Session 1

Introduction to Computational Chemistry

Introduction to Computational Chemistry

Computational (chemistry education)

and/or

(Computational chemistry) education

- First one: Use computational tools to help increase student understanding of material already covered in various courses
- Second one: Teach students about computational chemistry (molecular modeling) itself, in both courses and research projects

Outline

- Why Molecular Modeling?
  - Definitions; New way of doing chemistry
- Molecular Modeling Methods
  1. Molecular Mechanics
  2. Hartree-Fock and post-HF
  3. Semiempirical
  4. Density Functional Theory
- Method Comparisons
  - Accuracy
  - Expense
- Units, Lab Exercise overview

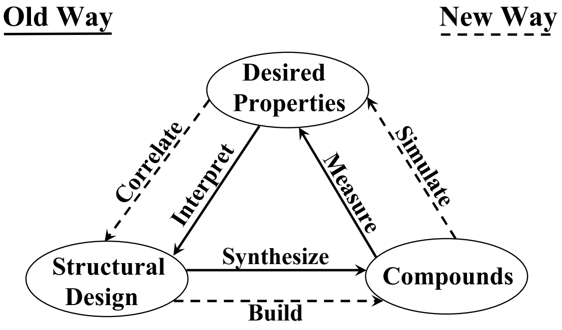
Why Computational Chemistry??

- In 1929, P.A.M. Dirac wrote:
  - “*The underlying physical laws necessary for the mathematical theory of . . .* ***the whole of chemistry*** *are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble.”*
  - Dirac didn’t have access to digital computers, but we do!

Molecular Modeling Defined

- Provides information that is complementary to experimental data on the structures, properties, and reactions of substances
- Largely based on a few algorithms: Schrödinger and Kohn-Sham Equations
- Used to require the use of high performance computers (architecture)
- Modern desktop machines now do what supercomputers did ten years ago
- **Everyone** now has access to this tool!!

Chemistry Today: A Different View



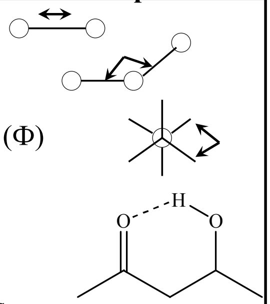
1. Molecular Mechanics

- Apply classical mechanics to molecules
  - → No electrons, no orbital interactions!!
  - – Atoms are spheres with element dependent mass
  - – Bonds are springs that obey Hooke's Law:

    F = -*k*x

    where *k* is the force constant (for a specified bond type between certain atoms)
  - – Other types of springs represent bond angles, dihedral angles, etc.

Molecular Mechanics: Some Components



- Bond stretching (l)
- Bond Angle bending (θ)
- Dihedral Angle rotation (Φ)
- Van der Waals forces
- Hydrogen bonding
- Electrostatic interactions
- Cross terms (stretch-bend, etc.)

Molecular Mechanics: Mathematics

- Bond stretching (MM2):

$$V_r = \tfrac{1}{2} k_l (l - l_0)^2 [1 - B(l - l_0)]$$

- Angle bending:

$$V_\theta = \tfrac{1}{2} k_\theta (\theta - \theta_0)^2 [1 + C(\theta - \theta_0)^4]$$

- Dihedral angle rotation (torsion):

$$V_\varphi = \tfrac{k}{2}(1 + \cos\varphi) + \tfrac{k'}{2}(1 + \cos 2\varphi) + \tfrac{k''}{2}(1 + \cos 3\varphi)$$

- Van der Waals:

$$V_{vdW} = Ae^{-B\rho} - \tfrac{C_6}{\rho^6}$$

Molecular Mechanics: Overall Energy

- Also called “steric” energy
  - Summation of all the terms:

$$V_{steric} = V_{stretch} + V_{bend} + V_{torsion} + V_{vdW} + \ldots$$

- Collection of functional forms and associated constants is called a ***force field***
- BEWARE: “Energies” reported by MM are meaningless (not externally referenced)
  - These values may be useful when comparing conformers of the same molecule

Parameters

~100 elements: N(N+1)/2 = 5050 single bonds

- Multiple bonds: Define atom hybridizations

  ~300 atom types: → 45,150 k_l values!

  - Also need l_0, θ_0, *k’* & *k”* (torsions), and numerous other values for other terms

→To be thorough, would need ~10^8 parameters gathered from experimental data, or from higher level theories!

- Force fields for particular types of molecules
- MM2, MM3 (organics); Amber, CHARMM (biomolecules); others for inorganics, etc.

Molecular Mechanics

- **Advantages:** Very fast, excellent structural results (if compound matches parameter set), works for large molecules)
  - Used to produce a starting geometry
  - Geometry optimization: Move all atoms until sum of all forces on each = 0
- **Disadvantages:** Many compounds do not have good parameters available
  - No orbital information
  - Can’t study reactions, transition states, etc.

Quantum Mechanics

- Electronic structure based on: $\hat{H}\,\Psi = E\,\Psi$

  $\hat{H}$ is known exactly

  Ψ is unknown, except for some simple systems

  - Hydrogen, and hydrogen-like atoms
  - Particle in a box, etc.

$$\hat{H} = -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_k \frac{\hbar^2}{2m_k}\nabla_k^2 - \sum_i\sum_k \frac{e^2 Z_k}{r_{ik}} + \sum_{i<j}\frac{e^2}{r_{ij}} + \sum_{k<l}\frac{e^2 Z_k Z_l}{r_{kl}}$$

→Want to get Ψ, but have to make approximations

Approximations Used

1. **Born-Oppenheimer**: Compared to electrons, nuclei are stationary
   - Electrons move in field of fixed nuclei
2. **Hartree-Fock**: Separate Ψ (many electron wavefunction) into series of one electron spin orbitals
3. **LCAO** (Linear Combination of Atomic Orbitals): MO's expressed as linear combinations of single electron atomic orbitals, represented by basis functions

Ab Initio Methods

- Use complete Ĥ and estimated Ψ (symbol ϕ)
- Molecular orbital construction (LCAO):

$$\boldsymbol{\phi} = \sum_{i=1}^{N} a_i \boldsymbol{\varphi}_i$$

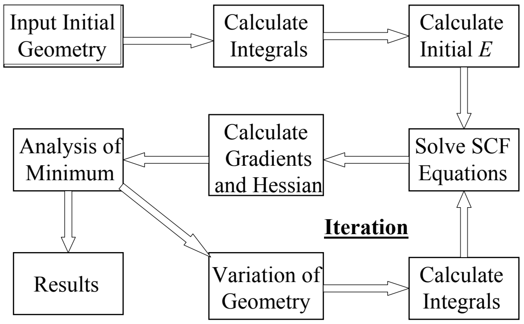
Basis set = set of N functions (GTO's) $\boldsymbol{\varphi}_i$, each associated with a molecular orbital expansion coefficient a_i

- Variational Principle: $(E_\phi \geq E_\psi = E_{\text{experimental}})$
  - Process: Iteratively adjust a_i values until lowest energy (ground state) is found

    → "Self-Consistent Field" approach

Optimization Procedure

- Overview:



Input Initial Geometry

Calculate Integrals

Calculate Initial *E*

Solve SCF Equations

Calculate Gradients and Hessian

Analysis of Minimum

Results

Variation of Geometry

Calculate Integrals

Iteration

Goal

→ Mathematical representation of chemical reality

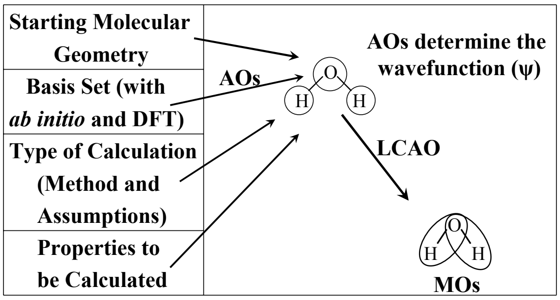
- Basis set should approximate actual wave function sufficiently well to give chemically meaningful results
- Using more complex basis sets improves results at the cost of added computational expense

- Question: *Quantitative* or *qualitative* results?
  - Always a trade-off between accuracy and computational cost

Computational Chemistry Map

Chemist Decides:

Computer calculates:



Starting Molecular Geometry

Basis Set (with *ab initio* and DFT)

Type of Calculation (Method and Assumptions)

Properties to be Calculated

AOs

AOs determine the wavefunction (ψ)

LCAO

MOs

2. Hartree-Fock Method

- ϕ given by Slater determinant
  - Each electron moves in an average electric field produced by all the other electrons
    - No instantaneous e^-/e^- repulsion is included
    - Result:
    - → e^-/e^- repulsion is overestimated
    - → Energy results are generally not accurate
  - ϕ produced is often "close enough" that some useful information can be obtained
  - DFT methods are now much more popular
  - Method serves as a starting point for more advanced theories

Post HF Methods

- Deals more explicitly with e^-/e^- interactions
  - **Configuration Interaction (CI)**:
    - Improve the wave function by adding in contributions from unoccupied orbitals
    - Extreme computational cost
  - **Møller-Plesset Perturbation Theory**:
    - Simplify $\hat{H}$ so exact wavefunctions and energies are found, then use these values to estimate the wave function and energies for the complete $\hat{H}$
    - Computational cost increases rapidly

Problems with HF Approach

→ Calculated energies are not good

→ Iterative solution process is time consuming

→ N^4 total integrals need to be evaluated

(N = number of basis functions used)

Overcoming HF Problems:

1. **Semiempirical** approach: Ignore part of $\hat{H}$

– Replace some integrals with *parameters* so that calculations better match experimental results

2. **Density Functional Theory** approach
   - Get rid of troublesome wave function altogether
   - Use electron density instead

3. Semiempirical Methods

- Simplifications:
  1. Only look at valence electrons
     - Core electrons subsumed into nucleus
  2. Neglect certain integrals (distance cut-off)
  3. Parameterize other integrals using experimental results
     - Different semiempirical methods are parameterized to reproduce different properties
  4. Use a minimal basis set (3-21G)
  5. Employ a non-iterative solution process

Semiempirical Methods: AM1, PM3, RM1

- Differ in how e^-/e^- repulsion is handled
- Advantages:
  - Fast → Can handle fairly large molecules
  - Good qualitative, some ~ quantitative results
  - Parameters available for solution phase
- Disadvantages:
  - Parameters not available for all atoms
  - Molecule/parameterization set similarity
  - Only properties that are parameterized for
  - Limited to ground state equilibrium geometries

4. Density Functional Theory

- A type of *ab initio* method, but may include some parameterization
  - Hohenberg and Kohn (1964)

    "*The ground state energy E of an N-electron system is a functional of the electronic density* **ρ***, and E is a minimum when evaluated with the exact ground state density*"
  - A function whose argument is also a function is called a *functional*
  - A functional enables a function to be mapped to a number

DFT Process

- The energy is minimized with respect to variations in $\boldsymbol{\rho}$, subject to the constraint of charge conservation:

$$N = \int \boldsymbol{\rho}(\mathbf{r})d\mathbf{r}$$

  - HF-SCF Theory is $4N$ dimensional
  - Electron density is $3N$ dimensional (N = # of e^{-}'s)
    - Get rid of one dimension (spin)
  - Easier to deal with electron density than with wave functions
  - DFT is also an iterative process, as was HF

Differences Between HF and DFT

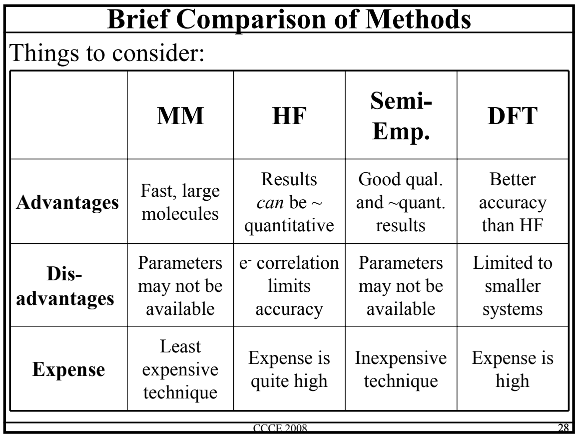
- DFT contains no approximations. It is exact.
  - All we need to know is the *exchange-correlation energy*, E_{xc}, as a function of ρ
    - We must approximate E_{xc}
- HF is a deliberately approximate theory so that we can solve the equations exactly
- So, with DFT our theory is exact and the equations are solved approximately, while with HF the theory is approximate so we can solve the equations exactly

DFT

- Some common functionals:
  - B3LYP, B3PW91, VWN#5, B88LYP, etc.
    - Differ in the way E_{xc} is approximated

→With few exceptions, DFT is the most cost-effective method to achieve a given level of quantitative accuracy

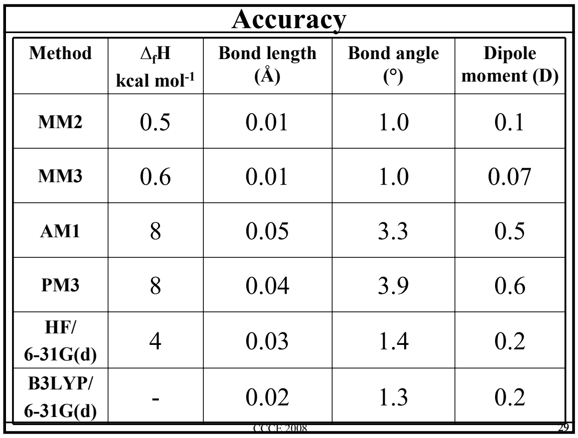
  - Electron correlation included with less expense
  - Basis functions are still used to adjust the electron density



Brief Comparison of Methods

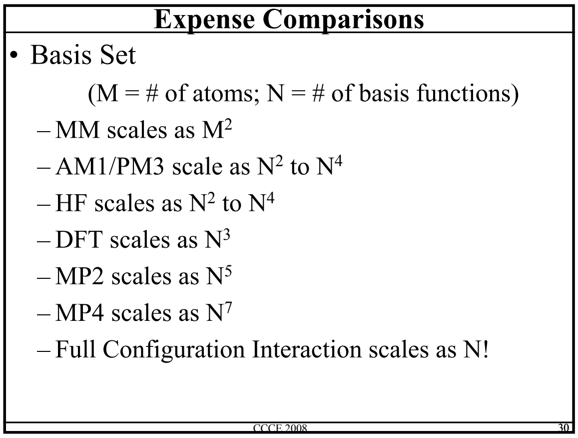
Things to consider:

	MM	HF	Semi-Emp.	DFT
Advantages	Fast, large molecules	Results *can* be ~ quantitative	Good qual. and ~quant. results	Better accuracy than HF
Dis-advantages	Parameters may not be available	e^- correlation limits accuracy	Parameters may not be available	Limited to smaller systems
Expense	Least expensive technique	Expense is quite high	Inexpensive technique	Expense is high



Accuracy

Method	$\Delta_f H$ kcal mol^{-1}	Bond length (Å)	Bond angle (°)	Dipole moment (D)
MM2	0.5	0.01	1.0	0.1
MM3	0.6	0.01	1.0	0.07
AM1	8	0.05	3.3	0.5
PM3	8	0.04	3.9	0.6
HF/ 6-31G(d)	4	0.03	1.4	0.2
B3LYP/ 6-31G(d)	-	0.02	1.3	0.2



Expense Comparisons

- Basis Set

  (M = # of atoms; N = # of basis functions)

  - MM scales as M^2
  - AM1/PM3 scale as N^2 to N^4
  - HF scales as N^2 to N^4
  - DFT scales as N^3
  - MP2 scales as N^5
  - MP4 scales as N^7
  - Full Configuration Interaction scales as N!

Units in Computational Chemistry

- Bohr: Atomic unit of Length (a_0)
  - Equal to the radius of the first Bohr orbit for a hydrogen atom
    - 5.29×10^{-11} m (0.0529 nm, 52.9 pm, 0.529 Å)
- Hartree: Atomic unit of Energy
  - Equal to twice the energy of a ground state hydrogen atom
    - 627.51 kcal/mole
    - 2625.5 kJ/mole
    - 27.211 eV
    - 219474.6 cm^{-1}

Lab Exercises

- Using WebMO

  Web-based GUI for computational chemistry
  - Directions for other software at the CCCE site: (http://www.computationalscience.org/ccce/)
  - Drawing and viewing molecules
  - Running various calculations
    - Determining bond distances and angles
    - Potential energy surface calculations
    - Viewing orbitals