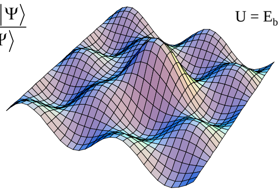
Molecular Simulation I

Quantum Chemistry

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



Classical Mechanics

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

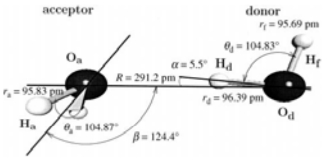
Jeffry D. Madura

Department of Chemistry & Biochemistry

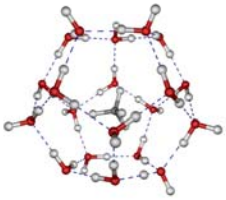
Center for Computational Sciences

Duquesne University

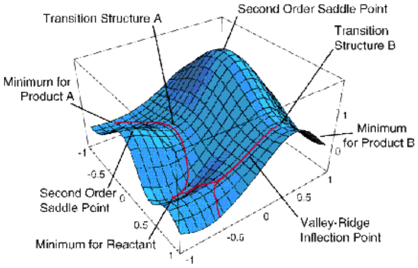
Water



Phys. Chem. Chem. Phys., 2000, 2, 2227-2234



Potential Energy Surface



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

In the beginning…

- 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

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

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_ν 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)

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

- Gaussian type orbitals (GTO)
  - – functional form

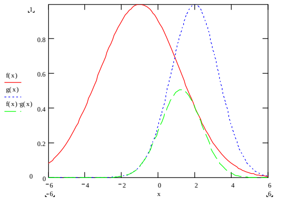
$$x^a y^b z^c e^{-\alpha r^2}$$

  - – zeroth-order Gaussian function

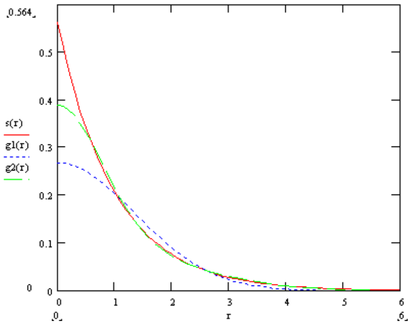
$$g_s(\alpha, r) = \left(\frac{2\alpha}{\pi}\right)^{3/4} e^{-\alpha r^2}$$

- 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

$$e^{-\alpha_m r_m^2} e^{-\alpha_n r_n^2} = e^{-\frac{\alpha_m \alpha_n}{\alpha_m + \alpha_n} r_{mn}^2} e^{-\alpha r_c^2}$$



- STO vs. GTO



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

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

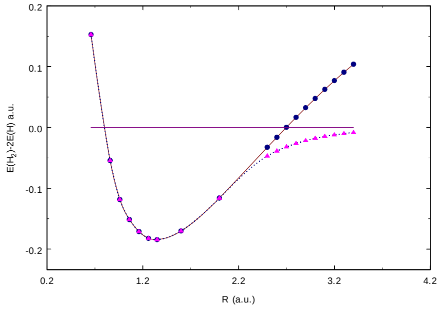
- Minimal basis set
  - STO-NG
- Double zeta basis set
  - linear combination of a ‘contracted’ function and a ‘diffuse’ function.
- Split valence
  - 3-21G, 4-31G, 6-31G

- 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

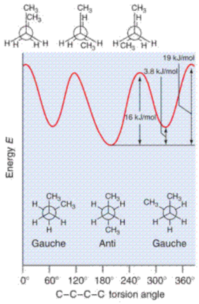
- Derivatives of the energy

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

  - 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.

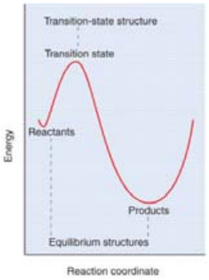
Potential Energy Surfaces



Copyright © 2006 Pearson Education, Inc., publishing as Benjamin Cummings

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

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



Copyright © 2006 Pearson Education, Inc., publishing as Benjamin Cummings

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

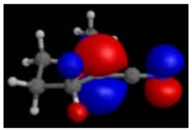
- Electron density at a point r

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

- Number of electrons is

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

- 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

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

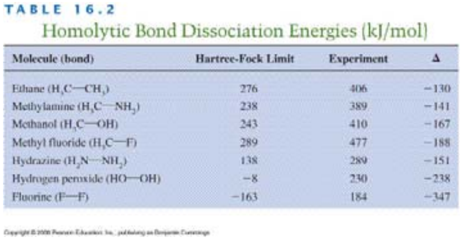


TABLE 16.2

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

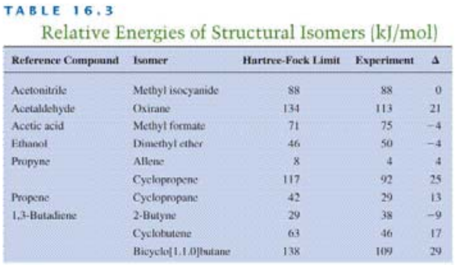


TABLE 16.3

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

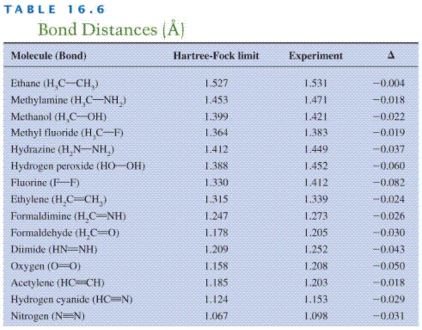


TABLE 16.6

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

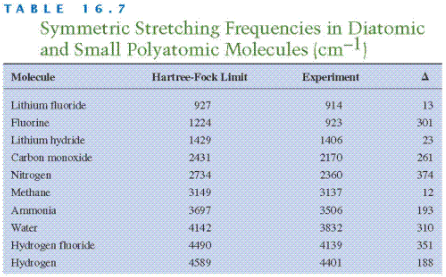


TABLE 16.7

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

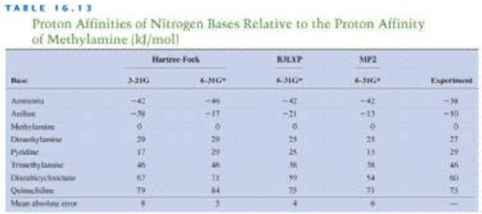


TABLE 16.13

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

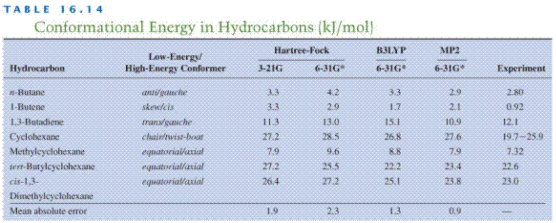


TABLE 16.14

Conformational Energy in Hydrocarbons (kJ/mol)

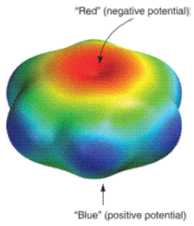
Hydrocarbon	Low-Energy/ High-Energy Conformer	Hartree-Fock 3-21G	Hartree-Fock 6-31G*	B3LYP 6-31G*	MP2 6-31G*	Experiment
n-Butane	*anti/gauche*	3.3	4.2	3.3	2.9	2.80
1-Butene	*skew/cis*	3.3	2.9	1.7	2.1	0.92
1,3-Butadiene	*trans/gauche*	11.3	13.0	15.1	10.9	12.1
Cyclohexane	*chair/twist-boat*	27.2	28.5	26.8	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

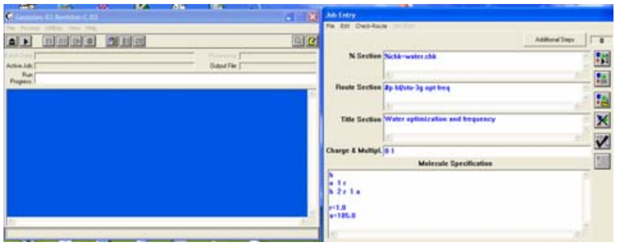
$$\phi_{nucl}\left(r\right)=\sum_{A=1}^{M}\frac{Z_A}{\left|r-R_A\right|} \qquad \phi_{elec}\left(r\right)=-\int\frac{dr'\rho\left(r\right)}{\left|r'-r\right|}$$

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

Electrostatic Potential



G03W - Water Example



% Section: %chk=water.chk

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

Title Section: Water optimization and frequency

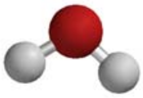
Charge & Multipl.: 0 1

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

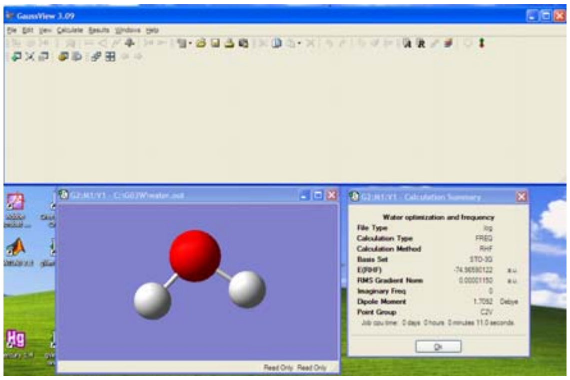
```
E(RHF) =  -74.9658952265     A.U.
```

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

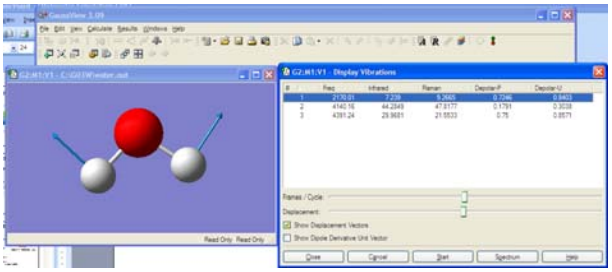


Water optimization and frequency

File Type	log	
Calculation Type	FREQ	
Calculation Method	RHF	
Basis Set	STO-3G	
E(RHF)	-74.96580122	a.u.
RMS Gradient Norm	0.00001150	a.u.
Imaginary Freq	0	
Dipole Moment	1.7052	Debye
Point Group	C2V	

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

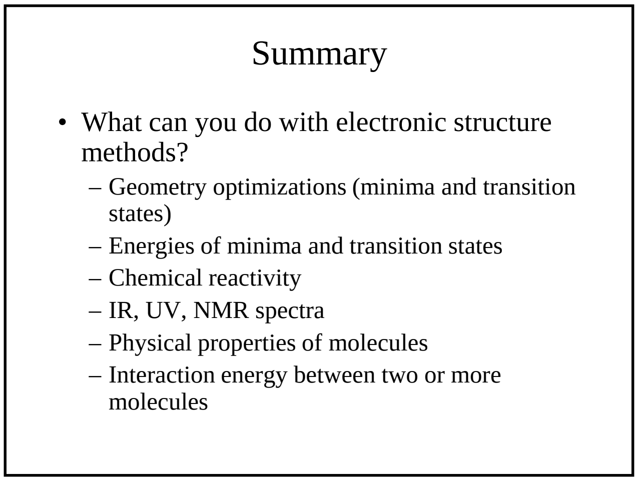
GaussView – Water Example



#	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	25.9681	21.5533	0.75	0.8571

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