to the relationship of the steric energy to the Gibb's Free Energy. In statistical mechanics, we find that the probability of a given state of a system occurring is proportional to the Boltzman weighting factor:

$$\text{probability of occurrence} \ \alpha \ e^{-E/RT} \qquad 2$$

where E is the total energy of the system, Eq. 1. In other words, states with low total energy are more likely to occur than states with high energy. A state of the system is determined by the conformation and motion of the molecule. The conformation determines the steric energy and the motions determine the kinetic energy.

In perturbation theory, we look at the effect of a small change in the structure of a molecule on its energy. To do the perturbation, the total energy is divided into two parts

$$E = E_o + E_1 \qquad 3$$

where E_o is a reference structure and E_1 is a small perturbation from the reference structure. The perturbation is a small change that we place upon the system, say a small change in bond angle or a small change in the charge on an atom. The corresponding change in free energy of the system caused by the perturbation is given as[5,6]

$$G - G_o = -RT \ln \langle e^{-E_1/RT} \rangle_o \qquad 4$$

where $\langle \rangle_o$ denotes the time average over the motion of the reference structure from a molecular dynamics run. The $e^{-E_1/RT}$ term is the probability of occurrence for the small change in energy caused by the perturbation, from Eq. 2. The free energy then depends on the time average of the probability of occurrence of the perturbed structure. In other words, if the perturbation produces a small change in energy, that change will contribute to the Gibb's Free energy.

In our case however, we wish to find the change in free energy for large changes in a molecule. These changes, or mutations, include changing the conformations of bonds, or attaching a hydrogen ion, or changing a hydrogen to a methyl group or even a phenyl group. For example, we might like to mutate glycine into alanine[7] for a study of site-specific mutagenesis of an enzyme. How do we apply Eq. 4 to such large changes? Assume that we wish to mutate molecule B into a different molecule A. First we define a total energy for mutating molecule B to A as

$$E_\lambda = \lambda E_A + (1 - \lambda) E_B \qquad 5$$

where E_A is the total energy for A and E_B is that for B, and λ is the coupling parameter. When $\lambda = 1$ the energy corresponds to molecule A, and when $\lambda = 0$ the energy corresponds to molecule B. When λ is at intermediate values, the system is a hypothetical superposition of A and B. It might seem quite strange to have such a combination of two molecules, in fact it is very unphysical; however, the theory is well-behaved and very useful none-the-less.

For the complete mutation to take place we vary λ from 0 to 1 over the course of the dynamics run. We divide this full range into short time slices, which are short enough that we can treat the change in each time slice as a perturbation. Then we apply Eq. 4 to each time slice and then add up the result for all the time slices. Let the λ value at each time slice be numbered λ_1, λ_2, λ_3, etc. Then the difference in Eq. 4 is $\Delta G(\lambda_i)$ for each time slice, i=1, 2 ,3,...n, for n total time slices. Then the total change in ΔG for the perturbation is

$$\Delta G_{B->A} = \sum_{i=1}^{n} \Delta G(\lambda_i) \qquad 6$$

Since each time slice in the mutation is a small change, we can simplfy Eqs. 4 and 6. We do the mutation is small steps; therefore $E_1 << RT$ for each time slice in the perturbation. Remembering that $e^{-x} \approx 1-x$, we can expand the exponential in the Boltzman distribution:

$$e^{-E_1/RT} \approx 1 - E_1/RT \qquad 7$$

Then Eq. 4 simplifies to:

$$G - G_o = -RT \ln \langle 1 - E_1/RT \rangle_o = -RT \ln (1 - \langle E_1 \rangle_o/RT) \qquad 8$$

Next remember that $\ln(1-x) \approx -x$, when x is small. This approximation on Eq. 8 gives:

$$G - G_o = -RT (- \langle E_1 \rangle_o/RT) = \langle E_1 \rangle_o \qquad 9$$

In words, this simple result means that the change in Gibb's Free Energy for a perturbation is just the time average of the total perturbation energy. Now applying Eq. 9 to each time slice in the total mutations simplifys Eq. 6 to:

$$\Delta G_{B->A} = \sum_{i=1}^{n} < E(\lambda_i) >_o \qquad 10$$

where $E(\lambda_i)$ is the total energy for the time slice in the mutation from Eq. 5. This very simple result makes FEP studies easy to do. The time average in Eq. 9 is automatically calculated during trajectory calculations. All we need do is to change λ in small steps during the trajectory. This approach to FEP simulations is called the slow-growth method.

Our initial efforts to use molecular dynamics are frustrated, however, because molecular dynamics is a classical theory, which gives too high a weight to high frequency vibrations. We must be careful to account for the difference between classical theories and the true distribution of vibrational energies in molecules. We can do this by always calculating the difference between our system and a reference system. In calculating differences, errors tend to cancel, and in so doing, classical molecular dynamics is a suprisingly useful tool for understanding complex systems. The success of classical dynamics is due in part to the observation that the major contributions to ΔG for solvation and binding interactions are low frequency vibrations, especially torsions, which are handled adequately by classical theory. In addition, these low frequency vibrations tend to change the most in systems of interest; high frequency vibrations change little, therefore the high frequency vibrations cancel out in comparisions.

For example, to study the Gibb's Free Energy of solvation of molecule B, $\Delta_{sol}G_B$, we will choose molecule A as the reference structure. The mutation will then be from B to A. To determine the difference in Free Energy of solvation between B and A, we will construct the following thermodynamic cycle:

$$\begin{array}{ccc} & \Delta_{sol}G_A & \\ A\ (aq) & \text{---->} & A\ (g) \\ \uparrow & & \vert \\ \Delta G^{aq}_{B->A} \vert & & \vert\ -\Delta G^{g}_{B->A} \\ \vert & \Delta_{sol}G_B & \downarrow \\ B\ (aq) & \text{---->} & B\ (g) \end{array} \qquad 11$$

where $\Delta G^{aq}_{B->A}$ is the Free Energy of perturbation of B to A in the solution phase, and $\Delta G^{g}_{B->A}$ is the Free Energy of perturbation in the gas phase. Adding contributions around the cycle gives:

$$\Delta_{sol}G_B = \Delta G^{aq}_{B->A} + \Delta_{sol}G_A - \Delta G^{g}_{B->A} \qquad 12$$

We then determine the difference

$$\Delta_{sol}G_B - \Delta_{sol}G_A = \Delta G^{aq}_{B->A} - \Delta G^{g}_{B->A} \qquad 13$$

These kinds of differences are often called $\Delta\Delta G$ values:

$$\Delta\Delta G = \Delta_{sol}G_B - \Delta_{sol}G_A = \Delta G^{aq}_{B->A} - \Delta G^{g}_{B->A} \qquad 14$$

We choose a reference system, A, where $\Delta_{sol}G_A$ is known from experiment. We can then predict our final result:

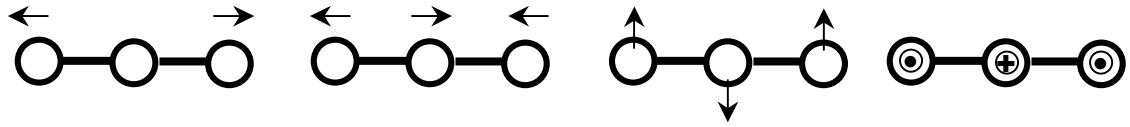
$$\Delta_{sol}G_B = \Delta_{sol}G_{A(experimental)} + \Delta\Delta G \qquad 15$$

Literature Cited

1. B. G. Rao, U. C. Singh, *J. Amer. Chem. Soc.* **1989**, *111*, 3125.
2. P. D. J. Grootenhuis, P. A. Kollman, *J. Amer. Chem. Soc.* **1989**, *111*, 4046.
3. P. L. Cummins, K. Ramnarayan, U. C. Singh, and J. E. Greadt, *J. Amer. Chem. Soc.* **1991**, *113*, 8247.
4. W. L. Jorgensen, *Science*, **1991**, *254*, 954.
5. D.A. McQuarrie, *Statistical Mechanics*, Harper and Row, New York, NY, 1976. Section 14-1,p 302.
6. R. W. Zwanzig, *J. Chem. Phys.* **1954**, *22*, 1420.
7. U. C. Singh. F. K. Brown, P. A. Bash, P. A. Kollman, *J. Amer. Chem. Soc.*, **1987**, *109(6)*, 1607.

Section 8
Classical Normal Mode Analysis: Harmonic Approximation

The vibrations of a molecule are given by its normal modes. Each absorption in a vibrational spectrum corresponds to a normal mode. The four normal modes of carbon dioxide, Figure 1, are the symmetric stretch, the asymmetric stretch and two bending modes. The two bending modes have the same energy and differ only in the direction of the bending motion. Modes that have the same energy are called degenerate. In the classical treatment of molecular vibrations, each normal mode is treated as a simple harmonic oscillator.



Symmetric stretch Asymmetric stretch Bend Bend

Figure 1. Normal Modes for a linear triatomic molecule. In the last bending vibration the motion of the atoms is in-and-out of the plane of the paper.

In general linear molecules have 3N-5 normal modes, where N is the number of atoms. The five remaining degrees of freedom for a linear molecule are three coordinates for the motion of the center of mass (x, y, z) and two rotational angles. Non-linear molecules have three rotational angles, hence 3N-6 normal modes.

The characteristics of normal modes are summarized below.

Characteristics of Normal Modes

1. Each normal mode acts like a simple harmonic oscillator.
2. A normal mode is a concerted motion of many atoms.
3. The center of mass doesn't move.
4. All atoms pass through their equilibrium positions at the same time.
5. Normal modes are independent; they don't interact.

In the asymmetric stretch and the two bending vibrations for CO_2, all the atoms move. The concerted motion of many of the atoms is a common characteristic of normal modes. However, in the symmetric stretch, to keep the center of mass constant, the center atom is stationary. In small molecules all or most all of the atoms move in a given normal mode; however, symmetry may require that a few atoms remain stationary for some normal modes. The last characteristic, that normal modes are independent, means that normal modes don't exchange energy. For example, if the symmetric stretch is excited, the energy stays in the symmetric stretch.

The background spectrum of air, Figure 2, shows the asymmetric and symmetric stretches and the bending vibration for water, and the asymmetric stretch and bending vibrations for CO_2. The symmetric stretch for CO_2 doesn't appear in the Infrared; a Raman spectrum is needed to measure the frequency of the symmetric stretch. These absorptions are responsible for the vast majority of the greenhouse effect. We will also use CO_2 as an example, below.

The normal modes are calculated using Newton's equations of motion.[1-4] Molecular mechanics and molecular orbital programs use the same methods. Normal mode calculations are available on-line.[5]

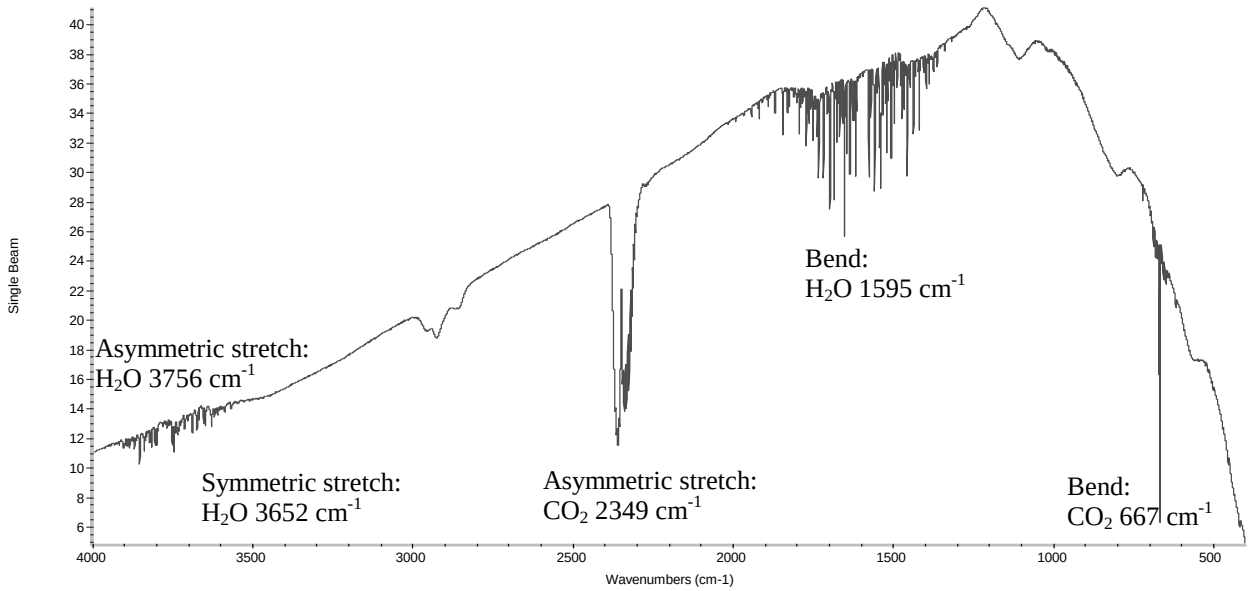


Figure 2. The Infrared spectrum of air. This spectrum is the background scan from an FT-IR spectrometer.

Harmonic Oscillator Review

Lets first review the simple harmonic oscillator. Consider a mass m, supported on a spring with force constant k. Hooke's Law for the restoring force for an extension, x, is F = -kx. In other words, if the spring is stretched a distance x>0, the restoring force will be negative, which will act to pull the mass back to its equilibrium position. The potential energy for Hooke's Law is obtained by integrating

$$F = -\frac{dV}{dx} = -kx \qquad (1)$$

to give $V = \frac{1}{2} k x^2$ (2)

In molecular mechanics and molecular orbital calculations, the force constant is not known. However, the force constant can be calculated from the second derivative of the potential energy.

$$k = \frac{d^2 V}{dx^2} \qquad (3)$$

The Hooke's Law force is substituted into Newton' Law:

$$F = ma \qquad \text{or} \qquad m\frac{d^2 x}{dt^2} = -kx \qquad (4)$$

and solved to obtain the extension as a function of time:

$$x(t) = A \sin(2\pi\nu t) \qquad (5)$$

where ν is the fundamental vibration frequency and A is the amplitude of the vibration. Taking the second derivative of the extension gives

$$\frac{d^2 x}{dt^2} = -4\pi^2\nu^2 x \qquad (6)$$

Substituting Eq 6 back into Eq 4 gives:

$$-4\pi^2\nu^2 m x = -kx \qquad (7)$$

which is the basis for the classical calculation of the normal modes of a molecule.

Normal Mode Analysis

For molecules the x, y, z coordinates of each atom must be specified. The coordinates are:
Atom 1: X_1, Y_1, Z_1, Atom 2: X_2, Y_2, Z_2, etc. ……
The extensions are the differences in the positions and the equilibrium positions for that atom:

$$\begin{array}{lll} \text{Atom 1: } x_1 = X_1 - X_{1,eq} & y_1 = Y_1 - Y_{1,eq} & z_1 = Z_1 - Z_{1,eq} \\ \text{Atom 2: } x_2 = X_2 - X_{2,eq} & y_2 = Y_2 - Y_{2,eq} & z_2 = Z_2 - Z_{2,eq} \\ \text{Atom i: } x_i = X_i - X_{i,eq} & y_i = Y_i - Y_{i,eq} & z_i = Z_i - Z_{i,eq} \end{array} \quad (8)$$

Where $X_{i,eq}$, $Y_{i,eq}$, and $Z_{i,eq}$ are the equilibrium (energy minimized) positions for atom i. For example, if x_1, y_1, and z_1 are all zero, then atom 1 is at its equilibrium position. Molecular mechanics or molecular orbital calculations are used to find the potential energy of the molecule as a function of the position of each atom, $V(x_1, y_1, z_1, x_2, y_2, z_2, x_3, y_3, z_3,...,x_N,y_N,z_N)$.
The second derivative of the potential energy can then be used to calculate the force constants, Eq 3. However, there are now 3Nx3N possible second derivatives and their corresponding force constants. For example,

$$\frac{\partial^2 V}{\partial x_1^2} = k_{xx}^{11} \quad (9)$$

is the change of the force on atom 1 in the x-direction when you move atom 1 in the x-direction. Similarly,

$$\frac{\partial^2 V}{\partial x_1 \partial y_2} = k_{xy}^{12} \quad (10)$$

is the change of the force on atom 1 in the x-direction when you move atom 2 in the y-direction. The various types of force constants are shown in Figure 3.

$\frac{\partial^2 V}{\partial x_1^2} = k_{xx}^{11}$	same atom same direction
$\frac{\partial^2 V}{\partial y_1^2} = k_{yy}^{11}$	same atom same direction
$\frac{\partial^2 V}{\partial x_1 \partial y_1} = k_{xy}^{11}$	same atom different directions
$\frac{\partial^2 V}{\partial x_1 \partial x_2} = k_{xx}^{12}$	different atom same direction
$\frac{\partial^2 V}{\partial x_1 \partial y_2} = k_{xy}^{12}$	different atom and direction

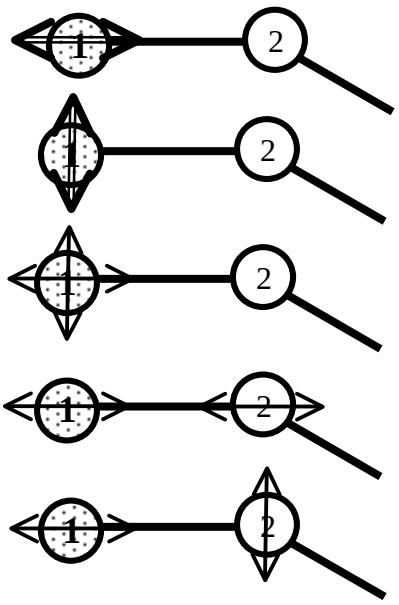


Figure 3. Types of second derivatives and force constants

These force constants are not the force constants for individual bonds, they are force constants for the motion of a single atom subject to all its neighbors, whether directly bonded or not. The

complete list of these force constants is called the Hessian, which is a 3Nx3N matrix. Eq 7 is then applied for each force constant.[1,2]

$$-4\pi^2\nu^2\, m_1x_1 = -k_{xx}^{11}x_1 - k_{xy}^{11}y_1 - k_{xz}^{11}z_1 - k_{xx}^{12}x_2 - k_{xy}^{12}y_2 - \ldots - k_{xz}^{1N}z_N \qquad (11)$$

$$-4\pi^2\nu^2\, m_1y_1 = -k_{yx}^{11}x_1 - k_{yy}^{11}y_1 - k_{yz}^{11}z_1 - k_{yx}^{12}x_2 - k_{yy}^{12}y_2 - \ldots - k_{yz}^{1N}z_N$$

$$\vdots$$

$$-4\pi^2\nu^2\, m_2x_2 = -k_{xx}^{21}x_1 - k_{xy}^{21}y_1 - k_{xz}^{21}z_1 - k_{xx}^{22}x_2 - k_{xy}^{22}y_2 - \ldots - k_{xz}^{2N}z_N$$

$$\vdots$$

$$-4\pi^2\nu^2\, m_Nz_N = -k_{zx}^{N1}x_1 - k_{zy}^{N1}y_1 - k_{zx}^{N1}z_1 - k_{zx}^{N2}x_2 - k_{zy}^{N2}y_2 - \ldots - k_{zz}^{NN}z_N$$

In words, the right-hand sides of the above equations simply state that the total force on atom i is the sum of the forces of all the atoms on atom i. In addition, we need to keep track of the x, y, and z directions for each atom. There are a total of 3Nx3N terms on the right. All these terms are confusing. A simple example will help at this point.

For our example consider a symmetrical linear triatomic molecule that can only vibrate along the x-axis, Figure 4. CO_2 is a good example of a symmetrical linear triatomic.

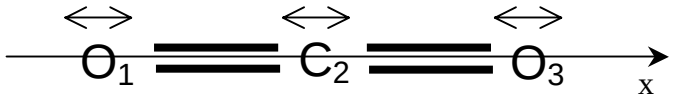


Figure 4. A symmetrical triatomic molecule with vibrations limited along the internuclear axis.

Because we have limited the vibrations to the x-axis, which is the internuclear axis, this model will provide the symmetric and asymmetric stretching modes, only. Eqs 11 then reduce to

$$-4\pi^2\nu^2\, m_1x_1 = -k_{xx}^{11}x_1 - k_{xx}^{12}x_2 - k_{xx}^{13}x_3 \qquad (12)$$

$$-4\pi^2\nu^2\, m_2x_2 = -k_{xx}^{21}x_1 - k_{xx}^{22}x_2 - k_{xx}^{23}x_3 \qquad (13)$$

$$-4\pi^2\nu^2\, m_3x_3 = -k_{xx}^{31}x_1 - k_{xx}^{32}x_2 - k_{xx}^{33}x_3 \qquad (14)$$

since we only need to keep the x-terms. Several numerical techniques are available to solve linear sets of simultaneous equations such as this. Conventionally, however, the problem is simplified by converting to mass weighted coordinates, for example:

$$\tilde{x}_1 = \sqrt{m_1}\, x_1 \qquad \tilde{x}_2 = \sqrt{m_2}\, x_2 \quad \text{, etc.} \qquad (15)$$

and mass weighted force constants:

$$\tilde{k}_{xx}^{12} = \frac{k_{xx}^{12}}{\sqrt{m_1}\sqrt{m_2}} \qquad (16)$$

In the new mass weighted coordinates, Eqs 12-14 become:

$$-4\pi^2\nu^2 \tilde{x}_1 = -\tilde{k}^{11}_{xx}\tilde{x}_1 - \tilde{k}^{12}_{xx}\tilde{x}_2 - \tilde{k}^{13}_{xx}\tilde{x}_3 \quad (17)$$

$$-4\pi^2\nu^2 \tilde{x}_2 = -\tilde{k}^{21}_{xx}\tilde{x}_1 - \tilde{k}^{22}_{xx}\tilde{x}_2 - \tilde{k}^{23}_{xx}\tilde{x}_3 \quad (18)$$

$$-4\pi^2\nu^2 \tilde{x}_3 = -\tilde{k}^{31}_{xx}\tilde{x}_1 - \tilde{k}^{32}_{xx}\tilde{x}_2 - \tilde{k}^{33}_{xx}\tilde{x}_3 \quad (19)$$

For example, we can show that Eq 17 is equivalent to Eq 11, by substituting Eqs 15 and 16 into Eq 17.

$$-4\pi^2\nu^2 \sqrt{m_1}\, x_1 = -\frac{k^{11}_{xx}}{\sqrt{m_1}\sqrt{m_1}}\sqrt{m_1}\, x_1 - \frac{k^{12}_{xx}}{\sqrt{m_1}\sqrt{m_2}}\sqrt{m_2}\, x_2 - \frac{k^{13}_{xx}}{\sqrt{m_1}\sqrt{m_3}}\sqrt{m_3}\, x_3 \quad (20)$$

Canceling mass terms and multiplying both sides by $\sqrt{m_1}$ gives Eq 11.

Eq 17-19 are most easily written in the equivalent matrix form:

$$-\begin{pmatrix} \frac{k^{11}_{xx}}{\sqrt{m_1}\sqrt{m_1}} & \frac{k^{12}_{xx}}{\sqrt{m_1}\sqrt{m_2}} & \frac{k^{13}_{xx}}{\sqrt{m_1}\sqrt{m_3}} \\ \frac{k^{21}_{xx}}{\sqrt{m_2}\sqrt{m_1}} & \frac{k^{22}_{xx}}{\sqrt{m_2}\sqrt{m_2}} & \frac{k^{23}_{xx}}{\sqrt{m_2}\sqrt{m_3}} \\ \frac{k^{31}_{xx}}{\sqrt{m_3}\sqrt{m_1}} & \frac{k^{32}_{xx}}{\sqrt{m_3}\sqrt{m_2}} & \frac{k^{33}_{xx}}{\sqrt{m_3}\sqrt{m_3}} \end{pmatrix} \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{pmatrix} = -4\pi^2\nu^2 \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{pmatrix} \quad (21)$$

The mass weighted force constants give a symmetric matrix. In other words, the corresponding off diagonal elements are equal. Eq 21 is an eigenvalue-eigenvector equation. The eigenvalues are the negative of the squared normal mode frequencies. The eigenvectors are the mass weighted normal coordinate displacements (see Appendix). Many efficient algorithms exist for solving eigenvalue equations.[6]

The Hessian and Energy Minimization The matrix of force constants is the matrix of the second derivatives of the potential energy. This matrix is also called the Hessian. The Hessian also plays a central role in energy minimization techniques. The equations in Section 4: "Energy Minimization" apply to one-dimensional systems. For molecules, we must find the x, y, z coordinates of each atom, for a total of 3N coordinates. To minimize the energy for these 3N coordinates, the equations in Section 4 are actually written in terms of the Hessian, instead of a single force constant or the second derivative of the energy for the x-coordinate alone. The use of the Hessian is necessary to minimize the energy of all the atoms in the molecule.

Numerical Example for Carbon Dioxide

The CO_2 example will provide some insight for understanding Eq 21. First, we need to discuss units. The fundamental vibration frequency for a harmonic oscillator is

$$\nu_o = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \quad \text{or} \quad 4\pi^2\nu^2 = \frac{k}{m} \quad (22)$$

with k in N m^{-1} an m in kg $molecule^{-1}$. Normally, vibrational spectra are plotted verses wavenumber, instead of frequency. To convert to wavenumbers, $\tilde{\nu}$:

$$\tilde{\nu} = \frac{1}{\lambda} \qquad \text{or} \qquad \nu = \frac{c}{\lambda} = c\tilde{\nu} \tag{23}$$

If $\tilde{\nu}$ is in cm^{-1}, c should be given in cm s^{-1}. Using $\tilde{\nu}$ in cm^{-1} and m in g mol^{-1}, Eq 22 becomes:

$$\frac{4\pi^2 c^2 \tilde{\nu}^2}{1000 \text{ g/kg } N_A} = \frac{k}{m} \tag{24}$$

or solving for the frequency squared in wavenumbers gives a convenient conversion factor

$$\tilde{\nu}^2 = \frac{k/m}{5.8921 \times 10^{-5}} \tag{25}$$

Now for our example. The CO_2 stretches are experimentally measured to be 1340 cm^{-1} for the symmetric stretch and 2349 $cm^{-1}$for the asymmetric stretch, Fig. 2. Lets roughly see if we can calculate these values through a normal mode analysis using our simplified one-dimensional model. First we will need all the force constants. However, some force constants are related by symmetry, since the left and right hand sides of the molecule are the same.

By symmetry : $k_{xx}^{11} = k_{xx}^{33}$ $\qquad k_{xx}^{12} = k_{xx}^{23}$ (26)

The terms that exchange the atom labels are also equivalent, since atom 1 interacting with atom 2 gives the same result as atom 2 interacting with atom 1. In matrix terms, these corresponding off-diagonal terms are equivalent for a symmetric matrix.

Symmetric matrix: $k_{xx}^{12} = k_{xx}^{21}$ $\qquad k_{xx}^{23} = k_{xx}^{32}$ (27)

These equivalences leave four force constants that we need to guess. First focus on atom 1. By trial an error, a good guess for

$$k_{xx}^{11} = 1600 \text{ N m}^{-1} \tag{28}$$

This force constant gives the restoring force as atom 1 is moved. The resorting force, F = -kx, will be negative, pulling the atom back to its equilibrium position. Another way to state this is if atom 1 is moved forward to shorten the bond length then atom 1 will try to move back to keep the bond length constant. A reasonable guess for

$$k_{xx}^{12} = -k_{xx}^{11} \tag{29}$$

Here the 12-force constant is negative, and the restoring force, F = -kx, is positive. This positive force results because as you move atom 1's neighbor, atom 1 will try to follow along in the same direction to keep the bond length constant. The absolute value of the two force constants is the same since moving either atom 1 or atom 2 has the same effect on the bond length and, therefore, the force on atom 1. Now focus on atom 2. Lets guess that it is twice as hard to move atom 2 as it is to move atom 1, since moving atom 2 effects two bonds:

$$k_{xx}^{22} = 2\, k_{xx}^{11} = 3200 \text{ N m}^{-1} \tag{30}$$

Finally, we will assume that

$$k_{xx}^{13} = 0. \tag{31}$$

We assume that atom 3 doesn't affect atom 1 significantly because the two atoms aren't directly bonded. Substituting Eqs 26-31 into Eq 21 gives the mass weighted force constant matrix. The row and columns correspond to the three different atoms, O_1, C_2, and O_3, respectively.

$$\begin{array}{c} \\ O_1 \\ C_2 \\ O_3 \end{array} \begin{array}{c} \begin{array}{ccc} O_1 & C_2 & O_3 \end{array} \\ \begin{pmatrix} \frac{1600}{\sqrt{16}\sqrt{16}} & -\frac{1600}{\sqrt{16}\sqrt{12}} & 0 \\ -\frac{1600}{\sqrt{12}\sqrt{16}} & \frac{3200}{\sqrt{12}\sqrt{12}} & -\frac{1600}{\sqrt{12}\sqrt{16}} \\ 0 & -\frac{1600}{\sqrt{16}\sqrt{12}} & \frac{1600}{\sqrt{16}\sqrt{16}} \end{pmatrix} \end{array} = \begin{pmatrix} -100 & 115.47 & 0 \\ 115.47 & -266.67 & 115.47 \\ 0 & 115.47 & -100 \end{pmatrix} \qquad (30)$$

The "eigen" Web applet is available to solve the eigenvalue problem.[7] Computer algebra programs like Maple and Mathematica are also handy for solving eigenvalue problems. The output of the "eigen" applet is shown below. The eigenvalues are listed with "E=." The normal mode frequencies are easily calculated using the units conversion factor from Eq 25.

Eigenvector 1: E=-0.000976903 ≈ 0
0.603024
0.522229
0.603024

Eigenvector 2: E=-100
-0.707107
0
0.707107

Eigenvector 3: E=-366.669
-0.369272
0.852805
-0.369272

Symmetric stretch:

$$\tilde{\nu} = \sqrt{\frac{100}{5.892\text{x}10^{-5}}} = 1303 \text{ cm}^{-1}$$

Asymmetric stretch:

$$\tilde{\nu} = \sqrt{\frac{366.67}{5.892\text{x}10^{-5}}} = 2495 \text{ cm}^{-1}$$

(for about 5% errors)

The three numbers below each eigenvalue are the normal coordinates. For example, the normal coordinates for the second eigenvector show atom 1 (-0.707) moving in the opposite direction as atom 3 (0.707), while atom 2 remains stationary (0). For the CO_2 example we have motion only in the x-direction, so there are only three coordinates listed, one for each atom. In general to display the motion of the atoms during the vibration, the atom coordinates are calculated for atom i as:

$$X_i = X_{i,eq} + \frac{\tilde{x}_1}{\sqrt{m_i}} q \qquad Y_i = Y_{i,eq} + \frac{\tilde{y}_1}{\sqrt{m_i}} q \qquad Z_i = Z_{i,eq} + \frac{\tilde{z}_1}{\sqrt{m_i}} q \qquad (33)$$

where $q = \sin(2\pi\nu t)$. For example, for the asymmetric stretch for CO_2 for the first O atom,

$$X_1 = X_{1,eq} + -0.369 \sin(2\pi\nu t) \qquad (34)$$
$$Y_1 = Y_{1,eq} + 0.853 \sin(2\pi\nu t)$$
$$Z_1 = Z_{1,eq} + -0.369 \sin(2\pi\nu t)$$

The first eigenvalue is zero, because it corresponds to the motion of the center of mass of the molecule in the x-direction. You can also tell that the first eigenvector is for the motion of the molecule as a whole because all the normal coordinates have the same sign, that is all the atoms are traveling in the same direction. For fully three-dimensional problems, the first 5 eigenvalues,

for linear molecules, or 6 eigenvalues, for nonlinear molecules, will correspond to translation and rotation. (Spartan, however, doesn't show you these first eigenvalues, but other programs do.)

You can tell that eigenvalue 2 is for the symmetric stretch, since the normal coordinates for the oxygen atoms are opposite to each other (i.e –0.707 and 0.707 respectively) and the carbon atom doesn't move. In the asymmetric stretch, eigenvalue 3, the oxygen atoms move backward while the carbon atom moves forward.

How well did our simplified model work? The symmetric stretch is a little low and the asymmetric stretch is a little too high for a combined error of about 5%. It doesn't make sense to try to get the results to agree any better. We have neglected the bending vibration in our treatment, and using a molecular mechanics or molecular orbital program is much more accurate. However, you should try changing the force constant guesses a little to see the effects of each force constant. If you make a change that is not consistent with the force field in a real molecule, then the first eigenvalue will increase. Better sets of guesses give a smaller first eigenvalue.

Normal Mode Analysis and Molecular Mechanics and Molecular Orbital Calculations

Our simple example of CO_2 is not representative of the accuracy available for predicting normal mode frequencies. Molecular mechanics and molecular orbital calculations can quite accurately predict the frequencies for the vibrations of complex molecules. Results for CO_2 are given in Table I. If you haven't gotten to molecular orbital theory yet, suffice it to say that you can calculate normal mode frequencies quite accurately.[8,9]

Table I. Molecular Mechanics and Molecular Orbital Based Normal Mode Analysis for CO_2.

Literature	MMFF	AM1	PM3	HF/ 6-31G*	MP2/ 6-311G**	pBP/DN*	BP/DN*	B3LYP/ 6-311G(d)
667	538	526	522	744	656	637	638	666
667	538	526	523	744	656	637	638	666
1340	912	1480	1408	1518	1344	1323	1319	1377
2349	1746	2565	2387	2585	2461	2363	2349	2438
error %	24.1%	15.5%	12.5%	11.6%	2.1%	2.7%	2.5%	1.7%

The MMFF molecular mechanics calculation poorly represents the accuracy for molecular mechanics in general, since the force field parameters aren't optimized for the unusual C=O bonds in CO_2. Molecular mechanics calculations are common and very useful for large biomolecules. Semi-empirical calculations at the AM1 or PM3 level are more accurate. Hartree-Foch, HF, calculations are even better, especially when MP2 electron-electron correlations are taken into account. Density functional methods like pBP, BP or B3LYP are now the best choice for careful analysis. Molecular orbital calculations are indispensable for helping to assign the vibration bands in Infrared and Raman spectroscopy.

Anharmonicity

The proceeding discussions assume all the vibrations are purely harmonic. Our treatment of molecular mechanics force fields showed that anharmonic corrections are often important for real molecules. What is the effect of anharmonicity on vibrational spectra and normal mode calculations? For weak anharmonicity, vibrational spectra also show overtones and sum and difference bands. Overtones are at integer multiples of the fundamental frequency, $n\tilde{\nu}_A$. Sum and difference bands occur at $\tilde{\nu}_A + \tilde{\nu}_B$, and $\tilde{\nu}_A - \tilde{\nu}_B$, respectively. Frequencies from *ab initio* molecular orbital calculations are normally multiplied by 0.9 to correct for anharmonicity. In Table I, if the HF/6-31G* values are multiplied by 0.9, the average deviation drops to 1%. Frequencies from

molecular mechanics are usually too approximate to warrant anharmonicity corrections when comparing with vibrational spectra.

For strong anharmonicity, such as occurs for very loose and floppy vibrations, a more refined treatment is necessary.[10] Such vibrations include bond torsions that have low energy barriers, ring vibrations in large ring systems, and vibrations in hydrogen-bonded systems and molecular complexes. Unfortunately, such vibrations are often the most interesting, especially in studies of proteins and nucleic acids. Treating very flexible, low energy vibrations in biomolecules is an active area of current study.[11-15]

Vibrations and Thermodynamics

Vibrations increase the Gibbs Free Energy of a substance. Vibrational enthalpy and entropy calculations are very useful in drug discovery for assessing the Gibbs Free Energy of binding.[16] Vibrations also play a central role in protein folding and protein flexibility.[13-15] The contribution of a vibration to the enthalpy and entropy of a substance is given by[17]

$$H_{vib} = \frac{1}{2} N_A h\nu_o + \frac{N_A h\nu_o \, e^{-h\nu_o/kT}}{1-e^{-h\nu_o/kT}} \qquad (35)$$

$$S_{vib} = -R \ln(1-e^{-h\nu_o/kT}) + \frac{N_A h\nu_o \; e^{-h\nu_o/kT}}{T \; (1-e^{-h\nu_o/kT})} \qquad (36)$$

where N_A is Avogadro's number, ν_o is the frequency of the normal mode, h is Planck's constant, and k is Boltzmann's constant = R/N_A. The $\frac{1}{2} N_A h\nu_o$ term in the enthalpy is the zero-point vibrational energy, which is the energy of the vibration at absolute zero temperature, $H_{vib}(0)$. Eqs 35 and 36 are summed for each normal mode vibration. Following a normal mode analysis, then, it is very easy to calculate the Gibbs Free Energy of a substance.

A specific example will help to clarify the importance of normal mode analysis in thermodynamic considerations. Consider two different conformations of a molecule, A and B:

$$A \rightleftarrows B \qquad (37)$$

Examples include the trans and gauche isomers of butane or two conformations of a large protein. For low frequency vibrations Eq 36 simplifies and the entropy difference reduces to[13]

$$\Delta S_{vib,conf} = R \ln\left(\frac{2\pi\nu_{A1} \, 2\pi\nu_{A2} \, 2\pi\nu_{A3} \ldots}{2\pi\nu_{B1} \, 2\pi\nu_{B2} \, 2\pi\nu_{B3} \ldots}\right) \qquad (38)$$

This entropy difference is called the configurational entropy difference. The numerator is the product of the low frequency normal modes for A, and the denominator is the product of the low frequency normal modes for B. Therefore, if B has lower frequency modes, the entropy of B will be larger and the entropy difference will favor B. In other words, the lower the mode frequencies, the more the conformation can rattle around, and the more that conformation is favored.

In molecular mechanics the enthalpy of formation of a molecule is given as (see Section 2: "Enthalpy of Formation"):

$$\Delta_f H° = {}^3/_2 RT + {}^3/_2 RT + RT + \text{bond energy} + \text{steric energy} + \text{vibrational contributions} \qquad (41)$$

Normal mode analysis gives us the tools to calculate the vibrational contributions directly using Eq 35. However, as mentioned in Section 2, for MM2 calculations a series of approximations are made for Eq 41. The zero point energy is often neglected in classical simulations, leaving the temperature dependent contribution from the second term of the vibrational enthalpy, Eq 35. This contribution to the enthalpy is plotted as a function of vibrational frequency in Figure 5.

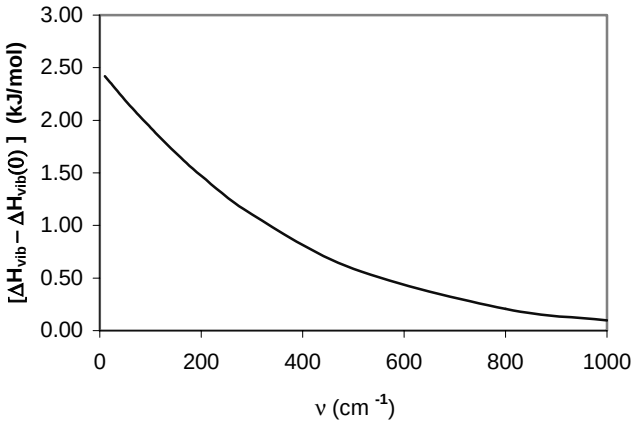


Figure 5. Contribution of a vibration to the Enthalpy of formation of a molecule above the zero point energy.

The contribution of vibrations becomes negligible for frequencies greater than about 500 cm^{-1}. Therefore, only low frequency vibrations contribute strongly. Torsional motions around freely rotating bonds are often the lowest frequency normal modes in molecules. Other low frequency vibrations are often ignored. The vibrational contributions can then be approximated by torsional increments for each freely rotating bond, giving the result presented in Section 2:

$$\Delta_f H° = {}^3/_2RT + {}^3/_2RT + RT + \text{bond energy} + \text{steric energy} + \text{torsional increments} \qquad (42)$$

Our treatment of normal modes now will allow us to discuss these approximations in detail. Examples of low frequency vibrations are bending vibrations and ring vibrations as well as freely rotating bond torsions. Clearly for careful calculations more contributions than just the torsional increments for freely rotating bonds are necessary. In addition, Eq 42 completely neglects the zero point energies. Molecular orbital and molecular mechanics programs readily provide these thermodynamic contributions when normal mode analyses are done, so we don't need to make the extreme approximations inherent in Eq 42.

Molecular Dynamics and Normal Mode Analysis

Molecular dynamics and normal mode analysis are really quite similar. Both include the kinetic and potential energy for the molecule. The force field is the same. They both calculate the Hessian and then integrate Newton's Laws of motion. The motions that you see in molecular dynamics simulations are in fact the normal modes of the molecule. The fluctuations of the atom positions in a molecular dynamics run can be used to extract the normal mode frequencies.[14,18]

The difference between molecular dynamics and normal mode analysis is that the equations of motion are integrated numerically in dynamics simulations, but sinusoidal solutions are assumed for normal mode analysis. In addition, in molecular dynamics the motions of all the normal modes are studied simultaneously, while in normal mode analysis one mode is studied at a time. The techniques have their strengths and weaknesses. Eqns 35 and 36 show that the link between normal mode analysis and thermodynamics is direct and straightforward. Thermodynamic properties can be calculated from dynamics runs, but particular care must be taken to ensure adequate statistical sampling (i.e. using long time simulations). On the other hand, molecular dynamics more easily handles anharmonicity and explicit solvation.

Valence Force Field Solutions

Normal mode analysis is particularly important in molecular spectroscopy. As a consequence, valence force field solutions have been worked out for many small molecule geometries. These solutions take a different approach to the problem. The force constants that are used are the force constants for individual bonds, rather than the force constants for moving atoms, e.g. Eq 9. Focussing on the bond force constants more closely corresponds to our "chemical intuition." Another advantage of valence force field calculations is that algebraic solutions can be written. For example, for a symmetric triatomic molecule, where $m_1 = m_3$, the internal coordinates are defined as

$$q_1 = (r_{12}-r_o)^2 \qquad (43)$$
$$q_2 = (r_{23}-r_o)^2$$
$$\delta = (\theta-\theta_o)^2$$

The q's are bond stretching terms and δ is the bond bending term; r_{12} is the distance between atoms 1 and 2, r_o is the equilibrium bond length, θ is the bond angle, and θ_o is the equilibrium bond angle. The potential energy is chosen as:

$$V = \frac{1}{2} k_1 q_1^2 + \frac{1}{2} k_1 q_2^2 + k_\delta \delta^2 \qquad (44)$$

The k_1 force constant is for stretching the 1-2 or 2-3 bond. For CO_2 this is the C=O stretch. The force constant for bond bending is k_δ. The Hessian second derivatives can be obtained by taking explicit derivatives of Eq 44. For this potential energy form the normal mode frequencies are given by[3,4]

$$4\pi^2 \nu_{asym}^2 = \left(1 + \frac{2m_1}{m_2} \sin^2 \frac{\theta_o}{2}\right) \frac{k_1}{m_1} \qquad (45)$$

$$4\pi^2 (\nu_{sym}^2 + \nu_{bnd}^2) = \left(1 + \frac{2m_1}{m_2} \cos^2 \frac{\theta_o}{2}\right) \frac{k_1}{m_1} + \frac{2}{m_1}\left(1 + \frac{2m_1}{m_2} \sin^2 \frac{\theta_o}{2}\right) \frac{k_\delta}{r_o^2} \qquad (46)$$

$$16\pi^4 (\nu_{sym}^2 \nu_{bnd}^2) = 2\left(1 + \frac{2m_1}{m_2}\right) \frac{k_1}{m_1^2} \frac{k_\delta}{r_o^2} \qquad (47)$$

Eqs 46 and 47 show that the frequency of the symmetric stretch depends on the bending force constant. As mentioned above, our example for one-dimensional CO_2 didn't include this effect.

The disadvantage of algebraic solutions is that they depend critically on the details of the potential energy function, e.g. Eq 44. If a stretch-bend interaction or Van der Waals terms are included, as in many molecular mechanics force fields, then Eqs 45-47 are no longer valid. In the early decades of vibrational spectroscopy, it was hoped that solutions to the normal mode problem could be used to determine the force constants for individual bonds, as in Eq 44. However, the dependence of the force constants on such over-simplified potential energy functions causes large errors. The attempt to determine bond force constants directly from spectra has therefore been abandoned. Equations such as 45-47 can still be useful in building our intuition about bond strengths, however the derived force constants must be treated as very approximate and can sometimes be misleading.

Appendix

We wish to show more clearly the relationship between Eqs 17-19 and the normal coordinates, for the curious. First note that substituting Eq 5 into Eq 7 gives:

$$-4\pi^2\nu^2 \, m \, A \sin(2\pi\nu t) = -k \, A \sin(2\pi\nu t) \quad (48)$$

Dividing both sides by the sin gives

$$-4\pi^2\nu^2 \, m \, A = -k \, A \quad (49)$$

In other words, the equation applies to the time dependence of the vibration and also to the amplitude of the vibration separately. Therefore Eqs 12-14 and 17-19 allow us to solve for the amplitudes of the vibrations, where x_i, y_i, z_i can be read as the amplitudes of the waves in the x, y, and z directions for atom i. Similarly, $\tilde{x}_i$, $\tilde{y}_i$, $\tilde{z}_i$ can be considered to be the corresponding mass weighted amplitudes. The time dependent values are then:

$$\tilde{x}_i(t) = \tilde{x}_i \sin(2\pi\nu t) \qquad \tilde{y}_i(t) = \tilde{y}_i \sin(2\pi\nu t) \qquad \tilde{z}_i(t) = \tilde{z}_i \sin(2\pi\nu t) \quad (50)$$

Dropping the "(t)" for convenience and converting back into non-mass weighted coordinates gives:

$$x_i = \frac{\tilde{x}_i}{\sqrt{m_i}} \sin(2\pi\nu t) \qquad y_i = \frac{\tilde{y}_i}{\sqrt{m_i}} \sin(2\pi\nu t) \qquad z_i = \frac{\tilde{z}_i}{\sqrt{m_i}} \sin(2\pi\nu t) \quad (51)$$

Converting from extensions into final coordinates using Eq 8 gives Eq 33.

Now you may have noted that Eqs 17-19 involve four unknowns (ν, $\tilde{x}_i$, $\tilde{y}_i$, and $\tilde{z}_i$) but only three equations. So to obtain unique solutions, some more information is necessary. We must add the requirement that the center of mass can't move:

$$m_1x_1 + m_2x_2 + m_3x_3 = 0 \quad (52)$$

or equivalently in mass weighted coordinates:

$$\sqrt{m_1}\tilde{x}_1 + \sqrt{m_2}\tilde{x}_2 + \sqrt{m_3}\tilde{x}_3 = 0 \quad (53)$$

As we solve for each successive normal mode we also need to ensure that the vibrations don't interact. Mathematically this requires that the normal modes are orthogonal. For each pair of normal modes A and B, with normal coordinates $\tilde{x}_{iA}$ and $\tilde{x}_{iB}$, respectively:

$$\tilde{x}_{1A}\,\tilde{x}_{1B} + \tilde{x}_{2A}\,\tilde{x}_{2B} + \tilde{x}_{3A}\,\tilde{x}_{3B} = 0 \quad (54)$$

Taken together, Eqs 17-19 and Eq 53 and 54 provide the unique set of normal modes satisfying the desired characteristics set out in the introduction. Solving these equations as a linear set of simultaneous equations is difficult. Luckily, solving the problem as an eigenvalue-eigenvector equation using Eq 21 automatically satisfies the requirement for orthogonality.

Literature Cited:

1. Moore, W. J., *Physical Chemistry, 4th Ed.*, Printice-Hall: Englewood Cliffs, NJ, 1972, Chapter 17, Sec. 14., pp 775-776.

2. Wilson, Jr., E. B.; Decius, J. C.; Cross, P. C., *Molecular vibrations; the theory of infrared and Raman vibrational spectra*, McGraw-Hill, New York,1955

3. Herzberg, G., *Molecular Spectra and Molecular Structure II. Infrared and Raman Spectra of Polyatomic Molecules,* Van Nostrand, Princeton, N. J., 1945.

4. Shoemaker, D. P.; Garland, C. W.; Steinfeld, J. I.; Nibler, J. W., *Experiments in Physical Chemistry, 4th Ed.,* McGraw-Hill, New York, NY, 1981, Exp. 40.

5. ComSpec3D: Gasteiger, J., Computer-Chemie-Centrum, Universität Erlangen-Nürnberg, http://www2.ccc.uni-erlangen.de/services/vrmlvib/index.html (accessed 2/2003).

6. Distance geometry and Hückel molecular orbital theory are other examples of eigenvalue equations. Hartree-Foch molecular orbital calculations, such as you do in Spartan for HF methods, are solved as eigenvalue equations.

7. Colby College Physical Chemistry Home Page, *eigen*, http://www.colby.edu/chemistry/PChem/eigen.html (accessed 2/2003).

8. After correcting the experimental frequencies for anharmonicity.

9. The symmetric stretch of CO_2 is shifted slightly because of a Fermi resonance with the first overtone of the bending vibration. So exact agreement isn't expected without the resonance correction.

10. Pitzer, Kenneth S., Quantum Chemistry, Prentice-Hall, New York, NY, 1953, pp 239-243, Appendix 18, pp 492-500.

11. Walther, M.; Plochocka, P.; Fischer, B.; Helm H.; Uhd Jepsen, P., "Collective vibrational modes in biological molecules investigated by terahertz time-domain spectroscopy" *Biopolymers(Biospectroscopy)*, 2002, *67(4-5)*, 310-313.

12. Hamm, P; Hochstrasser, R M., "Structure and dynamics of proteins and peptides: femtosecond two-dimensional infrared spectroscopy." *Pract. Spectrosc*., **2001**, *26*, 273-347.

13. Karplus, M.; Kushick, J. N., "Method for Estimating the Configurational Entropy of Macromolecules," *Macromolecules*, **1981**, *14*, 325-332.

14. Van Vlijmen, H. W. T.; Karplus, M.,. "Analysis of Calculated Normal Modes of a Set of Native and Partially Unfolded Proteins," *J. Phys. Chem. B*, **1999**, *103(15)*, 3009-3021.

15. Levy, R. M.; Srinivasan, A. R.; Olson, W. K.; McCammon, J., "A Quasi-harmonic method for studying very low frequency modes in proteins," *Biopolymers,* **1984**, *23(6)*, 1099-112.

16. Schwarzl, S. M.; Tschopp, T. B.; Smith, J. C.; Fischer, S., "Can the Calculation of Ligand Binding Free Energies Be Imporved with Continuum Solvent Electrostatics and an Ideal-Gas Entropy Correction?" *J. Comp. Chem*, **2002**, *23*,1143-1149.

17. McQuarrie, D. A., *Statistical thermodynamics***,** Harper & Row, New York, 1973.

18. Rempe, S. B.; Jonsson, H., "A Computational Exercise Illustrating Molecular Vibrations and Normal Modes," *Chemical Educator,* **1998**, *3(4)*, 1-17.

QUANTA/CHARMm

Introduction

QUANTA is a molecular modeling program, which is specifically designed to handle large biological molecules. CHARMm is a molecular mechanics and dynamics package that QUANTA uses for its mechanics and dynamics calculations. QUANTA can also be used to setup input files for MM2 molecular mechanics and MOPAC molecular orbital calculations.

General Notes:

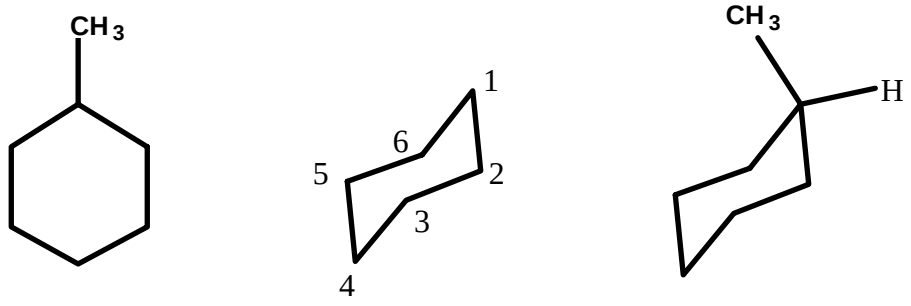
The following exercises are designed to be done in order. Detailed instructions given in earlier exercises will not be repeated in later exercises. If you have questions, turn to this tutorial or use the QUANTA manuals. The QUANTA manuals have many interesting examples that extend well beyond the skills taught here.

The user interface is very similar to the Macintosh, with several notable exceptions. First there are three mouse buttons. **Usually only the left one is used**. In this manual, use the left button unless instructed otherwise. In order to type input to a dialog box, the mouse cursor must be in the same window. This is because, unlike the Macintosh, many windows can be active at the same time. **Never use the "go-away" button** in the upper left-hand corner of a palette, always use the Quit or Exit entry on the palette menu. To move a covered window into the foreground, click anywhere on the frame of the hidden window.

QUANTA is actually very easy to learn. Follow these instructions carefully until you get the feel of the program. Then try new things. Don't hesitate to explore QUANTA on your own.

Chapter 1. Building and Minimizing.

The following will illustrate a few of the options available for structural input, minimization and display using QUANTA. We will begin with axial-methyl cyclohexane, Figure 1.1. We will use the ChemNote Application, where structures may be drawn on the screen in essentially the same way as they would in ChemDraw. The minimum energy configuration will then be calculated using CHARMm.



methyl cyclohexane chair cyclohexane axial methyl cyclohexane

Figure 1.1. Axial-methylcyclohexane

ChemNote Model Building Pull down the Applications menu, slide right on "Builders" and choose "2-D sketcher." to start ChemNote. Click on the cyclohexane ring in the middle of the bonds palette, move the curosr to the middle of the sketch pad and click again. A green cyclohexane ring should appear. We now want to add the axial methyl group. Click on the icon for a single bond coming out of the plane: the solid triangle in the bonds palette. Move the cursor to the right most carbon on the cyclohexane. Hold the mouse button down and drag the bond away from the ring to the right. Don't worry about the hydrogens, ChemNote will add those

automatically. To finish adding bonds, double click on the selection tool icon, , in the Edit Icons palette in the upper left hand portion of the window.

To save this molecule, pull down the File menu and choose 'Return to Molecular Modeling..' Click Yes to the question 'Save Changes First.' The file librarian dialog box will appear; first choose the small_molecules/ directory with a single click, second type in the name amecyc6. **Remember-- don't use punctuation or spaces in file names.** Click on 'Save.' Next the charge of the molecule is calculated using standard values for each atom type. This charge will be -0.180 for methyl cyclohexane, but we wish to have a neutral molecule. We need to smooth the charge over the atoms to yield a net charge of 0.000. Choose 'CT, CH1E, CH2E, CH3E, C5R, C6R, C5RE, C6RE, HA type', and then click OK. This choice is for all carbon atoms and non-polar hydrogens. All the carbons in our molecule are type CT, which stands for tetrahedral carbon.

Upon returning to QUANTA, hydrogen atoms are added automatically and the 3D structure is constructed using tabulated values of bond lengths and angles. The program then displays: Which molecule do you want to use? Choose the 'Use the new molecule only' option and click OK. Verify that you have constructed the axial isomer by reorienting the molecule on the screenusing the following instructions.

Rotations,Translations, and Scaling To change the orientation, size, and position of the molecule, you can use either of two methods, (1) using the mouse or (2) using the Dial palette. To use the mouse, position the cursor in the main window and hold down the center mouse button. Moving the mouse reorients the molecule. If you wish to rotate the molecule only around the axis perpendicular to the screen, hold down the right mouse button and move the cursor left and right.

Alternatively, you can use the Dial palette. The Dial palette is in the lower right corner of the screen. If the Dials are not visible, pull down the Views menu, slide right on "Show Windows" and choose "Palettes." Clicking on the dial controls causes the listed action. If you hold the mouse button down, the action occurs smoothly, with the rate depending on the horizontal distance from the center of the control. Clicking on 'Reset' will allow you to start fresh with a centered molecule. The Dials allow you to rotate, translate, and scale (enlarge) the molecule. The Clipping control adjusts the size of the box in the z direction where atoms will be displayed. Atoms outside this box, both front and back, will not be displayed. Decreasing the Clipping control allows you to "drive" inside the molecule.

CHARMm Minimization The structure made by ChemNote will not be the lowest energy structure. To prepare to minimize the structure, pull down the CHARMm menu and choose 'Minimization Options.' For small molecules that are close to the minimum geometry, choose the Conjugate Gradient Method. For rough starting geometries Steepest Decent is faster, less likely to fail, but less accurate. Use Adopted-Basis Newton Raphson for large molecules like proteins. Choose the Conjugate Gradient Method. Enter the default values for the parameters given below, if they are not already shown:

Number of Minimization Steps	50
Coordinate Update Frequency	5
Energy Gradient Tolerance	0.0001
Energy Value Tolerance	0
Initial Step Size	0.02
Step Value Tolerance	0

Pull down the CHARMm menu, slide right on "CHARMm mode" and choose 'PSF's'. (The RTF options are for biopolymers.) You need only set the Minimization options and RTF options once. These choices will be used for all subsequent modeling, until changed. To actually do the minimization, choose 'CHARMm Minimization' in the Modeling palette. The calculation will stop after 50 steps, but the energy won't necessarily be minimized. Click on 'CHARMm Minimization' repeatedly until the energy listed in the main window no longer changes. You can monitor the progress of all CHARMm calculations in the TextPort at the bottom of the screen. The final result should be 6.3047 kcal/mol. Select 'Save Changes' in the Modeling palette to save your minimized structure to a file. The program asks you to 'Choose the MSF Saving Option.' A MSF is a molecular structure file, which QUANTA uses as the principal means of saving 3D information. Choose "Overwrite amecyc.msf' and click OK.

We often need to find the contribution to the total energy for each degree of freedom, i.e. bond stretching, bond angle bending, dihedral torsions, Lennard-Jones-Van der Waals energies, and electrostatic interactions. To find these contributions, select 'CHARMm Energy' in the Modeling palette. The results are listed in the TextPort by the keywords underlined above. Use the scroll bar to scan the results. The conformation of the molecule remains unchanged during this calculation. The "Improper" torsions entry is an additional term in the force field to get the proper conformation for small rings.

DISPLAY OPTIONS

Is there a 'hole' in the middle of a cyclohexane ring? We will construct solid models using literature values for atomic radii to answer this question. Van der Waals radii are usually used for this purpose, and refer to average covalent interactions. There are a wide variety of solid modeling options to use. Try them all. But, to get you started: reorient the molecule so that the methyl group points away from you as in Figure 1.2.

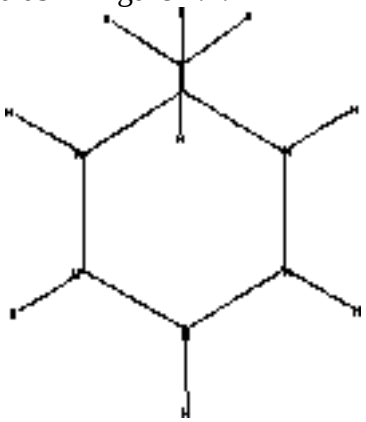


Figure 1.2. Axial methyl cyclohexane

Surface Rendering Stick structures of molecules are easy to visualize, but they present a very distorted view of molecular structure. Various techniques for displaying the surface of molecules are designed to present a more realistic model of what molecules "look like" to other molecules.

Pull down the Draw menu, slide right on 'Solid Models' and choose "Van der Waal's". You can reorient the molecule using the mouse by holding down the center mouse button. Is there a 'hole' in the middle of the ring? When the solid model is displayed, a new window appears in the lower right portion of the screen. Click on the "No" box underneath the "Delete"

column to remove the object. You should also try other options including "Ball and Stick" models in "Solid Models" and several of the "Raster Models." "Ray Trace" gives the best quality, but rotations aren't possible in ray trace mode. Click a mouse button to exit "Raster Models" or "Ray Trace" mode.

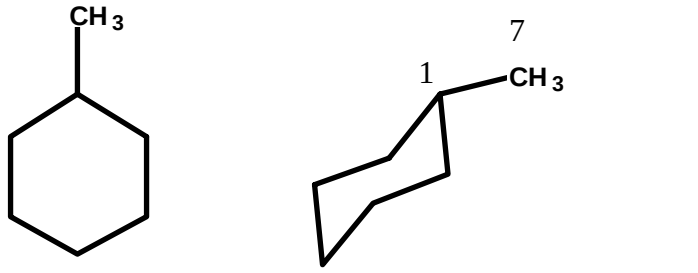
Another method of surface rendering that is especially popular for biomolecules is a dot surface. Pull down the Draw menu and choose 'Dot Surfaces.' Dot surfaces provide the fastest reorientation. Choose "Big Dots" and otherwise use the default settings, then click "OK." A Selection Palette is then displayed so that you can specify which atom's surfaces are to be modeled with dots. Make sure "Include" is highlighted, click on "All Atoms," and then click "Finish." To remove the dot surface pull down the Draw menu and choose "Dot Surfaces" again, select "Delete Dot Surface" , and click OK..

Problem 1.1: Using the printout in the TextPort after using the 'CHARMm Energy' option, decide which term in the force field dominates the steric energy of axial methyl cyclohexane. (e.g. bond stretch, bond bending, etc.) Compare absolute values for each term.

Problem 1.2: Start fresh in the ChemNote application, build axial methyl cyclohexane again (or Open your old ChemNote file) but this time minimize the structure using the Steepest Descents method. What happens? How do your results compare between Steepest Descents and the Conjugate Gradient minimization you did before? With rough beginning geometries, it is often best to minimize first with Steepest Descents, and then switch to Conjugate Gradient and minimize again. This two step process gives the best of both techniques. After you are finished with this problem, remember to switch back to Conjugate Gradient in the CHARMm "Minimization Options" dialog and then reminimize. Then "Save Changes" and once again choose the "Overwrite" option.

Chapter 2: Conformational Preference of Methylcyclohexane

Does methylcyclohexane prefer the axial or equatorial conformation? Do Problem 1.1 first. If your axial structure is not on the screen, pull down the File menu and 'Open' your file amecyc6.msf. We now need to create the structure of the equatorial isomer (Figure 2.1). You could do this by using ChemNote, but we will use the Molecular Editor to give you some practice with this powerful tool. Pull down the Edit menu and choose 'Molecular Editor.' Choose the Swap Bonds option in the Molecular Editor Palette. Notice that instructions are often printed in the lower left hand corner of the main window. To pick the first bond to swap, click on the bond to the equatorial hydrogen on carbon 1 (Figure 1.1). Now click on the C-C bond to the methyl group. The equatorial isomer will be produced. Click on 'Save and Exit' in the Molecular Editor palette. In the Save Options dialog box that follows, choose 'Reassign atom types,' 'Reassign atom charges,' and then OK. Adjust and smooth the charge using 'Carbon and non-polar hydrogen' types and click OK. In the MSF Saving Options choose "Save to New Filename" so that your old axial- file is not overwritten. Click OK to exit. In the file librarian dialog box, type emecyc6 as the new file name.



methyl cyclohexane equatorial methyl cyclohexane

Figure 2.1. Equatorial methyl cyclohexane

To minimize the structure, check to make sure that the same options as in Chapter 1 are chosen. Click on "CHARMm Minimization" until the energy is minimized. The result should be 4.5006 kcal/mol. Select "Save Changes" and then "Overwrite emecyc6.msf." After you minimize the structure, use the 'CHARMm Energy' command can be used to list the contributions to the total steric energy of the molecule to allow comparison with the results from Problem 1.1.

Problem 2.1 Compare the two conformers of methylcyclohexane. Record the various contributions to the steric energy in the table below. Calculate the difference in energy for each contribution in the 4th column. In the 5th column record which conformer is favored by each contribution. Finally, from the difference column, decide which contribution dominates the conformational preference. Which changes most, the ring strain (as measured by the bond stretch, bond angle, and dihedral torsion terms) or the through-space Lennard-Jone terms?

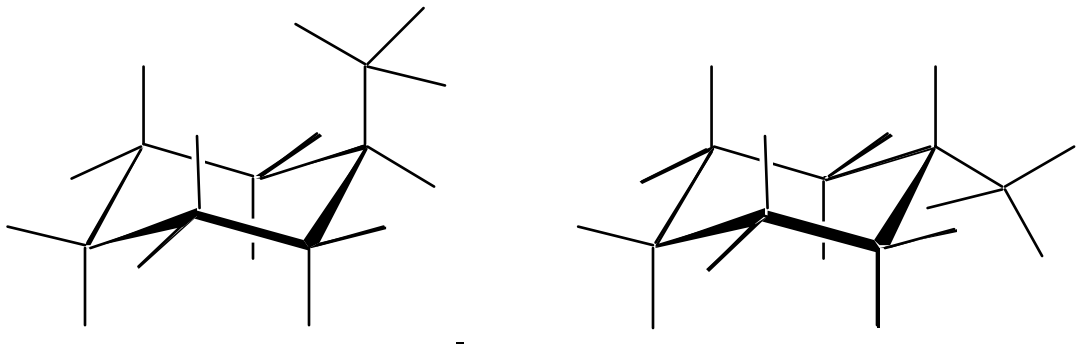
Contribution	equitorial (kcal/mol)	axial (kcal/mol)	difference (kcal/mol)	favored conformer
bond energy				
angle energy				
dihedral energy				
Lennard-Jones				
Electrostatic				
total				

Chapter 3. Geometry (or How Does Molecular Mechanics Measure Up?)

In this chapter you will learn how to measure distances, bond lengths and angles from your minimized structures. We will make our measurements on axial- and equatorial-methylcyclohexane, so do Chapter 2 first. General Chemistry texts list the C-H bond length as 1.09Å and the C-C bond length as 1.54Å for sp^3 hybridized systems. The ideal bond angle around tetrahedral carbon is the tetrahedral angle, 109.5°. How close to these values do real molecules come?

Make sure your axial-methylcyclohexane is on the screen. If it isn't, from the main QUANTA screen, pull down the File menu and choose "Open." In the File Librarian select your axial file, click on the "Replace" button on the bottom of the screen, and then click on "Open." Bring the "Geometry" palette to the front. To bring a window forward, click on the border of its window (notice that the cursor changes to a >| symbol when you are on the border of a window). Make sure "Show distance monitors," "Show angle monitors," and "Show dihedral monitors" are highlighted.

Bond distances: To find bond distances, first click on "Distance" in the Geometry palette. Now whenever you click on any two atoms, the distance between those two atoms will be displayed. Measure the bond distances in your compound. Record the values on the structure below. Don't measure every bond length, only the ones that are not related by symmetry. You can also measure the distances between atoms that are not attached. Find the shortest distance between a methyl hydrogen and a ring hydrogen. Include this distance on the structure below.

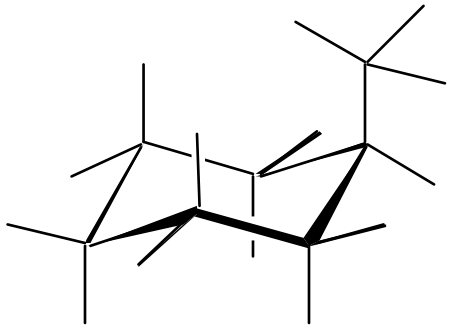


When you are finished click on "Distance" again to deselect it. To remove the atom labels click on "Clear ID" at the top of the Geometry palette. Finally, click on "Delete distance monitors."

Bond Angles: To find bond angles, first click on "Angles" in the Geometry palette. Now whenever you click on three atoms in a row, the bond angle will be displayed. Make sure that the central atom in the angle is the second atom that you click on. Measure the bond angles in your compound. Record them in the structure above. When you are finished click on "Angles" again to deselect it. To remove the atom labels click on "Clear ID" at the top of the Geometry palette. Finally, click on "Delete angle monitors."

Dihedral Angles: To find dihedral angles, first click on "Dihedrals" in the Geometry palette. Now whenever you click on four atoms in a row, the dihedral angle will be displayed. Make sure that you click on the four atoms in the order in which they are connected. For example, to find the ring dihedral angle for adjacent C-H bonds, click on the atoms in the order: ring-H, the attached ring-C, the adjacent ring-C, and finally the attached ring-H.. Measure the dihedral angles

in your compound, including the ring C-C-C-C dihedral. Record them in the structure below. When you are finished click on "Dihedrals" again to deselect it. To remove the atom labels click on "Clear ID" at the top of the Geometry palette. Finally, click on "Delete dihedral monitors." Note that you can leave dihedral monitors on while you do other tasks, which include minimization, Conformational Searches, and dynamics.



Atom Charges: To display the charge on each atom, pull down the Draw menu, slide right on "Label Atoms," and select "Atomic Charge." Hydrocarbons don't have large charges on the atoms, so the Electrostatic" contribution to the total steric energy is expected to be small. In compounds with heteroatoms, however, the Electrostatic contribution can dominate the steric energy. To remove the atom charges from the screen, pull down the Draw menu, slide right on "Label Atoms," and select "Off."

The CHARMm Force Field

The CHARMm force field recognizes that bond lengths and angles change depending on hybridization and bonding partners even in normal-unstrained molecules. In Table 3.1 is listed the "normal" bond parameters that CHARMm uses in its force field for a few bond types. These parameters are the starting point for energy minimizations. Any deviations from these "normal" values will be reflected in increases in steric energy. These parameters are derived by finding the "best fit" to experimental data for a reference set of compounds. This reference set of compounds is often called the learning set. The experimental data is from electron and x-ray diffraction studies.

Table 3.1. CHARMm force field parameters.

Bond	Distance(Å)	Angle	Angle (degrees)
$C(sp^3)$-$C(sp^3)$	1.529	$C(sp^3)$-$C(sp^3)$-$C(sp^3)$	112.70
$C(sp^3)$-O(alcohol)	1.405	$C(sp^3)$-$C(sp^3)$-O(alco)	110.5
$C(sp^2)$*-$C(sp^3)$	1.502	$C(sp^3)$-$C(sp^2)$-$C(sp^3)$	114.2
		$C(sp^2)$-$C(sp^3)$-$C(sp^3)$	112.90
C(carbonyl)-$C(sp^3)$	1.530	$C(sp^3)$-C(=O)-$C(sp^3)$	117.0
		C(=O)-$C(sp^3)$-$C(sp^3)$	109.9
C(carbonyl)=O	1.207	$C(sp^3)$-C=O	124.80
H-$C(sp^3)$	1.090	H-$C(sp^3)$-H	107.8
		H-$C(sp^3)$-$C(sp^3)$	110.7
H-O(alcohol)	0.948	H-O(alcohol)-$C(sp^3)$	106.7

* sp^2 hybridized but not conjugated.

Problem 3.1

Measure the shortest distance between a methyl hydrogen and a ring hydrogen in equatorial-methylcyclohexane. Include this distance on the equatorial structure above. Do these shortest distances in the axial and equatorial conformers correlate with the change in Lennard-Jones energy that you found in Problem 2.1?

Problem 3.2

Compare the bond distances and angles in axial-methylcyclohexane to the "normal" CHARMm values of the C-H bond length of 1.090Å, the C-C bond length of 1.529Å, and the angles listed in Table 3.1. Deviations from the normal values cause bond strain. Which C-C bonds differ most from the normal values? Is it easier to deform the bond length or the bond angle; that is, do the bond lengths or bond angles deviate more from the normal values?

Problem 3.3

Build ethanol in ChemNote (pull down the Applications menu, slide right on "Builders, and select "2-D Sketcher). Minimize ethanol, and then display the atom charges. Compare the magnitude of these charges to the charges for methylcyclohexane. In which molecule will the Electrostatic contribution to the steric energy be greatest? Measure the C-C bond length in ethanol. By what % does this bond length differ from the C-C bonds in methylcyclohexane?

Chapter 4: Building More Complex Structures: 1-Methyl trans Decalin

We will next illustrate how to build up complex structures from simple structures in ChemNote, by converting the axial methyl cyclohexane into 1-methyl trans decalin. Start up ChemNote by pulling down the Applications menu, sliding right on "Builders," and choosing "2-D Sketcher." Begin by Opening amecyc6: pull down the File menu and choose "Open." . Our final structure will look like Figure 4.4. We will append the ring from carbon 2 to carbon 3 (see Figure 1.1 and 2.1 for atom numbering). Click on the cyclohexane ring in the middle of the bonds palette. Position the cursor on carbon 2, hold down the mouse button, and move the cursor to carbon 3. The rings should be fused as shown in Figure 4.1. You can now pull down the File menu and choose 'Return to Molecular Modeling' to minimize the structure. Choose to 'Save' the changes and type in a unique name in the file librarian. Remember, don't use punctuation or spaces in file names. The minimized structure is shown in Figure 4.2.

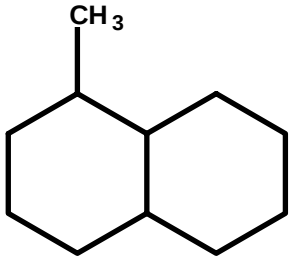


Figure 4.1. 1-methyl trans-decalin

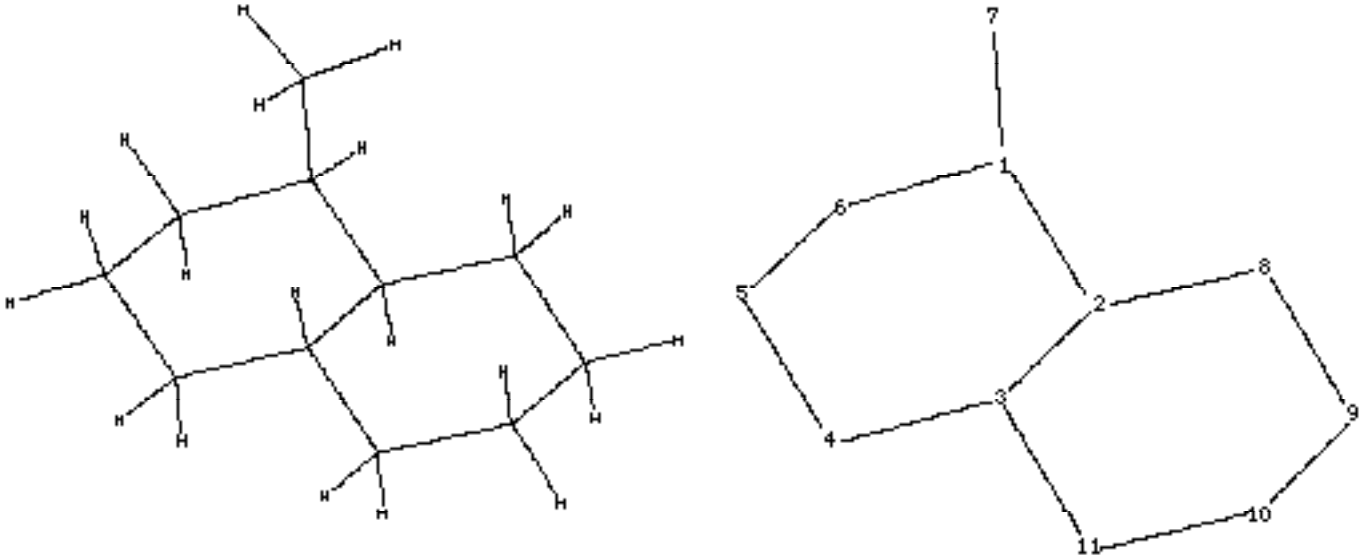


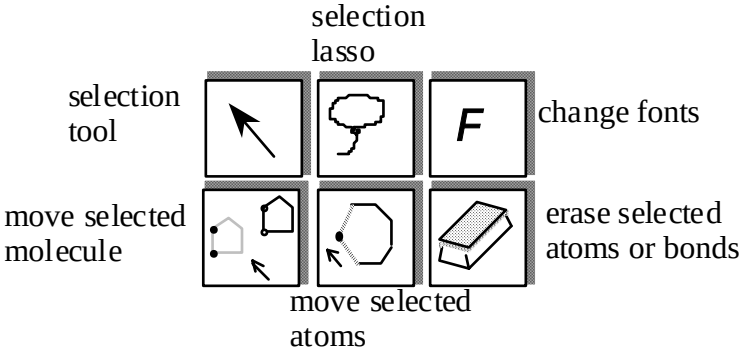
Figure 4.2: 1-methyl-transdecalin. Left: the structure, Right: without hydrogens

Editing Molecules In Chem Note

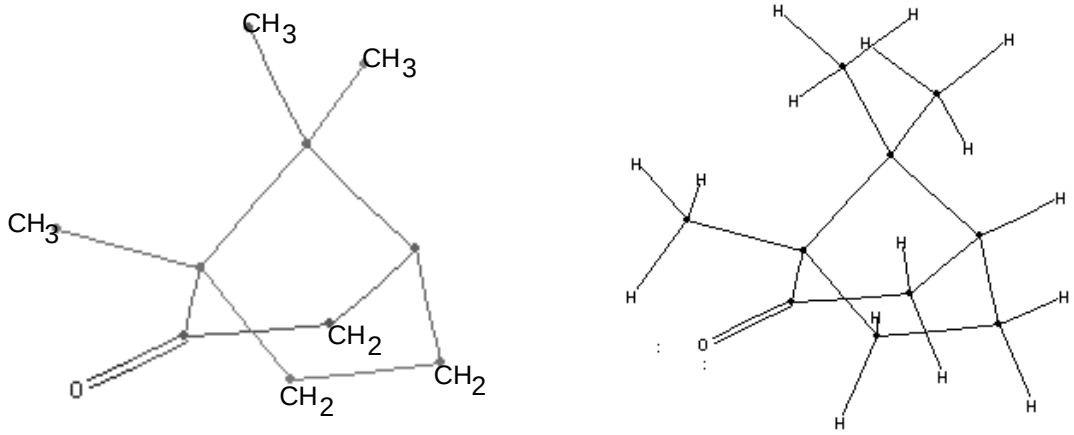
To illustrate how to edit atoms in ChemNote, we will now change axial methyl cyclohexane to chlorocyclohexane. Begin again by Opening amecyc6 in ChemNote. Click on the Cl icon in the atoms palette and then click on the methyl carbon in the sketch pad window. The methyl group

should change to a Cl. Playing with ChemNote is the best way to learn how to build molecules. Try other changes to your structure. If you make a mistake, click on the selection tool icon, ↑, in the Edit Icons palette in the upper left-hand portion of the window. Select the incorrect atoms, and then click on the eraser.

The edit tools palette is shown below with the action of each button:



Problem 4.1 Use ChemNote to build the structure for camphor, Figure 4.3. In ChemNote build the molecule starting with cyclohexane, and add the other bonds as if you were looking from above, Figure 4.4. Use Conjugate Gradient minimization to refine the structure. Report the final steric energy and the various energy contributions. Which term dominates the energy of camphor? Compare your results with methylcyclohexane, or better 2-methyl, 5-isopropylcyclohexanone. Does this comparison bare out the expectation that camphor is a highly strained molecule?



a. b.

Figure 4.3. (a). Camphor. (b) Structure of camphor from molecular mechanics.

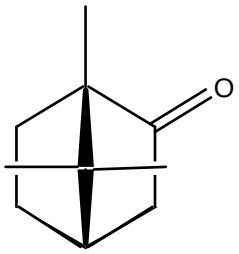


Figure 4.4. The appearance of camphor in ChemNote.

Chapter 5. Conformational Preference for Butane

We will determine the conformational preference and corresponding equilibrium constant for butane, which is an important and experimentally well-studied system. We will also learn how to use the Conformational Search application.

First consider ethane. Two possible conformations of ethane are shown in Figure 5.1.

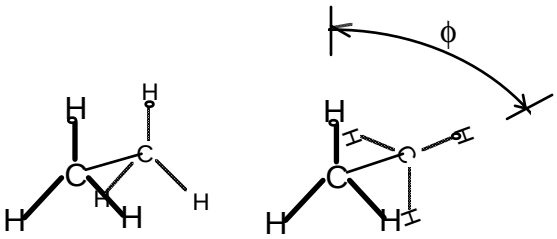


Figure 5.1. Eclipsed and staggered ethane.

The eclipsed conformer is higher in energy than the staggered form. The increase in **dihedral energy** of the eclipsed form is caused by the repulsion of the electrons in the C-H bonds on different ends of the molecule. In the staggered form, the bonds are further apart thus reducing the electron-electron repulsion between the bonds. A plot of the dihedral energy of ethane is shown in Figure 5.2. The energy penalty of having eclipsed bonds rather than staggered bonds is seen to be 2.7 kcal/mol (11.3 kJ/mol). The energy curve has three minima because the three atoms attached to each end of the molecule are the same. Therefore, the conformations with ϕ = 0°, 120°, and 240° are all identical eclipsed conformations. The conformations with ϕ = 60°, 180°, and 300° are all identical with staggered, low energy conformations. Locate these energies in Figure 5.2.

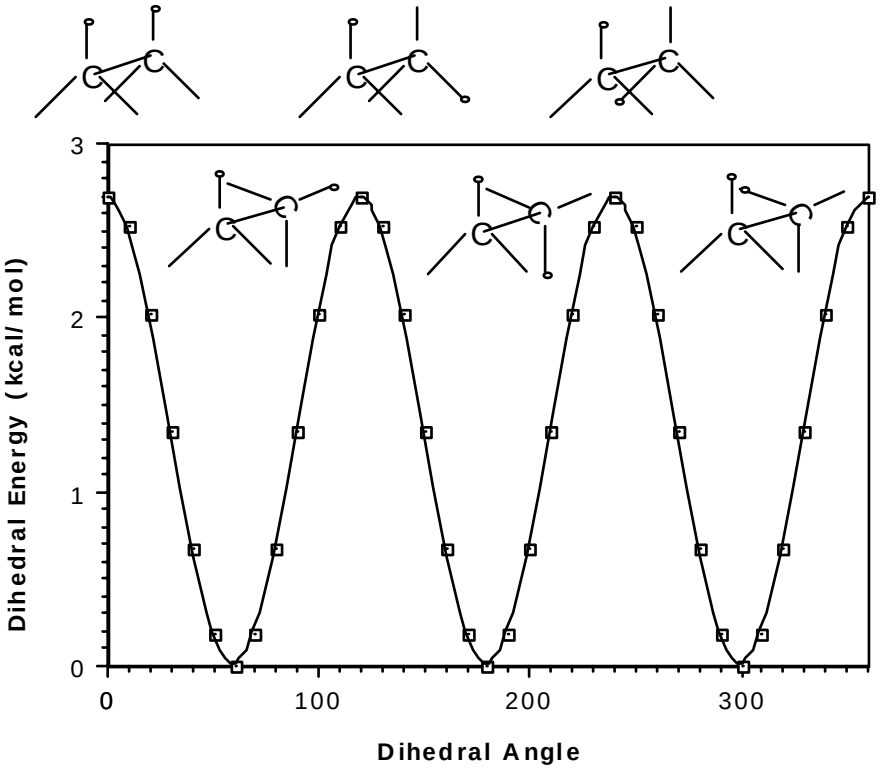
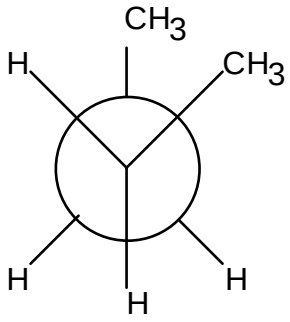


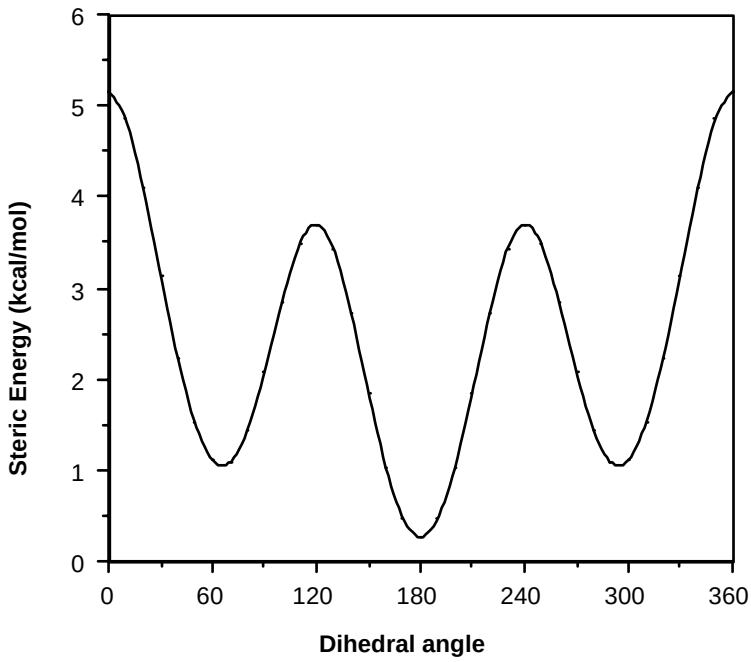
Figure 5.2. Dihedral energy in ethane. In the structures all hydrogens are equivalent, however one particular hydrogen on the front of the molecule and one on the back are shown with a dot so that you can follow the change in the dihedral angle over a full 360°.

Figure 5.2 is a plot of the dihedral, or torsional, potential energy for a 3ϕ, three-fold torsional barrier. Remember that the full torsional potential energy is given by:

$$E_{tor} = 1/2\ k_{tor,1}\ (1 - \cos \phi\ ) + 1/2\ k_{tor,2}\ (1 - \cos 2\ \phi\ ) + 1/2\ k_{tor,3}\ (\ 1 - \cos 3\ \phi\ ) \qquad 1$$

Butane, Figure 5.3a, will also have a large term for the one-fold potential. The CHARMm steric energy as a function of dihedral angle is shown in Figure 5.3b.





a. b.

Figure 5.3. (a.) Butane, in the gauche conformation. (b) Steric energy for butane.

In butane, the difference in energy between the anti and gauche forms is -0.8 kcal/mol. Also note that the minimum energy dihedral angle is 67° and not the ideal 60°. The equilibrium constant for the ratio of anti to gauche forms can be estimated from this energy difference. First, we will assume that there are no significant changes in vibrations between the two conformers. The steric energy difference is then ΔU. Remember $\Delta H = \Delta U + \Delta n_g\ RT$, where Δn_g is the change in the number of moles of gas. Since we are calculating the difference in energy between two conformers:

butane (gauche) -> butane (anti) 2

$\Delta n_g = 0$. Therefore, ΔU = ΔH. Next we need to calculate the change in entropy for the conformational change. Since there are two equivalent gauche conformers and only one anti conformer:

$$\Delta S \text{ (anti-gauche)} = R \ln (1/2) = -1.38 \text{ cal/mol K} = -5.76 \text{ J/mol K} \qquad 3$$

Then $\Delta G \text{ (anti-gauche)} = \Delta H - T\Delta S$ 4

in calories:

$$\Delta G = -0.8 \text{ kcal/mol} - (298.2 \text{ K})(-1.38 \times 10^{-3} \text{ kcal/mol K}) = -0.39 \text{ kcal/mol} \qquad 5$$

and in kJ:

$$\Delta G = -3.35 \text{ kJ/mol} - (298.2 \text{ K})(-5.76\text{x}10^{-3} \text{ kJ/mol K}) = -1.63 \text{ kJ/mol} \quad 6$$

and the equilibrium constant can be obtained from:

$$\Delta G = -RT \ln K \quad 7$$

giving:

$$K = \frac{[\text{anti}]}{[\text{gauche}]} = 1.93 \quad 8$$

In other words, there are two molecules in the anti-conformation for every molecule in the gauche conformation at 25°C.

The following instructions will show you how to repeat the above calculations for the energy minimum structures for the anti and gauche forms of butane and also how to generate the energy plot in Figure 5.3b.

Conformational Preference for Butane

Build butane in ChemNote: pull down the Applications menu, slide right on "Builders", and choose "2-D sketcher" Click on the single bond icon in the bonds palette. Drag the three C-C bonds of butane out on in the ChemNote window. Don't worry about the hydrogens, ChemNote will add them automatically. Pull down the File menu and choose "Return to Molecular Modeling..." Proceed as you did in Chapter 1, using conjugate gradient minimization. After minimization, select "Save changes " from the Modeling palette, and select the "Overwrite.." option. This conformation should be the anti-conformer. Click on "CHARMm Energy..." in the Modeling palette. The contributions to the total steric energy will be listed in the window at the bottom left side of the screen. Record these energies and the total steric energy.

To find the energy minimized structure for the gauche isomer: select "Torsions..." from the Modeling palette. The "Torsions" palette will appear. Pick the first atom defining the torsion by clicking on a -CH_2- carbon. Pick the second atom defining the torsion by clicking on the second -CH_2- carbon atom. Select "Finish" in the torsions palette. The dials palette will now change to show only one dial, that for torsion 1. If the dials palette isn't completely visible, click on the border of its window (notice that the cursor changes to a >| symbol when you are on the border of a window). Click on the "torsion 1" dial until the dihedral angle is near 60°. Then select "CHARMm Minimization..." from the Modeling palette. Remember to click on "CHARMm Minimization..." repeatedly to make sure the structure is completely minimized. Click on "CHARMm Energy..." in the Modeling palette. The contributions to the total steric energy will be listed in the window at the bottom left side of the screen. Record these energies and the total steric energy. Select "Save changes " from the Modeling palette, and select the "Overwrite.." option.

Problem 5.1

Record the various contributions to the steric energy in the table below. Calculate the difference in energy for each contribution in the 4th column. In the 5th column record which conformer is favored by each contribution. Finally, from the difference column, decide which contribution dominates the conformational preference in butane.

Contribution	anti (kcal/mol)	gauche (kcal/mol)	difference (kcal/mol)	favored conformer
bond energy				
angle energy				
dihedral energy				
Lennard-Jones				
Electrostatic				
total				

The Boltzman Distribution: An Alternative Viewpoint

The Boltzman distribution describes the probability of occurrence of a structure with energy E_i:

$$\text{probability of occurrence} = \frac{e^{-E_i/RT}}{q} \qquad 9$$

where $e^{-E_i/RT}$ is called the Boltzman weighting factor, R is the gas constant 8.314 J $mol^{-1}K^{-1}$, T is the temperature in degrees K, and q is the sum of the probabilities over all possible states. The q term, which is called the partition function, just assures that the probabilities sum to 1.0. The effect of a temperature increase is to increase the probability of high energy structures. For example, at a low temperature most molecules will be found in the lowest energy state, but as the temperature increases molecules gain energy through collisions and are promoted into higher energy states, Figure 5.4a. Alternatively, if the temperature is constant, systems with large energy differences have few molecules in high energy states. Systems with small energy differences between their levels have many molecules in upper energy states, Figure 5.4b.

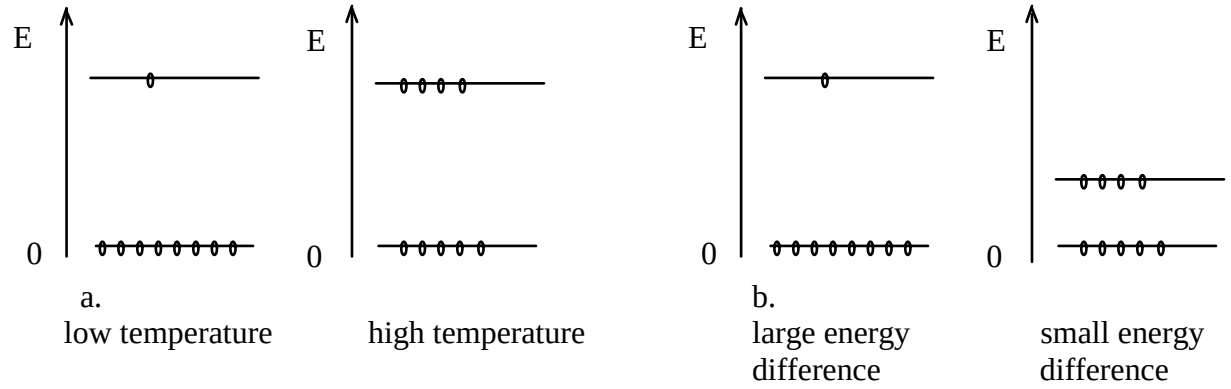


Figure 5.4 The Boltzman distribution determines the probability of occurrence of a given energy state of a molecule. a. High temperatures favor higher energy states. b. Small energy differences favor higher energy states.

What determines the energy difference between energy states? A good example is the conformational energy of butane. The difference in energy between the gauche and anti forms is

0.8 kcal/mol. The Boltzman distribution will tell us the relative numbers of molecules in the anti and in the higher energy gauche states. Another example is the conformational preference of axial and equatorial methylcyclohexane. The CHARMm steric energy of axial-methylcyclohexane is 1.8 kcal/mol higher than the equatorial isomer (Chapter 2).

If there is more than one structure at a given energy, then we must multiply the probability by the number of structures at the same energy. The number of structures at the same energy is called the degeneracy and is given the symbol g. For example, butane has one anti-conformer, $g_{anti}=1$, and two gauche-conformers, $g_{gauche}=2$. The Boltzman distribution with degeneracy is:

$$\text{probability of occurrence} = \frac{g\, e^{-E_i/RT}}{q} \qquad 10$$

and

$$q = \sum_{\text{all states}} g\, e^{-E_i/RT} \qquad 11$$

Take butane as an example. The anti-conformer has the lowest energy, which we can assign as $E_{anti}= 0$. Then the gauche-conformer has an energy $E_{gauche}= 0.8$ kcal/mol = 3.35 kJ/mol above the anti-state. Table 5.1 shows how to calculate the probabilities from Eq. 10 and 11. The probabilities are in the last column.

Table 5.1. Calculation of the Boltzman factors for gauche and anti-conformations of butane at 298.2K.

Conformation	Energy, $E_{i\ (kJ)}$	E_i / RT	$e^{-E_i/RT}$	$g\, e^{-E_i/RT}$	$g\, e^{-E_i/RT} / q$
gauche	3.35	1.35	.2589	0.5178	0.3411
anti	0	0	1	1	0.6588
sum=q=				q=1.5178	

To calculate q we sum the weighting factors in column 5. Then we use q to calculate the probabilities in column 6. Notice that if we take the ratio of the probabilities of the anti and gauche states we get the same result as Eq. 8, above, which was calculated from Gibb's Free Energy:

$$\frac{\text{probability for anti}}{\text{probability for gauche}} = \frac{0.6588}{0.3411} = 1.93 = K \qquad 12$$

The Gibb's Free Energy and Boltzman approach are equivalent but take slightly different points of view.

Dihedral Angle Conformation Searches

Pull down the Applications menu and choose "Conformational Search." The Conformation Search palette will then appear. Select "Torsions..." and then a new palette will appear. Make sure "Define all torsions." and "Use Default Names." is selected. Next click on "Pick torsions..." If a dialog box appears that says "Torsions already defined" then click on "Define New Torsion

Angles" and then "OK." The "Pick Torsions" palette will appear. Click on the four carbons in the order they appear in the chain: in other words, click on the first CH_3 , then the $-CH_2-$, the second $-CH_2-$, and finally the last CH_3. The "Torsion Name" window will then be displayed with "tor 1" as the default name and with "This is a backbone torsion" already selected; click "Apply." Next click on "Finish" in "Pick Torsions" palette, then "Exit torsions." Now click on "Setup search" and then select "Grid Scan." A grid scan search will change the dihedral angle in equal steps. In the "Define Ranges of Values" setup window choose "Absolute values", and change the settings to "From 0 to 350" and "Step size 10." Click on "OK" and then the "Grid Search Options" dialog will be displayed. Select "CHARMm minimization for each structure", "Constrain Grid torsions during minimization", "OK", and finally select "Do search" in the "Conformational Search" palette. The File Librarian will be displayed; type in a file name, for example "butanesrch." Click on "New." To see the results of the search, choose "ANALYSIS." In the "Plots" palette select "Trace...". Choose the plotting and sorting property as the "Torsion Angle." Click on "OK." The trace plot will be displayed. To set the trace plot x-axis to 0 to 360°, pull down the Trace tools menu and choose "Set 360 deg Scale."

To see the structure that corresponds to a given point in the trace plot: Pull down the Trace tools menu in the Trace plot window and choose "Select Structure." Drag the Trace window to the right so that you can see your structure in the normal modeling window. Now when you double click in the Trace plot window at various angles, the corresponding structure will be displayed. To exit the "Select Structures" mode **Press the F1 key** at the top of the keyboard. Pull down File in the Trace plot window and choose "Quit." Next click on "Exit Plots," "Exit Analysis" in the Analysis palette, and finally "Exit Conformational Search."

Please note that for the 'Torsions...' tool in the Modeling palette, you mark only two atoms. In other parts of Quanta, for example in the Geometry palette and for Conformational Searches, you need to specify all four atoms of the dihedral.

Problem 5.2

Calculate the equilibrium constant for the anti to gauche conformers for dichloroethane. Find the dihedral angle in the gauche conformer. Why is this angle different from butane? Also, use the Conformational Search application to plot the steric energy as a function of dihedral angle.

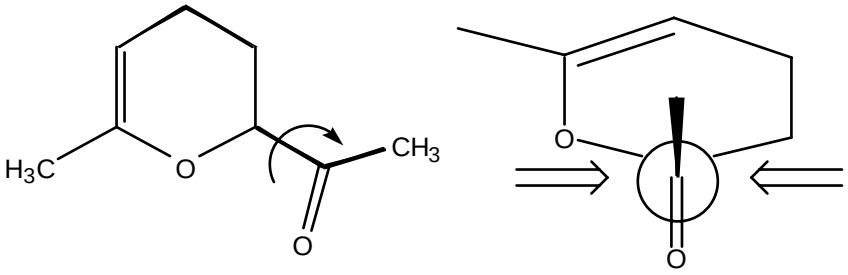
Problem 5.3

Using the energy difference from Problem 5.2, calculate the probabilities of occurrence of the gauche and anti forms for dichloroethane. Follow Table 5.1.

Conformation	Energy, $E_{i\ (kJ)}$	E_i / RT	$e^{-E_i/RT}$	g $e^{-E_i/RT}$	g $e^{-E_i/RT}$ / q
gauche					
anti	0	0	1	1	
sum=q=				q=	

Problem 5.4

The dimer of methylvinylketone is shown in Figure 5.4. For this problem we will study just the axial conformer for the -CO-CH_3 side chain. An interesting question is which face of the carbonyl is more susceptible to nucleophilic attack? Nucleophilic attack will be perpendicular to the trigonal plane of the sp^2 hybridized carbon, Figure 5.4b. According to Cram's rule, the less hindered side is likely to be most susceptible. Make sure that you build the axial conformer. To begin this study we need to know the low energy conformers about the side-chain C-C bond to the ring. Do a conformational search around this bond. What are the low energy conformers? Draw these low energy conformers and note the less hindered side. Van der Waals solid models will be helpful in looking at steric influences.



a. b.

Figure 5.4 (a). Methylvinylketone dimer. The bond with free rotation is marked. (b) Newman projection. Which side of the carbonyl is attacked by nucleophiles? The favored direction of attack will change with conformation angle. Only one possible conformation is shown here.

Chapter 6: MM2

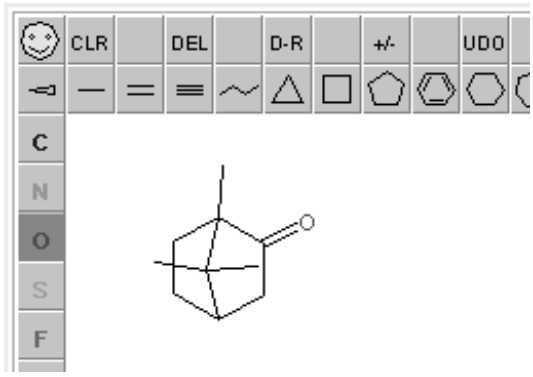
Energy minimization using CHARMm and MM2 are very similar. The force fields are a little different, but the calculations do the same thing. One reason for using MM2 is to calculate enthalpies of formation, which most other molecular mechanics programs can't do. MM2 also treats conjugated pi-electron systems better than MOE. You can't, on the other hand, use MM2 for large molecules. In this chapter you will calculate the enthalpy of formation of camphor, so do Problem 4.1 first.

MM2 Minimization

The easiest way to access MM2 is to use the "Distance Geometry with MM2 and SCF" applet. This applet uses the Java Molecular Editor for structure input. You will find the link for this Web applet on the PChem home page (or iris12.colby.edu/~www/jme/dgmmp2.html). The distance geometry part of the applet is a method for constructing 3D structures from 2D line drawings (Please read Introduction Section 6). Distance geometry also is used when you want to build a molecule with distance constraints. Distance constraints are ranges of allowable distances between pairs of atoms. These ranges can be values that you choose to help specify the conformation of the final molecule. For example, you may want a conformation that puts two parts of a long molecule close to each other. Another example is that you can specify that two atoms are to be within a normal hydrogen bond distance of each other. NMR spectra are very useful for experimentally determining distance constraints. We don't need distance constraints for our project on camphor, so we will just skip the distance constraints section for now.

Building Camphor:

To build camphor, start by clicking on the cyclohexane ring button and then click in the molecule window. Next select the single bond button, [-] . Drag the single bonds as shown at right. Clicking on a bond will change it to a double bond. Finish by putting in the oxygen. Click the Help button for more information on using the Java Molecular Editor. Scroll down until you see the options button. The Add hydrogens button should be selected.



The "SCF pi calculation" option won't make any difference since camphor has no conjugated double bonds. The "Set-up Gaussian 98" input button is used when the applet is run to prepare the input files for molecular orbital calculations, which we will ignore for now. Click Submit to start the calculation.

A new window will be opened with the output from the MM2 run. An example printout for cyclohexane is shown at right.

```
FINAL STERIC ENERGY IS      6.5510 KCAL
          bond               0.3375
          bend               0.3652
          stretch-bend       0.0826
          Lennard-Jones,add:
            1,4 energy       4.6734
            + other         -1.0633
          dihedral           2.1556
BOND ENTHALPY =             -38.48
SIGMA STRAIN ENERGY =         2.61
FINAL STERIC ENERGY IS      6.5510 KCAL
HEAT OF FORMATION =         -29.53
```

.The different contributions the energy force field are listed, with the Van der Waals energy listed as the Lennard-Jones energy and the torsional energy listed as the dihedral energy. The enthalpy of formation is listed on the line labeled HEAT OF FORMATION. The line labeled SIGMA STRAIN ENERGY is a very useful measure of the total strain in the molecule.

You can print this information by pulling down the File menu and choosing Print. The energy for the starting structure and the final minimized structure are listed, so that you can see the difference. You will also see your minimized structure in a Chime window. You can use all of the normal Chime display options. Below the Chime window is a section that is used to prepare the input files for molecular orbital calculations, which we will ignore for now.

Problem 6.1 The Enthalpy of Formation of Camphor

Camphor is an interesting molecule because of its many uses and because it is a highly strained molecule. Because of the rings in the molecule, there are no torsional increments other than for methyl groups. There is also only one low energy conformation. Calculate the enthalpy of formation from bond energies (Table I in the Introduction to Molecular Mechanics or from your General Chemistry or Physical Chemistry text). The bond energy calculation for camphor from MMX is reproduced below for comparison. MMX uses a very similar force field to that in MM2.

#	Bond or Structure	Each	Total
9	C-C SP3-SP3	- 0.004	- 0.036
16	C-H ALIPHATIC	- 3.205	-51.280
1	C=O	-25.000	-25.000
2	C-C SP3-SP2 C=O	- 3.000	- 6.000
1	ISO (ALKANE)	0.078	0.078
1	NEO (ALKANE)	- 0.707	- 0.707
3	C(SP3)-METHYL	- 1.510	- 4.530
1	TERT-CARBONYL	- 1.300	- 1.300
		bond energy =	-88.775 kcal

Report your bond energy calculation, using Table I or data from your text, MMFF94 steric energy, MM2 steric energy, MM2 bond enthalpy, the strain energy, and the enthalpy of formation of camphor.

Compare the calculated results with the literature by completing the following calculations. The enthalpy of combustion of camphor is -1411.0 kcal/mol. But we must also add the enthalpy of sublimation since our MM2 calculation is for the gas phase. The enthalpy of sublimation of camphor is 12.8 kcal/mol. From the enthalpy of combustion and the enthalpy of sublimation calculate the enthalpy of formation of gaseous camphor and compare with the MM2 value. How close did you come?

Problem 6.2. Comparisons with Literature Values (or How Good is MM2?)

How well do MM2 enthalpies of formation match literature values? The monoterpenes are an important group of natural products, Figure 6.1. Determine the enthalpy of formation for each. The literature values are from Lange's Handbook or the CRC. Remember, you must add in the enthalpy of vaporization for liquids or the enthalpy of sublimation for solids, since molecular mechanics energies are for the gas phase. The enthalpy of vaporization or sublimation values in

kJ/mol are: camphene, 43.9; α-pinene, 45.2; β-pinene, 46.4; limonene, 43.9; α-terpineol, 52.3; menthol, 56.5kJ/mol. Even though you've been given the phase transition enthalpies, it is still a good idea to learn how to find them for yourself.

Enthalpies of Vaporization or Sublimation: The Clausius Clapeyron equation describes the change in vapor or sublimation pressure with temperature:

$$\ln \frac{P_2}{P_1} = -\frac{\Delta_{tr}H}{R}\left(\frac{1}{T_2}-\frac{1}{T_1}\right) \quad \text{or equivalently} \quad \ln P = -\frac{\Delta_{tr}H}{RT} + \text{cst} \qquad 6.1$$

where P_1, and P_2 are the vapor pressures at temperatures T_1 and T_2, respectively, and $\Delta_{tr}H$ is the enthalpy of vaporization or sublimation. Comparing with Eq. 6.1, if you use the CRC for enthalpies of vaporization from the vapor pressure verses temperature tables, the listed "a" parameters are equal to the enthalpy of vaporization in kJ/mol. Remember to change to kcal/mol (1cal = 4.184 J). If you use the sublimation pressure verses temperature table from the CRC then the enthalpy of sublimation = 2.303 R B, as listed in the table caption. If you use R in J/mol K the result will be in J. If you use R=1.987 cal/mol K the results will be in cal. If the enthalpy of vaporization isn't available from tables directly, Eq. 6.1 shows that a plot of the ln (vapor pressure) versus 1/T gives a straight line with slope $-\Delta_{tr}H/R$. The CRC has tables of vapor pressure versus temperature for many organic compounds.

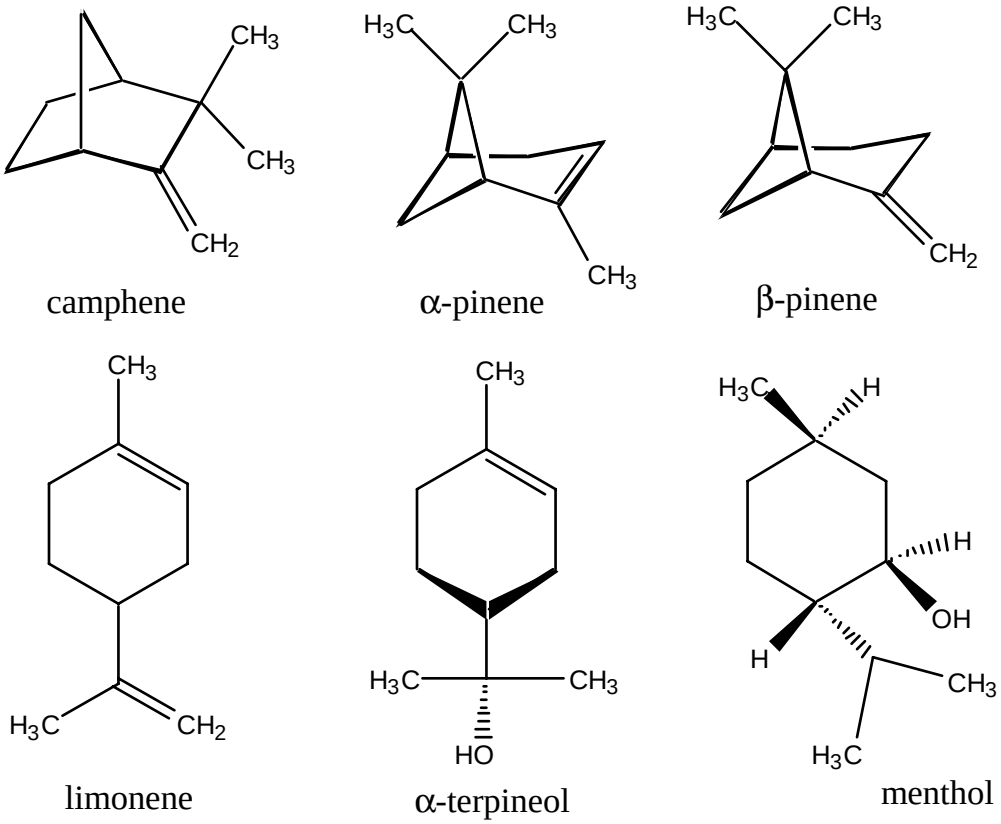


Figure 6.1 Some monoterpine natural products.

Report your results in the following table. The literature values are listed in kcal/mol. You can share data with other students, but make sure to do at least two of the molecules yourself. Label the two that you did.

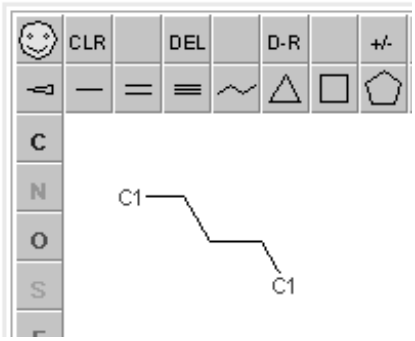
	Literature (kcal/mol)			Calculated (kcal/mol)				
compound	$\Delta_f H_{(l \text{ or } s)}$ from tables	$\Delta H_{sublim.}$ or vaporiz.	$\Delta_f H^\circ(g)$	MMFF94 steric energy	MM2 steric energy	MM2 Bond enthalpy	$\Delta_f H^\circ$ MM2	Error (kcal/mol)
Camphene	-18.22	10.49						
α-pinene	-4.04	10.80						
β-pinene	-1.84	11.09						
limonene	-23.51	10.49						
α-terpineol	-85.84	12.50						
menthol	-114.86	13.50						

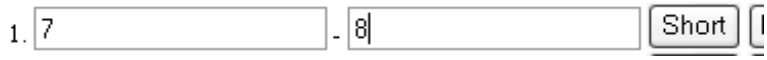
Problem 6.3 Torsional Increments to the Enthalpy of Formation

If there are unconstrained or free bond rotations in a molecule, the MM2 $\Delta_f H°$ should be low. In this problem we wish to determine if the addition of torsional increments (see Introduction) improve the agreement of the calculated values with literature values. Remember the torsional increment is estimated as 0.36 kcal/mol or 1.51 kJ mol^{-1} for each internal rotation. For example, butane, CH_3-CH_2-CH_2-CH_3 , has one additional internal rotation other than the methyl group rotations; so the torsional increment for butane would be 0.36 kcal/mol. Complete the following table (the data is from the CRC 46th Ed.):

compound	Literature $\Delta_f H^\circ(g)$ kcal/mol	Calculated $\Delta_f H^\circ$MM2	Δ kcal/mol	torsional increments		Calculated $\Delta_f H^\circ$MM2 + total increment	new Δ kcal/mol	% improv-ment
				# free rotations	total increment kcal/mol			
n-pentane	-35.00							
1-pentene	-5.00							
cis-2-pentene	-6.71							
trans-2-pentene								
2-methyl-1-butene								

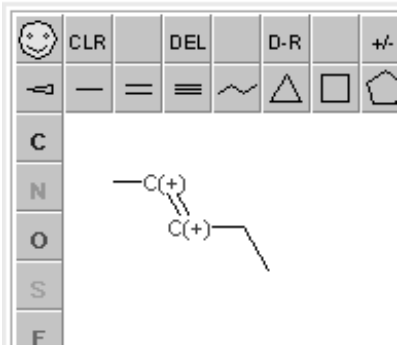
Hints: Finding the lowest energy conformation can take many tries. You can bias the calculation by setting a long distance constraint. Use the [X] tool to set in the C1 element names. For example, for pentane you might try:





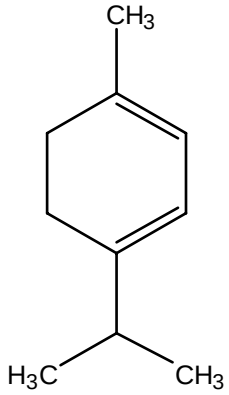
This constraint is actually longer than pentane can stretch, but that's OK (6-7Å) works well too. For more information on constraints, see the instructions in the applet and Introduction Section 7.

If you have problems getting trans-2-pentene you might try the following:



Conjugated Pi-Electron Systems

α-Terpinene is an important mono-terpene (see Problem 6.2). However, the pi-electrons in the two double bonds are conjugated. MM2 in its simplest form does not do a good job on calculations of conjugated pi-electron systems. The MM2 $\Delta_f H^\circ(g)$ as calculated in the same fashion as above is 22 kcal/mol whereas the experimental value is -4.89 kcal/mol. We must account for the extra stability of the conjugated pi-system and also the extra barrier to rotation about the bond between the two double bonds. This extra barrier to rotation is also caused by conjugation. MM2 accounts for these factors by doing a molecular orbital calculation on the conjugated pi-system. This molecular orbital calculation is called a self-consistent-field calculation, which is abbreviated SCF. The calculation only covers the pi-electrons.



α-terpinene

Figure 6.1

1,3-Butadiene, Figure 6.2, is a simple conjugated system that will serve as a good first example. The printout from the calculation on butadiene is shown in Figure 6.3.

The MO orbital diagrams and the energy diagram are not normally part of the printout, but they are included to help you learn how to interpret the molecular orbital portion of the results.

The MO diagrams are only shown for the lowest two orbitals, since only these two are filled with electrons. The molecular orbital coefficients are listed in columns. At the bottom of each column is the energy of the MO, in kcal/mol. For example, the coefficients for the lowest energy orbital are all positive; therefore all the p atomic orbitals have their positive lobes in the same direction. The energy diagram, at right, shows that the two filled orbitals have significantly lower energy than the empty orbitals. The bond order portion of the printout shows that the end double bonds have a pi-bond order of 0.9662, which is less than a full double bond. However, the single bond between the two double bonds takes on some double bond character, with a pi-bond order of 0.2576. The bond energy in the pi-electron system is -118.06 kcal/mol and the total bond energy, sigma and pi, is -356.71 kcal/mol.. The final $\Delta_f H^\circ$ with the pi-calculation included is listed as the "HEAT OF FORMATION" and is calculated to be 25.09 kcal/mol. The experimental $\Delta_f H^\circ$ is 26.75 kcal/mol. The default mode for the dgmmp2.html applet is to do SCF pi-calculations for all conjugated pi systems. If you don't want to do the pi-calculation deselect the "SCF pi calculation" checkbox.

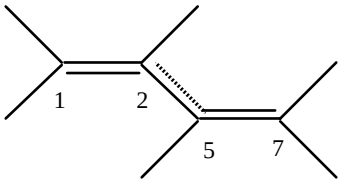


Figure 6.2. 1,3-butadiene. (The atom numbers correspond to the printout in Figure 6.3.)

Problem 6.5 MM2 Calculations with SCF Pi Calculations

Calculate the enthalpy of formation of α-terpinene. The MM2 $\Delta_f H^\circ(g)$ as calculated without the SCF molecular orbital calculation is 22 kcal/mol; the experimental value is -4.89 kcal/mol (from the CRC).

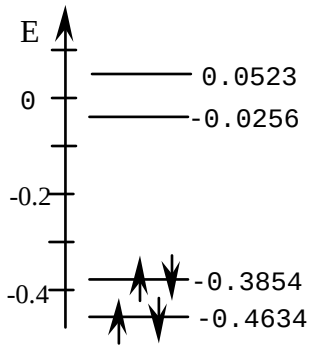
Figure 6.3: See the next page

Figure 6.3 The MM2 printout for 1,3-butadiene. The calculation is done with the SCF pi-molecular orbital calculation.

```
Butadiene MM2 calculation with SCF calculation

  MOLECULAR ORBITALS
 0.4308          0.5607          0.5607          0.4308
 0.5607          0.4308         -0.4308         -0.5607
 0.5607         -0.4308         -0.4308          0.5607
 0.4308         -0.5607          0.5607         -0.4308

EIGENVALUES
-0.4634         -0.3854         -0.0256          0.0523
```




```
BOND ORDERS (P) AND RESONANCE INTEGRALS (B) FOR PI-BONDS
          ---NONPLANAR GEOMETRY---          ----PLANAR GEOMETRY-----
     BOND    P(W)     B(W)    P*B(W)        P(0)     B(0)    P*B(0)
    1-  2   0.9662   0.9868   0.9534       0.9662   0.9868   0.9534
    2-  5   0.2576   0.8032   0.2069       0.2576   0.8032   0.2069
    5-  7   0.9662   0.9868   0.9535       0.9662   0.9868   0.9535
BOND ENERGY FROM SCF CALCULATION
SIGMA-BOND  -238.65     PI-BOND  -118.06      TOTAL ENERGY -356.71
------------------------------------------------------------------
  FINAL STERIC ENERGY IS     -2.4960 KCAL.
     COMPRESSION       0.0466
     BENDING           0.1723
     STRETCH-BEND      0.0113
     VANDERWAALS
        1,4 ENERGY      1.5450
        OTHER          -0.0312
     TORSIONAL        -4.2400
     CHARGE-DIPOLE     0.0000
------------------------------------------------------------------
HEAT OF FORMATION AND STRAIN ENERGY CALCULATIONS (KCAL/MOLE)
  BOND ENTHALPY (BE) AND STRAINLESS BOND ENTHALPY (SBE) CONSTANTS AND SUMS
     #   BOND OR STRUCTURE        --- NORMAL ---     --STRAINLESS--
     6   C-H OLEFINIC             -3.205  -19.23     -3.125  -18.75
     3   C-C     DELOCALIZED     152.750  458.25    152.850  458.55
     2   SEC-    DELOCALIZED     -28.540  -57.08    -29.167  -58.33
                                ----------------   ----------------
                                   BE =  381.94      SBE =  381.47
  PARTITION FUNCTION CONTRIBUTION (PFC)
          TRANSLATION/ROTATION TERM (T/R)    PFC =   2.40

  STERIC ENERGY (E)                                          -2.50
  SIGMA-STRETCHING (ECPI)                                     0.04
  CORRECTED STERIC ENERGY (EC) = E-ECPI                      -2.54
  ENERGY FROM PLANAR SCF CALCULATION  (ESCF)                -356.71
  HEAT OF FORMATION = EC + BE + PFC + ESCF                    25.09
  STRAINLESS HEAT OF FORMATION FOR SIGMA SYSTEM (HFS)
            HFS = SBE + T/R + ESCF - ECPI                     27.11
  INHERENT SIGMA STRAIN (SI) = E + BE - SBE                   -2.02
  SIGMA STRAIN ENERGY (S) = POP + TOR + SI                    -2.02
```

Chapter 7: Comparing Structures

Changes in a molecule's structure not only affect the local environment, but can have affects on the structure many bonds away. In this section you will compare the structures of axial- and equatorial- methylcyclohexane from Chapters 1 and 2. The "Molecular Similarity" application is used to calculate the differences in two structures and to produce an overlaid view of the two structures. Pull down the File menu and choose 'Open'. Click on amecyc6.msf, at the bottom of the dialog box choose 'Append' (rather than 'Replace') so that both molecules will be displayed, and click on "Open." In the Molecule Management window, in the lower right portion of the screen, you can control which molecules are displayed by clicking in the 'Visible' column for the molecle of interest. Pull down the Applications menu and choose 'Molecular Similarity.' A new palette will appear. We now need to move one of the molecules to the right so they are no longer overlapped. Choose 'Move Molecule', click on an atom in the equatorial isomer, and move it to the right so that the isomers no longer overlap. To move the molecule use the Dials palette (lower right hand corner) or hold down the shift key and use the mouse. Click on "Move molecules" again to finish up. If the molecules aren't in orientations where you can see all the carbon atoms, choose 'Rotate Molecules in Place' and reorient the molecules. To rotate only one molecule, use the Molecule Management window to select the molecule you wish to rotate by clicking in the 'Active' column. Use the Dials palette to rotate the molecule. Make sure both molecules are active in the "Molecule Management" window, before proceeding.

We must now choose atom pairs that we wish to superimpose in the two isomers. Select 'Match Atoms.' A new palette will appear; choose 'Pick Equivalent Atoms'. Click on carbon 1 (Figure 1.1 and 2.1) in each isomer. A dotted line will be drawn between the two equivalent atoms. Do the same for carbon 2 in each structure (carbon 2 is the secondary ring carbon adjacent to the tertiary carbon). Also choose the equatorial hydrogens on carbon 2. If you make a mistake, choose 'undo last' and choose again. You can reorient the molecules at anytime using the center mouse button as before. You should now have three dotted lines between equivalent atoms. When the three pairs are choosen click on 'End Atom Picking' and then 'Exit Match Atoms.'

To do the comparison choose 'Rigid Body Fit to Target.' The target molecule is listed with an asterisk in the Molecule Management window. In this rigid body option, no diherdral angles are changed, the algorithm simply does a least squares fit by adjusting the position of the center of mass and orientation of the molecules. The root mean square (rms) differences are displayed in the TextPort

Notice first that the C-C bond to the methyls doesn't align with the C-H bond from the other isomer on the same tertiary carbon. The methyl groups are bent away from their respective ring to minimize repulsions. These bond angle changes are local differences. Also notice that the secondary carbon on the opposite side of the ring, carbon 4, and its attached hydrogen don't exactly align. In other words, local changes can have an effect many bonds away. This may be caused by ring strain or through-space Van der Waals (Lennard-Jones) interactions. Choose 'Exit Molecular Similarity.'

Color Atoms To make the two molecules easier to tell apart, use the 'Color Atoms' option. Pull down the Draw menu, slide right on "Color Atoms," and choose "By molecule." After you finish remeber to return to normal colors by pulling down the Draw menu, sliding right on "Color Atoms," and choose "By element."

Problem 7: tert-butylcyclohexane

Compare axial and equatorial tert-butylcyclohexane. Which conformer is more stable this time? Is the ring more or less distorted than in the methylcyclohexane case?

Chapter 8: Printing Structures

Structures can be plotted on the HP 870 printer. Orient your molecule in a good position on the QUANTA window, then follow the directions below.

1. Pull down the File menu, slide right on "Plot Molecules," and choose "Generate."
2. In the Plot Dialog box, choose "Artist Plot," and one of the styles listed. "Ball and stick" works well. The "Van der Waals" option is good for small molecules.
5. Enter a title for the plot in the "Title" edit box at the bottom of the screen. Choose the default option to "Plot with Titles and Border."
6. Click on "OK."
7. In the next dialog box select "Preview Plot," and click "OK."
8. A new window will appear with your plot. To continue, click the left mouse button in the Preview window.
9. The Plot Disposition window will be displayed. If the plot looked good, click on "Postscript Format," "Translate as color," "OK.," and go to step 11. If the plot wasn't sized properly, click on "Regenerate plot," and click on "OK."
10. The Plot dialog box will be displayed again. Select the same options as before. To change the plot scaling, change the number in the "Plot Scale" dialog box. The units are in mm/Å, so a bigger number increases the size of the molecule on the screen. Click on "OK" and continue at step 7.
11. The File Librarian window is displayed. Type in a file name. Click on "Save." The ".ps" postscript format file will then be generated. Next the "Plot Disposition" dialog box will be displayed, again. This time click on "Cancel."
12. To actually print the file, double click on the "quanta" folder icon on the desktop. Scroll the directory window until you find your file. The file should have the ".ps" suffix applied. Drag the file to the "HP" printer icon (on the desk top) for color printing or to the "Schupf Lab" printer icon for black and white. If there are no problems, the printer should begin printing within 10 sec. You are now finished. If there were problems go to the next step.
13. If the file didn't print: a) make sure the tray has paper loaded; b) make sure that the file name was correct. The ".ps" is added to your file name by QUANTA, so even though you didn't type it in, it is still necessary.

Fancier Plots

You can copy the current screen to the printer. This copy includes any solid surfaces. However, since these plots print with a black background, much ink is used. Therefore, please only use screen copies for special purposes like papers. To make a screen copy, in step 2 just choose "Color Screen Image."

Another possiblity for good looking plots is to try the "stick plots" option. This type prints with a white background.

Chapter 9: Conformational Preference of Small Peptides

The purpose of this lab is to determine the lowest energy conformation of alanylalanine and to compare this to the value found in the alpha helix in proteins. In particular, we wish to ask if the alpha-helix is the lowest energy conformation of the backbone, or rather is it a higher energy conformation that must be stabilized by hydrogen bond interactions in large systems. The backbone angles are defined in the Figure 9.1. ψ is defined by the N-C-C-N dihedral and ϕ is defined by the carbonyl carbons in the dihedral C-N-C(α)-C. The normal values in the alpha helix are ψ= -47°and ϕ=-57°. The structure of alanylalanine is shown in Figure 9.1. The protein is shown in its non-ionized form. At neutral pH the N-terminus would be a $-NH_3^+$ and the C-terminus would be $-COO^-$. However, in our current work the attraction of the charged end-groups would dominate the conformation. Since we want to study the conformational preference of the backbone, we will build the non-ionized form to avoid the charged end-group attraction, which does not play an important role in large proteins.

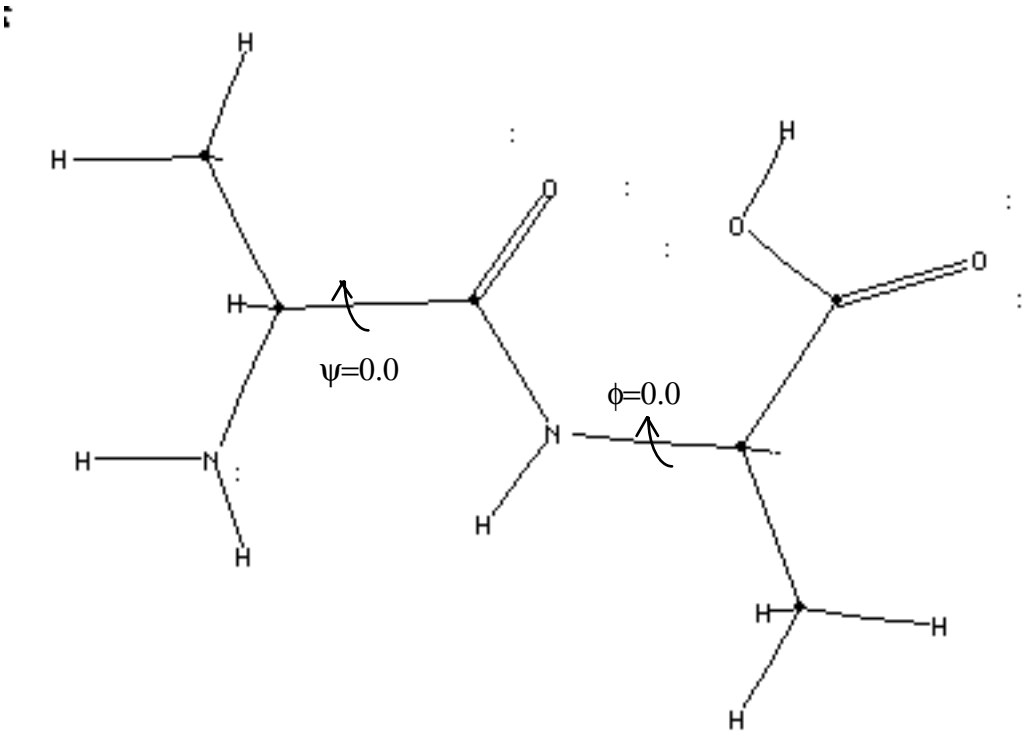


Figure 9.1. The backbone dihedral angles in alanylalanine.

First, Set up CHARMm for normal operation in RTF mode by: pulling down the CHARMm menu and sliding right on "CHARMm Mode" and choosing "RTF." RTF mode is the normal mode for working with polymers, which are composed of repeating units.

Only use RTF mode with proteins and nucleic acids where QUANTA has RTF files that include all of your monomers. For small molecule work always use PSF mode.

Next pull down the CHARMm menu and select "Minimization options." and set the following options:

Number of Minimization Steps	500
Coordinate Update Frequency	5
Energy Gradient Tolerance	0.0001
Energy Value Tolerance	0
Initial Step Size	0.02
Step Value Tolerance	0

Next we need to build the dipeptide. Pull down Applications, slide right on "Builders," and choose "Sequence Builder." The File Librarian will be displayed for you to "Select a residue library." Choose the "AMINOH.RTF" line in the scroll box. Click on "Open." The "Sequence Builder" window will now be displayed with a list of available amino acids in the upper left corner. Click on "ALA" twice. The main window show now show "-ALA ALA." The peptide is built in the default zwitter ion form, which we must now changes. Changes to the sequence are made with "Patches." Pull down the Edit menu and choose "Apply Patches to Residues." The buttons in the upper left corner will change to the available patches. Click on "NH2" and then on the left hand "ALA." Next click on "COOH" and then the right hand "ALA." To set the initial conformation, we will choose the all-trans structure. Then we will check to see if the minimized structure changes much. Pull down the Conformation menu and select "Set Secondary Conformation." You are then instructed to "Pick the residue or range of residues." Click on the two "ALA" residues, and then click on "OK." A dialog box will be displayed, choose the "Extended Backbone (180.0)" option. Select the "OK" button. To exit the builder, pull down the Sequence Builder menu and choose "Return to Molecular Modeling." You will be asked: "Do you wish to save changes," click on "Yes." The File Librarian will then be displayed: type in a file name for your sequences and click on "Save. Next you will be asked "Which molecule do you want to use"; click on "Use new molecule only." The dipeptide is produced in the all-trans conformation. Now we can minimize the structure: choose "CHARMm minimization" repeatedly until the structure is at an energy minimum. What dihedral angles and energy did you get?

To measure the dihedral angles go to the "Geometry" palette. Make sure "Show dihedral monitors" is highlighted. Click on "Dihedrals" to begin selection of your angles. To select the ψ dihedral, start from the N-terminus and click on the backbone atoms: N-C-C-N in turn. To select the ϕ dihedral, start with the carbonyl-carbon on the N-terminus end, and then select the backbone atoms: N-C(α)-C(carboxyl) in turn. Compare these values to the "ideal" alpha helix values. Leave these dihedral monitors on.

What hydrogen bonding exists for this conformation? Go back to the "Modeling Palette" and click on "Hydrogen bonds." How do these hydrogen bonds stabilize the conformation? Are the hydrogen bonds that form similar to those in an alpha helix? Select "Reject changes" to return to the all trans structure. (A file is available with 10 alanines in a right-handed helix, AAAAAAAAAA.msf, so you can compare to a regular helix. Alternately, you can build a short alanine polypeptide in the sequence builder to see what the normal hydrogen bonding pattern looks like. Just make sure the peptide is at least four ALA's long, and choose the "Right-Handed Alpha Helix" secondary conformation option.)

Problem 9.1

Adjust the torsional angles in your dipeptide to give $\psi = -60$ and $\phi = -60$. To accomplish this do the following. Select "Torsions..." from the Modeling palette. The "Torsions" palette will appear. Pick the first atom defining the ψ torsion by clicking on the C(α)- carbon at the N-terminus end. Pick the second atom defining the torsion by clicking on the adjacent carbonyl-carbon atom. Select "Finish" in the torsions palette. The dials palette will now change to show only one dial,

that for torsion 1. If the dials palette isn't completely visible, click on the border of its window (notice that the cursor changes to a >| symbol when you are on the border of a window). Click on the "torsion 1" dial until the dihedral angle is near -60°. Next repeat the above procedure for the ϕ angle, which should be set to -60. Then select "CHARMm Minimization..." from the Modeling palette. Remember to click on "CHARMm Minimization..." repeatedly to make sure the structure is completely minimized. Measure the new dihedral angles and record the energy. After you are finished, select "Reject changes" in the "Modeling palette" before going on to Chapter 10. Which conformation is lowest in energy, the 180,180 or this one? Which structure is better stabilized by hydrogen bonds?

Please note that for the 'Torsions...' tool in the Modeling palette, you mark only two atoms. In other parts of Quanta, for example in the Geometry palette and for Conformational Searches, you need to specify all four atoms of the dihedral.

Chapter 10: Dynamics in Small Peptides.

Purpose: The purpose of the chapter is to use molecular dynamics to find low energy conformations for the alanylalanine dipeptide. This chapter is a continuation of Chapter 9.

Introduction

Molecular dynamics is useful for visualizing the motions of macromolecules. Motional flexibility of enzymes plays a role in binding interactions and in catalytic events. In this Chapter we will study the alanylalanine dipeptide, which you built in Chapter 9. We choose such a small system so that the calculations will run quickly. However, the same procedures are used routinely for large enzymes and oligonucleotides. Molecular dynamics is also a good way to find low energy conformations. Often, energy minimization alone catches the molecule in conformations that are not the lowest energy conformation. Molecular dynamics helps the molecule explore other conformations that may be lower in energy. The take home message from dynamics simulations is that there is more motion than we expect from viewing static textbook models. The motion of molecules is exceedingly important in determining the energetics and course of chemical reactions.

Molecular mechanics minimization corresponds to the structure the molecule would have at zero degrees K. Dynamics calculations are done in three steps. We first do a "heating" run to warm the molecule to room temperature. Next we "equilibrate" the molecule at room temperature to ensure that all the degrees of freedom are at the same temperature. Finally, we do a "simulation" run that generates the trajectory of the molecule at room temperature. The "simulation" run is used to answer questions about the motion of the molecule.

Procedure:

To begin, complete Chapter 9 and leave the alanylalanine molecule in the all-trans, extended backbone conformation. We will see if moleculear dynamics is successful in finding the lowest energy conformation of the dipeptide. Make sure any dihedral monitors are off, but that hydrogen bonds are showing (click on Hydrogen Bonds in the Modeling palette if you haven't already done so).

Make sure the "Shake" option is on. This option keeps the C-H bonds in the molecule from gaining energy. By damping these high frequency vibrations, a longer time step may be chosen, thus decreasing the computation time. To turn "Shake" on, pull down the CHARMm menu and choose "SHAKE Options." Select the following options:

- SHAKE On
    - BonH

Run dynamics with Shake ON.
Shake tolerance: 1e-09
Maximum number of iterations: 500
Use Parameter-specified Geometry

Pull down the CHARMm menu and choose " Dynamics options." Select "Setup Heating," and click on "OK." In the heating setup dialog choose the following options:

Dynamics steps:	3000
Restart Read File:	_____
Restart Write File:	heat
Coordinate trajectory file:	heat
Energy values file:	heat
Output frequency:	10
Time step:	0.001
Initial temperature:	0
Final temperature:	300.0

•Start Heating From the Beginning

The time step is in picoseconds. Therefore, the time step of 0.001psec is $1x10^{-15}$sec or 1 femtosecond! Click on "OK." Next we set up the equilibration run: click on "Setup Equilibration." In the Equilibration setup dialog enter:

Dynamics steps:	3000
Restart Read File:	heat
Restart Write File:	equil
Coordinate trajectory file:	equil
Energy values file:	equil
Output frequency:	10
Equilibration Frequency:	200
Time step:	0.001
Temperature:	300.0

Restart Equilibration From the Restart File

Click on "OK." Finally, we setup the simulation run:click on "Setup Simulation." In the Simulation setup dialog enter:

Dynamics steps:	3000
Restart Read File:	equil
Restart Write File:	simul
Coordinate trajectory file:	simul
Energy values file:	simul
Output frequency:	10
Time step:	0.001
Temperature:	300.0

Resart Simulation From the Restart File

Click on "OK." In the "Dynamics Setup" dialog select "Done." To run the dynamics trajectories select "CHARMm dynamics.." from the "Modeling" palette. The structures are displayed as they are generated by CHARMm. Notice that the total time of the dynamics run is 9 psec. After the run is complete, choose "Save changes.." (so that you can later minimize this structure).

Now we want to see the simulation trajectories: pull down the Applications menu and choose "Dynamics Animation." In the "Dynamics Animation" palette, click on "Select trajectories..." Click on the "Initialize Dynamics Files" check box and then click on "OK." The File Librarian will be displayed; click on the file "simul.DCD," and then "Open." Click on "Exit" to return to

the "Dynamics Animation" palette. Next click on "Set Up Animation..." In the setup dialog verify the following settings:

•Use CHARMm Header	
Dataset range from:	6040
to:	9000
Step size:	40
Clock speed	1
Number Steps to Average Over:	0
•Regenerate hydrogen bonds	
•Display geometry monitors	
•Display trials	
•Do not display dipole	

Click on "OK." From the "Dynamics Animation" palette, select "Create Animation." To see the animation, click on "Cycle." The trajectory will cycle through all the time steps and then repeat. To change the speed of the animation, bring the "Dials" palette forward, and click on the vertical "speed" dial. Click on "Exit Dynamics Animation," to return to the main QUANTA palettes and window.

Minimize the current structure, to see the minimized structure that corresponds to the final dynamics conformation.

Analysis of Dynamics Trajectories

Does the dynamics trajectory find any new conformations that are significantly different from those you have already found? What do we mean by significantly different? We expect fluctuations about a local minimum energy conformation; these small changes in dihedral angles don't indicate a new conformation. The presence of a new conformation is shown by the trajectory traveling to and then fluctuating about a new local minimum, with distinctly different dihedral angles. These new angles, when energy minimized, give a new low energy structure. The presence of new conformations may be difficult to determine by watching the dynamics animation. The Analysis application can be used to produce a scatter plot of the ψ and ϕ angles during the trajectory calculation. From this plot you can determine if any significantly different conformations are part of the trajectory. You may already have used the Analysis application for looking at the results of conformational analysis for butane in an earlier chapter; the procedure here is very much the same as before.

1. Pull down the Analysis menu and choose "Analysis." A dialog box will appear asking, "Choose the Type of Input File." Choose "Dynamics File (.DCD)" and click on "OK." In the File Librarian, select "simul.DCD," and then "Open." The Analysis and Plots palette will be displayed.
2. Choose "Torsions..." in the Analysis palette. We need to define the ψ and ϕ dihedral angles for the Analysis plots. The "Torsions" palette will be displayed. Click on "Define Peptide Backbone Torsions." Click on "EXIT TORSIONS."
3. In the "Plots" palette select "Trace...". Choose the plotting and sorting property as the "Torsion Angle." Click on "OK." A dialog box will be displayed with the default torsion name "psi(1)" displayed. Click on "OK" to choose this as the plotted torsion. The trace plot will be displayed. To set the trace plot x-axis to 0 to 360°, pull down the Trace tools menu and choose "Set 360 deg Scale."

To see the structure that corresponds to a given point in the trace plot: Pull down the Trace tools menu in the Trace plot window and choose "Select Structure." Drag the Trace window to the right so that you can see your structure in the normal modeling window. Now when you double click in the Trace plot window at various angles, the corresponding structure will be displayed. To exit the "Select Structures" mode **Press the F1 key** at the top of the keyboard. Pull down File in the Trace plot window and choose "Quit."

To see the variation in energy with the phi angle; in the "Plots" palette select "Trace...". Choose the plotting and sorting property as the "Torsion Angle." Click on "OK." A dialog box will be displayed with the default torsion name "psi(1)" displayed. You must change this to read "phi(2)." Click on "OK" to choose this as the torsion. The Trace plot will be displayed.

4. To set the scatter plot x-axis to 0 to 360°, pull down the Trace tools menu and choose "Set 360 deg Scale." Also pull down the Trace tools menu and choose "Full Torsion Scale."
5. To see the structure that corresponds to a given point in the scatter plot: Pull down the Trace tools menu in the scatter plot window and choose "Select Structure." Now when you double click in the scatter plot window at various angles, the corresponding structure will be displayed. To exit the "Select Structures" mode **Press the F1 key** at the top of the keyboard. Pull down "File" in the Trace plot window and choose "Quit." Next click on "Exit Plots."
6. In the Analysis palette click on “lowest energy structure,” to load in the most stable structure from the dynamics run. Finally click on "Exit Analysis" in the Analysis palette.
7. In the Modeling palette click on CHARMm minimization repeatedly until the energy is minimized. Record the energy. Click “Hydrogen Bonds” twice to locate any hydrogen bonds.

Problem 10.1

Does the dynamics trajectory find any new conformations that are significantly different from those you have already found? What hydrogen bonds form to stabilize the structure? Are these hydrogen bonds different than in Chapter 9? Is there more or less motion than you expected? Compare the energy of the minimized structure from the end of the simulation with the $\psi = -60$ and $\phi = -60$ and 180,180 structure from Chapter 9.

Chapter 11: Solvation and β-Cyclodextrin

For small molecules, molecular orbital theory is more accurate and more useful than molecular mechanics. However, molecules with even as few as 20 atoms require large amounts of computer time for molecular orbital calculations. For large molecules or small molecules in a solvent, molecular mechanics is still the only practical computational technique. The real power of molecular mechanics calculations is in the ability to handle explicit solvent molecules. The conformation of a molecule can depend strongly on the presence of solvent. In addition, studies of molecular recognition and binding require careful consideration of solvent effects. In this chapter we study the interaction of water with the cyclic saccharide host, β-cyclodextrin.

Cyclodextrins are often used as active site analogs for enzymes[1]. Cyclodextrins are used to aid the absorption of drugs in the body. Other uses for cyclodextrins include the petroleum industry for separating aromatic hydrocarbons and in agriculture to reduce volatility of insecticides. Cyclodextrins are natural products produced by bacteria from starch. CD is made from seven D(+)-glucopyranose units linked through α-(1->4) glycosidic bonds[2], Figure1.

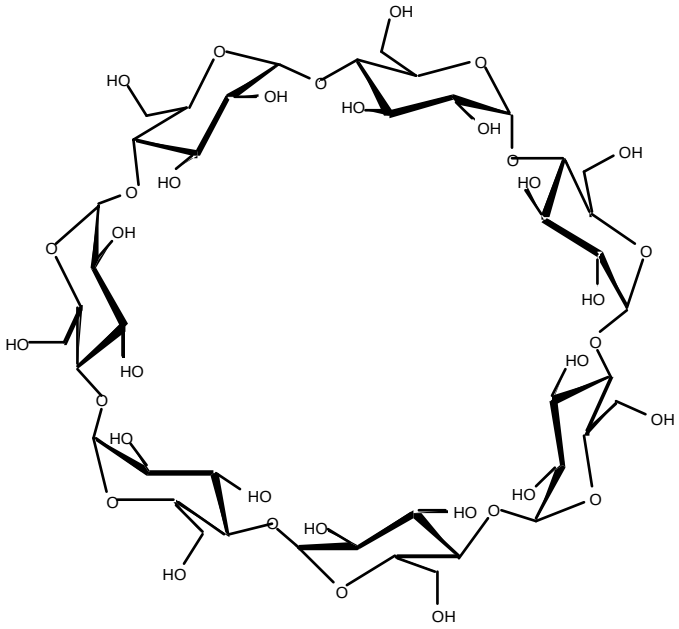


Figure 1. ß-cyclodextrin (cycloheptaamylose).

When dissolved in water, water molecules will fill the cavity of the host. Then when a guest interacts with the cavity of the host, water molecules are displaced. The binding affinity depends on the interactions of guest with the host and the difference in the interactions of bound water and water with the bulk solvent. The cavity of cyclodextrin holds around 11 water molecules. Most or all of these are excluded from the cavity upon binding with a guest. This process is called desolvation:

$$\text{host}\bullet 11\ H_2O \rightarrow \text{host(g)} + 11\ H_2O\ \text{(bulk)}$$

In addition, the guest interacts with water before it binds to the host, and these waters also must be released to the bulk of the waters in solution. The number of interacting waters is difficult to predict, so lets call the number y:

$$guest \bullet y\ H_2O \rightarrow guest(g) + y\ H_2O\ (bulk)$$

The change in Gibbs Free energy in this process is often unfavorable and is called the "desolvation" penalty. The net process is then:

$$host \bullet 11\ H_2O + guest \bullet y\ H_2O \rightarrow host \bullet guest + 11\ H_2O\ (bulk) + y\ H_2O\ (bulk)$$

In summary, solvation plays a very important role in molecular binding. In this chapter we will minimize β-cyclodextrin in aqueous solution and in vacuum to determine the number of waters in the cavity and any changes that occur upon solvation/desolvation.

Procedure:

Solvate β-Cyclodextrin: The β-cyclodextrin file is in the quanta home directory and is called bcyclodextrin.msf. Open this file using the "File" menu and "Open." Please rename this structure so that the original file is not changed by pulling down the "File" menu and choosing "Save As" Give a new name in the File Librarian dialog.

To add solvent, pull down the CHARMm menu, pull right on Solvate Structure and select " 15Å radius sphere." Choose CENTROID in the next dialog box to center the solvent sphere on your molecule. If the Saving dialog box appears, choose the MSF saving option "Overwrite." (Always choose "Overwrite" to keep the number of files down.) The aquated structure needs to be minimized as follows (alternately, instead of doing a minimization, if time is short your instructor may allow you to open a pre-minimized structure: bcyclodexaq.msf and you can skip to the next paragraph). Do a quick, rough minimization by pulling down the CHARMm menu and selecting "Minimization options.." Select "Steepest descents" for 50 iterations, energy gradient tolerance 0.01, and click OK. (See Chapter 1 for a quick discussion of these minimization options.) Minimize. Return to the "Minimization options.." dialog and select "Adopted Basis Newton Raphson " for 500 steps. Adopted Basis Newton Raphson is particularly good for very large systems. Minimize. Click on Save Changes in the Modeling palette.

Measure the diameter of the top and bottom of the cyclodextrin cavity, using the "Geometry" palette. See Chapter 3 for instructions. You can temporarily remove the waters from the display by pulling down the Draw menu, sliding right on Display Atoms, and choosing All except solvent. Average your values. You can then make your measurements. To display the waters, just return to Display Atoms and choose All Atoms.

Count the number of waters in the cavity. This is most easily done by adjusting the clipping plane, or using proximity tools to select waters in the cavity.

Clipping . To help you see things better, try reducing the clipping plane to avoid displaying water molecules that are in the foreground. To reduce the clipping plane use the Clip dial in the Dials palette. To get an idea of the dense nature of solutions and the close interactions of the solvent with the solute, pull down the Draw menu and slide right on Solid Models and choose Van der Waal's. To reset the clipping planes to their default distances, click on the Reset dial in the Dials palette.

Proximity Tools: Pull down the Display Atoms menu, slide right and choose Selection Tools. Next choose Proximity tools… In the Proximity Tools dialog box click on “Around Residues.” Next click on “whole residues.” This choice means that all the atoms in each water will be

selected, otherwise only close atoms will be selected. Next click on "Set Radius/Length," and type in 4.5 for the radius (i.e. half the diameter of the cavity). The units are Å. Click OK. Now rotate the cyclodextrin to locate a water molecule that is in the center of the cavity. Click on that water molecule. Several waters should now be colored red, showing that they are selected along with the cyclodextrin. These waters should fill the cyclodextrin cavity. If things didn't go well, you can always click on Undo, and try again. When you have waters in the cavity selected, click on Exit Proximity, and then click on Finish in the Display Atoms dialog to return to the main screen. Now only the waters in the cyclodextrin cavity should be visible. To get an idea of the dense nature of solutions and the close interactions of the solvent with the solute, pull down the Draw menu and slide right on Solid Models and choose Van der Waal's. To finish, in the Object Management window, click on the "No" in the Delete column to remove the Van Der Waal's surface.

Delete the Solvent Molecules: To remove the solvent from the actively displayed molecule on the screen, pull down the Edit menu, slide right on Active Atoms and choose All Except Solvent. Then Minimize. Measure the diameter of the top and bottom of the cyclodextrin cavity. Average your values.

Problem 11.1: Report the number of water molecules in the cavity. Report the change in the average diameter of the cyclodextrin cavity. Also report the change in average diameter as a percentage, i.e. give a statement like "the cavity enlarged by ~5% with waters present."

Literature Cited

(1) Furuki, T.; Hosokawa, F.; Sakurai, M.; Inoue, Y.; and Chûjô, R. *J. Am. Chem. Soc.* **1993**, *115*, 2903-2911.
(2) Diaz, D.; Vargas-Baca, I.; and Garcia-Mora, J., *J. Chem. Ed.* , **1994**, *71*, 708.

Rulers in QUANTA and β-Cyclodextrin

A ruler is available to measure distances in QUANTA. These rulers are useful if you want to know distances between Van der Waals radii instead of distances between nuclei.

β-Cyclodextrin: The β-cyclodextrin(aq) file is in the CH141 directory and is called bcyclodextrin.msf.

Clipping: The β-cyclodextrin structure has been minimized including solvent water molecules to give more realistic conformations. You can count the number of water molecules in the cyclodextrin cavity. To help you see things better, try reducing the clipping plane to avoid displaying water molecules that are in the foreground. To reduce the clipping plane use the Clip dial in the Dials palette. To reset the clipping planes to their default distances, click on the Reset dial in the Dials palette.

Delete the Solvent Molecules: To remove the solvent from the actively displayed molecules on the screen, pull down the Edit menu, slide right on Active Atoms and choose All Except Solvent.

Ruler: To display the ruler pull down the Draw menu and select Ruler. Use the Dials palette to enlarge, move, and reorient the ruler. When the ruler is in a good spot click on Finish in the Ruler palette.

To display Van der Waals radii, you can choose either Dots or Solid Spheres, as you wish.

Dot Van der Waals Spheres: To add dots, select Dot surface in the Modeling palette. In the next dialog click on OK. In the Select Atoms palette select "all atoms." Click on Finish.

Remove Dots: Select Dot surface from the Modeling palette. In the Dots dialog click on Delete Existing Dots and then click on OK.

Solid Van der Waals Spheres: Pull down the Draw menu and slide right on Solid Models and choose Van der Waal's.

Remove Solid Spheres: Click on the "No" under the Delete header in the Object Management box in the lower right hand portion of your screen.

Chapter 12: Docking: β-Cyclodextrin and β-Naphthol

In Chapter 14 we discussed some of the influences of the solvent on guest-host binding. In this chapter we will focus on the interactions of the guest and host. We want to find the conformation and energies of the interaction of the guest and host. β-Cyclodextrin and β-naphthol form a guest-host complex.

ß-Naphthol is representative of a wide variety of guests, Figure 1. Many compounds have the same bifunctional nature. ß-Naphthol is expected to bind to CD because it has a hydrophobic group that fits into the cyclodextrin cavity, while the OH group participates in hydrogen bonds with the sugar OH groups. The reaction stoichiometry is 1:1.

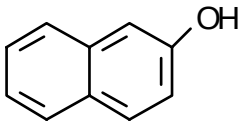


Figure 1 ß-Naphthol

We will use the docking features of QUANTA to predict the conformation of the complex. Docking monitors the Van der Waals interactions of the guest and host as you change the position and orientation of the guest in the cavity of the host. We will then do a full molecular mechanics minimization do estimate the binding energy for the guest-host interaction.

Procedure:

1. **β-Cyclodextrin:** The β-cyclodextrin(aq) file is in the CH141 directory and is called bcyclodextrin.msf. Open this file using the "File" menu and "Open."

2. **Delete the Solvent Molecules:** To remove the solvent from the actively displayed molecule on the screen, pull down the Edit menu, slide right on Active Atoms and choose All Except Solvent.

3. **β-Naphthol**: The β-Naphthol file is in the CH141 directory and is called b-naphthol.msf. Open this file using the "File" menu and "Open." When you use the "Open" dialog box, make sure to click on "Append," which is at the bottom of the dialog box. Please rename these structures so that the original files will not be changed by pulling down the "File" menu and choosing "Save as…" Give a new name in the File Librarian dialog.

4. In the "Modeling" palette click on "Move Fragment." Click on an atom in β-naphthol near the center of the molecule. A new "Dial Box" will appear at the lower right hand side of the screen. You can use these controls to move the guest around in the host cavity. Alternately you can use the mouse, which is probably more convenient. You can move the guest by holding down the "Ctrl" key. The mouse keys are then:

> "Ctrl" alone: x-y translation
> "Ctrl"+middle button: x-y rotation
> "Ctrl"+right button: z-rotation

Turn the guest around so that the OH group is close to the OH's that are directly attached to the rings. Center the guest in the cavity. You can switch between moving the guest only and moving both the host and guest together by either holding down the "Ctrl" key or not.

5. **Docking:** To turn on manual docking, click on "Continuous", "Energy", and "Bumps" in the "Modeling" palette. Now as you move the guest around, the Van der Waals energy will be

continuously displayed in the upper right hand corner of the modeling window. In addition, close contacts will be labeled on the screen. Position the guest to get as low an energy as possible. Pay close attention to the possibility of hydrogen-bonding between the guest and the host. Minimize and record the total steric energy, also click on CHARMm Energy and record the energy contributions listed in the TextPort window.

6. **Auto-Docking**: There is also an automated docking procedure. For β-cyclodextrin and β-naphthol, you need a pretty good starting geometry for auto-docking to be successful. To start the automated procedure, click on "Sensitivity" in the "Modeling" palette. Set the "Translation" sensitivity to 0.01 and then click OK. In the next dialog box make sure both "Translation" and "Rotation" are checked, and then click OK. Now click on "Solid Docking" in the "Modeling" palette. The program will seek to minimize the Van der Waals energy of the complex. Turn off the automatic calculation by clicking on "AutoDock" again. If the automated procedure wasn't successful try another starting geometry and click on "AutoDock" again. You can also "tug" the guest around while Solid Docking is active, by holding down the "Ctrl" key as before. However, the tight fit of β-cyclodextrin and β-naphthol makes it difficult to "tug" the β-naphthol into the cavity. Solid Docking may or may not be successful. Only Van der Waals (Lennard-Jones) terms are used for this minimization. Therefore, if electrostatic interactions or hydrogen bonding is important for the staility of your complex, Solid Docking may not find a reasonable conformation for the complex.

7. Alternately load in just β-cyclodextrin and just β-naphthol by themselves, minimize and record the total steric energy, also click on CHARMm Energy and record the energy contributions listed in the TextPort window. **PLEASE**: Don't "Save Changes" to avoid changing the original files.

Problem 12.1: Describe the conformation of the guest in the cavity of cyclodextrin. Find the binding energy for the complex by taking differences in the total steric energy of the reactants and product. Also, find the differences in each of the contributions to the steric energy. Which contributions to the steric energy dominate the binding interaction (eg. Bond stretch, bond bending, Van der Waals (Lennard-Jones), electrostatic (Coulomb), etc.).

Chapter 13. FEP and Henry's Law Constants and Gibb's Free Energy of Solvation

Read Section 7, Free Energy Perturbation of the Introduction to Molecular Mechanics

Introduction

The fate of organic molecules in the environment is determined in part by their solubility in water. For example, an oil spill or a leaking underground gasoline tank introduce organics into surface and ground water. The long term damage done to the environment is determined by the solubility of the organic contaminants in the water[1,2]. Soluble organics can travel long distances and allow the spread of the contamination over wide areas. Less soluble organics quickly evaporate and cause less of a problem. Henry's Law governs the solubility of compounds in dilute solution[3]:

$$P_B = X_B K \qquad 1$$

where P_B is the partial pressure of dilute solute B above the solution, X_B is the mole fraction, and K is the Henry's Law constant for B. Compilations of K values are limited; many thousands of compounds are of concern in the environment and in the laboratory. The purpose of this Chapter is to calculate K values from Free Energy perturbation studies (FEP). If FEP calculations are successful, considerable time, effort, and money can be saved in screening compounds for their environmental hazards.

We wish first to establish the connection of the Henry's Law constant to the Gibb's Free Energy of solvation. The equilibrium described by Eq. 1 can be written as:

$$B\ (aq)\ -\ >\ B\ (g) \qquad 2$$

The equilibrium constant for reaction 2 is:

$$K_{eq} = \frac{P_B}{X_B} \qquad 3$$

if we measure the concentration of B in mole fraction. Comparing Eqs. 1 and 3 shows that the equilibrium constant K_{eq} is the same as K, the Henry's Law constant. The Gibb's Free Energy change for Eq. 2 is the Gibb's Free energy of solvation, $\Delta_{sol}G_B$. Therefore, since K is the equilibrium constant for Eq. 2:

$$\Delta_{sol}G_B = -\ RT \ln K \qquad 4$$

Therefore, FEP calculations of $\Delta_{sol}G_B$ can be directly used to find Henry's Law constants.

The units of K as defined above are in atm. We will call this constant K_{PX} to help us remember the units. Environmental chemists often prefer to deal with unitless Henry's law constants, K_{CC}, where the gas phase pressure is replaced by the gas phase concentration and the solution mole fraction is replaced by the concentration:

$$K_{CC} = \frac{K_{PX}}{kRT} \qquad 5$$

where k is the conversion factor from mole fraction to concentration in mol L^{-1}. Here, k= 1000 mL x d_{H2O} / M_{H2O}, where d_{H2O} is the density of water and M_{H2O} is the molar mass of water. At 25°C, k = 55.35 mol L^{-1}. Also common in the literature is the Henry's Law constant with pressure for the gas phase and concentration for the aqueous phase:

$$K_{PC} = \frac{K_{PX}}{RT} \qquad 6$$

Table 1 lists values for the various K's and the $\Delta_{sol}G$ values derived from them using Eq. 4. The proper parameters for comparison with FEP calculations are K_{CC} and $\Delta_{sol}G_{CC}$.

Table 1. Henry's Law constants and Free Energies of solvation[1]. The number in parenthesis is the source for that substance and following values. $\Delta_{sol}G$ = - RT ln K. The units are indicated as subscripts: p=pressure, x=mole fraction, and c=molarity.

substance	Kpx	K_{CC}	Kpc	$\Delta_{sol}G$px
	Units: atm	unitless	atm /M	kJ/mol
benzene(8)	294.	0.216	5.32	-14.1
toluene	358.	0.263	6.47	-14.6
ethylbenzene	433.	0.318	7.83	-15.1
m,p-xylene	406.	0.298	7.34	-14. 9
o-xylene	278.	0.204	5.02	-13.9
1,1,1-trichloroethane	978.	0.718	17.7	-17.1
trichloroethylene	572.	0.42	10.3	-15.7
tetrachloroethylene	950.	0.697	17.2	-17.0
methyl-t-butyl ether	29.4	0.0216	0.532	-8.38
tetrachloroethylene(9)	985.	0.723	17.8	-17.1
trichloroethylene	534.	0.392	9.65	-15.6
1,1-dichloroethylene	1457.	1.069	26.3	-18.1
cis-1,2-dichlororethylene	228.	0.167	4.11	-13.5
trans-1,2-dichloroethylene	523.	0.384	9.45	-15.5
vinylchloride	1549.	1.137	28.0	-18.2
1,1,1-trichloroethane	958.	0.703	17.3	-17.0
1,1-dichloroethane	313.	0.23	5.66	-14.2
chloroethane	621.	0.456	11.2	-15.9
carbon tetrachloride	1695.	1.244	30.6	-18.4
chloroform	204.	0.15	3.69	-13.2
dichloromethane	122.	0.0895	2.20	-11.9
chloromethane	492.	0.361	8.89	-15.4
methane (10)	413.	0.303	7.46	-14.9
oxygen	4.34E+04	31.84	784.	-26.5

Free Energy Perturbation Calculation of K_{CC}

Before we use FEP on a new system, we should determine the accuracy of FEP methods by running some known compounds and comparing with literature values in Table 1. Let's work through an example of a FEP study. The comparison we will discuss is the mutation of 1,1,1-

trichloroethane to 1,1-dichloroethane. This mutation is a relatively simple one where a Cl atom is mutated to a H atom. However, we also need to adjust the charges on the attached C atom. We choose 1,1-dichloroethane as our reference structure, and from that reference calculate $\Delta_{sol}G$ of 1,1,1-trichloroethane. From Table 12.1, we expect $\Delta\Delta G = \Delta_{sol}G$(1,1,1-trichloroethane)-$\Delta_{sol}G$(1,1-dichloroethane) = 0.87 - 3.64 kJ/mol = -2.77 kJ/mol or equivalently -0.66 kcal mol. To calculate $\Delta\Delta G$ we need to run two FEP studies, one Cl -> H mutation in the gas phase to determine $\Delta G^{g}_{B->A}$ and the same Cl -> H mutation in the aqueous phase for $\Delta G^{aq}_{B->A}$. The difference in Eq. 14 gives $\Delta\Delta G$.

Perturbation Setup Files

For CHARMm there are currently three setup files for FEP studies. The first: "pert1to2" manages the perturbation between two structures produced by QUANTA. This file can be used for atom changes, for example Cl to H, or for dihedral angle changes, for example axial to equatorial or cis to trans. In general, "pert1to2" can be used whenever both structures have the same number of atoms and the mutation is not too large. The second setup file, "pertCltoH," mutates the chloro group to a H atom. To use "pertCltoH" you must build your structure in ChemNote starting with the chlorine to be mutated as the first atom. The reverse process is handled by "pertHtoCl." Here the hydrogen to be mutated must be the first atom.

Neither of these FEP methods use the SHAKE option. Recent studies have shown that results are more accurate without SHAKE. You should always start dynamics runs with a minimized structure, so don't forget to minimize before calling these files.

Procedure

In this example, we run the mutation discussed above. However, these instructions will work for any Cl->H mutation. The procedure we use follows the outline:

1. Build and minimize 1,1,1-trichloroethane.
2. Copy the structure files to new files named PERT1.
3. Run the gas phase FEP.
4. Solvate 1,1,1-trichloroethane and minimize using periodic boundary conditions.
5. Copy the solvated structure files to new files named PERT1.
6. Run the aqueous phase FEP.

The detailed instructions follow.

1. Build and minimize 1,1,1-trichloroethane: Pull down the Applications menu, slide right on "Builders" and choose "2D Sketcher." Build 1,1,1-trichloroethane with a chlorine as atom 1 by clicking on the single bond icon and placing the 4 C-C or C-Cl bonds in the window. Remember the bond that you first drew. You should now have a structure that is similar to a "+". Now click on the Cl atom icon and click on three of the ends of the "+" starting with the first bond that you drew. Pull down the File menu and choose "Return to Molecular Modeling." "Save changes" and choose the default smoothing option, i.e. all carbon atoms and non-polar hydrogens. In the QUANTA window, pull down the CHARMm menu, slide right on "CHARMm Mode", and choose "PSF." Pull down the CHARMm menu and choose "Minimization Options." Change the method to Adopted Basis Newton Raphson. This minimization method works well for large systems. Choose "800 steps", energy gradient tolerance 0.0001 and energy value tolerance "0." Click on "OK." Now we need to set-up to do a constant pressure dynamics run. To operate at constant pressure, we need to specify periodic boundry conditions. In this way we don't have to worry about what happens at the sides of the box. Mirror images of the box will be stacked next

to each other in all directions so that there are no surfaces to the solution. Pull down the CHARMm menu and choose Periodic Boundaries. In the new dialog box click on "Turn ON Periodic Boundries." Also, click on "Enter Lattice Constants." Then enter the lattice parameters to give a cubic box as follows:

a: 15
b: 15
c: 15
alpha: 90
beta: 90
gamma: 90
orthogonalization code: 1

and then click on OK. Minimize the structure: remember to click on "CHARMm Minimization" repeatedly until the energy no longer changes. In the Modeling palette, choose "Save Changes," and select "Overwrite ____." Make sure that one of your chlorines is atom 1. To do this, use the mouse to click on the chlorines. One of them should be listed as Cl1. If not repeat this step, making sure that the first atom you draw in ChemNote is a Cl. Now stop the current CHARMm session by pulling down the CHARMm menu, pull right on CHARMm Process, and select "Terminate Interactive."

2. Copy the structure files to new files named PERT1: Pull down the File menu and choose "Open System Window." In the system window type: mkpert1.

3. Run the gas phase FEP: Type: pertCltoHp to begin the FEP dynamics run. At the conclusion, about four minutes later, type: jot p.out; the "jot" application will open and display the final results. The output file has many details, however at this point you only need record the final total $\Delta G_{B->A}$ value. Start at the end of the file and scroll backwards. Find the heading "perturbation energy total=". The result is labeled "Parameter: 1 <-". Here is an example from a similar run with the final energy of 1.35753 kcal/mol (your number will be different):

```
CHARMM>    ! perturbation energy total=
CHARMM>    set 1 ?SLTOT
RDCMND substituted energy or value “?SLTOT” to “1.35753”
Parameter: 1  <-  “1.35753”
```

Also record the hysteresis energy correction, which follows similarly on the succeeding lines. This energy is a correction that is applied to your results to account for the estimated numeric difference on running the perturbation forwards and backwards. It is an approximation in the uncertainty in your perturbation energy total.

Pull down the File menu in "jot" and choose "Quit." At this point you can close the system window by typing: exit, or you can minimize the window using the minimize button on the title bar of the system window.

4. Return to the 1,1,1-trichloroethane and solvate and minimize: Pull down the CHARMm menu, slide right on "Solvate," and choose "15Å length box water." Next you will be asked how you want to center the water solvation box, choose "Centroid." Back in the QUANTA screen

choose "Save Changes" from the Modeling palette. Choose "Overwrite ___," and then "OK." Now we need to set-up to do a constant pressure dynamics run. Pull down the CHARMm menu and choose Periodic Boundaries. In the new dialog box click on "Turn ON Periodic Boundries." Also, click on "Enter Lattice Constants." Then enter the lattice parameters to give a cubic box as follows:

a: 15
b: 15
c: 15
alpha: 90
beta: 90
gamma: 90
orthogonalization code: 1

and then click on OK. Pull down the CHARMm menu and choose "Minimization Options." Set the method to Steepest Decent. Change the number of minimization steps to "100," and energy gradient tolerance "0.001," and then click "OK." Minimize your structure. Pull down the CHARMm menu and choose "Minimization Options." Set the method to Adopted Basis Newton Raphson. Change the number of minimization steps to "800," and energy gradient tolerance "0.001," and then click "OK." Minimize your structure. The minimization will require about 20 minutes. We set the number of steps to 800, so that you won't have to repeatedly click on "CHARMm Minimization." Make sure your structure is minimized to better than 0.1 kcal. Choose "Save Changes" from the Modeling palette, "Overwrite_____," and click "OK." Now stop the current CHARMm session by pulling down the CHARMm menu, pull right on CHARMm Process, and select "Terminate Interactive."

5. Copy the solvated structure files to new files named PERT1: If you closed the system window in step 5, pull down the File menu and choose "Open System Window." In the system window type: mkpert1.

6. Run the aqueous FEP:: Because the aqueous run takes much longer we want to do the CHARMm run in the background. This way others can use the computer if need be. To run a job in the background we just follow the command with an "&." Type: "pertCltoHp &" to begin the FEP dynamics run. At the conclusion, about six hours later, type: jot p.out; the "jot" application will open and display the final results. Record the final total $\Delta G_{B->A}$ value. Pull down the File menu in "jot" and choose "Quit." Close the system window by typing: exit. You can view the p.out file at any time, as long as no one else has done a perturbation calculation in the mean time.

Problem 13.1

Calculate $\Delta\Delta G$, using Eq. 14. Calculate a rough uncertainty using the EHYS values. Compare your $\Delta\Delta G$ with the literature value. Since these numbers are expected to be small, the % error will be large; instead just report the difference. Is the sign correct on your value? Using Eqs. 15 and 19, calculate the K for your compound using the $\Delta_{sol}G$ literature value for 1,1-dichloroethane (or whatever reference compound you chose). The % error is expected to be large in this final value. But at least your result should be in the correct direction. To shorten the length of the simulation, we chose too small a number of dynamics steps for each time slice (150). Better results should be obtained with longer simulation times at each time slice.

Literature Cited

1. G. A. Robbins, S. Wang, J. D. Stuart, *Anal.Chem.*. **1993**, *65*, 3113.
2. J.M. Gossett, *Environ. Sci. &Tech.*, **1987**, *21*, 202.
3. P. W. Atkins, *Physical Chemistry, 5th. ed.*, W. H. Freeman,Co, New York,1994.Table C18.

Chapter 14 Distance Geometry

Please read Section 6 “Distance Geometry” in the Introduction

The rings in progesterone can take many conformations. Building a model of progesterone with the proper ring conformations can be difficult. Applying constraints based on NMR data can greatly ease the construction of the proper conformation of the molecule. Distance geometry is an ideal technique for constructing molecules subject to constraints. The purpose of this exercise is to compare distance geometry with template based construction methods. We will use the EMBED program for distance geometry and the CORINA program for the template based approach, with access from the Web for both applications. The structure of progesterone with the conventional atom numbering is shown in Figure 1. The X-ray structure from the Cambridge Structure Database, CSD, is shown in Figure 2, to indicate the proper ring conformations.

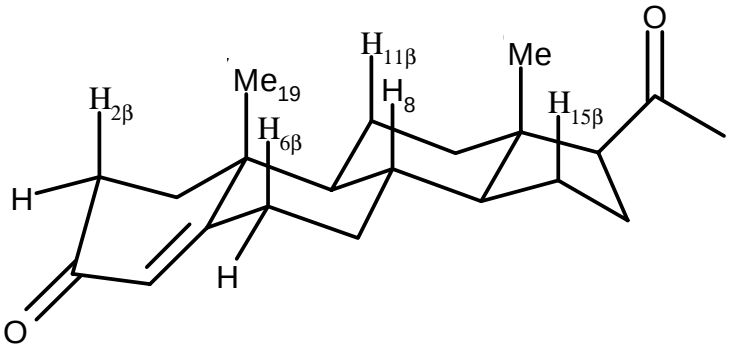


Figure 1. Progesterone

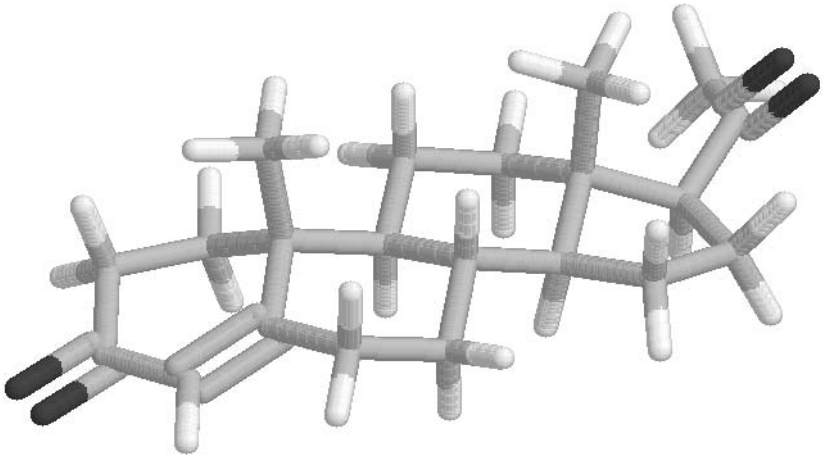
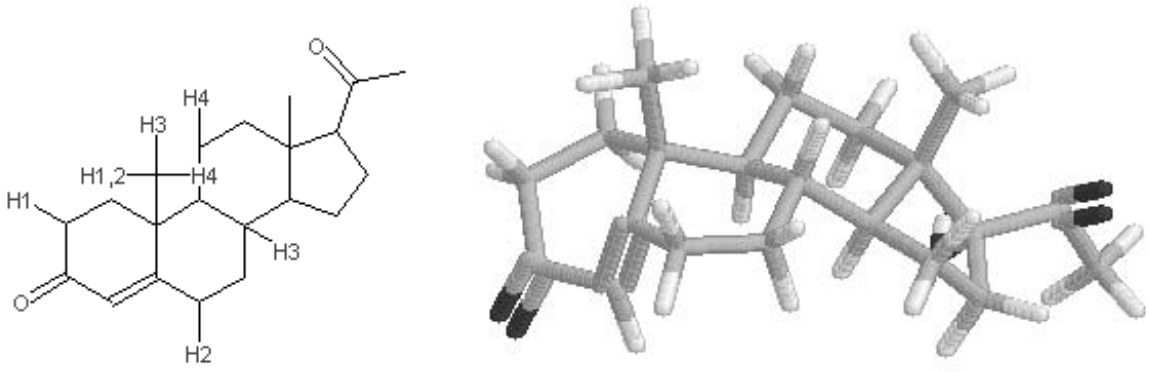


Figure 2. X-ray structure for progesterone from the CSD.

The 19-methyl of progesterone is reported to have short-range nOe couplings with four hydrogens on one face of the molecule[1], $H_{2\beta}$, $H_{6\beta}$, $H_{11\beta}$, and H_8. These nOe constraints are shown as input using the JME applet for EMBED, Figure 3a. A successful EMBED structure is shown in Figure 3b. Of course distance geometry is most useful when you don’t have an X-ray structure. But for this beginning exercise comparison with the X-ray structure is very instructive.



a. b.

Figure 3: Progesterone with short nOe distance constraints. (a) Distance constraints as input using the JME applet. The distance ranges were specified as 2.5-3.5Å. (b) EMBED results. The Smiles notation is CC(=O)C3CCC4C2CCC1=CC(=O)CCC1(C)C2CCC34C.

Instructions

Distance Geometry: The distance geometry structure will be built using the "Distance Geometry with MM2 and SCF" applet. This applet uses the Java Molecular Editor for structure input. You will find the link for this Web applet on the PChem home page.[1] Instructions for adding constraints are given below the JME editor window.

Building Progesterone:

To build progesterone, start by clicking on the cyclohexane ring button and then clicking in the molecule window to add the three fused cyclohexane rings. With the ring buttons, clicking on an existing ring bond adds a fused ring and clicking on a ring atom adds a spiro ring. If you make a mistake click on the UDO (undo) button. Click the cyclopentane ring and add it to the right-most cyclohexane. Next select the single bond button, ▭ . Drag the single bonds as shown at right. Clicking on a bond will change it to a double bond. Finish by putting in the oxygens.

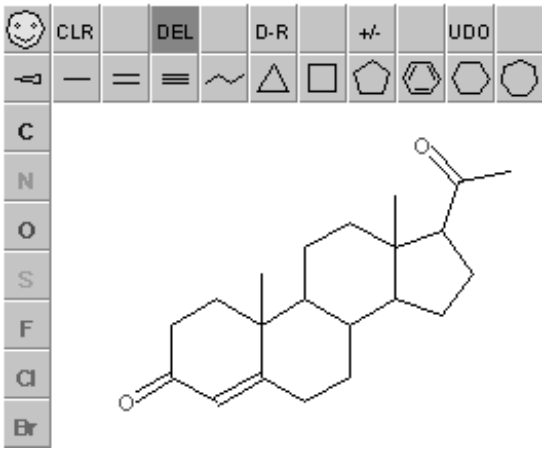


Figure 4. Progesterone in JME.

Click the Help button for more information on using the Java Molecular Editor. You would be finished at this point if you didn't have any experimental information. We now need to add the nOe based distance constraints. Use the single bond button, ▭ to add bonds for the seven hydrogens as shown in Figure 3a. Use the [X] tool to set in the element names and constraint numbers. For example for H1, click on the [X] button, enter H1, and then click on the end of the bond at the H1 position. (For Macitosh systems make sure to highlight the entire dialog box when you change the label in the X tool, otherwise part of the old label may remain.)

When you have added all the atom labels, scroll down to the Distance Constraints section and click on the "Short" button for constraints 1-4. The typical distance range for a strong nOe will be entered automatically for you, as in Figure 5.

Figure 5. Select Short nOe distance constraints for all four distances.

Scroll down until you see the options button. The Add hydrogens button should be selected. The "SCF pi calculation" option won't make any difference since progesterone has no conjugated double bonds. The "Set-up Gaussian 98" input button is used when the applet is run to prepare the input files for molecular orbital calculations, which we will ignore for now. Click Submit to start the calculation.

Write down the FINAL STERIC ENERGY and compare your results with Figure 2 and 3b. You will probably need to repeat the calculation several times to get the proper ring conformations. Just use the browser Back arrow and click on Submit again. Keep track of the FINAL STERIC ENERGY for each run. You should get the proper ring conformation for the rings around your distance constraints within two or three calculations. Getting the other ring conformations correctly requires a few more calculations.

In preparation for your CORINA calculations, return to the JME page. Click on the Delete button and remove the hydrogens from your structure. Your structure should now look again like Figure 4. Press the yellow "smiley face" button. A dialog box will appear with the Smiles string that corresponds to progesterone. Highlight and copy this Smiles string to the clipboard.

Template Based Builders: The CORINA program is available on-line through the University of Erlangen-Nürnberg at J. Gasteiger's Group Web site. A link to this Web site is on the PChem Home Page. On this page scroll down to the JME applet. Paste your Smiles string into the Smiles string dialog box and click on the Generate 3D Structure button. (Alternatively you can draw in the structure using the JME editor.) Compare the CORINA generated structure to Figure 2. How many centers are inverted from the X-ray structure? Generating the structure a second time isn't useful, since template based builders produce the same result every time. To get all the rings correctly requires the stereochemistry to be included in the Smiles string.[4] (Don't bother to do the calculation with the stereochemistry specified.)

MOE can also interpret Smiles strings. If you have time try inputting the progesterone Smiles string into the Builder Smiles dialog box.

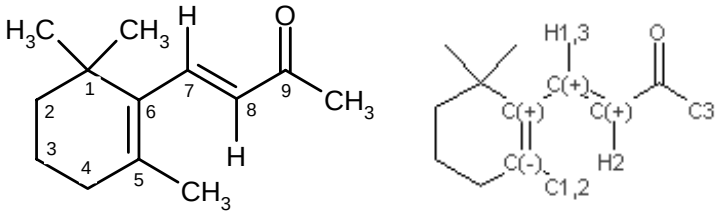
Summary

The ring conformations of progesterone highlight the advantage of distance geometry. Template based model builders like CORINA give the same result every time. Distance geometry allows the various ring conformations of progesterone to be studied because each calculation has the potential to give a different result. Which is the better approach? CORINA has been highly optimized to give structures that are very close to X-ray structures. So if you can, using CORINA is a better approach. However, distance geometry is the best choice if you don't know the proper conformation and need to generate several options, and distance geometry is especially useful if distance constraints are available from experiment.

Problem 14.1

a. How may calculations were required to get the ring conformations for progesterone correctly around the nOe constraints?
b. How many more calculations were necessary to get all the ring conformations correctly?
c. Was the X-ray conformation (Figure 2) the lowest energy conformation?
d. How many centers are inverted in the CORINA structure from the X-ray structure?
e. Without looking at the structure of progesterone, use the Smiles string form the caption of Figure 3 to draw the structure of progesterone. Give you best effort, but don't cheat and compare to the Figures.

Problem 14.2 Build β-Ionone, Figure 8, without using constraints and do the MM2 calculation. Make sure the SCF pi calculation is not selected. What conformation and steric energy do you get? Now use your nOe constraints from your NOESY spectra. However, make sure to use the applet that only does the distance geometry calculation[5] (not the MM2 version). Some of these possible nOe constraints are shown as input using the JME applet, Figure 8b.[6] In Figure 8b, the constraints are shown to the carbon atoms since the nOe distances are average distances and the average position of a methyl group hydrogen is near the corresponding carbon atom. You can also give distance constraints to explicit hydrogens as we did above. You can also use constraints to the methyl groups attached to the top of the ring. Do several repeat calculations and compare the side chain conformations.



a. b. c.

Figure 8: β-Ionone with short nOe distance constraints. (a) Atom numbering, (b) Distance constraints as input using the JME applet.

Literature Cited:

1. Williams, D. H., Fleming, I., Spectroscopic Methods in Organic Chemistry, 4th Ed, McGraw-Hill, London, UK, 1987. p 118.

2. Distance Geometry and MMP2: http://iris12.colby.edu/~www/jme/dgmmp2.html (accessed Jan 2003).

3. Gasteiger, J. Research Group, *Corina*, http://www2.chemie.uni-erlangen.de/software/corina/free_struct.html (accessed Jan 2003).

4. [H]C1CC2[C@]3([H])CC[C@H](C(C)=O)[C@@]3(C)CC([H])[C@]2([H])[C@@]4(C)CC([H])C(=O)C=C14

5. Distance Geometry: http://iris12.colby.edu/~www/jme/dg.html (accessed Jan 2003).

6. Honig, B.; Hudson, B.; Sykes, B. D.; Karplus, M., "Ring orientation in β-ionone and retinals," Proc. Nat. Acad. Sci. U. S., **1971**, *68(6)*, 1289-93.

Chapter 15. Protein Structure and Gramicidin-S

One of the most active and interesting areas in bio-physical chemistry is the study of protein structure. The problem is simply this: given the uncountable number of possible conformations for a protein, how can we determine the lowest energy structure. In this exercise we tackle a relatively simple problem, which retains the flavor of the more complicated problems under current study. We will model the structure of the antibiotic gramicidin-S.

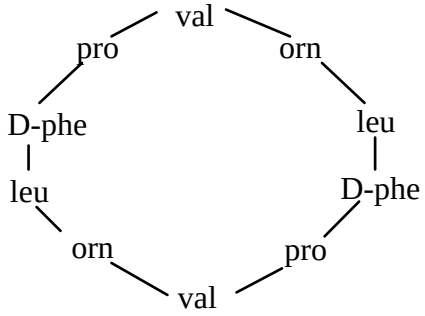
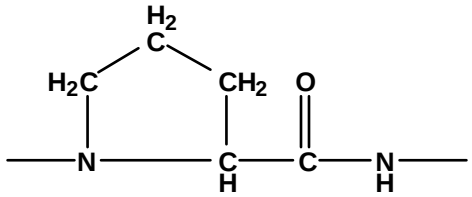


Figure 1 Gramicidin-S

Gramicidin-S is a cyclic decapeptide, Figure.1, produced by the soil fungus *Bacillus Brevis*. The protein is unusual for several reasons. First it includes D-phenylalanine, rather than the normal L-isomer. Secondly, the unusual amino acid ornithine is used. Thirdly, the protein is hydrophobic. Most proteins have a hydrophilic exterior, to enhance interaction with water, and a hydrophobic interior. Gramicidin-S has just the opposite. Its hydrophobic exterior suggests that the mode of action is through a strong membrane interaction[1]. The linear gramicidins form ion channels in cell membranes.

Even a small peptide like gramicidin-S has too many possible conformations for each conformation to be exhaustively studied. To find the global minimum structure, we must rely on experimental information and some intuition. The NMR spectrum shows that gramicidin-S is symmetrical; the like-amino acid pairs have the same chemical shifts. Therefore, we really only need to worry about five amino-acids, the other five are related by symmetry. In our modeling we must make sure that this symmetry is maintained. In lab you will be determining NMR constraints on the dihedral angles for some of the amino-acids. The spin-spin J coupling between the α -CH and the backbone NH proton is about 4 Hz for a alpha-helix type structures and 9Hz for beta-pleated sheet structures[2]. The presence of alpha-helical or beta-pleated sheet type-regions will help to constrain our modeling. Of course just a few monomers with the proper dihedral angles aren't sufficient to establish a "real" alpha-helix or beta-pleated sheet, but the NMR dihedral constraints can be used to point us in the right direction for molecular modeling.

We can also use some intuition. Prolines have a cyclic structure that is formed by the side chain and the backbone N, Figure 13.2a. Prolines often occur at turns, because of the kink caused by the cyclic structure, and proline can't assume the backbone dihedral angles necessary for alpha-helix or beta-pleated sheet structures[3]. QUANTA/ CHARMm has two types of proline based secondary structures, which can be used to establish proline turns. The proline-1 secondary structure is based on the unusual formation of *cis*- peptide bonds, Figure 13.2b.



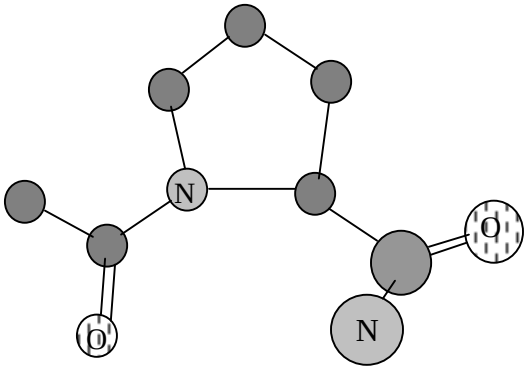


Figure .2 a. Proline in a protein

b. Proline in a proline-1 structure based on *cis*-peptide linkages. (H's not shown.)

Since complete turns require two amino acids, only three residues remain in the "body" of the protein for us to worry about.

The average backbone angles for some regular secondary structures are shown in Table 1.

Table 1. Dihedral Angles for Regular Secondary Protein Structures[3]

	Bond Angle (degrees)		
	ϕ	ψ	ω
Antiparallel β–sheet	-139	+135	-178
Right-handed α-helix	-57	-47	180
Polyproline I	-83	+158	0
Polyproline II	-78	+149	180

Applying Constraints

Experimental constraints are necessary to help us narrow down the number of possible conformations for proteins. Constraints may be applied on dihedral angles or on distances. For example, an alpha-helix ϕ dihedral should be about -60° and a beta-pleated sheet ϕ dihedral should be about -140°. As mentioned above, we can measure these dihedral angles using spin-spin J coupling constants and then use these measured values as constraints.

Distance constraints can be determined from nOe measurements. The normal value for nOe based constraints is 3.0Å. QUANTA uses 3.0Å as the default distance constraint. Distance constraints can also be inferred from secondary structure assignments. Examples for such inferences include N-H···O=C hydrogen bond distances. For alpha-helices, strong hydrogen bonds form between residue i and residue i+4. For beta-pleated sheets, N-H···O=C distances between strands can be constrained. Hydrogen bond lengths are in the range from 1.8-3.0Å, with 2Å being normal for strong hydrogen bonds[4]. The hydrogen bond distance between residues i and i+4 in the alpha-helix is about 1.86Å[4], and about 1.96Å between beta-pleated sheet strands. Comparing dihedral and distance constraints, distance constraints limit the conformational flexibility of the molecule more, and are preferable if known.

Instructions

Using the Sequence Builder to set the Sequence and Secondary Structure

In this section, you will find out how to specify the sequence for your protein,. You will also find out how to change L-amino acids to D- amino acids and how to make cyclic proteins by applying patches. You will also specify secondary structures like alpha-helical or beta-pleated sheet regions and specify turns.

1. Pull down the Applications menu, slide right on "Builders," and choose "Sequence Builder." In the File Librarian you must choose the "RTF" file for the type of polymer you want to build. Click the "AMINOH.RTF" entry and then "OK." (RTF stands for residue topology file.)
2. In the Sequence Builder, you click on the amino acids in your structure in the order that they appear in the protein. Start with ORN and continue around the ring, Figure 13.1.
3. Next we need to apply "patches" to phenylalanine to convert to the D-stereo isomer. Pull down the Edit menu and choose "Apply Patches to Residues."
4. Click on the "LTOD" button in the left hand portion of the screen, and then click on one of the PHE's in your structure. Then click on "LTOD" again, and then click on the other PHE.
5. Next we need to set the secondary structure, if you know it. If you don't have guesses for the secondary structure proceed to step 6. From your NMR spectra, determine the type of secondary structure that you have. Hint: two of the amino acids will have the same dihedral angles. Use these two only. Leave the third to be determined by the minimization. Pull down the

Conformation menu and choose "Set Secondary Conformation." Select the two amino acids that you have determined and then on "OK." A scroll box will appear, choose the secondary structure for these two residues, and click "OK." Repeat this process for the corresponding residues on the other side of the ring (remember that the structure must be symmetric).
6. We need to specify turns for the prolines. Select the first PHE and PRO and click "OK." In the scroll box choose PROLINE 1, and click on "OK." Now select the second PHE and PRO pair, click "OK," and the select PROLINE 1 again. Click on "DONE" to return to the main screen.
7. Now we need to create the cyclic structure. Pull down the Edit menu and choose "Apply Patches to Residues." This time click on "LINK" and then the first ORN and then the last VAL. The sequence should now read:

```
          -ORN [ LINK 10 ] -  LEU
3         -PHE [ LTOD] - PRO - VAL
6         -ORN - LEU - PHE [ LTOD ]
10        -PRO - VAL [ 1 LINK ]
```

8. Pull down the Sequence Builder menu and choose "Return to Molecular Modeling." When the system asks if you want to save your changes click on "YES." In the File Librarian type in a file name, remembering not to use any punctuation in the file name and **eight characters or less**. Click on "Save." In the next dialog box click on "Use the new molecule ___.msf only," and then "OK."

Minimizing Your Structure
After you build your sequence, you will note that one of the bonds is very long. This long bond was caused by specifying a cyclic structure. You must minimize your structure to get a reasonable starting conformation. You then will specify either dihedral or distance constraints and reminimize using Adopted Basis Newton Raphson techniques to attempt to find a minimum energy structure that is consistent with the experimental evidence.

9. Pull down the CHARMm menu, slide right on "CHARMm mode," and choose "RTF." Pull down the CHARMm menu and choose "Minimization Options." Select "Steepest Descents," and then if not already shown:

Number of Minimization Steps	200
Coordinate Update Frequency	5
Energy Gradient Tolerance	0.0001
Energy Value Tolerance	0
Initial Step Size	0.02
Step Value Tolerance	0

Click on "OK."
10. Click on "CHARMm minimization" in the "Modeling" palette. The energy of this structure is still high, however. To speed subsequent minimization steps, you can apply dihedral or distance constraints. If you wish to apply dihedral constraints go to step 11. If you wish to apply distance constraints, which are better, go to step 15.

Setting Dihedral Constraints
11. Pull down the CHARMm menu, slide right on "Constraints Options" and choose "Dihedral/Distance." Click on "Define Dihedral Constraint(s)." Click on the four consecutive atoms that define the dihedral you wish to constrain. A new window will appear, labeled "Constraints_Database.con." In this window, you specify the constraints. QUANTA takes the current value of the dihedral as the target value. You also need to set the allowable ranges for your dihedral. This range should be fairly broad for this exercise, since we expect some strain in

the ring to distort the angles from their ideal values. Remember to set the same constraints for both amino acids in the pair, since the structure should be symmetrical.
12. Repeat step 11 for each dihedral you wish to constrain.
13. Click on "Save Constraints" in the "Edit Constraints" palette. Then click on "Exit Edit Constraints" to return to modeling.
14. Pull down the CHARMm menu, slide right on "Constraints Options" and choose "Dihedral On." Minimization will now include your constraints. Continue with step 21.

Setting Distance Constraints
15. Pull down the CHARMm menu, slide right on "Constraints Options" and choose "Dihedral/Distance." Click on "Define Distance Constraint(s)." Make sure "Individual Atoms" is checked.
16. Click on the two atoms that you wish to constrain. A new window will appear, labeled "Constraints_Database.con." In this window, you specify the constraints. You also need to set the allowable ranges for your distance. This range should be fairly broad for this exercise, since we expect some strain in the ring to distort the angles from their ideal values. For hydrogen bonds try a target of 2.0Å with a range of 3.0Å to 1.9Å. Remember to set the same constraints for both amino acids in the pair, since the structure should be symmetrical.
17. Repeat step 16 for each distance you wish to constrain.
18. Click on "Save Constraints" in the "Edit Constraints" palette. Then click on "Exit Edit Constraints" to return to modeling.
19. Pull down the CHARMm menu, slide right on "Constraints Options" and choose "Distance On." Minimization will now include your constraints.
20. A dialog box will appear that allows you to choose an overall scale factor; just use the default value of 1.000, by clicking "OK."

Minimization with Constraints
20. The bond lengths should now be close enough to normal values that we can use conjugate gradient minimization. Pull down the CHARMm menu and choose "Minimization Options." Select "Conjugate Gradient," and then if not already shown:

Number of Minimization Steps	1000
Coordinate Update Frequency	5
Energy Gradient Tolerance	0.0001
Energy Value Tolerance	0
Initial Step Size	0.02
Step Value Tolerance	0

Click on "OK."

21. Click on "CHARMm minimization" in the "Modeling" palette.
22. Check to see that your final structure is symmetrical. If it is not, apply additional constraints, use the "Torsions..." option, or use the "Move Atoms" function in the "Modeling" palette to produce a symmetrical structure. For example, you might need to use "Torsions..." in the "Modeling" palette to rotate some side chains so they are symmetrical.
23. The constraints that you have applied are an artificial term in the potential energy function. For your **final minimization you should remove all constraints** by completing the following. Pull down the CHARMm menu, slide right on "Constraints Options" and choose "Dihedral off" or "Distance Off," depending on the constraint type(s) you used. Reminimize. Repeat minimization until the energy is minimized. This minimization may take 6000 steps or more.
24. Remember to click on "Save Changes" in the "Modeling" palette. Overwrite the data file.
25. Plot out several views of your structure. Measure some of the backbone dihedrals to see if the ϕ angles are close to the NMR determined values.

References

1. M. Waki, N. Izumiya, in *CRC Crit. Rev. Biotechnol.* , **1988**, *8*, 206.
2. K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY, 1986, pp 166-7.
3. T. E. Creighton, *Proteins: Structures and Molecular Properties*, W. H. Freeman, New York, NY, 1983. pp171-5.
4. J. S. Richardson, D. C. Richardson, Principles and Patterns of Protein Conformation, in *Prediction of Protein Structure and the Principles of Protein Conformation*, G. D. Fasman, ed., Plenum Press, New York, NY, 1989. p9-10.