Molecular Simulation III

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 molecules
- **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


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).

\[
U(r) = \frac{1}{2} \sum (b - b_o)^2 + \frac{1}{2} \sum (\theta - \theta_o)^2 + \frac{1}{2} \sum V_j [1 + (-1)^{j+1} \cos(n_j \phi - \delta)] + 4 \varepsilon [\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^6] + \frac{q_i q_j}{4 \varepsilon r_{ij}}
\]
Bond Stretch

\[ E_l = k_b (l - l_0)^2 \]

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

\[ E_l = k_l (l - l_0)^2 + k'_l (l - l_0)^3 + k''_l (l - l_0)^4 \]
Angle Bending

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

- Harmonic approximation
  - \( k_\theta \) is the bending force constant
  - \( \theta^0 \) is the reference angle
- Other forms include

\[ E_\theta = k_\theta \left( 1 + \cos \theta \right) \]
Torsion Interactions

\[ E_\phi = V_1 (1 + \cos \phi) + V_2 (1 - \cos 2\phi) + V_3 (1 + \cos 3\phi) \]

- 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

\[ E_\omega = k_\omega (\omega - \omega_0)^2 \]

• Harmonic approximation
  – \( k_\omega \) is the oop force constant
  – \( \omega_0 \) is the reference value

• Different methods in which to calculate \( \omega \)
  – 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

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

- Lennard-Jones 12-6 potential
  - \( \varepsilon \) is the well depth
  - \( R^*_i \) 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
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.

\[ F = -\nabla U \]
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, $x_1, x_2, \ldots$, 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 $x_k$

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

- 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: \[ R_k = R_{k-1} + l_k S_k \]

Where \( R_k \) is the new position at step \( k \),
\( R_{k-1} \) in the position at the previous step \( k-1 \),
\( l_k \) is the size of the step to be taken at step \( k \) and \( S_k \) is the direction.

**SD**: At each step the gradient of the potential \( g_k \) (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. \[ S_k = -g_k \]

**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.,
\[ S_k = -g_k + b_k S_{k-1} \]
**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

\[
|\text{grad}| = \sqrt{\sum_i \left(\frac{dU}{dx_i}\right)} \\
RMS = \frac{|\text{grad}|}{\sqrt{n}}
\]

• 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

Diagram showing the conformational analysis of butane, including eclipsed (A), eclipsed (C), eclipsed (C), and gauche (E) conformations, with a potential energy diagram indicating energy minima at angles of 0, 180, and 360 degrees, and energy maxima at angles of 60 and 240 degrees.
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

<table>
<thead>
<tr>
<th>Term</th>
<th>MM2</th>
<th>Dreiding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trans</td>
<td>Gauche</td>
</tr>
<tr>
<td>Stretch</td>
<td>0.15</td>
<td>0.16</td>
</tr>
<tr>
<td>Stretch-Bend</td>
<td>0.05</td>
<td>0.07</td>
</tr>
<tr>
<td>Bend</td>
<td>0.29</td>
<td>0.63</td>
</tr>
<tr>
<td>Torsion</td>
<td>0.01</td>
<td>0.44</td>
</tr>
<tr>
<td>VDW</td>
<td>1.68</td>
<td>1.75</td>
</tr>
<tr>
<td>Total</td>
<td>2.18</td>
<td>3.05</td>
</tr>
</tbody>
</table>

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 only 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

\[ E_{ie} = E_{ab} - (E_a + E_b) \]
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 retained.
- Consider hydracrylic acid vs. lactic acid.

\[
\text{HO} \quad \begin{array}{c}
\text{OH} \\
\text{O} \\
\text{C} \\
\text{HO}
\end{array} \quad \text{O} \quad \begin{array}{c}
\text{OH} \\
\text{HO} \\
\text{C}
\end{array}
\]

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

- The free energy can be written as

\[ G = -kT \ln Q + PV \]

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

- The difference in free energy can be written as

\[ \Delta G = -RT \ln \frac{P_1}{P_2} \]
Molecular Simulation IV

Jeffry D. Madura
Department of Chemistry & Biochemistry
Center for Computational Sciences
Duquesne University
Time and Length Scales

Tamar Schlick’s Biomolecular Structure and Modeling
Simulation Lengths and Complexity

Tamar Schlick’s Biomolecular Structure and Modeling
Molecular Dynamics
Crystal-Phospholipid Bilayer Interactions

- Pseudogout (human inflammatory disease) caused by presence of \textit{in vivo} crystals of calcium pyrophosphate dihydrate (CPPD).
- Molecular aspect of \textit{in vivo} crystal induced inflammation is unknown
- Rupture of the lysosome phospholipid membrane is a commonly accepted mechanism of inflammation.
- Important to elucidate the nature of crystal-phospholipid bilayer interactions
- The knowledge will aid in developing inhibitors to diminish the adhesion of CPPD to membranes
Solvated DMPC Bilayer in Absence and Presence of CPPD Crystal
MD Review

- Molecular dynamics is a numerical integration of the classical equations of motion

\[ \ddot{F} = m\ddot{a} = m \frac{d^2\vec{x}}{dt^2} \]

- assuming conservative forces….

\[ \ddot{F} = -\nabla \dot{U} \]

- …the integrated equations of motion become

\[ \vec{r} (t + \delta t) = \vec{r} (t) - \vec{r} (t - \delta t) + \frac{1}{m} \ddot{F} (t) \delta t^2 \]
The Basic MD Simulation Picture

- In the *initialization* the positions and momenta of the atoms are generated or loaded from a previous simulation.
- During the *equilibration* the simulation is run until equilibrium is reached and the memory of the initial configuration has been erased.
- During the *production* phase, the simulation is continued while data are stored on disk.
Molecular Dynamics on BPTI

- Steps involved in the simulation:
  1. heating up the system from the energy minimized structure at 0 K to a temperature for collecting results (300 K)
  2. initial equilibration with velocity rescaling.
  3. second equilibration without velocity rescaling.
  4. one longer run collecting the data.

*Note*: The energy minimized structures disulfide bridges are used.
Molecular Dynamics on BPTI

*Plot the total energy, potential energy and temperature vs simulation time.*

The different energies increase to a maximum after ~1/3 of the total steps (3 ps). **This plateau should be the systems energy at 300K.**

**Note:** total energy consists of both kinetic and potential energy.
The mean square displacement (vs. time) for protein's backbone, the side chains and the whole protein is calculated and plotted (above).
Mean Sq. Displacement vs. Time

• *How does the RMSD of the backbone and sidechains compare to the whole protein?*
  **Answer:** The plot shows the backbone is rigid compared to the protein, whereas the side chains are more flexible.

• *How do they vary with temperature?*
  **Answer:** The higher the temperature, the higher the values for RMSD.

• *Which parts of the protein are more flexible?*
  **Answer:** The loops of the protein are more flexible
Investigation of a Disulfide Bridge

*We will use the data from the MD calc. to investigate a particular disulfide bridge (cys14--cys38) in BPTI.*

![S-S distance vs. time](image1)

![C-S-S angle vs. time](image2)
Conclusions

- What is the average value and standard deviation of the S-S distance and the C-S-S angle?

**Answer:**

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>S-S distance</strong></td>
<td>1.916</td>
<td>2.007</td>
<td>2.033</td>
<td>2.032</td>
<td>2.058</td>
<td>2.156</td>
<td>0.03785765</td>
</tr>
</tbody>
</table>

- The distance between the sulfur atoms is rather constant.

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>C-S-S angle</strong></td>
<td>95.69</td>
<td>103.10</td>
<td>105.60</td>
<td>105.60</td>
<td>108.30</td>
<td>116.70</td>
<td>3.776578</td>
</tr>
</tbody>
</table>

The angle differs quite a lot within the total time. The orientation between the sulfur atoms might be achieved through frequent interactions of parts of the protein.
Velocity Auto-Correlation Function

We define the *velocity autocorrelation function* (VACF), as

\[ \psi(t) = \frac{\langle \sum_j v_j(0) \cdot v_j(t) \rangle}{\langle \sum_j v_j(0) \cdot v_j(0) \rangle} = \frac{1}{3Nk_BT} \langle \sum_j v_j(0) \cdot v_j(t) \rangle \]

Whose short time behavior reduces to,

\[ \psi(t) = 1 - \Omega_0^2 t^2 + \Omega_0^4 t^4 + \ldots \]

This function is a measure on how the velocity of one particle remains correlated with its initial value, and gives an idea of the local environment in which the particles move.

It is found that if the density is high enough, this function shows oscillations in the time domain, indicating the rattling motion of a given particle in the ``cage'' of its nearest neighbors.
Velocity Auto-Correlation Function

We show this effect (below) where the VACF for two state points of a Lennard-Jones fluid is shown.

Velocity autocorrelation function (left) and corresponding spectra (right) for the Lennard-Jones fluid at two thermodynamic points.

Solid line, $T^* = 0.77, \rho^* = 0.85$

Dashed line, $T^* = 1.35, \rho^* = 0.5$
Diffusion Coefficient

General Form: \( \langle \Delta d^2 \rangle \propto D \times \text{elapsed time} \),
Where \( D \) is the diffusion coefficient.

- Molecular dynamics simulations are useful for determining diffusion coefficients.
- Due to interactions (scatterings) particles random walk away from their original positions.
- The variance of the displacement is therefore proportional to elapsed time. The proportionality constant is the diffusion coefficient.
Diffusion Coefficient

The plot below shows the variance of the displacement averaged over all the particles in a 2D simulation. \((L = 20 \text{ and } N = 64)\). The average temperature was \(T = 8.6\).
Monte Carlo
Reaction:
Hydroxide Addition to Formaldehyde

• Fundamental reaction in chemistry and biochemistry
Monte Carlo Introduction

- **1953** Seminal paper by Metropolis, Rosenbluth, Rosenbluth, Teller and Teller.
  - Introduced the Monte Carlo method
  - Proposed a specific form of “importance sampling”
  - Applied the method to a system of hard disks

- **1957** W. W. Wood and F. R. Parker
  - Applied the Metropolis Monte Carlo technique to a 3D Lennard-Jones fluid
  - Periodic conditions

- **1969** J. A. Barker and R. O. Watts
  - First molecular Monte Carlo simulation of water
  - Potential based on experimental gas phase data
  - Obtained structural information of a real molecular fluid
Monte Carlo Integration

Consider evaluating a definite integral using Monte Carlo,

\[ E = \int_{a}^{b} f(x) \, dx \]

By using the mean value theorem of calculus, the integral may be approximated by

\[ E_N = \frac{(b - a)}{N} \sum_{i=1}^{N} f(x_i) \]
Monte Carlo Integration

• A conventional choice for the points $x_i$ would be a uniform grid.

• More accurate methods of numeric quadrature, such as Simpson's rule or Gaussian quadrature use a weighted average of the points:

$$E_N = \frac{(b - a)}{N} \sum_{i=1}^{N} w_i f(x_i) \sum_{i=1}^{N} w_i .$$

Computational cost of evaluating an integral, of dimensionality $d$ increases as $N^d$.

A more efficient approach is to select the points randomly, from a given probability distribution by Monte Carlo methods.
Define: Importance Sampling

- Many functions to be integrated have significant weight in only a few regions.
  - For example, most of the contributions to an integral of a simple Gaussian are located near the central peak.
- In a simple Monte Carlo integration scheme, points are sampled uniformly, wasting considerable effort sampling the tails of the Gaussian.
- Techniques for overcoming this problem act to increase the density of points in regions of interest. These techniques are called importance sampling.
Importance Sampling

Note that $p(x)$ is the Boltzmann distribution for our purposes.

\[
\frac{n_i}{N} = \frac{e^{-\beta E_i}}{\sum_i e^{-\beta E_i}}
\]

• Points are sampled over a distribution $w(x)$, where $w(x)$ is always positive and is chosen to approximate $f(x)$ over the region of interest.

• The integral can now be evaluated by selecting points from the probability distribution, $p(x)$.

\[
E = \int_a^b g(x)p(x)dx \approx \frac{1}{N} \sum_{i=1}^N g(x_i)
\]
Metropolis Algorithm

• *Generates a random walk of points* distributed according to the (Boltzmann) probability distribution.

• *From an initial "position" in phase or configuration space, a proposed "move" is generated* and the move either accepted or rejected according to the Metropolis algorithm.

• By taking a sufficient number of trial steps all of phase space is explored and the Metropolis algorithm *ensures that the points are distributed according to the probability distribution.*
Metropolis Algorithm

The Metropolis algorithm corresponds to choosing

\[ P(X \rightarrow X') = \min \left\{ 1, \frac{\rho(X')}{\rho(X)} \right\} . \]

*Points of Interest:*

1. For the Metropolis algorithm to be valid, the random walk must be **ergodic**, that is any point in configuration space may be reached from any other point.
2. In some applications of the Metropolis algorithm, **parts of configuration space may be difficult to reach**.
3. **Long simulations** or a modification of the algorithm are then necessary.
NpT Monte Carlo Method

1. Pick a configuration and calculation $E_0$
2. Randomly choose a molecule and randomly translate and rotate and/or change the volume
3. Calculate $E_n$
4. Calculate $\Delta W$ as follows

$$\Delta W = (E_n - E_o) + P(V_n - V_o) - Nk_B T \ln \left( \frac{V_n}{V_o} \right)$$

5. if

$$\Delta W \leq 0$$

$$\exp \left( - \frac{\Delta W}{k_B T} \right) > \text{rnd} (0,1) \right\} \text{ accept}$$

6. Update averages
7. Set $E_n$ to $E_0$
8. Goto 2
Statistical Mechanics

\[
\langle A \rangle = \frac{\int dr^N \exp[-\beta U(r^N)] A(r^N)}{\int dr^N \exp[-\beta U(r^N)]}
\]

Importance sampling

\[
\langle A \rangle \approx \frac{1}{L} \sum_{i=1}^{L} A(r^N_i) \exp[-\beta U(r^N_i)]
\]

Generate a sequence of configurations \( r^N_i \) with statistical distribution
Pure Liquids

• Water
  – Applied NpT Monte Carlo simulations to derive a water potential that reproduced experimental results.
  – Used the following effective pair potential function

\[
\epsilon_{mn} = \frac{A_{OO}}{r_{OO}^{12}} - \frac{C_{OO}}{r_{OO}^6} + \sum_{i}^{on} \sum_{j}^{on} \frac{q_i q_j e^2}{r_{ij}}
\]

<table>
<thead>
<tr>
<th></th>
<th>SPC</th>
<th>TIP3P</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A_{OO})</td>
<td>629.4 x 10^3</td>
<td>582.0 x 10^3</td>
</tr>
<tr>
<td>(C_{OO})</td>
<td>625.0</td>
<td>595.0</td>
</tr>
<tr>
<td>(q(O))</td>
<td>-0.82</td>
<td>-0.834</td>
</tr>
<tr>
<td>(q(H))</td>
<td>0.41</td>
<td>0.417</td>
</tr>
<tr>
<td>(A_{OO})</td>
<td>600.0 x 10^3</td>
<td></td>
</tr>
<tr>
<td>(C_{OO})</td>
<td>610.0</td>
<td></td>
</tr>
<tr>
<td>(q(O))</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>(q(H))</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>(q(M))</td>
<td>-1.04</td>
<td></td>
</tr>
</tbody>
</table>
Simulation Analysis

• Thermodynamics
  – Average potential energy, enthalpy
  – Heat capacity

\[ C_P = \frac{\langle H^2 \rangle - \langle H \rangle^2}{k_B T^2} \]

<table>
<thead>
<tr>
<th></th>
<th>SPC</th>
<th>TIP3P</th>
<th>TIP4P</th>
<th>Expt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>d (g/cm3)</td>
<td>0.971</td>
<td>0.982</td>
<td>0.999</td>
<td>0.997</td>
</tr>
<tr>
<td>-E_i (kcal/mol)</td>
<td>10.18</td>
<td>9.86</td>
<td>10.07</td>
<td>9.92</td>
</tr>
<tr>
<td>ΔH_{vap} (kcal/mol)</td>
<td>10.77</td>
<td>10.45</td>
<td>10.66</td>
<td>10.51</td>
</tr>
<tr>
<td>C_p (cal/mol deg)</td>
<td>23.4</td>
<td>16.8</td>
<td>19.3</td>
<td>17.99</td>
</tr>
</tbody>
</table>

Properties at 25 °C and 1 atm.
Simulation Analysis

• Structural
  – Radial distribution function
    \[ g_{xy}(R) = \frac{\left\langle N_y \left( R, R + dR \right) \right\rangle}{\rho_y 4\pi R^2 dR} \]
  – Coordination number
    \[ CN = \int_0^{R_c} g(R) \rho 4\pi R^2 dR \]
Simulation Analysis

• Energetics
  – Hydrogen bond analysis
  – Interaction energy of $\leq 2.25$ kcal/mol or less

<table>
<thead>
<tr>
<th></th>
<th>SPC</th>
<th>TIP3P</th>
<th>TIP4P</th>
</tr>
</thead>
<tbody>
<tr>
<td># of H-bonds</td>
<td>3.54</td>
<td>3.50</td>
<td>3.57</td>
</tr>
<tr>
<td>$\epsilon$ (H bond), kcal/mol</td>
<td>-4.34</td>
<td>-4.20</td>
<td>-4.17</td>
</tr>
<tr>
<td>$\epsilon$ (Coulomb), kcal/mol</td>
<td>-5.65</td>
<td>-5.39</td>
<td>-.5.2</td>
</tr>
<tr>
<td>$\epsilon$ (LJ), kcal/mol</td>
<td>1.31</td>
<td>1.19</td>
<td>1.35</td>
</tr>
<tr>
<td>$\theta$, deg (O–H...O)</td>
<td>156</td>
<td>155</td>
<td>158</td>
</tr>
<tr>
<td>$\phi$, deg (H...OH$_2$)</td>
<td>99</td>
<td>100</td>
<td>99</td>
</tr>
</tbody>
</table>
Solvation: Methanol in Water

- Does the rotamer population change in moving from the gas phase to solution?
- Sampling of rotational barriers is facilitated through the use of umbrella sampling.

\[
\langle \theta \rangle = \frac{\langle \theta / w \rangle_w}{\langle 1/w \rangle_w} \quad w = \exp\left( \beta \left( V(\phi) - V'(\phi) \right) \right)
\]

Surrogate rotation potential
Structural and Energetic Results

<table>
<thead>
<tr>
<th></th>
<th>Simulation</th>
<th>Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta V_{\text{sol}} (\text{A}^3)$</td>
<td>119 +/- 24</td>
<td>63.5</td>
</tr>
<tr>
<td>$\Delta H_{\text{sol}} (\text{kcal/mol})$</td>
<td>-17 +/- 5</td>
<td>-10.8</td>
</tr>
</tbody>
</table>
Reaction: Hydroxide Addition to Formaldehyde

- Fundamental reaction in chemistry and biochemistry

nucleophile

Tetrahedral intermediate
Gas Phase and Aqueous Solution Energy Profiles

\[ W(r) = -k_B T \ln g(r) \]

The origin of the activation barrier in solution stems from the weakening of the five to six hydrogen bonds in the transition state and product as opposed to the same strong hydrogen bonds to the reactants. This reflects the charge delocalization in the transition state and tetrahedral intermediate.
Other Monte Carlo Methods

• Other Ensembles
  – Canonical ensemble (NVT)
  – Grand Canonical Ensemble (N\(\mu\)T)
    • Equations of state
    • Adsorption isotherms
  – Gibbs “Ensemble”
    • Interfacial simulations

• Other Methods
  – Hybrid MD/MC
    • Stochastic collisions
    • Force-bias Monte Carlo
  – Free Energy Simulations
    • \(W(r) = -k_B T \ln \langle \exp(-\beta(V_1 - V_0)) \rangle_0\)
  – Parallel Tempering
    • Sampling of energy landscapes with many local minima
Docking
Ligand-Receptor Docking

- Deals with identification of suitable ("best") ligands for specific receptors in proteins.

- Ligands can act either as activators or as inhibitors of the biological function of the protein in the cell.

- Artificial ligands (i.e. drugs) can be used to up-regulate or down-regulate the activity of proteins that are associated with specific diseases.

- To the left, HIV-1 Protease complexed with an efficient inhibitor, TL-3-093.
Docking

• Three-dimensional molecular structure is one of the foundations of structure-based drug design.

• Often, data are available for the shape of a protein and a drug separately, but not for the two together.

• Docking is the process by which two molecules fit together in 3D space.
Docking

• Two general classes
  – “Unbiased”
    • Autodock
  – “Direct”
    • DOCK
    • LUDI

• Goals
  – Robust and accurate
  – Computationally feasible
Ligand-Receptor Docking Approach: Challenges

- Must screen **millions** of possible compounds that fit a particular receptor.
- Must specifically select those ligands that show a high affinity.
- The set of ligands selected can then be screened further by more involved computational techniques, such as free-energy perturbation theory ($\Delta G_{\text{bind}}$).
- We would like an automated, standard protocol to find the best Ligand-Receptor fit.
Docking

• Terms to consider in docking
  – Shape complementarity
  – Interaction specificity
  – Solvation/desolvation
  – Hydrophobic
  – Hydrogen bonding

• Terms considered in MOE-Dock (Autodock)
  – Van der Waals
  – Hydrogen bonding
  – Electrostatics
Docking

• Energy evaluation
  – Based on a Grid approach

• Search engine
  – Simulated Annealing (SA)
    • Autodock
    • MOE-Dock
  – Genetic Algorithms (GA)
    • Autodock 3.0
MOE-Dock Application

- We will look at a docking example of a TIBO-like inhibitor to HIV-1 Reverse Transcriptase (HIV-RT).
- Crystal structure to be used: HIV-RT with TIBO-R86183.
MOE-Dock Application

• Setting up the calculation.
  – *Prepare the protein.* Color the ligand, receptor, and metal ions distinctly. Add H atoms to the X-ray structure if none are given.
  – *Select ForceField.*
  – *Minimize.*

Here you can turn on solvation model; Place partial charges on on atoms
MOE-Dock Application

- MOE | Compute | Simulations | Dock

The docking box appears around the ligand. Graphic shows HIV-RT (red) and its ligand TIBO-R86183.
MOE-Dock Application
Docking Results

• Examine the docked structures compared to the crystal structure of the ligand and its receptor.

• In this database, columns contain the total energy of the complex, the electrostatic (U_ele) and van der Waals energies (U_vdw) between the protein and the ligand, and the energy of the (flexible) ligand (U_ligand).
MOE-Dock Application

- To find the best (lowest energy) docked structure, you will sort the database in ascending order with respect to the total energy (U_total)
Brownian Dynamics
Triose Phosphate Isomerase

- Enzyme that catalyzes the interconversion of D-glyceraldehyde phosphate (GAP) to dihydroxyacetone phosphate (DHAP)
- Rate-limiting step of TIM with GAP as substrate is diffusion-controlled ($k_d = 4.8 \times 10^8$ M$^{-1}$ s$^{-1}$)

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Calculated Rate Constant ($10^8$ M$^{-1}$ s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sphere</td>
<td>148.</td>
</tr>
<tr>
<td>Sphere (no electrostatics)</td>
<td>30.6</td>
</tr>
<tr>
<td>Dumbbell</td>
<td>1.664</td>
</tr>
<tr>
<td>Flexible loop / dumbell</td>
<td>4.2</td>
</tr>
</tbody>
</table>
Bimolecular Diffusion-Controlled Rate Constant

\[ \tilde{r}' = \tilde{r}_0 + \frac{D \tilde{F}(\tilde{r}_0)}{k_B T} \Delta t + \tilde{R} \]

Diffusion constant

Random vector

\[ k = k_D(b) \beta \left[ 1 - (1 - \beta) \Omega \right]^{-1} \]

\[ k_D(b) = 4\pi \int_b^\infty dr \frac{e^{(U(r)/k_B T)^{-1}}}{4\pi r^2 D(r)} \]

\[ \Omega = \frac{k_D(b)}{k_D(q)} \]

\[ \beta = \frac{\text{# of hits}}{\text{# of trials}} \]
Diffusional Encounter between GAP and TIM

• Snapshot of a ~11 ns trajectory of GAP diffusing to the active site of TIM. In the top figure the random nature of the substrate (shown in green) and the large volume of space sampled can be seen.

• The bottom figure illustrated 32 snapshots at intervals of 0.25 – 1 ns colored according to time (indigo to red corresponds to increasing time)
Brownian Dynamics Simulation of Lysozyme to a Charged Surface

- Schematic diagram showing the details of the simulation method. In this figure the protein molecule is represented as an arbitrarily shaped object with patches corresponding to both positively charged (blue) and negatively charged (red) amino acid residue collections.
Protein – Surface Interactions

Fraction of Successful Trajectories for Two Different Salt Concentrations

<table>
<thead>
<tr>
<th>I(M)</th>
<th>Successful Trajectories</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.71 ± 0.03</td>
</tr>
<tr>
<td>0.3</td>
<td>0.64 ± 0.02</td>
</tr>
<tr>
<td>No Electrostatics</td>
<td>0.62 ± 0.03</td>
</tr>
</tbody>
</table>
Poisson – Boltzmann Electrostatics
Application Areas of Electrostatics

- Electrostatic Energies
- Electrostatic Forces
- Electrostatic Binding Free Energy
- Electrostatic Solvation Free Energy
- pKa Shifts
- Protein Stability
- Conformational pH Dependence
- Redox
- Electrostatic Steering in Enzyme/Substrate Encounters
- Electrostatic Forces Coupled to Molecular Mechanics/Dynamics
Based on a suggestion by Born, the explicit solvent model may be very crudely approximated by a structureless continuum. In this continuum picture the solvent is represented by a dielectric constant, $\varepsilon_{\text{sol}}$, and the effect of ions by, $\kappa$. The solute is a set of embedded charges inside a cavity with a dielectric constant of, $\varepsilon_{\text{in}}$. 
Continuum Solvent Model

\[ \Delta G^{solv} = \Delta G^{np} + \Delta G^{elec} \]

\[ \Delta G^{np} \approx \gamma \ SA \]

\[ \Delta G^{elec} = \frac{1}{2} \sum_{i=1}^{N_{\text{atoms}}} q_i \left( \phi_i^s - \phi_i^v \right) \]
Poisson-Boltzmann Model of Molecular Electrostatics

\[-\nabla \cdot (\varepsilon \nabla \phi) = 4\pi \rho^f - \kappa^2 \phi \lambda\]

- \(\varepsilon\): permittivity
- \(\phi\): electrostatic potential
- \(\rho^f\): "fixed" charge density
- \(\kappa\): inverse Debye length
- \(\lambda\): "masking" function

\[\kappa^2 = \frac{8\pi e^2 N_A I}{1000 \varepsilon k_B T}\]
Solving the FDPB Equation

• In practice, one knows the
  – charge density (ρ) from the fixed charges in the receptor and substrate.
  – the permittivity (dielectric constant).
  – Kappa (κ), which is related to the ionic strength.
• Make a guess at the potential.
• Solve the equation for a new potential.
• Continue to solve until the change in potential is small.
Poisson-Boltzmann Electrostatic Forces

\[ \vec{f} = F^{\text{Coul}} + F^{\text{RF}} + F^{\text{DBF}} + F^{\text{IBF}} \]

\(F^{\text{Coul}}\) is the Coulombic force which is the interaction of all the solute atoms with each other and is referred to as the “qE” force.

\(F^{\text{RF}}\) is the reaction field force, \(F^{\text{RF}} = qE^{\text{RF}}\) where \(E^{\text{RF}}\) is the solvent reaction field acting at an atom.

\(F^{\text{DBF}}\) is the dielectric boundary force. This is due to the tendency of high dielectric medium to reduce the field energy by moving into regions of low-dielectric constant.

\(F^{\text{IBF}}\) is the ionic boundary force and is generally small in comparison with the other forces in the system. This force results from the tendency of mobile ions to reduce the field energy by moving into regions of zero ionic strength (i. e. the molecular interior).
Langevin Dynamics

\[ m \frac{d^2 x(t)}{dt^2} = \vec{F}(x) - m \gamma \frac{dx(t)}{dt} + \vec{R}(t) \]

- **mass**
- **position**
- Force which depends upon the position of the particle relative to the other particles

Random fluctuations due to interactions with the solvent

- Force due to the motion through the solvent
  \[ \gamma = \frac{k_B T}{mD} \]
  - diffusion constant
Dichloroethane

Summary of simulation parameters

\[ \varepsilon_i = 1 \]
\[ \varepsilon_s = 80 \]
\[ \gamma = 6.5 \text{ ps}^{-1} \]
\[ dt = 0.001 \text{ ps} \]
\[ T = 1000 \text{ K} \]
\[ \text{grid spacing} = 0.5 \text{ to } 1.2 \text{ Å} \]

<table>
<thead>
<tr>
<th>Atom Type</th>
<th>Charge (e)</th>
<th>Radius (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cl</td>
<td>-0.25</td>
<td>1.82</td>
</tr>
<tr>
<td>CH₂</td>
<td>0.25</td>
<td>1.99</td>
</tr>
</tbody>
</table>

Trans conformer dominates in the gas phase

Increased gauche conformer in liquid phase
Dichloroethane

Summary of simulation results

[Graphs showing normalized probability density vs dihedral angle for reference gas phase results and 2 ns stochastic dynamics simulations with different grid sizes and solvent boundary force conditions.]
Alanine “dipeptide”

Summary of simulation parameters

\( \varepsilon_i = 1 \)
\( \varepsilon_s = 80 \)
\( \gamma = 6.5 \text{ ps}^{-1} \)
\( dt = 0.001 \text{ ps} \)
\( T = 1000 \text{ K} \)
grid spacing = 0.7 to 1.7 Å

Conclusions

- Good equilibration
- Good agreement with other computational models
- Weak sensitivity to grid spacing
- No heating from numerical forces
Alanine “dipeptide”

in vacuo

Aqueous

Reference

2 ns stochastic dynamics
Thermodynamic Treatment of Ion-Solvent Interactions: 

*The Born Model*

- **Ion-Solvent interaction:** Consists of solvent dipoles interacting with the electric field of the ion.
- **Two cases to consider for the solvent:**
  - A structure-less *continuum* of dielectric $\varepsilon$ ("The Born Model")
  - *Discrete molecules* with dipoles, polarizability, etc.
The Born Model

- Consider: Continuum model of ion solvation.

If medium 1 is a vacuum, $\Delta G_{\text{born}}$ is just the free energy of solvation.

We will calculate the free energy of transfer of an ion from medium 1 ($\varepsilon_1$) to medium 2 ($\varepsilon_2$). This will be called $\Delta G_{\text{born}}$. 
The path for $\Delta G_{\text{born}}$ refers to:

First **discharging** the ion in medium 1 ($\Delta G^o_1$)

Transferring the ion from medium 1 to medium 2 ($\Delta G^o_2$)

Recharging the ion in solvent 2 ($\Delta G^o_3$)
The Charging Process

- Energies of charging/discharging:
  - computed by a model where \textit{infinitesimal pieces of charge} are brought from infinity,
  - and placed on the surface of the ion until the final charge is obtained

\[
\begin{align*}
 dq & \quad \text{ion of radius } a \\
r & = \infty \\
\Phi & = 0
\end{align*}
\]

\text{The charging process}
The Charging Process

What is the energy of bringing a charge $dq$ from infinity and placing it on the surface of a sphere with radius $a$?

$$dG = \Phi dq$$
The Charging Process

- Knowing the potential ($\Phi$) of a point charge, we have,

$$dG_{\text{charging}} = \Phi \, dq = \frac{q}{4\pi \varepsilon_0 \varepsilon_a} \, dq$$

Integrating this from 0 to the final charge on the ion, $Ze$ (where $Z$ is the valence)......(Next Slide)
The Charging Process

\[ \Delta G_{\text{charging}} = \frac{Z^2 e^2}{8 \pi \varepsilon_0 \varepsilon l} \]

Therefore, For \( \Delta G^o_1 \), \( \Delta G^o_2 \), and \( \Delta G^o_{\text{born}} \) we have...(Next Slide)

\[ \Delta G_{\text{dissolving}} = - \Delta G_{\text{charging}} \]
The Charging Process

If $\varepsilon_2 < \varepsilon_1$, then $\Delta G^o > 0$

It takes work to move an ion from water to a less polar solvent (such as vacuum or hydrocarbon)

\[
\Delta G_i^o = -\frac{Z^2 e^2}{8\pi \varepsilon_0 \varepsilon_1 a}
\]

\[
\Delta G_z^o = +\frac{Z^2 e^2}{8\pi \varepsilon_0 \varepsilon_2 a}
\]

\[
\Delta G_{\text{Born}}^o = \frac{Z^2 e^2}{8\pi \varepsilon_0 a} \left( \frac{1}{\varepsilon_2} - \frac{1}{\varepsilon_1} \right) + \Delta G_z^o
\]
Free Energy of Solvation

• Consider: Transferring an ion from a vacuum to a medium of $\varepsilon$.
  
  – Assume $\Delta G^o_2 = 0$. (No interaction between solvent and discharged ion).

$$\Delta G^o_{\text{solvation}} = \frac{Z^2 e^2}{8\pi \varepsilon_0 a} \left( \frac{1}{\varepsilon} - 1 \right)$$

Two points to note:
1. $\Delta G < 0$ if $\varepsilon > 1$
2. $\Delta G$ increases as ionic Radius increases. Why?

The field and the potential At the ion surface becomes Less.
Generalized Born

- Widely used to represent the electrostatic contribution to the free energy of solvation
- Model is comprised of a system of particles with radii $a_i$ and charges $q_i$
- The total electrostatic free energy is given by the sum of the Coulomb energy and the Born free energy of solvation in a medium of relative permittivity $\varepsilon$.

$$G_{elec} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{\varepsilon r_{ij}} - \frac{1}{2} \left(1 - \frac{1}{\varepsilon}\right) \sum_{i=1}^{N} \frac{q_i^2}{a_i}$$
Generalized Born

• The previous equation can be re-written into the generalized Born equation

\[ \Delta G_{elec} = -\frac{1}{2} \left( 1 - \frac{1}{\varepsilon} \right) \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{f(r_{ij}, a_{ij})} \]

• \( f(r_{ij}, a_{ij}) \) depends upon the interparticle distances \( r_{ij} \) and the Born radii \( a_i \).

\[ f(r_{ij}, a_{ij}) = \sqrt{r_{ij}^2 + a_{ij}^2 e^{-D}} \]

\[ a_{ij} = \sqrt{a_i a_j} \]

\[ D = \frac{r_{ij}^2}{\left(2a_{ij}\right)^2} \]
Generalized Born

• Note the following
  – When $i=j$ the equation returns the Born expression
  – When $r_{ij} \ll a_i$ and $a_j$ the expression is close to the Onsager result (I.e. a dipole)
  – When $r_{ij} \gg a_i$ and $a_j$ the result is very close to the sum of the Coulomb and Born expression

• A major advantage to this formulation is that the expression can be differentiate analytically, thereby enabling the solvation term to be included in gradient-based optimization methods
MacroModel GB/SA Solvation Model

- Accounts for solvation effects, especially in complex systems.
- Generalized Born/Surface Area (GB/SA) approach (continuum).
  - increases the speed of the calculation
  - avoids convergence problems, apparent in explicit models, where longer simulations or different solvent starting geometries yield different final energies.
- The GB/SA model can be used to calculate absolute free energies of solvation.
Application of GB/SA Solvation Model

• Hall group applied the GB/SA continuum solvation model to RNA hairpins with much success.

• Simulations of the UUCG tetraloop give average structures within 1.2 Å of the initial NMR model, in agreement with an explicit solvent simulation (Williams, D. J., Hall, K. B. 1999. Biophys J. 76:3192-3205).
Electrostatic Free Energy of Solvation Calculation

• In this calculation one computes the electrostatic energy difference between the molecule in the aqueous phase and in vacuum.
  – This is equivalent to computing the work in moving a charge from a low dielectric to a high dielectric.
  – This work is equivalent to a change in the free energy.
  – MOE-Electrostatics can be used by performing two calculations
    • Compute the electrostatic energy with both dielectric constants set to 1
    • Compute the electrostatic energy with the interior dielectric set to 1 and the exterior dielectric set to 80.
MOE-Electrostatics

For the chloride ion we have

\[ EE: \frac{1}{80} = 2472.76 \]
\[ EE: \frac{1}{1} = 2545.44 \]
\[ \Delta EE = \Delta G = -72.68 \]

From the Born equation we have

\[ \Delta G = -332 \frac{q^2}{2a} \left(1 - \frac{1}{\varepsilon} \right) \]

\[ \Delta G = -67.79 \text{ kcal/mol} \]
Binding Free Energy

• Consider the following noncovalent binding process

\[ R + S \iff R : S \]

• Where R represents the receptor, S represents the substrate, and R:S is the noncovalent complex.

• The binding free energy can be partitioned into

\[
\Delta G = \Delta G_s (R : S) - \Delta G_s (R) - \Delta G_s (S) + \Delta G_a + \Delta G_n
\]
Binding Free Energy

• Pictorially the previous equation is

\[ \Delta G(S) \]

\[ \Delta G(R) \]

\[ \Delta G(R:S) \]

\[ \Delta G_a + \Delta G_n \]
Binding Free Energy

- Relative binding free energies are best to compute ($\Delta\Delta G$)
- Results for sulphate-binding protein

<table>
<thead>
<tr>
<th>Protein</th>
<th>$\Delta\Delta G_s$</th>
<th>$\Delta\Delta G_a$</th>
<th>$\Delta\Delta G_{calc}$</th>
<th>$\Delta\Delta G_{expt}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S130G</td>
<td>-4.0</td>
<td>5.3</td>
<td>1.3</td>
<td>1.6</td>
</tr>
<tr>
<td>S130A</td>
<td>-2.4</td>
<td>5.3</td>
<td>2.9</td>
<td>2.7</td>
</tr>
<tr>
<td>S130C</td>
<td>-0.5</td>
<td>4.2</td>
<td>3.8</td>
<td>4.8</td>
</tr>
</tbody>
</table>
UHBD Capabilities

- The **UHBD, University of Houston Brownian Dynamics**, program is capable of computing a variety of properties for biomolecules
  - electrostatic binding free energy for an enzyme/substrate complex
  - bimolecular diffusion-controlled rate constant for an enzyme-substrate encounter with a flexible substrate
  - protein-protein association constants
  - perform a molecular mechanics / dynamics calculations using a continuum solvent
  - determining the pKa’s of ionizable groups in proteins and small molecules.