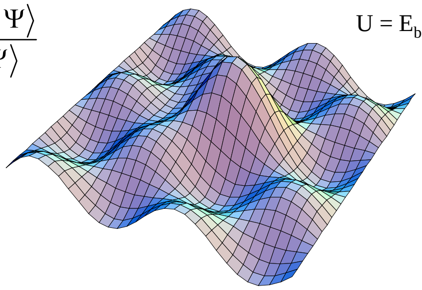
Molecular Simulation II

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

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

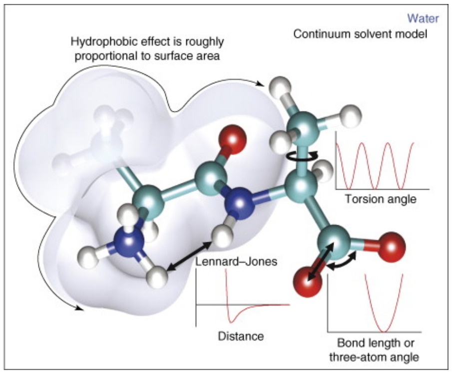
Comparison and Evaluation

- 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).

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



Curr. Opin. Struct. Bio. 2007, 17, 199

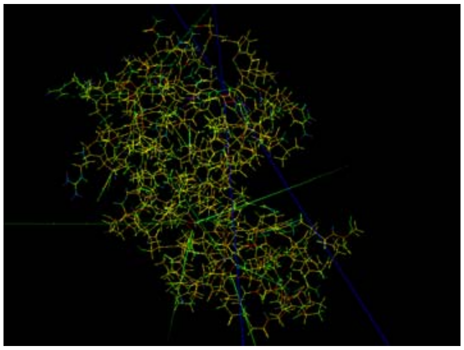
Molecular Mechanics

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

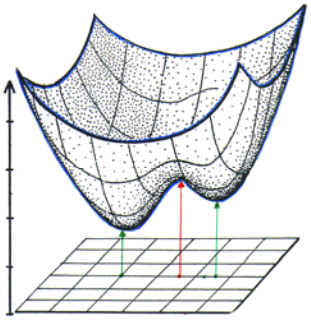
$$F = -\nabla U$$

Potential Energy



Potential Energy Minimizations

- **Potential Energy Surface**: Has minima (stable structures) and saddle points (transition states).
  Below: 2 minima & 1 Saddle Point.

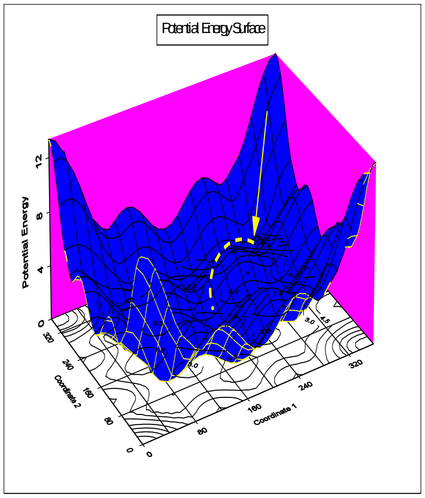


Energy Minimization

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.
- Common algorithms (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

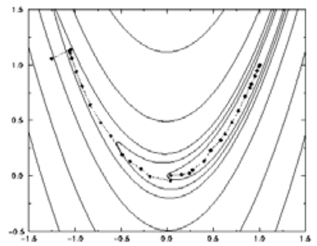


Figure 11: BFCS quasi-Newton minimization path for the two-dimensional Rosenbrock function.

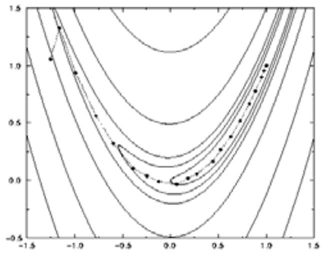


Figure 12: Truncated Newton minimization path for the two-dimensional Rosenbrock function.

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.,

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

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

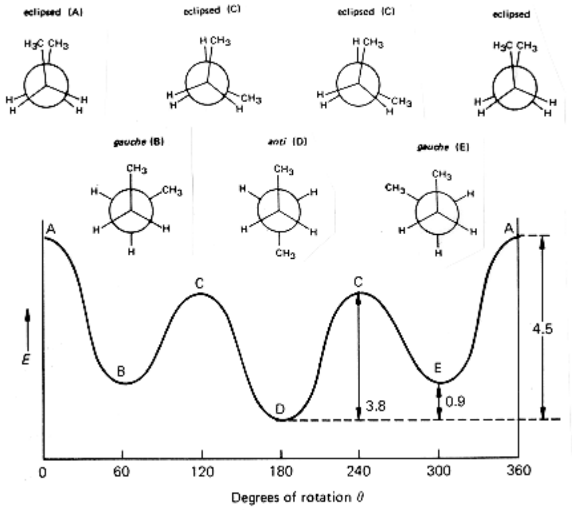
- Molecular conformations
  - term used to describe molecular structures that interconvert under ambient conditions.
  - this implies several conformations may be present, in differing conc., under ambient conditions.
  - a proper description of “the” molecular structure, “the” molecular energy, or “the”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

Boltzmann’s equation

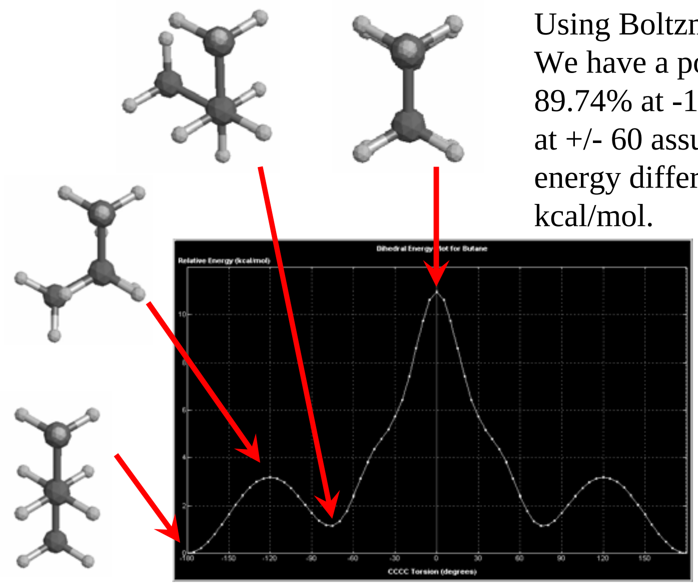
$$P_i = \frac{f_i e^{-\frac{E_i}{RT}}}{\sum_j f_j e^{-\frac{E_j}{RT}}}$$

- 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.