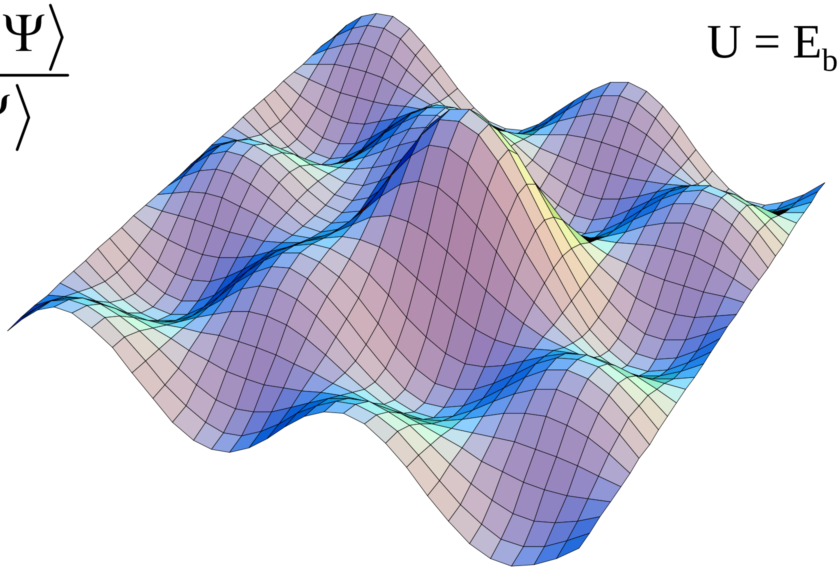
Molecular Simulation I

Quantum Chemistry

$$E = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle}$$

Classical Mechanics

$U = E_{bond} + E_{angle} + E_{torsion} + E_{non\text{-}bond}$



Jeffry D. Madura

Department of Chemistry & Biochemistry
Center for Computational Sciences
Duquesne University

Quantum Chemistry

- Molecular Orbital Theory
  - Based on a wave function approach
  - Schrödinger equation
- Density Functional Theory
  - Based on the total electron density
  - Hohenberg – Kohn theorem
- Semi-empirical
  - Some to most integrals parameterized
  - MNDO, AM1, EHT
- Empirical
  - All integrals are parameterized
  - Huckel method

The Beginning...

- Schrödinger equation

$$H\Psi = E\Psi$$

- Hamiltonian operator

$$H = -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)+V\left(r\right) = -\frac{\hbar^2}{2m}\nabla^2+V\left(r\right)$$

- Wave function (Ψ)
  - – characterizes the particles motion
  - – various properties of the particle can be derived

Quantum Chemistry

- Start with Schrödinger's equation

$$H\Psi = E\Psi$$

- Make some assumptions
  - Born-Oppenheimer approximation
  - Linear combination of atomic orbitals

$$\Psi = c_a\varphi_a + c_b\varphi_b$$

- Apply the variational method

$$E \leq \frac{\int \Psi^* H \Psi d\tau}{\int \Psi^* \Psi d\tau}$$

LCAO

- A practical and common approach to solving the Hartree-Fock equations is to write each spin orbital as a linear combination of single electron orbitals (LCAO)

$$\psi_i = \sum_{\nu=1}^{K} c_{\nu i}\phi_\nu$$

  - the ϕ_v are commonly called *basis functions* and often correspond to atomic orbitals
  - K basis functions lead to K molecular orbitals
  - the point at which the energy is not reduced by the addition of basis functions is known as the Hartree-Fock limit

Basis Sets

- Slater type orbitals (STO)

$$R_{nl}\left(r\right)=\left(2\zeta\right)^{n+1/2}\left[\left(2n\right)!\right]^{-1/2} r^{n-1}e^{-\zeta r}$$

- Gaussian type orbitals (GTO)
  - functional form

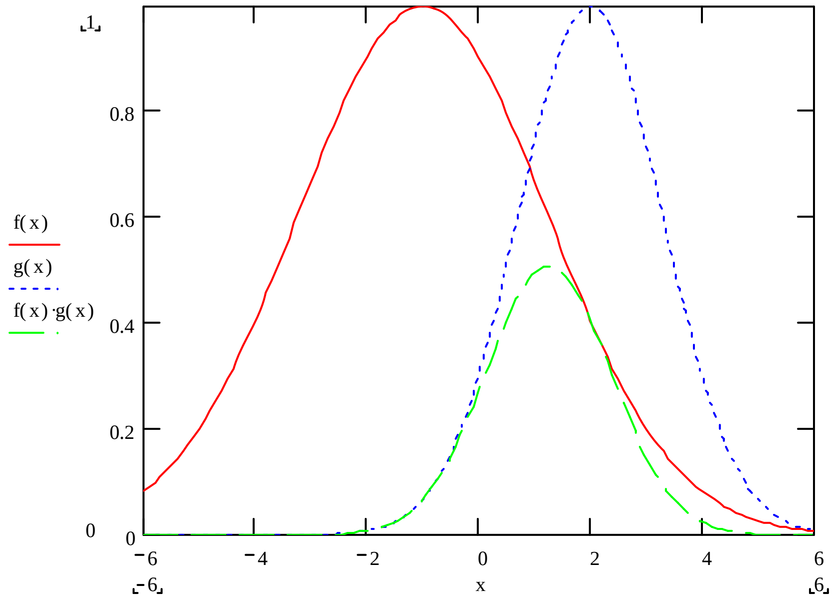
$$x^{a}y^{b}z^{c}e^{-a\,r^{2}}$$

  - zeroth-order Gaussian function

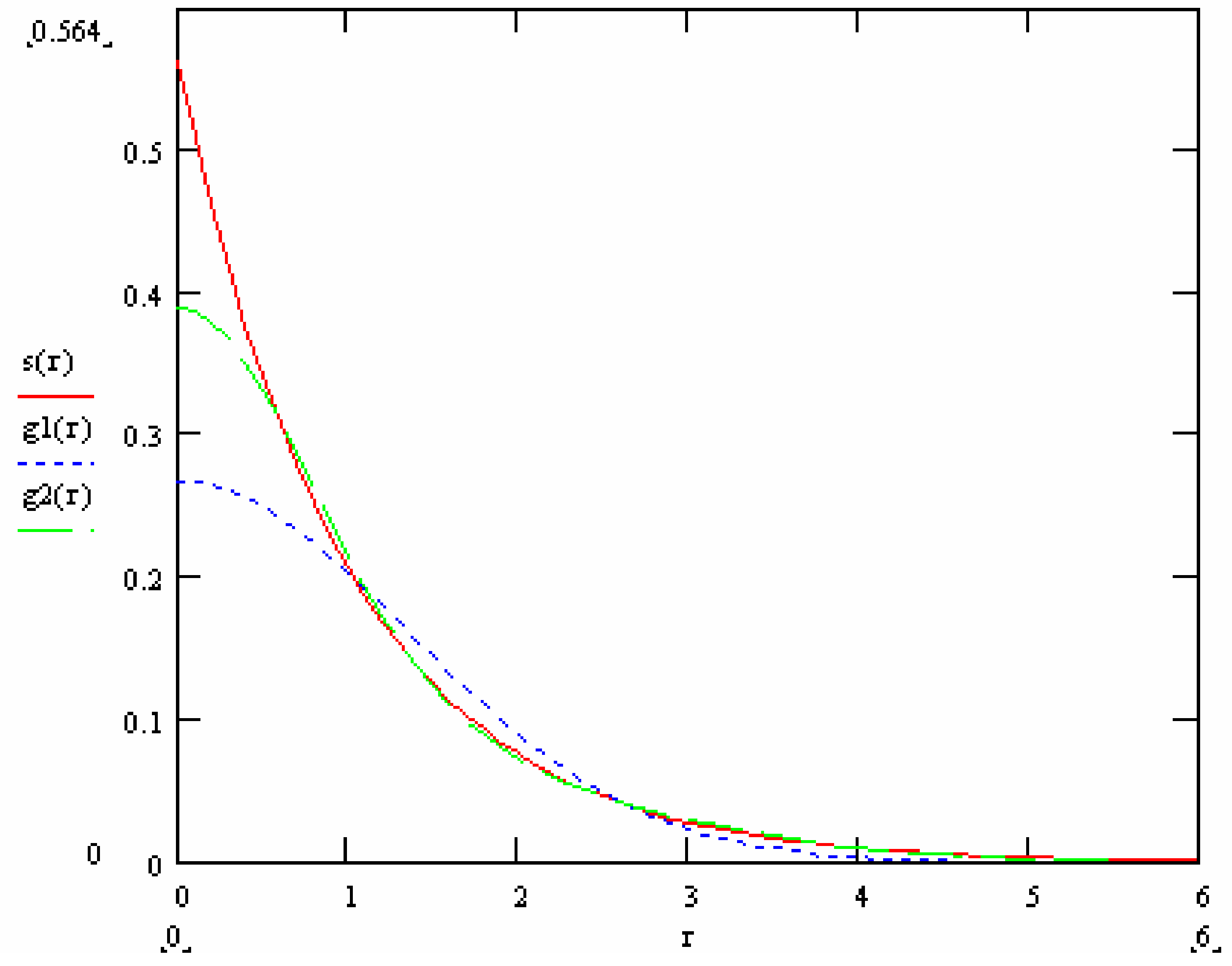
$$g_{s}\left(\alpha,r\right)=\left(\frac{2\alpha}{\pi}\right)^{3/4}e^{-\alpha r^{2}}$$

- Property of Gaussian functions is that the product of two Gaussians can be expressed as a single Gaussian, located along the line joining the centers of the two Gaussians

$$e^{-a_m r_m^2} e^{-a_n r_n^2} = e^{-\frac{a_m a_n}{a_m + a_n} r_{mn}^2} e^{-a r_c^2}$$



- STO vs. GTO



- Gaussian expansion
  - – the coefficient
  - – the exponent
  - – uncontratcted or primitive and contracted
  - – s and p exponents in the same shell are equal

$$\phi_\mu = \sum_{i=1}^{L} d_{i\mu} \phi_i \left(\alpha_{i\mu} \right)$$

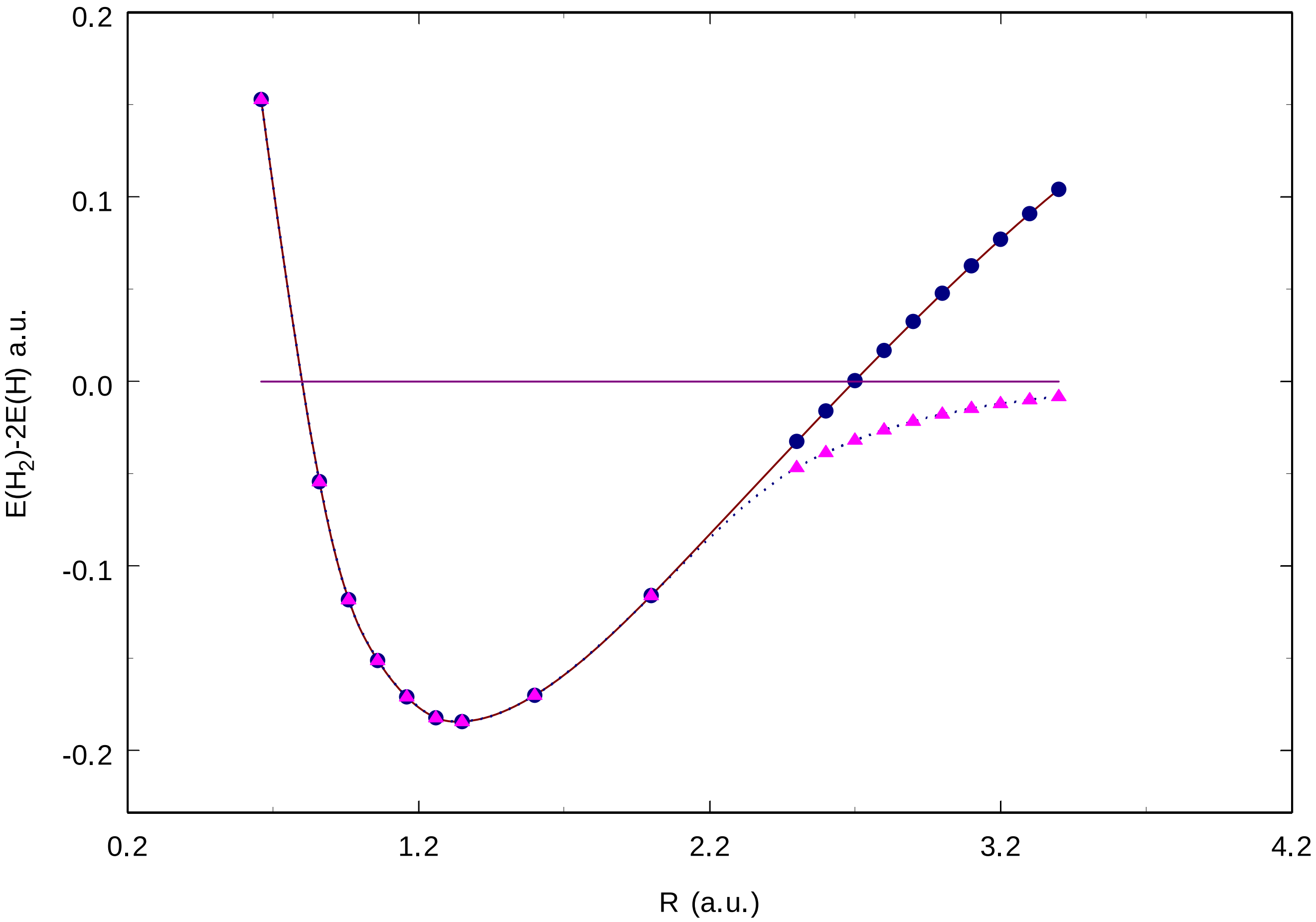
- Minimal basis set
  - – STO-NG
- Double zeta basis set
  - – linear combination of a ‘contracted’ function and a ‘diffuse’ function.
- Split valence
  - – 3-21G, 4-31G, 6-31G

- Polarization
  - to solve the problem of non-isotropic charge distribution.
  - 6-31G*, 6-31G**
- Diffuse functions
  - fulfill as deficiency of the basis sets to describe significant amounts of electron density away from the nuclear centers. (e.g. anions, lone pairs, etc.)
  - 3-21+G, 6-31++G

RHF vs. UHF

- Restricted Hartree-Fock (RHF)
  - closed-shell molecules
- Restricted Open-shell Hartree-Fock (ROHF)
  - combination of singly and doubly occupied molecular orbitals.
- Unrestricted Hartree-Fock (UHF)
  - open-shell molecules
  - Pople and Nesbet: one set of molecular orbitals for α spin and another for the β spin.

UHF and RHF Dissociation Curves for H_2



Electron Correlation

- The most significant drawback to HF theory is that it fails to adequately represent electron correlation.

$$E_{corr} = E^{NR} - E^{HF}$$

- Configuration Interactions
  - excited states are included in the description of an electronic state
- Many Body Perturbation Theory
  - based upon Rayleigh-Schrödinger perturbation theory

Configuration Interaction

- The CI wavefunction is written as

$$\Psi = c_0 \Psi_0 + c_1 \Psi_1 + c_2 \Psi_2 + \cdots$$

  - where Ψ_0 is the HF single determinant
  - where Ψ_1 is the configuration derived by replacing one of the occupied spin orbitals by a virtual spin orbital
  - where Ψ_2 is the configuration derived by replacing one of the occupied spin orbitals by a virtual spin orbital
- The system energy is minimized in order to determine the coefficients, c_0, c_1, etc., using a linear variational approach

Many Body Peturbation Theory

- Based upon perturbation concepts $H = H_0 + V$
- The correction to the energies are

$$E_i^{(0)} = \int \Psi_i^{(0)} H_0 \Psi_i^{(0)} d\tau$$

$$E_i^{(1)} = \int \Psi_i^{(0)} V \Psi_i^{(0)} d\tau$$

$$E_i^{(2)} = \int \Psi_i^{(0)} V \Psi_i^{(1)} d\tau$$

$$E_i^{(3)} = \int \Psi_i^{(0)} V \Psi_i^{(2)} d\tau$$

  - Perturbation methods are size independent
  - these methods are not variational

Theoretical Model

Theoretical Model = Level of Theory + Basis Set

Level of Theory = HF, MP2, DFT, CI, CCSD, etc

Basis Set = STO-3G, 3-21G, 6-31G*, 6-311++G(d,p), etc

Geometry Optimization

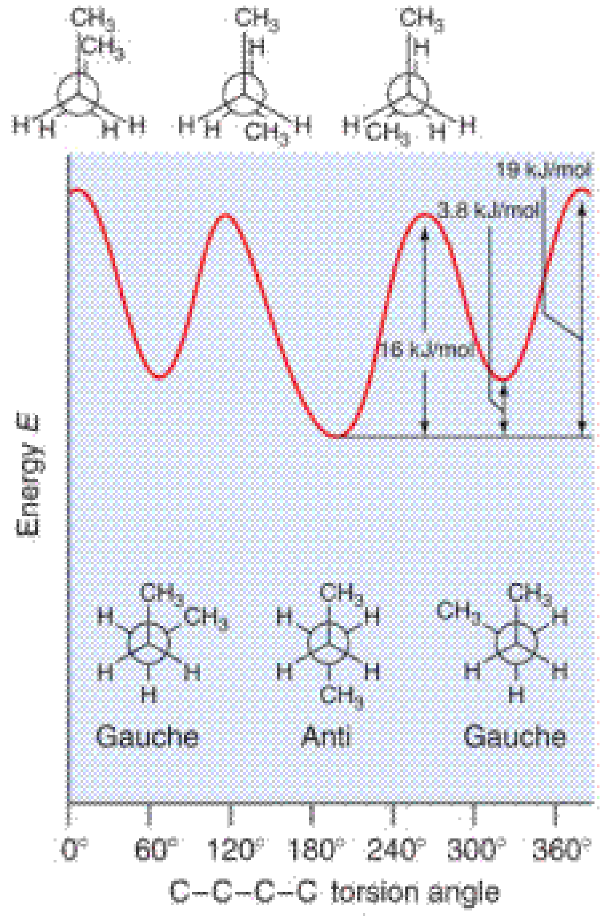
- Derivatives of the energy

$$E\left(x_f\right)=E\left(x\right)+\sum_i \frac{\partial E\left(x\right)}{\partial x_i}\left(x_{if}-x_i\right)+\frac{1}{2}\sum_i\sum_j \frac{\partial^2 E\left(x\right)}{\partial x_i \partial x_j}\left(x_{if}-x_i\right)\left(xj-x_j\right)+\cdots$$

  - the first term is set to zero
  - the second term can be shown to be equivalent to a force
  - the third term can be shown to be equivalent to a force constant

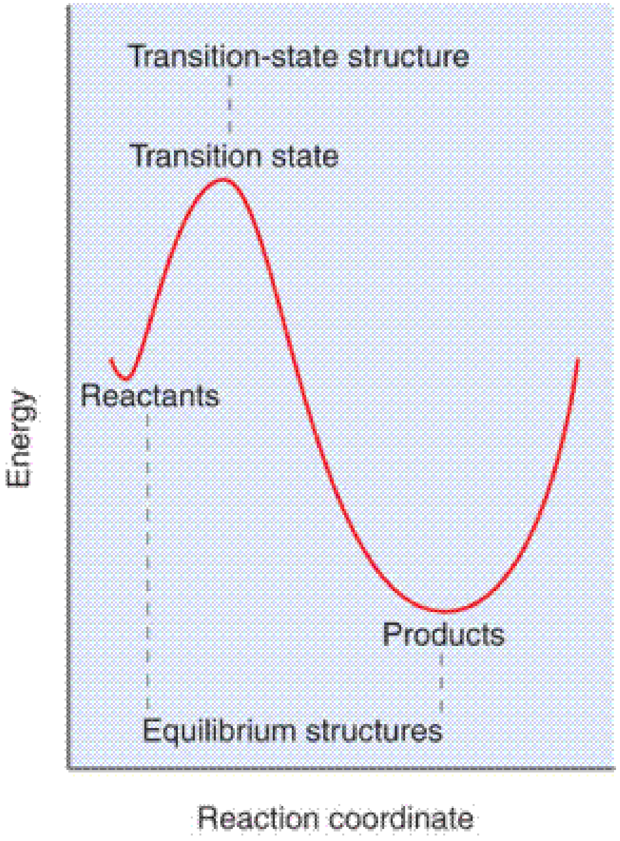
- Internal coordinate, Cartesian coordinate, and redundant coordinate optimization
  - choice of coordinate set can determine whether a structure reaches a minimum/maximum and the speed of this convergence.
  - Internal coordinates are defined as bond lengths, bond angles, and torsions. There are 3N-6 (3N-5) such degrees of freedom for each molecule. Chemists work in this world. Z-matrix...
  - Cartesian coordinates are the standard x, y, z coordinates. Programs often work in this world.
  - Redundant coordinates are defined as the number of coordinates larger than 3N-6.

Potential Energy Surfaces



$$f = -\frac{dV}{dr}$$

$$k = \frac{d^2V}{dr^2}$$



Frequency Calculation

- The second derivatives of the energy with respect to the displacement of coordinate yields the force constants.
- These force constants in turn can be used to calculate frequencies.
  - All real frequencies (positive force constants): local minimum
  - One imaginary frequency (one negative force constant): saddle point, a.k.a. transition state.
- From vibrational analysis can compute thermodynamic data

Molecular Properties

- Charges
  - Mulliken
  - Löwdin
  - electrostatic fitted (ESP)
- Bond orders
- Bonding
  - Natural Bond Analysis
  - Bader’s AIM method
- Molecular orbitals and total electron density
- Dipole Moment
- Energies
  - ionization and electron affinity

Energies

- Koopman’s theorem
  - equating the energy of an electron in an orbital to the energy required to remove the electron to the corresponding ion.
    - ‘frozen’ orbitals
    - lack of electron correlation effects

Dipole Moments

- The electric multipole moments of a molecule reflect the distribution of charge.
  - Simplest is the dipole moment

$$\mu = \sum_i q_i r_i$$

  - nuclear component

$$\mu_{nuclear} = \sum_{A=1}^{M} Z_A R_A$$

  - electronic

$$\mu_{electronic} = \sum_{\mu=1}^{K} \sum_{\nu=1}^{K} P_{\mu\nu} \int d\tau \phi_\mu \left(-r\right) \phi_\nu$$

Molecular Orbitals and Total Electron Density

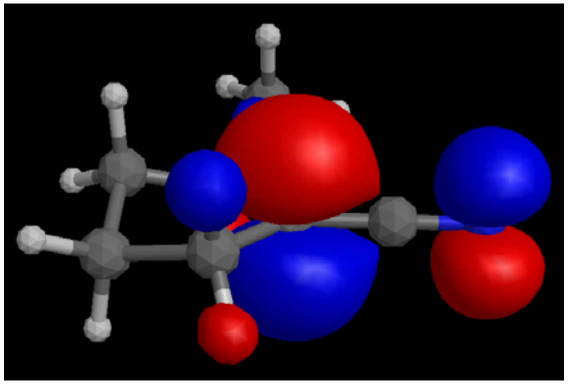
- Electron density at a point r

$$\rho\left(r\right)=2\sum_{i=1}^{N/2}\left|\psi_i\left(r\right)\right|^2=\sum_{\mu=1}^{K}P_{\mu\mu}\phi_\mu\left(r\right)\phi_\mu\left(r\right)+2\sum_{\mu=1}^{K}\sum_{\nu=\mu+1}^{K}P_{\mu\nu}\phi_\mu\left(r\right)\phi_\nu\left(r\right)$$

- Number of electrons is

$$N=2\sum_{i=1}^{N/2}\int dr\left|\psi_i\left(r\right)\right|^2=\sum_{\mu=1}^{K}P_{\mu\mu}+2\sum_{\mu=1}^{K}\sum_{\nu=\mu+1}^{K}P_{\mu\nu}S_{\mu\nu}$$

- Molecular orbitals
  - – HOMO
  - – LUMO



Bonding

- Natural Bond Analysis
  - a way to describe N-electron wave functions in terms of localized orbitals that are closely tied to chemical concepts.
- Bader
  - F. W. Bader’s theory of ‘atoms in molecules’.
  - This method provides an alternative way to partition the electrons among the atoms in a molecule.
  - Gradient vector path
  - bond critical points
  - charges are relatively invariant to the basis set

Bond Orders

- Wiberg

$$W_{AB} = \sum_{\mu\, on\, A} \sum_{\nu\, on\, B} \left| P_{\mu\nu} \right|^2$$

- Mayer

$$B_{AB} = \sum_{\mu\, on\, A} \sum_{\nu\, on\, B} \left(PS \right)_{\mu\nu} \left(PS \right)_{\nu\mu}$$

- Bond orders can be computed for intermediate structures which can be useful way to describe similarity of the TS to the reactants or to the products.

Charges

- Mulliken

$$q_A = Z_A - \sum_{\mu=1;\, on\, A}^{K} P_{\mu\mu} - \sum_{\mu=1;\, on\, A}^{K} \sum_{\nu=1;\, \nu\neq\mu}^{K} P_{\mu\nu} S_{\mu\nu}$$

- Löwdin
  - – atomic orbitals are transformed to an orthogonal set, along with the mo coefficients

$$\phi_\mu^{'} = \sum_{\nu=1}^{K} \left(S^{-1/2}\right)_{\nu\mu} \phi_\nu$$

$$q_A = Z_A - \sum_{\mu=1;\, \mu\, on\, A}^{K} \left(S^{1/2} P^{1/2}\right)_{\mu\mu}$$

Summary of Methods

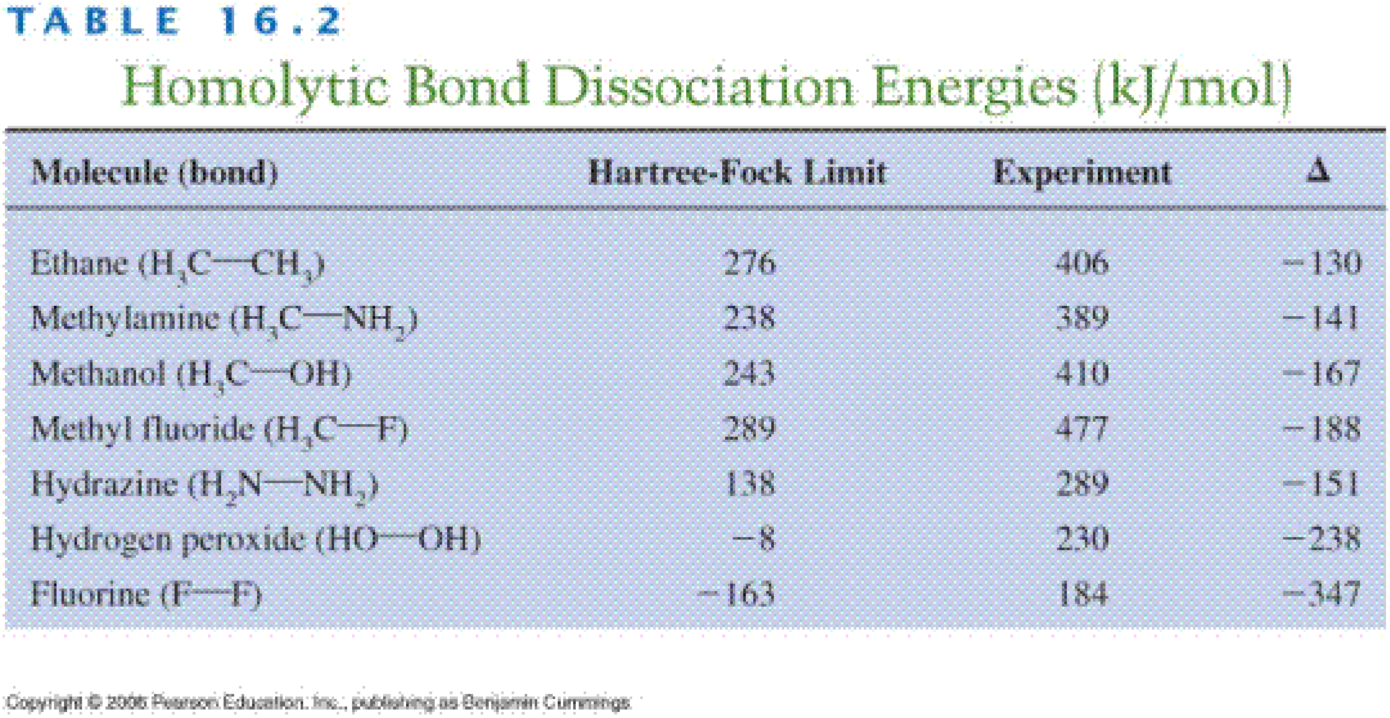


TABLE 16.2

Homolytic Bond Dissociation Energies (kJ/mol)

Molecule (bond)	Hartree-Fock Limit	Experiment	Δ
Ethane ($H_3C—CH_3$)	276	406	−130
Methylamine ($H_3C—NH_2$)	238	389	−141
Methanol ($H_3C—OH$)	243	410	−167
Methyl fluoride ($H_3C—F$)	289	477	−188
Hydrazine ($H_2N—NH_2$)	138	289	−151
Hydrogen peroxide (HO—OH)	−8	230	−238
Fluorine (F—F)	−163	184	−347

Summary of Results

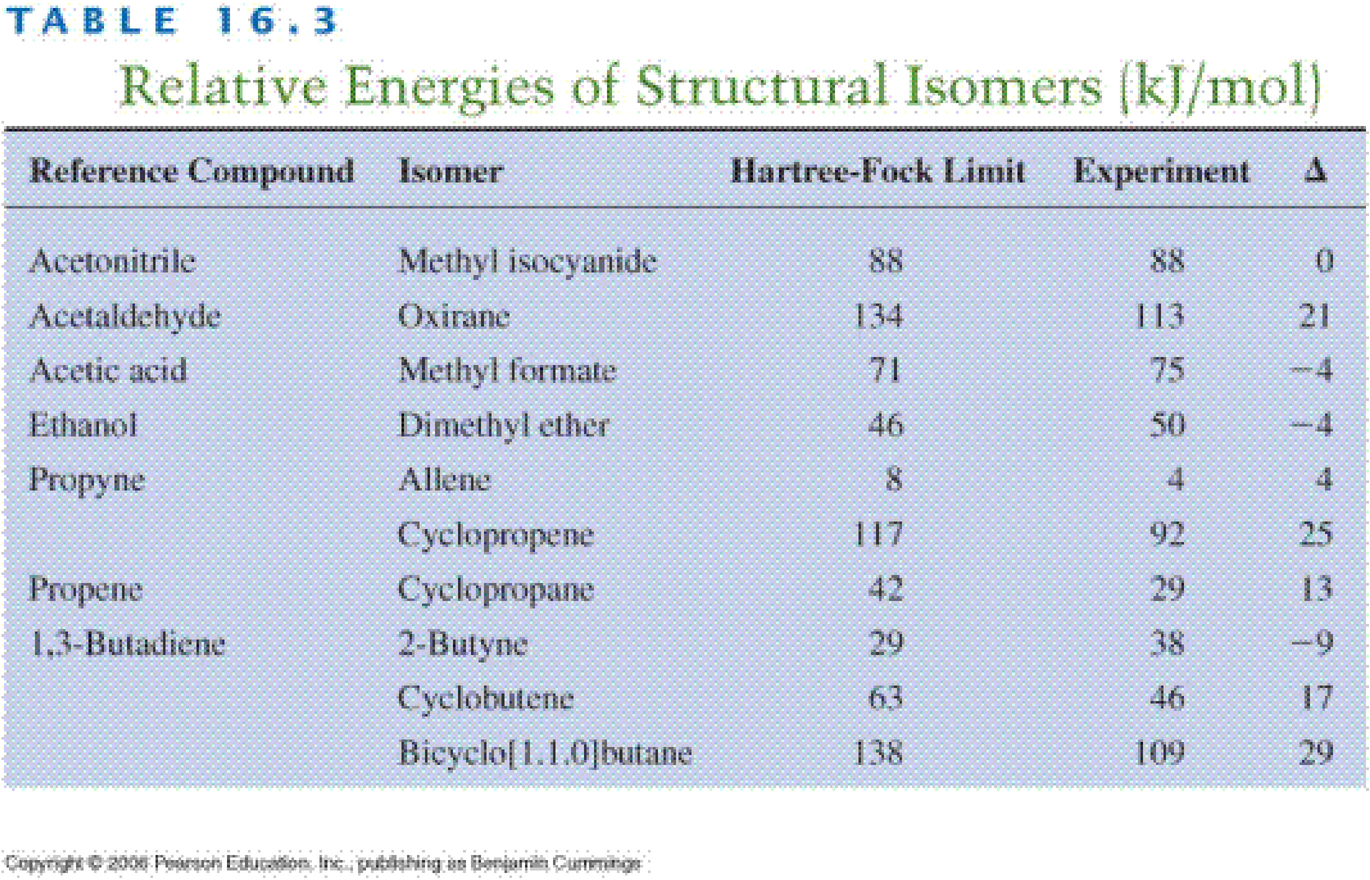


TABLE 16.3

Relative Energies of Structural Isomers (kJ/mol)

Reference Compound	Isomer	Hartree-Fock Limit	Experiment	Δ
Acetonitrile	Methyl isocyanide	88	88	0
Acetaldehyde	Oxirane	134	113	21
Acetic acid	Methyl formate	71	75	−4
Ethanol	Dimethyl ether	46	50	−4
Propyne	Allene	8	4	4
	Cyclopropene	117	92	25
Propene	Cyclopropane	42	29	13
1,3-Butadiene	2-Butyne	29	38	−9
	Cyclobutene	63	46	17
	Bicyclo[1.1.0]butane	138	109	29

Summary of Results

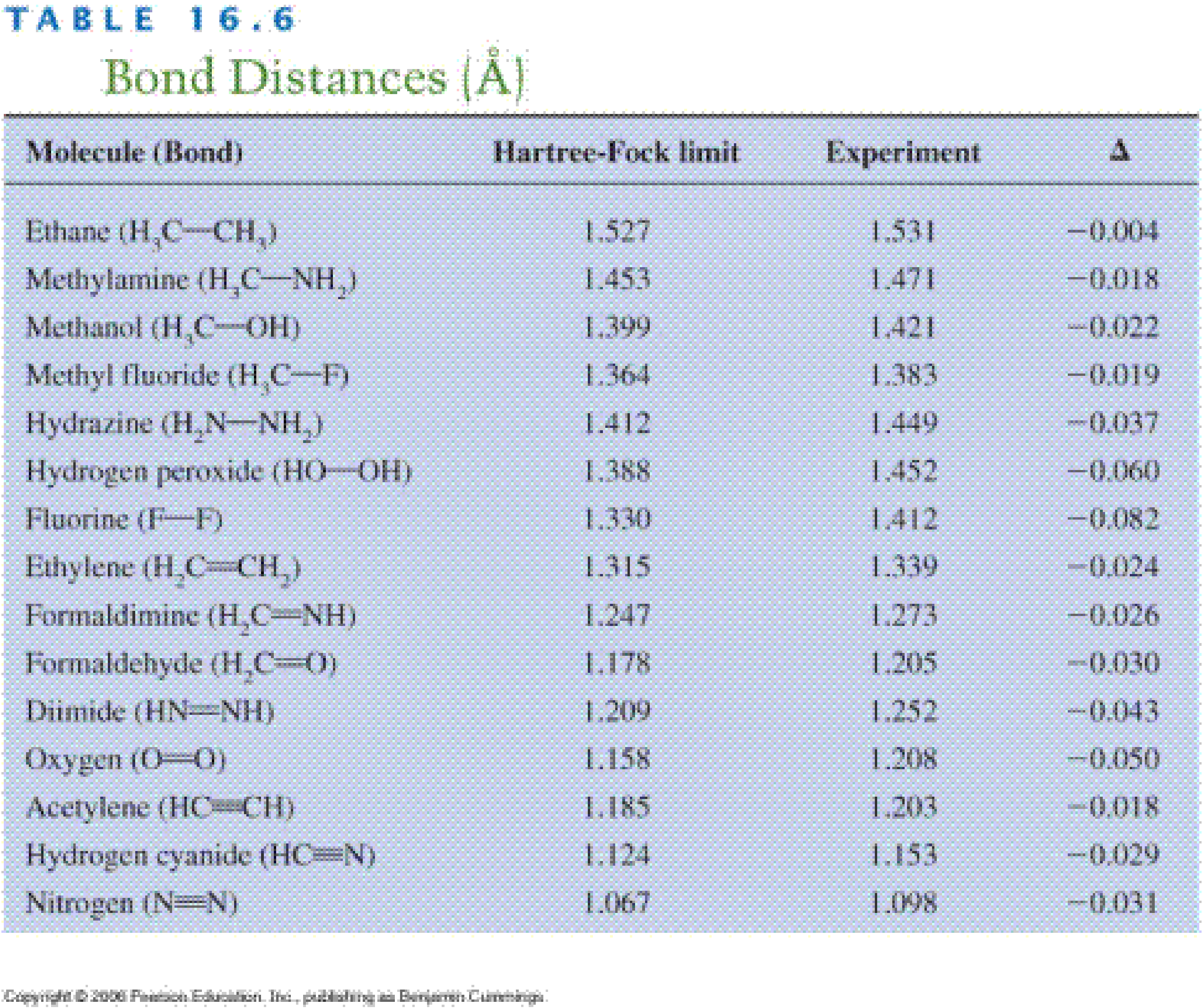


TABLE 16.6

Bond Distances (Å)

Molecule (Bond)	Hartree-Fock limit	Experiment	Δ
Ethane ($H_3C{-}CH_3$)	1.527	1.531	−0.004
Methylamine ($H_3C{-}NH_2$)	1.453	1.471	−0.018
Methanol ($H_3C{-}OH$)	1.399	1.421	−0.022
Methyl fluoride ($H_3C{-}F$)	1.364	1.383	−0.019
Hydrazine ($H_2N{-}NH_2$)	1.412	1.449	−0.037
Hydrogen peroxide ($HO{-}OH$)	1.388	1.452	−0.060
Fluorine ($F{-}F$)	1.330	1.412	−0.082
Ethylene ($H_2C{=}CH_2$)	1.315	1.339	−0.024
Formaldimine ($H_2C{=}NH$)	1.247	1.273	−0.026
Formaldehyde ($H_2C{=}O$)	1.178	1.205	−0.030
Diimide ($HN{=}NH$)	1.209	1.252	−0.043
Oxygen ($O{=}O$)	1.158	1.208	−0.050
Acetylene ($HC{\equiv}CH$)	1.185	1.203	−0.018
Hydrogen cyanide ($HC{\equiv}N$)	1.124	1.153	−0.029
Nitrogen ($N{\equiv}N$)	1.067	1.098	−0.031

Summary of Results

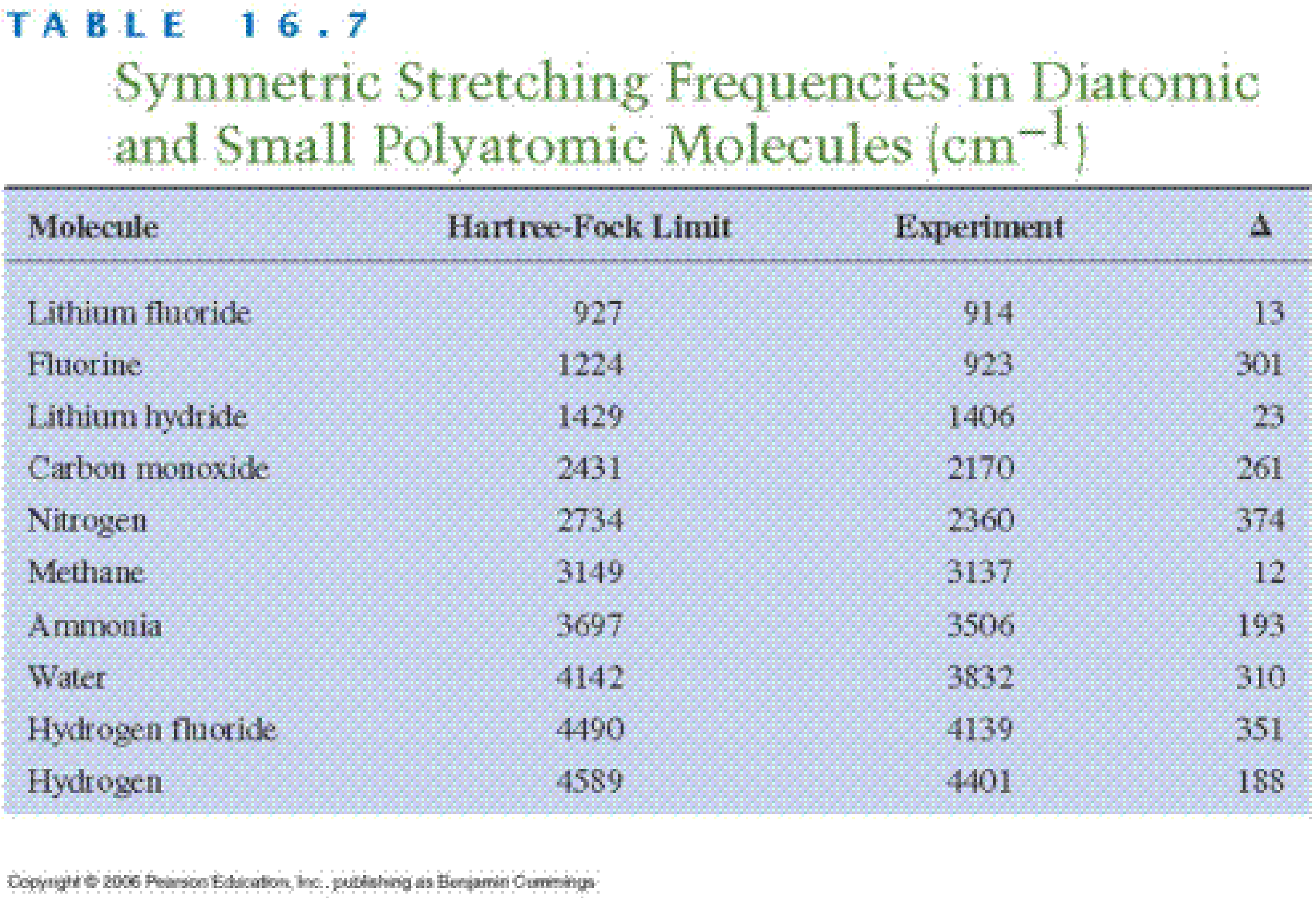


TABLE 16.7

Symmetric Stretching Frequencies in Diatomic and Small Polyatomic Molecules (cm^{-1})

Molecule	Hartree-Fock Limit	Experiment	Δ
Lithium fluoride	927	914	13
Fluorine	1224	923	301
Lithium hydride	1429	1406	23
Carbon monoxide	2431	2170	261
Nitrogen	2734	2360	374
Methane	3149	3137	12
Ammonia	3697	3506	193
Water	4142	3832	310
Hydrogen fluoride	4490	4139	351
Hydrogen	4589	4401	188

Summary of Results

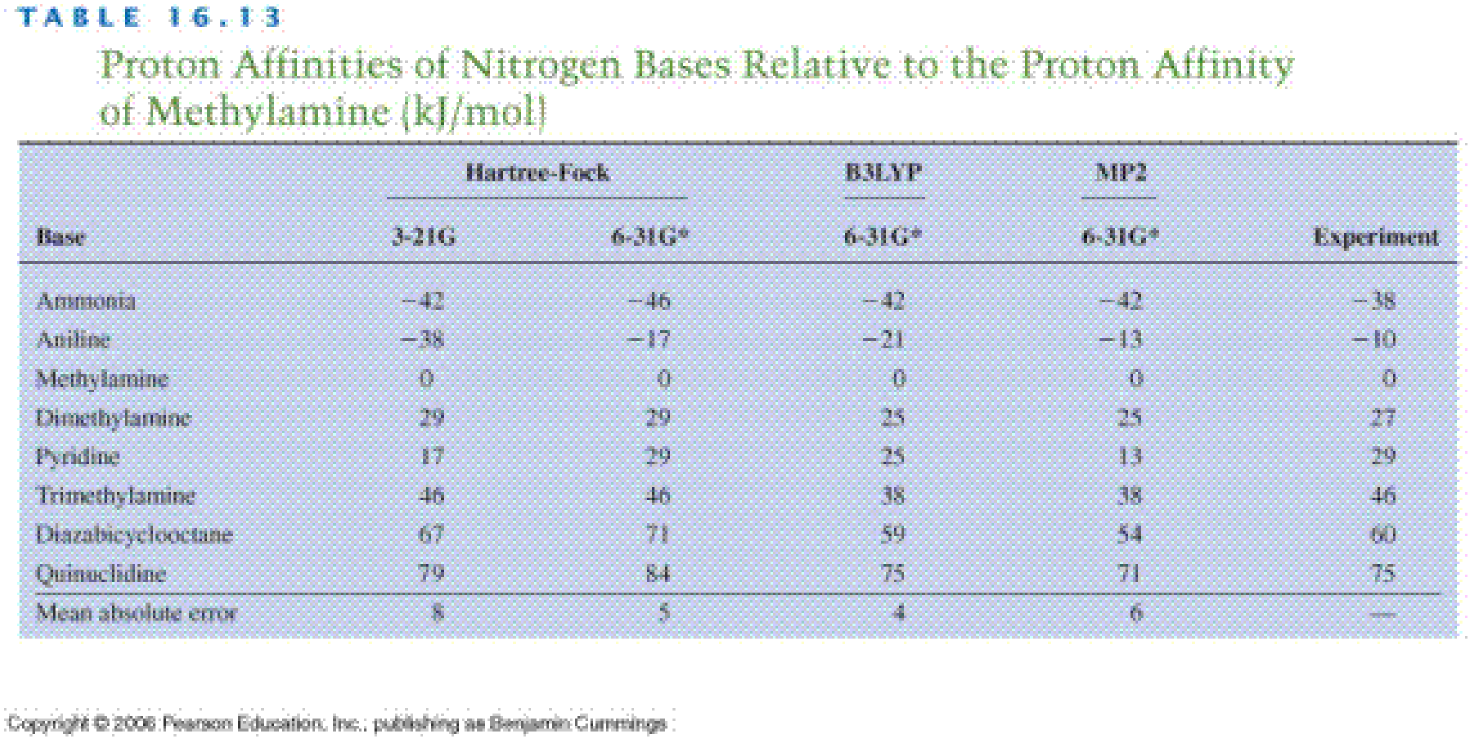


TABLE 16.13

Proton Affinities of Nitrogen Bases Relative to the Proton Affinity of Methylamine (kJ/mol)

	Hartree-Fock		B3LYP	MP2	
Base	**3-21G**	**6-31G***	**6-31G***	**6-31G***	**Experiment**
Ammonia	−42	−46	−42	−42	−38
Aniline	−38	−17	−21	−13	−10
Methylamine	0	0	0	0	0
Dimethylamine	29	29	25	25	27
Pyridine	17	29	25	13	29
Trimethylamine	46	46	38	38	46
Diazabicyclooctane	67	71	59	54	60
Quinuclidine	79	84	75	71	75
Mean absolute error	8	5	4	6	—

Summary of Results

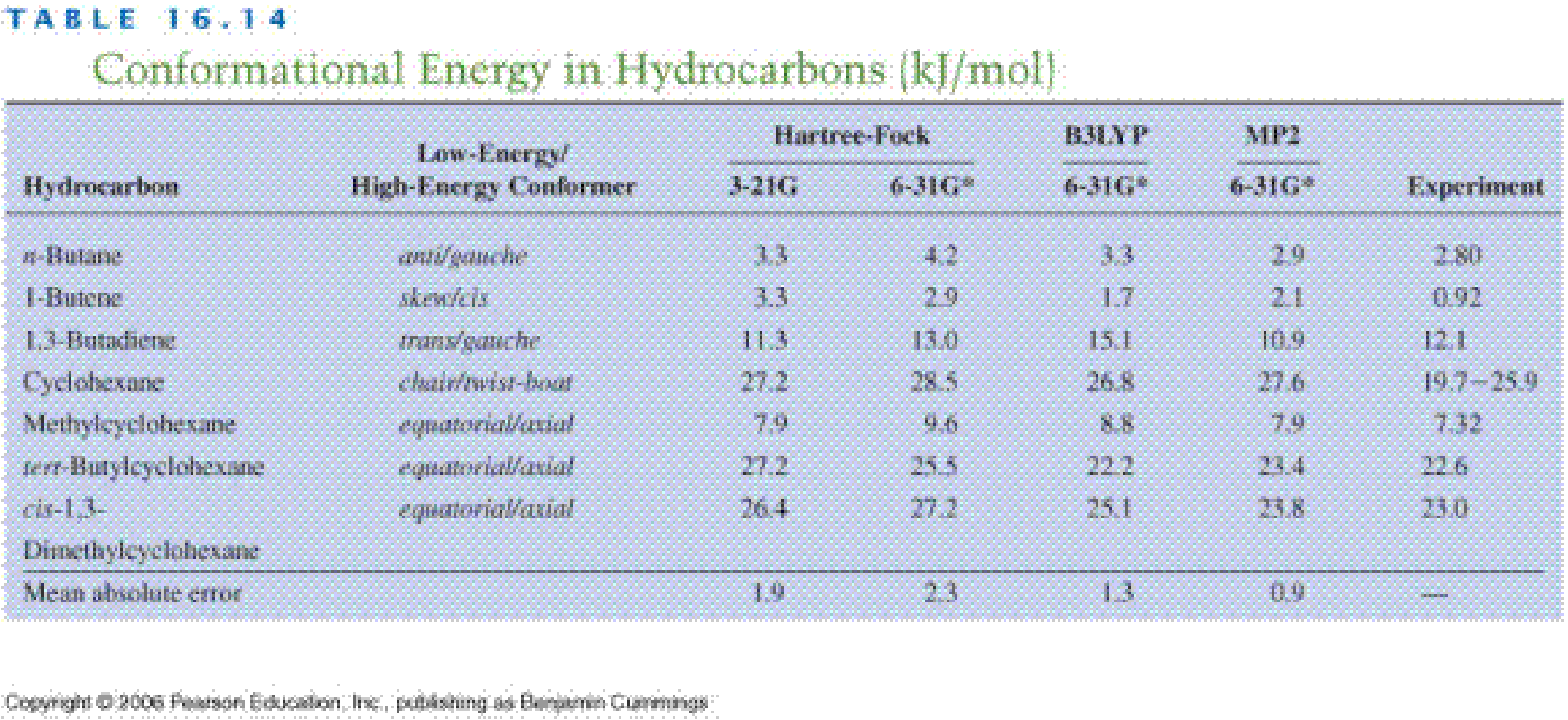


TABLE 16.14

Conformational Energy in Hydrocarbons (kJ/mol)

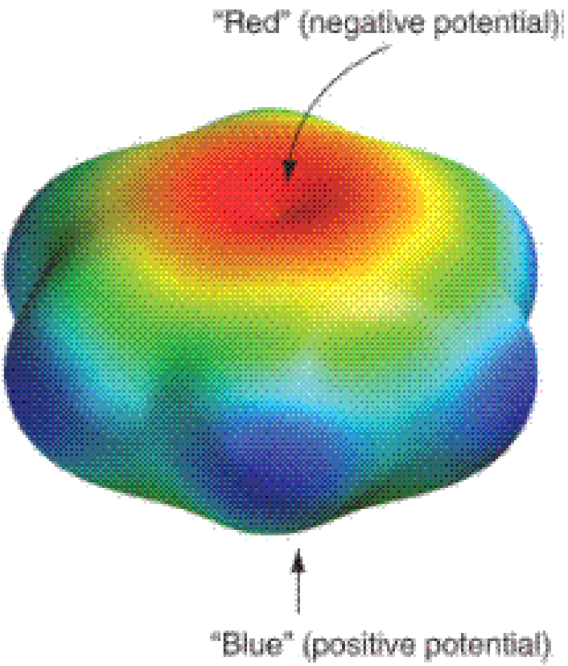
Hydrocarbon	Low-Energy/ High-Energy Conformer	Hartree-Fock 3-21G	Hartree-Fock 6-31G*	B3LYP 6-31G*	MP2 6-31G*	Experiment
n-Butane	*anti/gauche*	3.3	4.2	3.3	2.9	2.80
1-Butene	*skew/cis*	3.3	2.9	1.7	2.1	0.92
1,3-Butadiene	*trans/gauche*	11.3	13.0	15.1	10.9	12.1
Cyclohexane	*chair/twist-boat*	27.2	28.5	26.8	27.6	19.7–25.9
Methylcyclohexane	*equatorial/axial*	7.9	9.6	8.8	7.9	7.32
tert-Butylcyclohexane	*equatorial/axial*	27.2	25.5	22.2	23.4	22.6
cis-1,3-Dimethylcyclohexane	*equatorial/axial*	26.4	27.2	25.1	23.8	23.0
Mean absolute error		1.9	2.3	1.3	0.9	—

- Electrostatic potentials
  - the electrostatic potential at a point **r**, ϕ(**r**), is defined as the work done to bring a unit positive charge from infinity to the point.
  - the electrostatic interaction energy between a point charge *q* located at **r** and the molecule equals *q*ϕ(**r**).
  - there is a nuclear part and electronic part

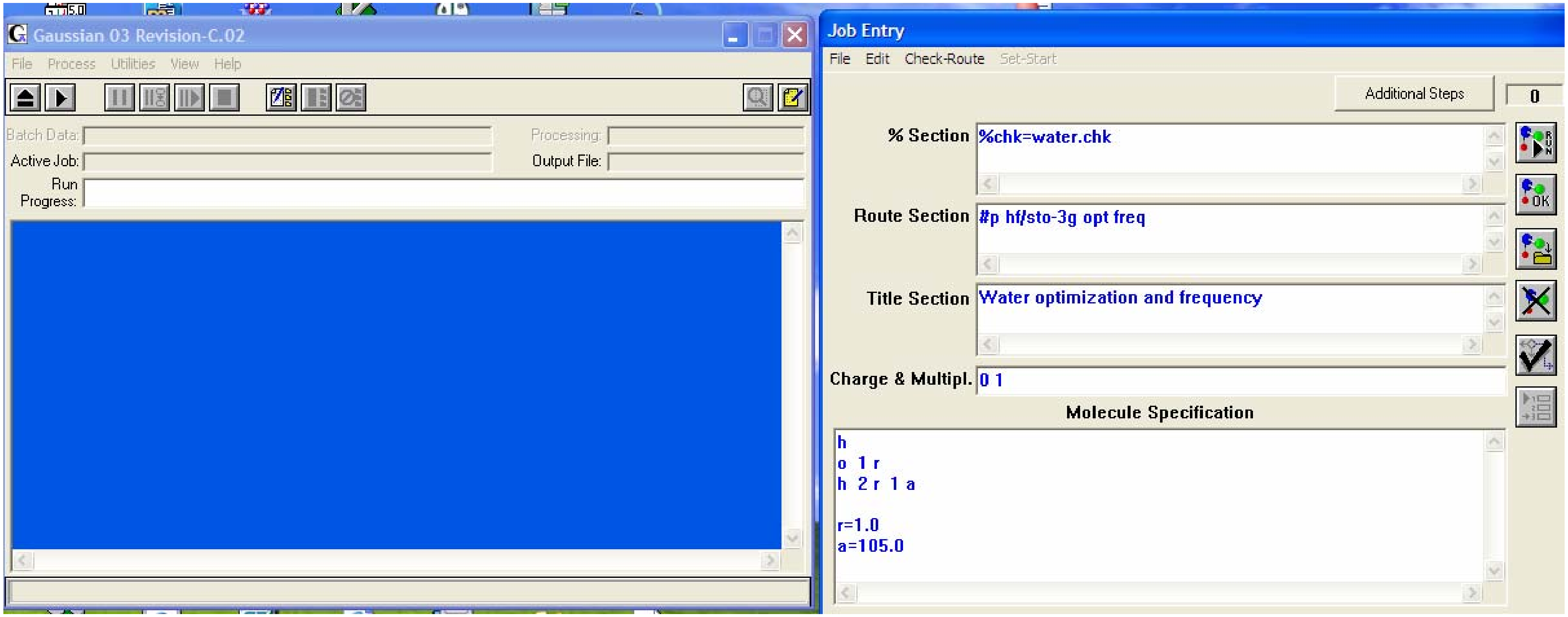
$$\phi_{nucl}\left(r\right)=\sum_{A=1}^{M}\frac{Z_A}{\left|r-R_A\right|} \qquad \phi_{elec}\left(r\right)=-\int\frac{dr^{'}\rho\left(r\right)}{\left|r^{'}-r\right|}$$

$$\phi\left(r\right)=\phi_{nucl}\left(r\right)+\phi_{elec}\left(r\right)$$

Electrostatic Potential



G03W - Water Example



% Section %chk=water.chk

Route Section #p hf/sto-3g opt freq

Title Section Water optimization and frequency

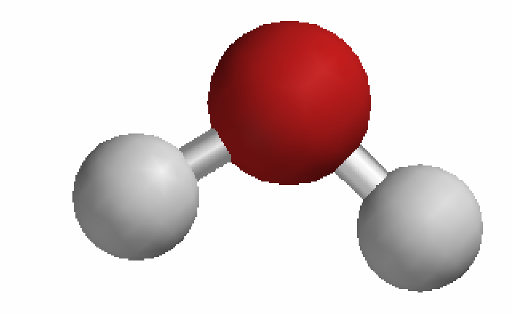
Charge & Multipl. 0 1

Molecule Specification

```
h
o  1 r
h  2 r  1 a

r=1.0
a=105.0
```

Water Example



Geometry

```
 ------------------------------------------------------------------------
                Z-MATRIX (ANGSTROMS AND DEGREES)
 CD    Cent  Atom  N1      Length/X           N2     Alpha/Y       N3     Beta/Z        J
 ------------------------------------------------------------------------
   1    1      H
   2    2      O      1       0.989400(  1)
   3    3      H      2       0.989400(  2)     1       100.028(  3)
 ------------------------------------------------------------------------
```

Energy

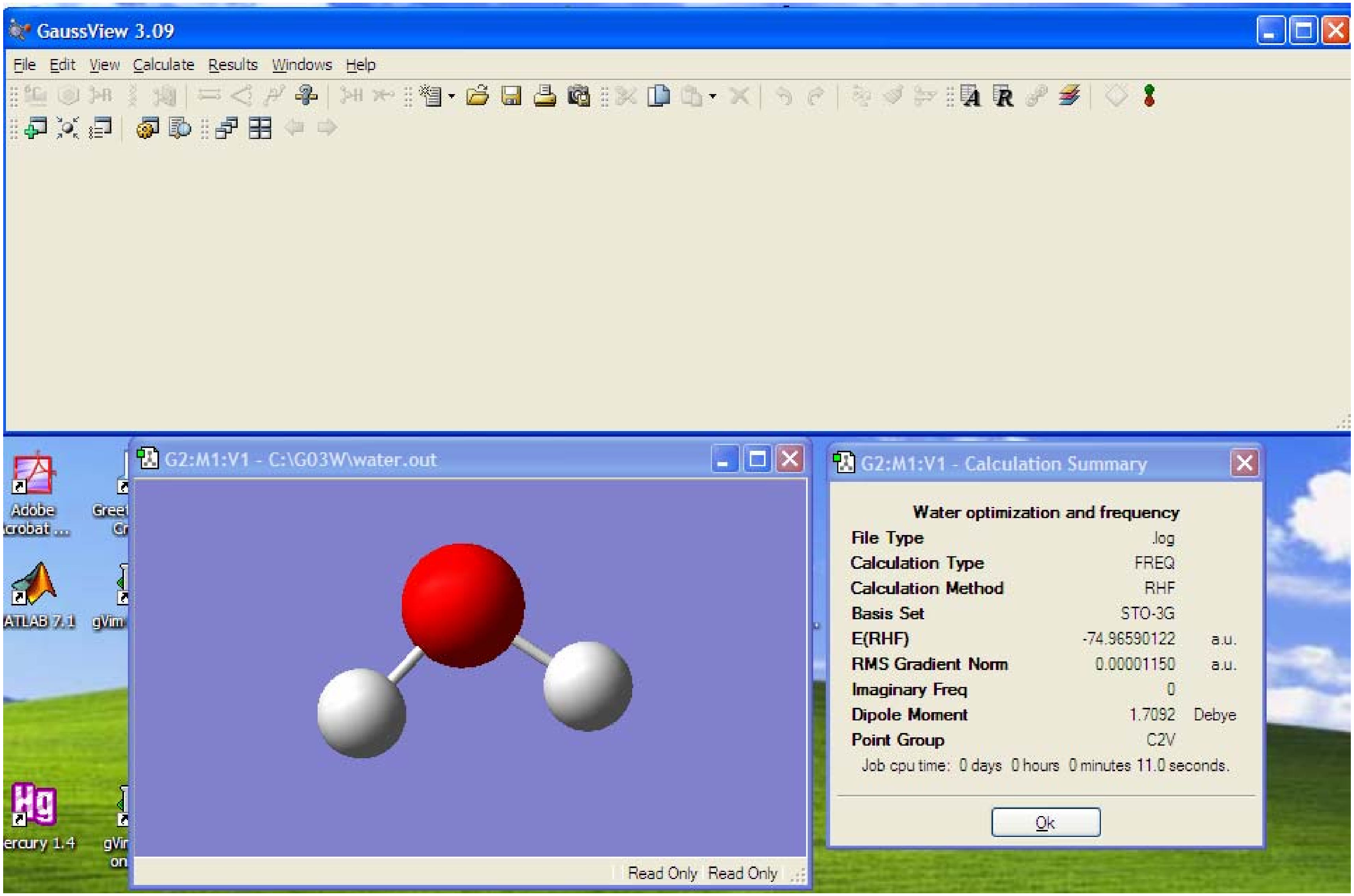
```
E(RHF) =  -74.9658952265     A.U.
```

Population Analysis

```
 **********************************************************************
           Population analysis using the SCF density.
 **********************************************************************
 Alpha  occ. eigenvalues --  -20.25226  -1.25780  -0.59411  -0.45987  -0.39297
 Alpha virt. eigenvalues --    0.58175   0.69242
          Condensed to atoms (all electrons):
                   1              2            3
   1  H    0.626190   0.253760  -0.045250
   2  O    0.253760   7.823081   0.253760
   3  H   -0.045250   0.253760   0.626190
```

```
 Total atomic charges:
              1
   1  H    0.165300
   2  O   -0.330601
   3  H    0.165300
```

GaussView – Water Example

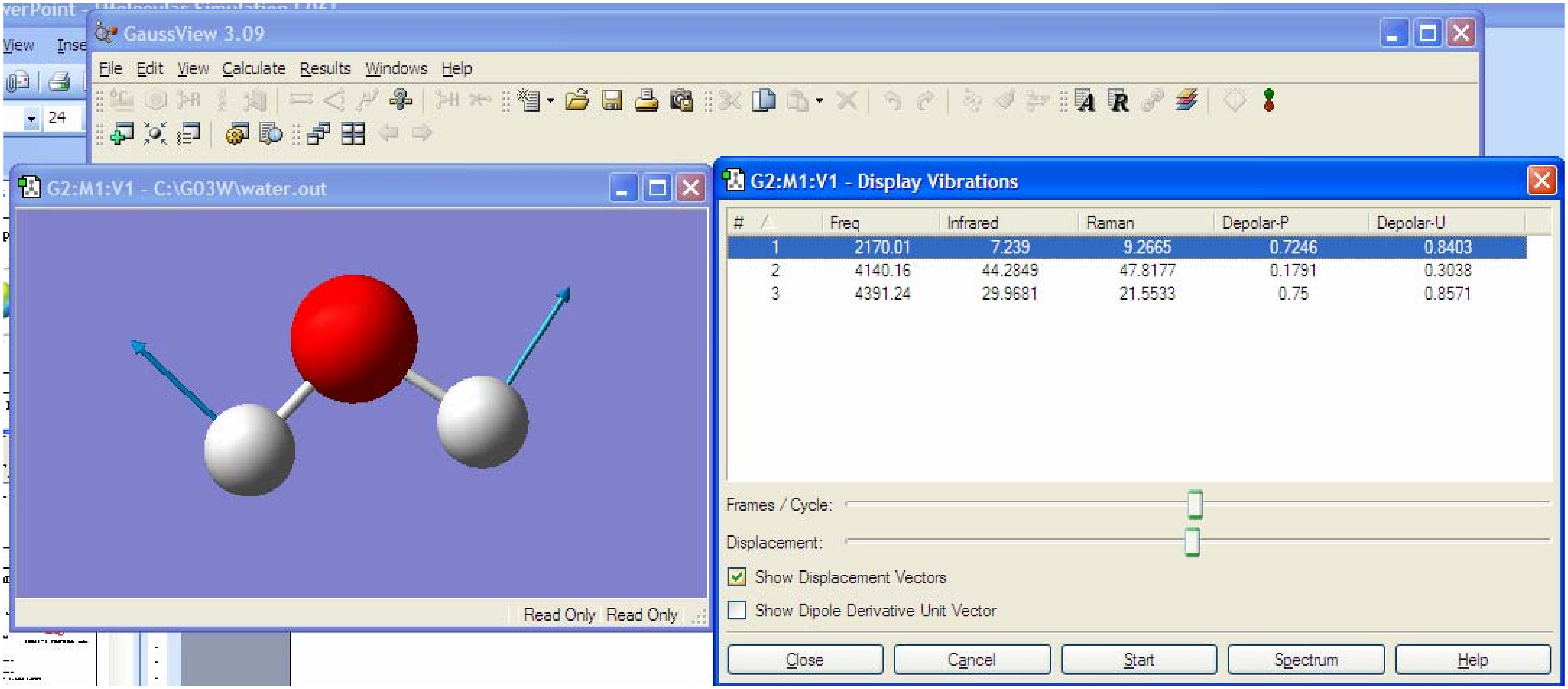


Water optimization and frequency

File Type	.log	
Calculation Type	FREQ	
Calculation Method	RHF	
Basis Set	STO-3G	
E(RHF)	-74.96590122	a.u.
RMS Gradient Norm	0.00001150	a.u.
Imaginary Freq	0	
Dipole Moment	1.7092	Debye
Point Group	C2V	

Job cpu time: 0 days 0 hours 0 minutes 11.0 seconds.

GaussView – Water Example



#	Freq	Infrared	Raman	Depolar-P	Depolar-U
1	2170.01	7.239	9.2665	0.7246	0.8403
2	4140.16	44.2849	47.8177	0.1791	0.3038
3	4391.24	29.9681	21.5533	0.75	0.8571

Frames / Cycle:

Displacement:

☑ Show Displacement Vectors

☐ Show Dipole Derivative Unit Vector

Limitations, Strengths & Reliability

- Limitations
  - Requires more CPU time
  - Can treat smaller molecules
  - Calculations are more complex
  - Have to worry about electronic configuration
- Strengths
  - No experimental bias
  - Can improve a calculation in a logical manner (e.g. basis set, level of theory,…)
  - Provides information on intermediate species, including spectroscopic data
  - Can calculate novel structures
  - Can calculate any electronic state
- Reliability
  - The mean deviation between experiment and theory for heavy-atom bond lengths in two-heavy-atom hydrides drops from 0.082 A for the RHF/STO-3G level of theory to just 0.019 A for MP2/6-31G(d).
  - Heats of hydrogenation of a range of saturated and unsaturated systems are calculated sufficiently well at the Hartree-Fock level of theory with a moderate basis set (increasing the basis set from 6-31G(d) to 6-31G(d,p) has little effect on the accuracy of these numbers).
  - Inclusion of electron correlation is mandatory in order to get good agreement between experiment and theory for bond dissociation energies (MP2/6-31G(d,p) does very well for the one-heavy-atom hydrides).
- http://www.chem.swin.edu.au/modules/mod5/limits.html

Summary

- What can you do with electronic structure methods?
  - Geometry optimizations (minima and transition states)
  - Energies of minima and transition states
  - Chemical reactivity
  - IR, UV, NMR spectra
  - Physical properties of molecules
  - Interaction energy between two or more molecules

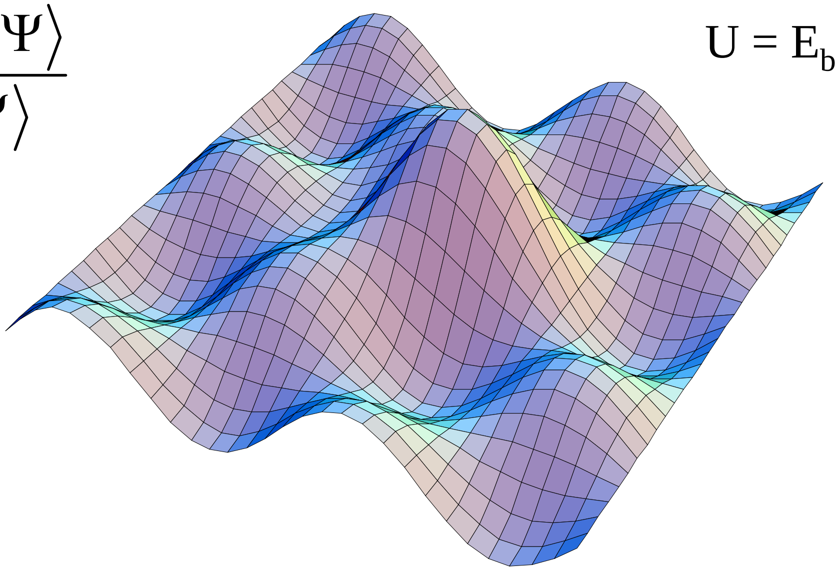
Molecular Simulation II

Quantum Chemistry

$$E = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle}$$

Classical Mechanics

$$U = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{non-bond}}$$



Jeffry D. Madura

Department of Chemistry & Biochemistry

Center for Computational Sciences

Duquesne University

Classical Mechanical Treatment

I. Classical Mechanics
  a. Implicit treatment of electrons
  b. Use simple analytical functions (i.e., harmonic springs)
  c. Use Cartesian coordinates, not the z-matrix

II. Force Fields
  a. Have evolved over time
  b. Use different analytical terms and parameters
  c. Are specific for classes of molecules (proteins, carbohydrates, nucleic acids, organic molecules, etc.)

Force Field

- What is a force field?
  - A mathematical expression that describes the dependence of the energy of a molecule on the coordinates of the atoms in the molecule
  - Also this sometimes used as another term for potential energy function.
- What are force fields used for?
  - Structure determination
  - Conformational energies
  - Rotational and Pyramidal inversion barriers
  - Vibrational frequencies
  - Molecular dynamics

Force Field History

- Pre-1970
  - Harmonic
- 1970
  - For molecules with less than 100 atoms one class of force fields went for high accuracy to match experimental results
  - The other class of force fields was for macromolecules.
- Present
  - There are highly accurate force fields designed for small molecules and there are force fields for studying protein and other large molecules

Force Field

- First force fields developed from experimental data
  - X-ray
  - NMR
  - Microwave
- Current force fields have made use of quantum mechanical calculations
  - CFF
  - MMFF94
- There is no single “best” force field

Force Fields

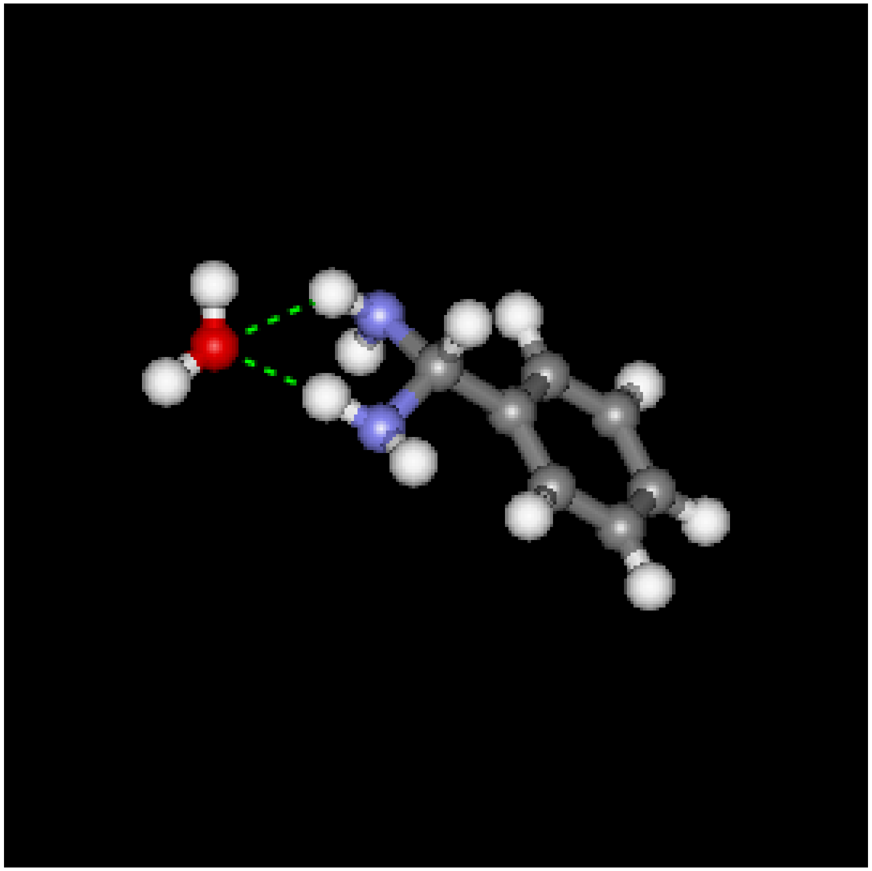
- MM2/3/4: Molecular Mechanic Force field for small moelcules
- CHARMM: Chemistry at Harvard Macromolecular Mechanics
- AMBER: Assisted Model Building with Energy Refinement
- OPLS: Optimized Parameters for Liquid Simulation
- CFF: Consistent Force Field
- CVFF: Valence Consistent Force Field
- MMFF94: Merck Molecular Force Field 94
- UFF: Universal Force Field

Comparison and Evaluation

- Engler, E. M., J. D. Andose and P. v. R. Schleyer (1973) "Critical Evaluation of Molecular Mechanics," J. Am. Chem. Soc. 95, 8005-8025. Gundertofte, K., J. Palm, I. Pettersson and A. Stamvik (1991) "A Comparison of Conformational Energies Calculated by Molecular Mechanics (MM2(85), Sybyl 5.1, Sybyl 5.21, ChemX) and Semiempirical (AM1 and PM3) Methods," J. Comp. Chem. 12, 200-208.
- Gundertofte, K., T. Liljefors, P-O Norrby, I. Pattesson (1996) "Comparison of Conformational Energies Calculated by Several Molecular Mechanics Methods," J. Comp. Chem. 17, 429-449 (1996).
- Hall, D., and N. Pavitt (1984) "A Appraisal of Molecular Force Fields for the Representation of Polypeptides," J. Comp. Chem. 5, 441-450.
- Hobza, P., M. Kabelac, J. Sponer, P. Mejzlik and J. Vondrasek (1997) "Performance of Empirical Potentials (AMBER, CFF95, CVFF, CHARMM, OPS, POLTEV), Semiemprical Quantum Chemical Methods (AM1, MNDO/M, PM3) and ab initio Hartree-Fock Method for Interaction of DNA Bases: Comparison of Nonempirical Beyond Hartree-Fock Results," J. Comp. Chem. 18, 1136-1150.
- Kini, R. M., and H. J. Evans (1992) "Comparison of Protein Models Minimized by the All-Atom and United Atom Models in the AMBER force Field," J. Biomol. Structure and Dynamics 10, 265-279.
- Roterman, I. K., Gibson, K. D., and Scheraga, H. A. (1989) "A Comparison of the CHARMM, AMBER, and ECEPP/2 Potential for Peptides I." J. Biomol. Struct. and Dynamics 7, 391-419.
- Roterman, I. K., Lambert, M. H., Gibson, K. D., and Scheraga, H. A. (1989) "A Comparison of the CHARMM, AMBER, and ECEPP/2 Potential for Peptides II." J. Biomol. Struct. and Dynamics 7, 421-452.
- Whitlow, M., and M. M. Teeter (1986) "A Empirical Examination of Potential Energy Minimization using the Well-Determined Structure of the Protein Crambin," J. Am. Chem. Soc. 108, 7163-7172.

Potential Energy Function

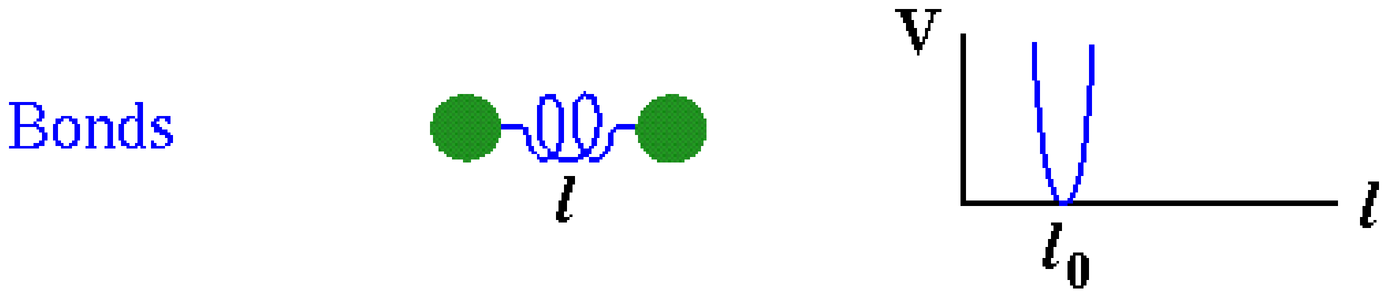
The potential energy function is a mathematical model which describes the various interactions between the atoms of a molecule or system of molecules. In general, the function is composed of *intramolecular terms* (1st three terms) and *intermolecular terms* (last two terms).



$$
\begin{aligned}
U(r) = {} & \frac{1}{2}\sum(b-b_{\circ})^2 + \\
& \frac{1}{2}\sum(\theta-\theta_{\circ})^2 + \\
& \frac{1}{2}\sum_j V_j[1+(-1)^{j+1}\cos(n_j\phi-\delta)]+ \\
& 4\varepsilon[(\frac{\sigma}{r_{ij}})^{12}-(\frac{\sigma}{r_{ij}})^{6}]+ \\
& \frac{q_i q_j}{4\varepsilon r_{ij}}
\end{aligned}
$$

Bond Stretch

$$E_l = k_l \left(l - l_0 \right)^2$$

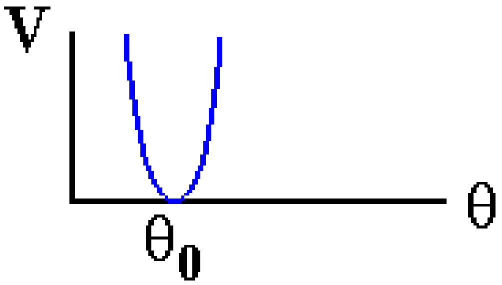


- Harmonic approximation is used
  - – k_b is the force constant
  - – l_0 is the reference bond length
- Higher order terms

$$E_l = k_l \left(l - l_0 \right)^2 + k_l^{'} \left(l - l_0 \right)^3 + k_l^{''} \left(l - l_0 \right)^4$$

Angle Bending

$$E_\theta = k_\theta\left(\theta - \theta^0\right)^2$$

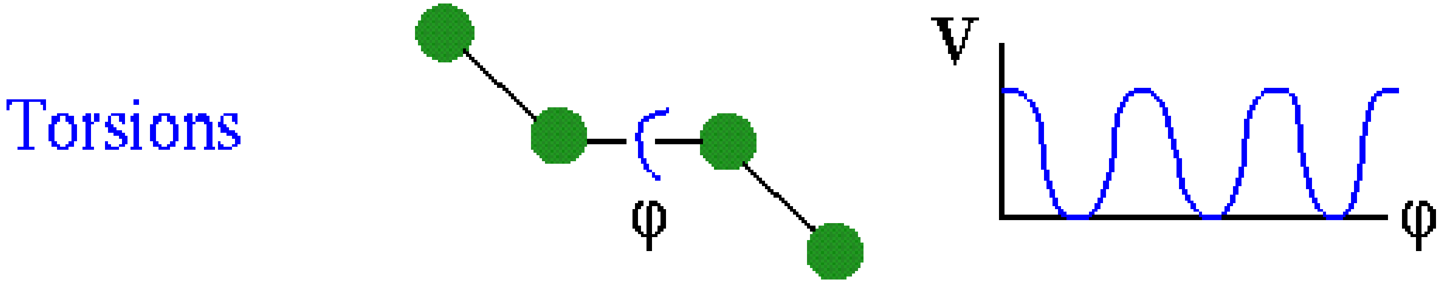


- Harmonic approximation
  - k_θ is the bending force constant
  - θ^0 is the reference angle
- Other forms include

$$E_\theta = k_\theta\left(1 + \cos\theta\right)$$

Torsion Interactions

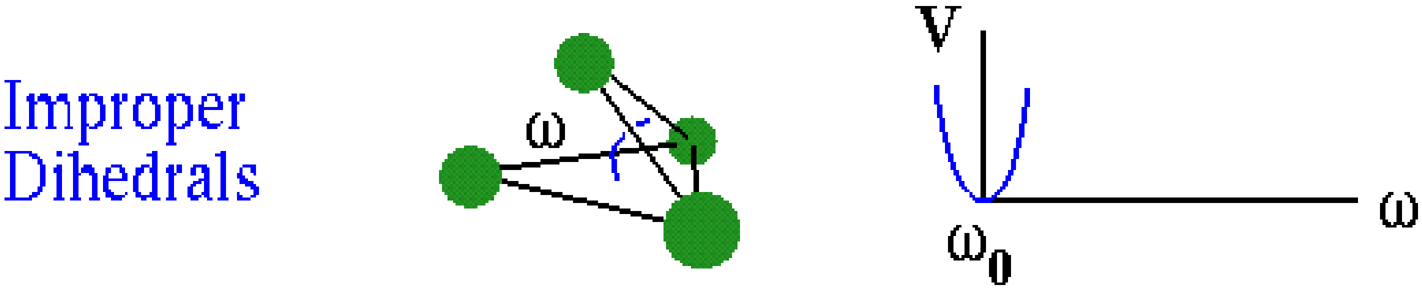
$$E_\varphi = V_1\left(1+\cos\varphi\right)+V_2\left(1-\cos 2\varphi\right)+V_3\left(1+\cos 3\varphi\right)$$



- Represented as a Fourier series
  - This term accounts for the energetics of twisting the 1-4 atoms
  - First term: important for describing the conformational energies (cis-trans)
  - Second term: important for determining the relatively large barrier to rotation about conjugated bonds
  - Third term: allows for accurate of the energy barrier for rotation about bonds where one or both of the atoms in the bond have sp^3 hybridization

Out-of-plane Bending

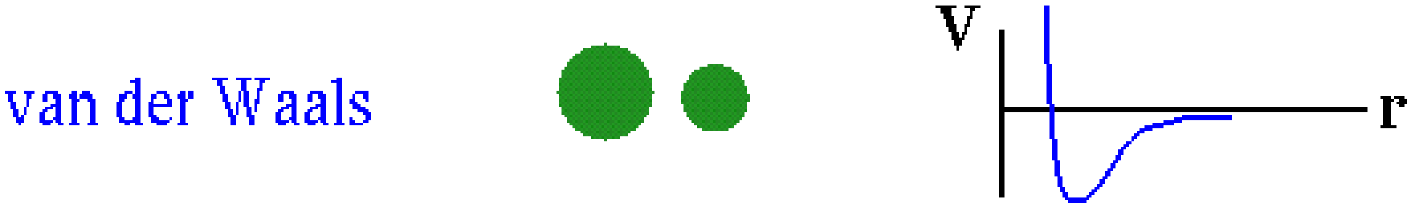
$$E_\omega = k_\omega \left(\omega - \omega_0\right)^2$$



- Harmonic approximation
  - k_ω is the oop force constant
  - ω^0 is the reference value
- Diffferent methods in which to calculate ω
  - MMF: angle between a bond *i*-*j* and a plane formed by *j*-*k*-*l* and *j* is the central atom
  - MM3: angle between a bond *i*-*j* and a point located in the place formed by *i*-*k*-*j*.

Van Der Waals Interactions

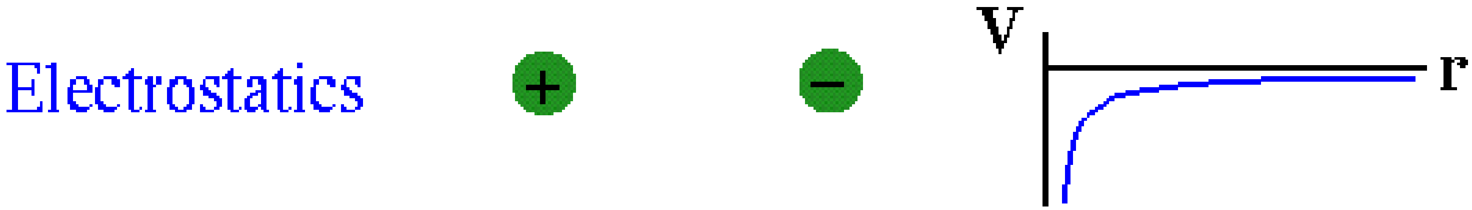
$$E_{vdw} = \varepsilon\left[\left(\frac{R_{ij}^{*}}{R_{ij}}\right)^{12} - 2\left(\frac{R_{ij}^{*}}{R_{ij}}\right)^{6}\right]$$



- Lennard-Jones 12-6 potential
  - ε is the well depth
  - R_{ij}^{*} is the sum of the van der Waals radii (of atoms *i* and *j*)
  - R_{ij} is the distance between interacting atoms

Electrostatics

$$E_{elec} = \frac{q_i q_j}{D R_{ij}}$$



- Coulomb’s law
  - q_i and q_j are the charges on atom i and j respectively
  - D is the dielectric constant
  - R_{ij} is the distance between atoms i and j
- Bond increment model (used in CFF and MMFF)

$$q_i = q_i^0 + \sum_j \delta_{ij}$$

Charge Classes

- Class I
  - Calculated directly from experiment
- Class II
  - Extracted from a quantum mechanical wave function (Mulliken analysis)
- Class III
  - Extracted from a wave function by analyzing a physical observable predicted from the wave function. (Electrostatic fitting)
- Class IV
  - A parameterization procedure to improve class II and III charges by mapping them to reproduce charge-dependent observables obtained from experiment

Cross Terms

$$E_{bb} = k_{bb}\left(b_1 - b_1^0\right)\left(b_2 - b_2^0\right)$$

$$E_{\theta\theta} = k_{\theta\theta}\left(\theta_1 - \theta_1^0\right)\left(\theta_2 - \theta_2^0\right)$$

$$E_{b\theta} = k_{b\theta}\left(b - b^0\right)\left(\theta - \theta^0\right)$$

- Bond/bond
  - Needed to get the correct splitting in the vibrational frequencies of the symmetric and asymmetric C-H bond stretching modes
- Angle/angle
  - Needed to determine correctly the extent of splitting in angular deformation modes for the cases in which the bending modes are centered on a single atom
- Bond/angle
  - Needed to predict the observed bond lengthening that often occurs when a bond angle is reduced

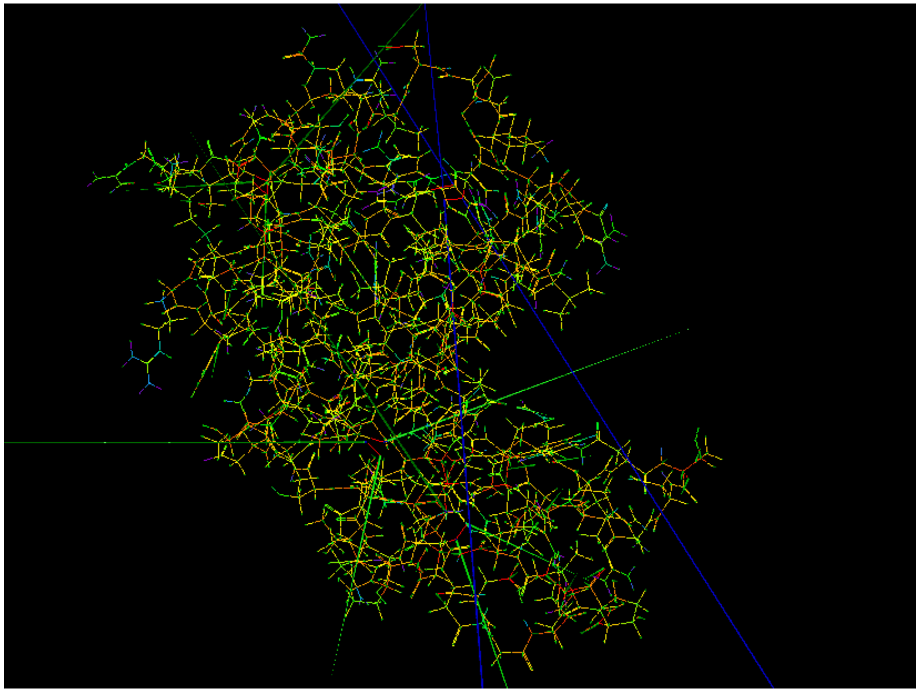
Molecular Mechanics

In the molecular mechanics model, a molecule is described as a series of point charges (atoms) linked by springs (bonds). A mathematical function (the force-field) describes the freedom of bond lengths, bond angles, and torsions to change. The force-field also contains a description of the van der Waals and electrostatic interactions between atoms that are not directly bonded. The force-field is used to describe the potential energy of the molecule or system of interest. Molecular mechanics is a mathematical procedure used to explore the potential energy surface of a molecule or system of interest.

Force

$$F = -\nabla U$$

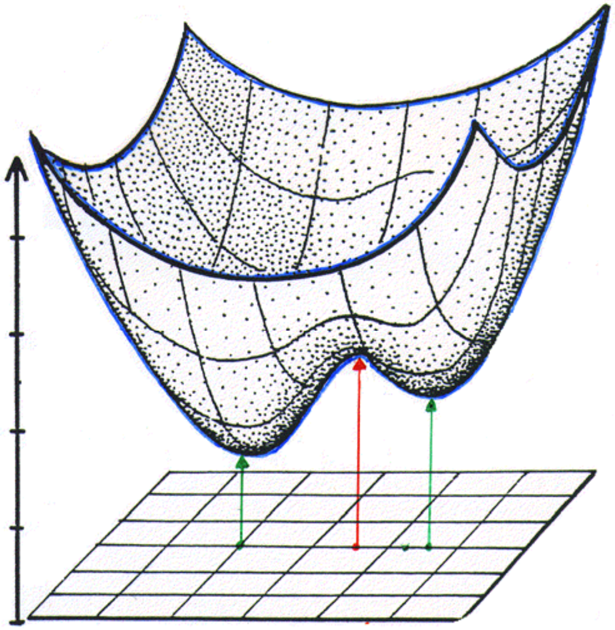
Potential Energy



Potential Energy Minimizations

- **Potential Energy Surface**: Has minima (stable structures) and saddle points (transition states).

  Below: 2 minima & 1 Saddle Point.

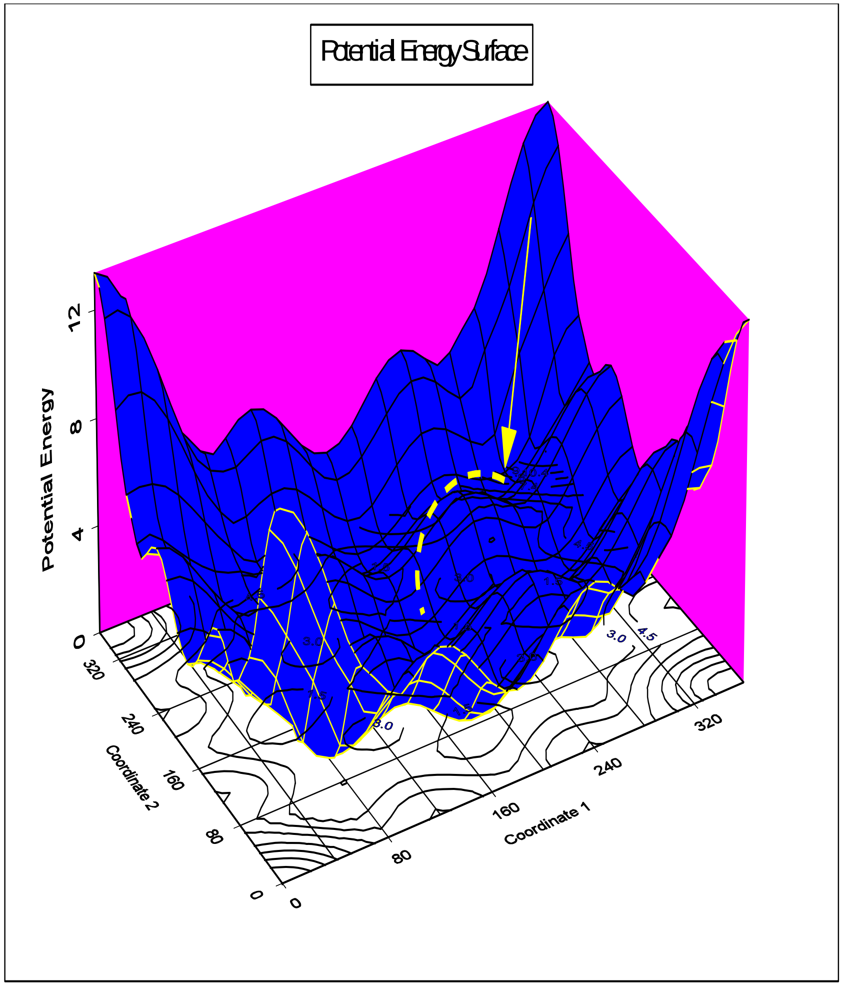


Energy Minimization

Given a function *f* which depends on one or more independent variables, x_1, x_2, …, find the values of those variables where *f* has a minimum value.

$$\frac{\partial f}{\partial x_i} = 0$$

$$\frac{\partial^2 f}{\partial x^2} > 0$$



Energy Minimization Methods

- Taylor series expansion about point x_k

$$U(x) = U(x_k) + (x - x_k)\frac{\partial U(x_k)}{\partial x_k} + \frac{(x - x_k)^2}{2}\frac{\partial^2 U(x_k)}{\partial x_k \partial x_j} + \cdots$$

  - the second term is known as the **gradient** (force)
  - the third term is known as the **Hessian** (force constant)
- Algorithms are classified by **order**, or the highest derivative used in the Taylor series.
- <u>Common algorithms</u> (1st Order): Steepest Descent (SD), Conjugated Gradients (CONJ)
- Non-derivative
  - Simplex
  - Sequential univariate method

Energy Minimization Methods

- Derivative
  - Steepest descents
    - Moves are made in the direction parallel to the net force
  - Conjugate gradient
    - The gradients and the direction of successive steps are orthogonal
  - Newton-Raphson
    - Second-order method; both first and second derivatives are used
  - BFGS
    - Quasi-Newton method (a.k.a. variable metric methods) build up the inverse Hessian matrix in successive iterations

Energy Minimization Methods

- Truncated Newton-Raphson
  - Initially follow a descent direction and near the solution solve more accurately using a Newton method.
  - Different from QN in that the Hessian is sparse allowing for a faster evaluation

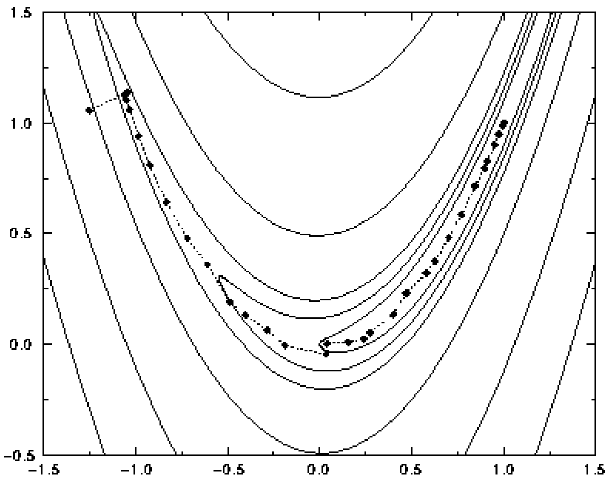


Figure 11: BFCS quasi-Newton minimization path for the two-dimensional Rosenbrock function.

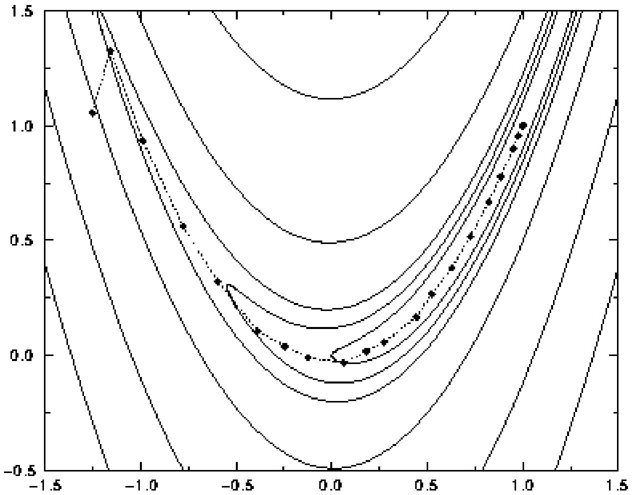


Figure 12: Truncated Newton minimization path for the two-dimensional Rosenbrock function.

Comparing 1st Order Algorithms

BOTH: iterate over the following equation in order to perform the minimization: $\mathbf{R}_k = \mathbf{R}_{k-1} + l_k\, \mathbf{S}_k$

Where $\mathbf{R}_k$ is the new position at step *k*,

$\mathbf{R}_{k-1}$ in the position at the previous step *k-1*,

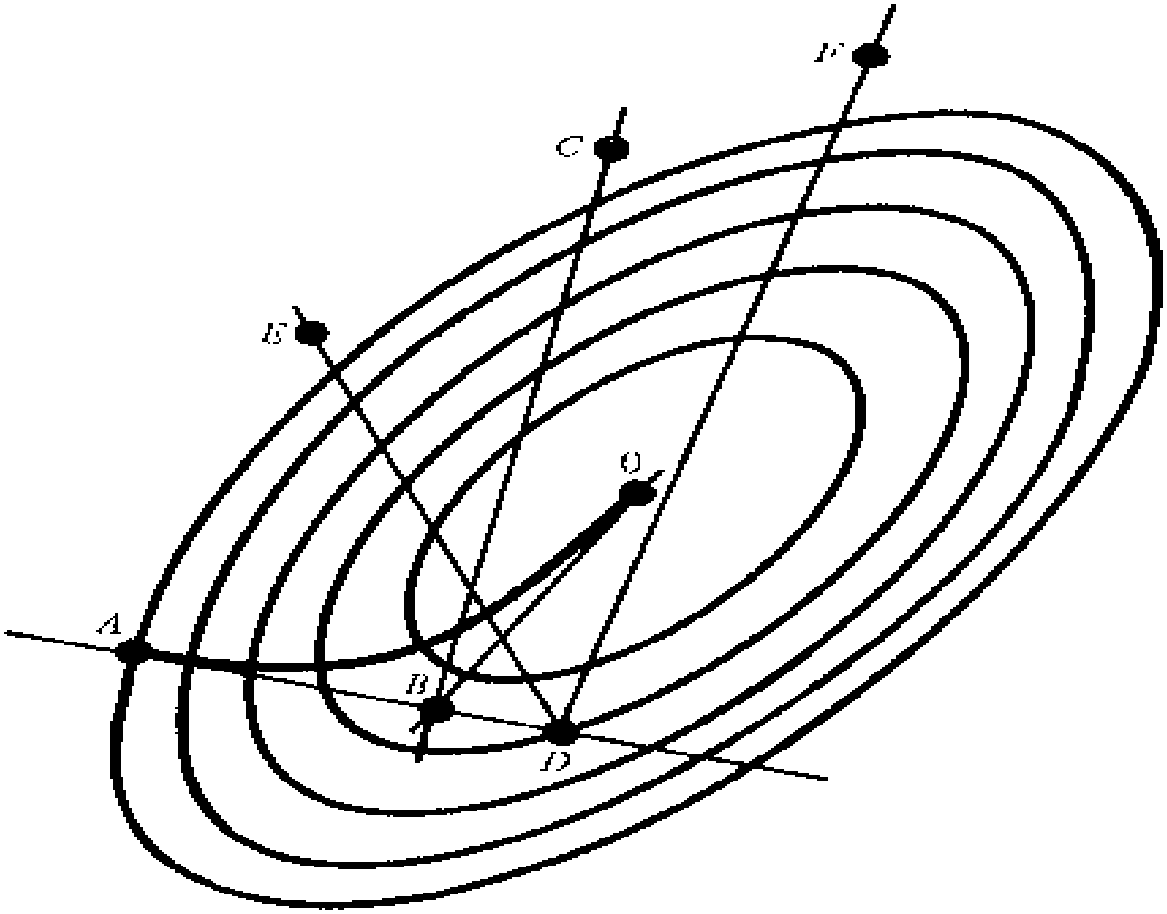
lk is the size of the step to be taken at step *k* and $\mathbf{S}_k$ is the direction.

SD: At each step the gradient of the potential $\mathbf{g}_k$ (i.e., the first derivative in multi-dimensions) is calculated and a displacement is added to all the coordinates in a direction opposite to the gradient. $\mathbf{S}_k = -\mathbf{g}_k$

CONJ: In each step, weighs in the previous gradients to compensate for the lack of curvature information.

For all steps *k* > 1 the direction of the step is a weighted average of the current gradient and the previous step direction, i.e.,

$$\mathbf{S}_k = -\mathbf{g}_k + b_k\, \mathbf{S}_{k-1}$$



***SD versus CONJ*.**

Starting from point A, **SD** will follow a path **A-B-C**. **CONJ** will follow **A-B-O** because it modifies the second direction to take into account the previous gradient along A-B and the current gradient along B-C.

Comparison of Methods

- Convergence
  - Small change in energy
  - Small norm of the gradient
  - RMS gradient

$$|grad| = \sqrt{\sum_i \left(\frac{dU}{dx_i}\right)}$$

$$RMS = \frac{|grad|}{\sqrt{n}}$$

- Number of steps vs. time
  - Steepest descents: 500 steps in 41.08 secs (not converged)
  - Conjugate gradient: 72 steps in 15.77 seconds
  - Newton-Raphson: 15 steps in 14.84 seconds

Which Method Should I Use?

- Must consider
  - Storage: Steepest descents little memory needed while Newton-Raphson methods require lots.
  - Availability of derivatives: Simplex, none are needed, steepest descents, only first derivatives, Newton-Raphson needs first and second derivatives.

- The following is common practice
  - SD or CG for the initial “rough” minimization followed by a few steps of NR.
  - SD is superior to CG when starting structure is far from the minimum
  - TN method after a few SD and/or CG appears to give the “best” overall and fastest convergence

Conformational Analysis

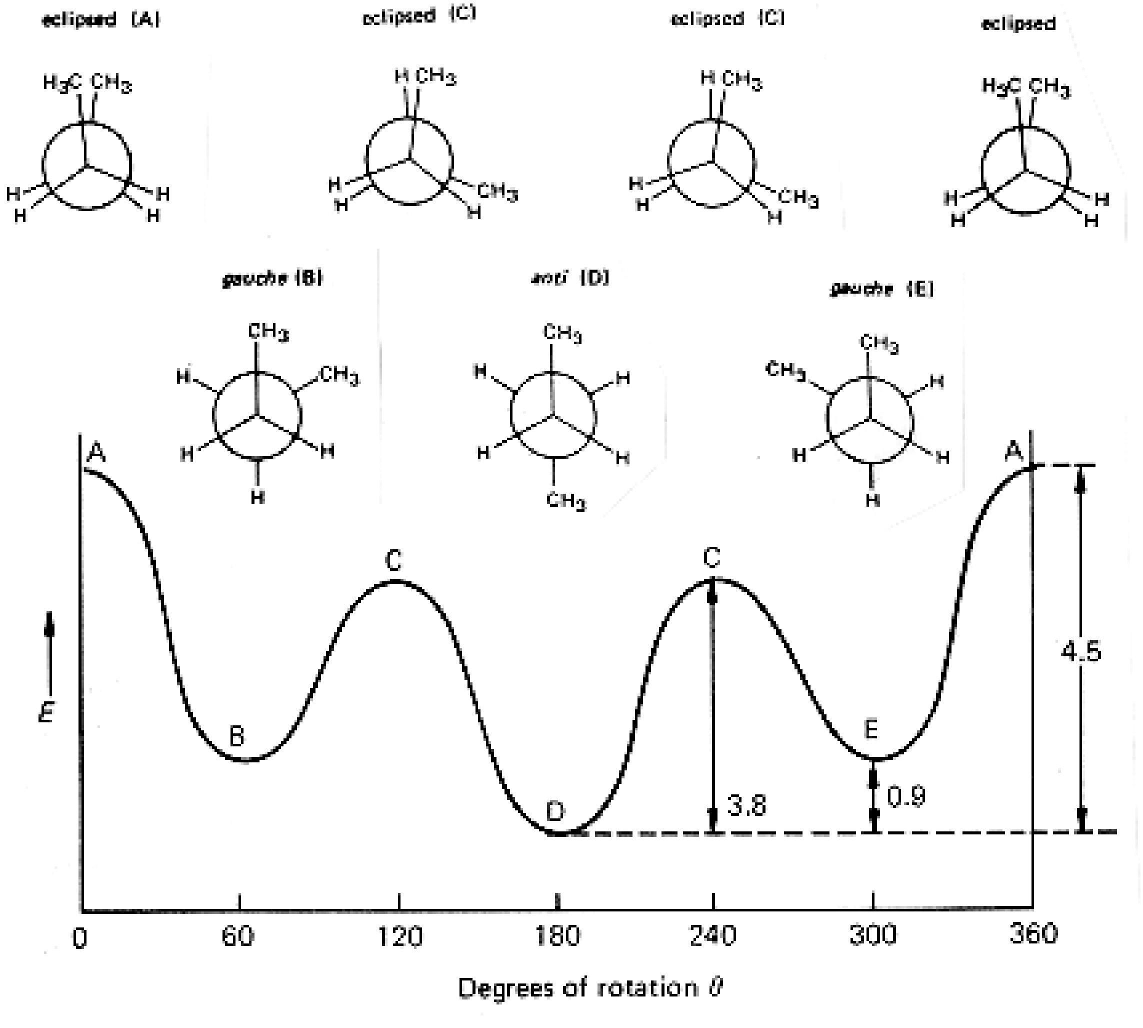
- Molecular conformations
  - term used to describe molecular structures that interconvert under ambient conditions.
  - this implies several conformations may be present, in differing conc., under ambient conditions.
  - a proper description of “the” molecular structure, “the” molecular energy, or “the” spectrum for a molecule with several conformations must comprise a proper weighting of all of the conformations.

Boltzmann’s equation

$$P_i = \frac{f_i e^{-\frac{E_i}{RT}}}{\sum_j f_j e^{-\frac{E_j}{RT}}}$$

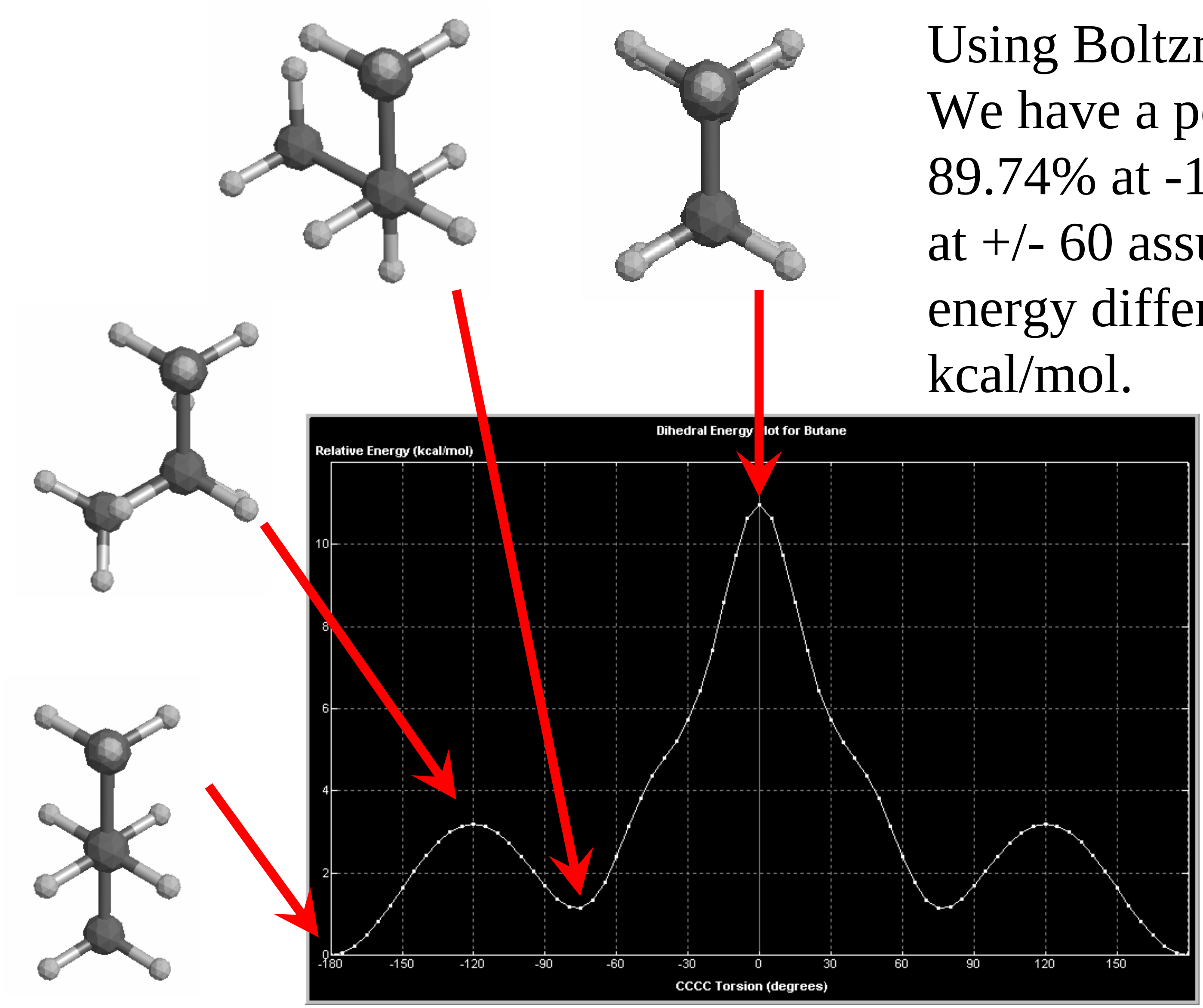
- f_i is the number of states or conf. of energy E_i
- R is 1.98 cal/mol-K (the ideal gas constant)
- T is the absolute temperature (K)
- j is the summation over all the conformations

Butane Conformational Analysis



Conformational Analysis Example

Using Boltzmann's equation We have a population of 89.74% at -180 and 10.26% at +/- 60 assuming a relative energy difference of 1.7 kcal/mol.



Conformational Analysis: A Cautionary Note

	MM2			**Dreiding**		
Term	*Trans*	*Gauche*	*ΔE*	*Trans*	*Gauche*	*ΔE*
Stretch	0.15	0.16	0.01	0.33	0.38	0.05
Stretch-Bend	0.05	0.07	0.02			
Bend	0.29	0.63	0.34	0.51	1.15	0.64
Torsion	0.01	0.44	0.43	0.01	0.11	0.01
VDW	1.68	1.75	0.07	3.59	3.59	0.00
Total	2.18	3.05	**0.87**	4.44	5.23	**0.79**

Even though the energetic difference given by the two models is similar, different contributions give rise to those differences.

Molecular Mechanics Energetics

- Steric energy
  - the energy reported by most molecular mechanics programs
  - energy of structure at 0 K.
  - correct for vibrational motion by adding the zero point energy

$$ZPE = \frac{1}{699.5}\sum_i \nu_i$$

  - MM energy is NOT equal to free energy!!
  - MM energy can be equivalent to enthalpy if one assumes the PV term can be ignored

Molecular Mechanics Energy

- Strain energy
  - Differences in steric energy are oly valid for different conformations or configurations.
  - Strain energy permits the comparison between different molecules.
  - A “strainless” reference point must be determined.
  - A particular reference point might be the all trans conformations of the straight-chain alkanes from methane to hexane (Allinger definition).
  - These compounds can be used to derive a set of strainless energy parameters for constituent parts of molecules.
  - Subtracting the strainless energy from the steric energy Allinger and co-workers concluded that the chair cyclohexane has an inherent strain energy due to the presence of 1,4 van der Waals interactions between the carbon atom within the ring.

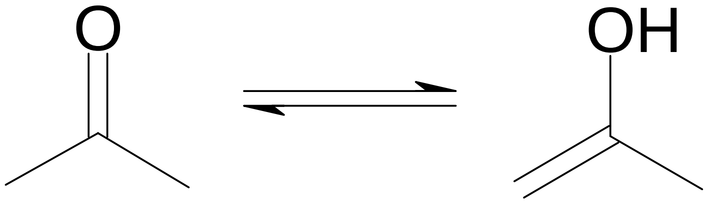
Molecular Mechanics Energy

- Interaction energy
  - This is the difference between the energies of two isolated species and the energy of the intermolecular complex
  - Mathematically this is represented as

$$E_{ie} = E_{ab} - \left(E_a + E_b \right)$$

Steric Energies

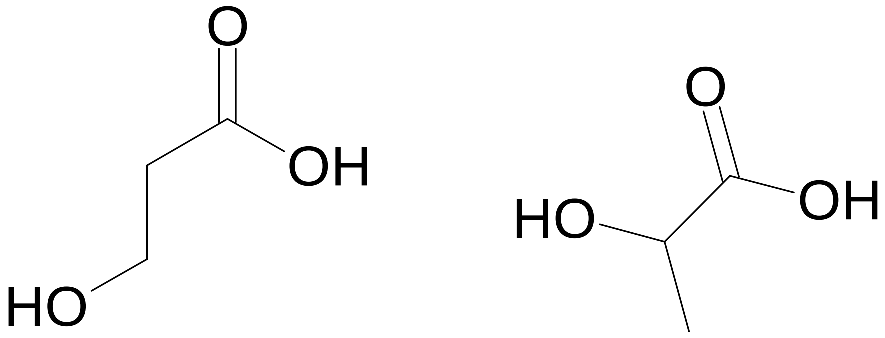
- Using steric energies to predict the thermodynamics of simple tautomerization



- The experimental heat of formation difference is approximately 8 kcal/mol
- MM2 steric energy difference is 2.3 kcal/mol
  - The large error is due to error in bond energy terms, I.e. the number/types of bonds broken and made are not precisely balanced.

Positional Isomers

- In this case the number and precise types of bonds are retianed.
- Consider hydracrlic acid vs. lactic acid.



- MM2 energy for hydracrylic acid is –13.23 kcal/mol while that for lactic acid is –2.37 kcal/mol yielding an energy difference of 10.86 kcal/mol in favor of hydracrylic acid.
- Experimentally the heat of formation difference is –4.4 kcal/mol in favor of lactic acid.

Conformers and Configurational Isomers

- Molecular mechanics can be employed to predict relative energies of conformers and configurational isomers.
- It should not be used on positional isomers or the relative energies of different molecules.
- Molecular mechanics can be used to study the binding between molecules is intermolecular interactions have been appropriately parameterized.

Ideal Gas Statistical Thermodyanmics

- The free energy can be written as

$$G = -kT \ln Q + PV$$

  - where

$$Q = \sum_i f_i e^{-\frac{E_i}{RT}}$$

- The difference in free energy can be written as

$$\Delta G = -RT \ln \frac{P_1}{P_2}$$