THE UNIVERSITY of TEXAS

HEALTH SCIENCE CENTER AT HOUSTON

SCHOOL of HEALTH INFORMATION SCIENCES

Molecular Dynamics Simulation: Overview

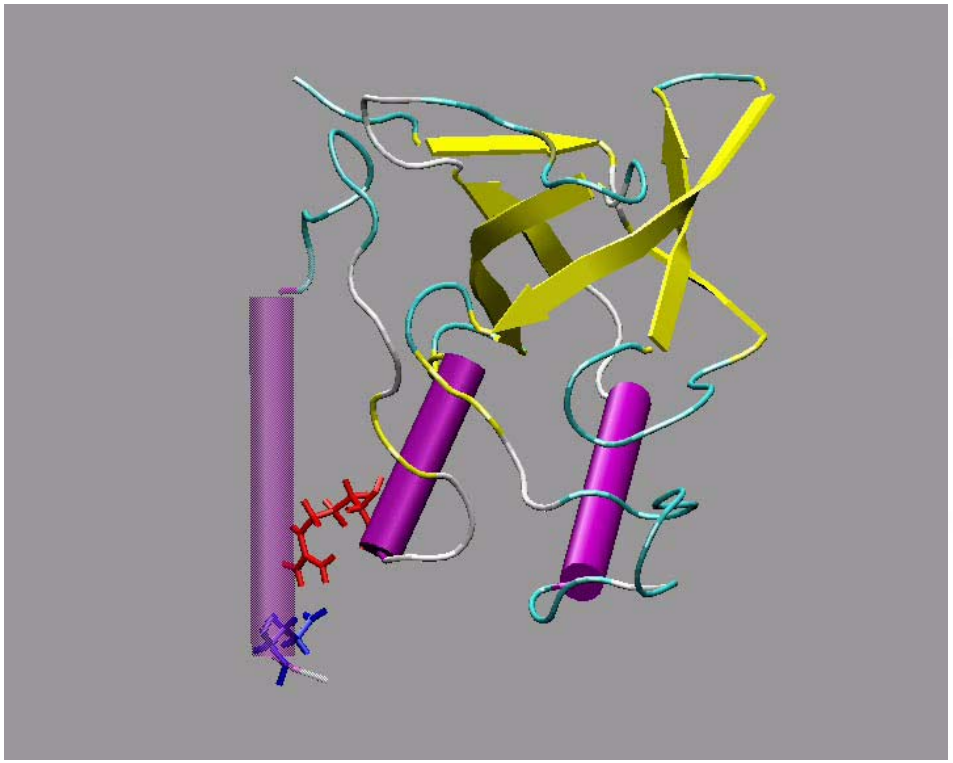
For students of HI 6327 “Biomolecular Modeling”

Willy Wriggers, Ph.D.

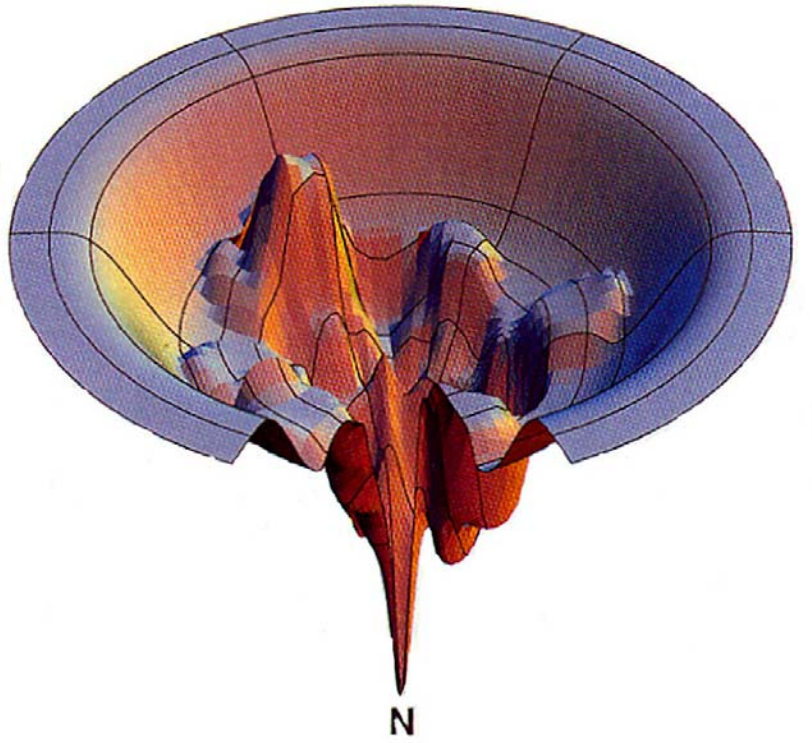
School of Health Information Sciences

http://biomachina.org/courses/modeling/04.html

Protein (Biomolecular) Folding



native structure N



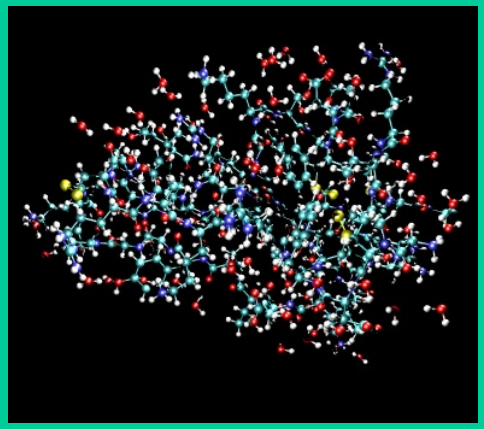
exploring the folding free energy (entropy + enthalpy) landscape

Use of Molecular Dynamics Simulation

- structure prediction / modeling
- flexibility
- solvent effects
- chemical reactions (with QM)
- conformational changes, allosteric mechanisms
- thermodynamics (free energy changes, binding)
- NMR/crystallography (refinement)
- Electron microscopy (flexible fitting)

Atomic Detail Computer Simulation

Model System

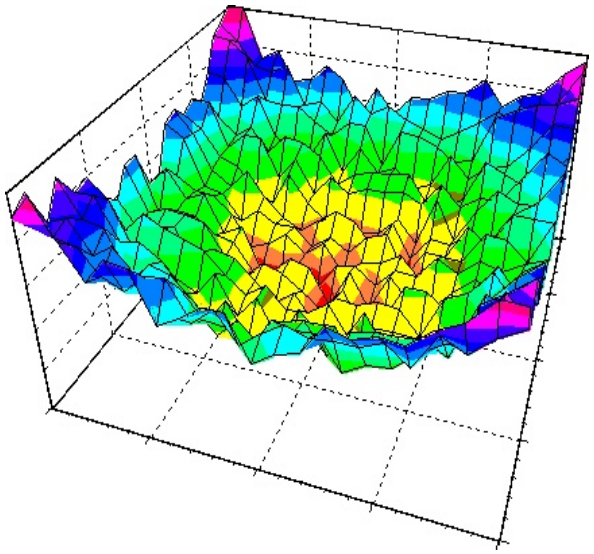


Molecular Mechanics Potential

$$V = \sum_{bonds} k_b (b - b_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 +$$

$$+ \sum_{dihedrals} \sum_{n=1}^{N} K_\phi^{(n)} [1 + \cos(n\phi - \delta)] + \sum_{impropers} K_\omega (\omega - \omega_0)^2$$

$$+ \sum_{i,j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \sum_{i,j} \left(\frac{q_i q_j}{D r_{ij}} \right)$$



Energy Surface $\rightarrow$
Exploration by Simulation..

Steric Energy

•A molecule can possess different kinds of energy such as bonded, nonbonded (both *enthalpic* contributions), and thermal energy (*entropy*).

•Molecular mechanics calculates the enthalpic or *steric* energy of a molecule - the energy due to the geometry or conformation of a molecule.

•Energy is *minimized* in nature, and the native conformation of a molecule that is favored is the lowest steric energy conformation.

•Studies of the conformation of proteins are difficult and therefore interesting, because their size makes many different conformations possible.

•In MD, steric energy allows the integration of the equations of motion following Newton's law.

Steric Energy

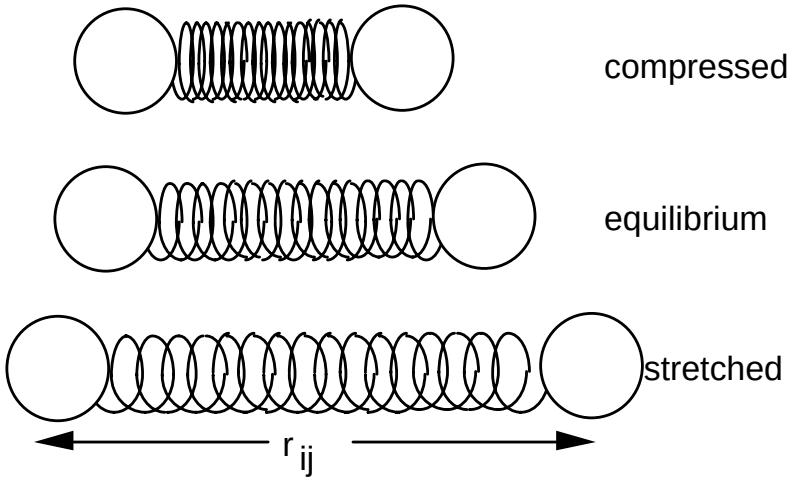
Molecular mechanics *assumes* the steric energy of a molecule to arise from a few, specific interactions within a molecule. These interactions include the stretching or compressing of bonds beyond their equilibrium lengths and angles, torsional effects of twisting about single bonds, the van der Waals (vdW) attractions or repulsions of atoms that come close together, and the electrostatic interactions (qq) between partial charges in a molecule. To quantify the contribution of each, these interactions can be modeled by a potential function that gives the energy of the interaction as a function of distance, angle, or charge.

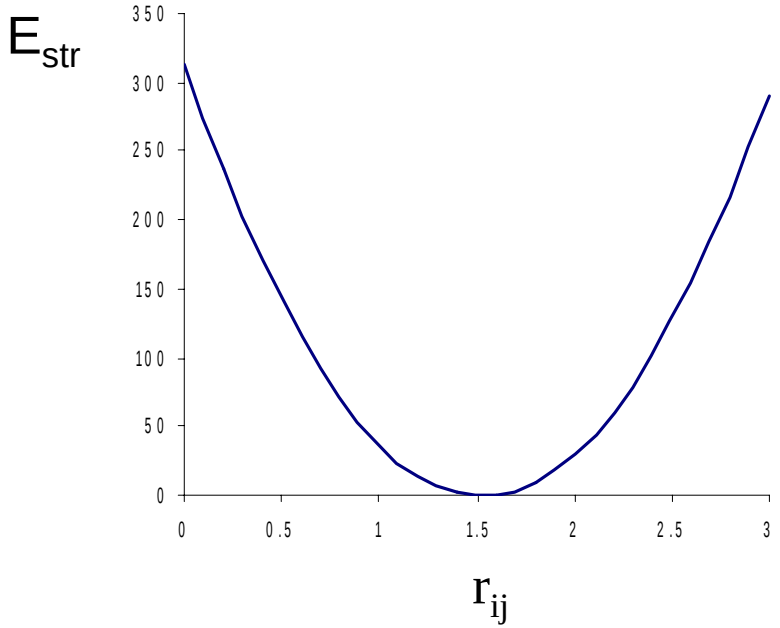
The total steric energy of a molecule can be written as a sum of the energies of the interactions:

$$E_{\text{steric energy}} = E_{\text{str}} + E_{\text{bend}} + E_{\text{improper}} + E_{\text{tor}} + E_{\text{vdW}} + E_{\text{qq}}$$

Bonded Interactions: Stretching

E_{str} represents the energy required to stretch or compress a covalent bond:



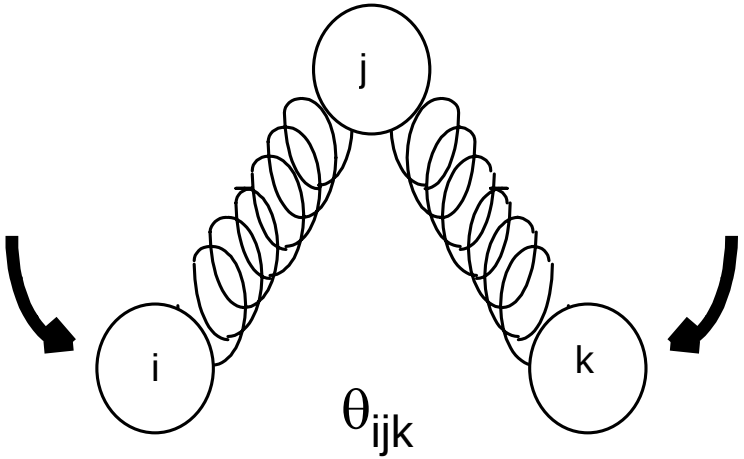


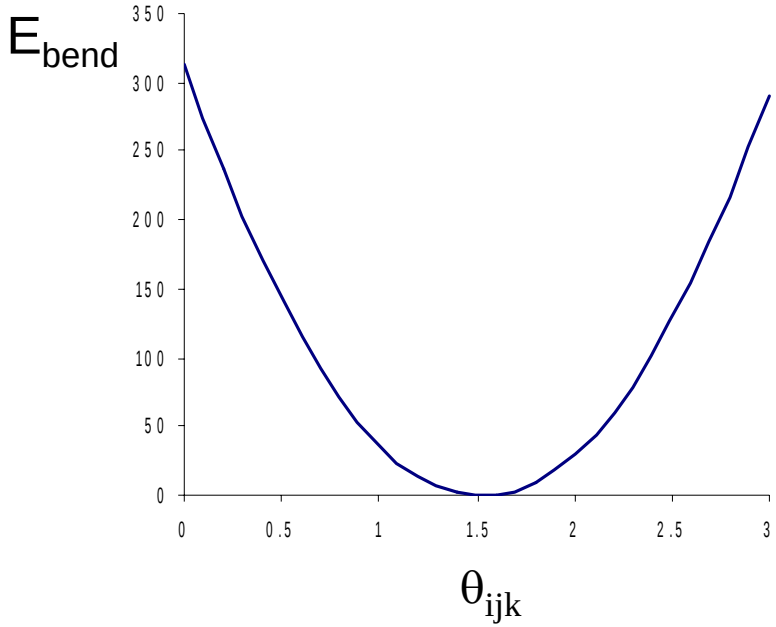
A bond can be thought of as a spring having its own equilibrium length, r_o, and the energy required to stretch or compress it can be approximated by the Hookean potential for an ideal spring:

$$E_{str} = \tfrac{1}{2}\, k_{s,ij} \left(r_{ij} - r_o \right)^2$$

Bonded Interactions: Bending

E_{bend} is the energy required to bend a bond from its equilibrium angle, θ_o:



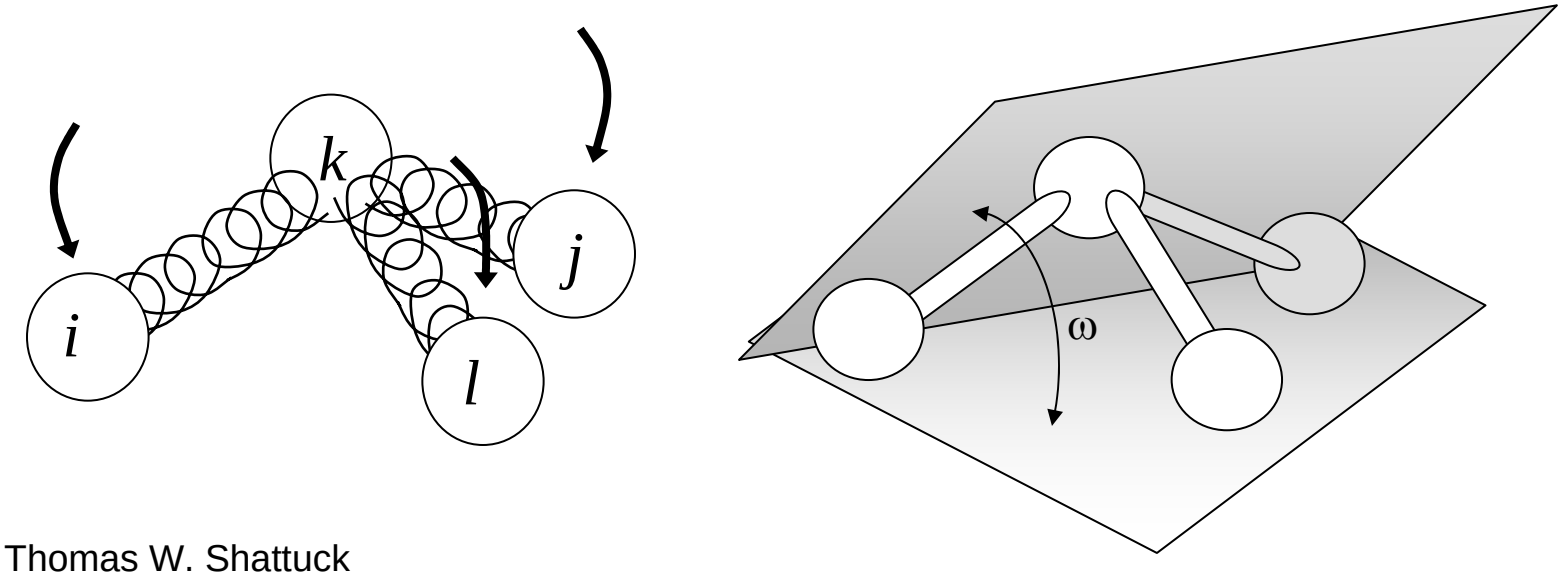


Again this system can be modeled by a spring, and the energy is given by the Hookean potential with respect to angle:

$$E_{bend} = \tfrac{1}{2}\, k_{b,ijk}\, (\theta_{ijk} - \theta_o)^2$$

Bonded Interactions: Improper Torsion

$E_{improper}$ is the energy required to deform a planar group of atoms from its equilibrium angle, ω_o, usually equal to zero:

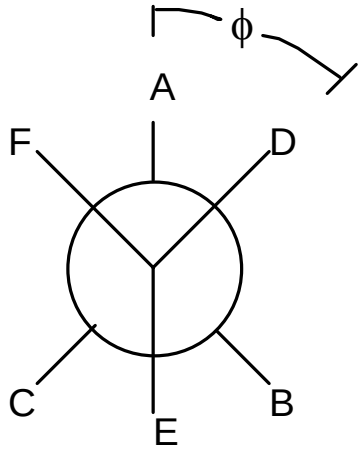


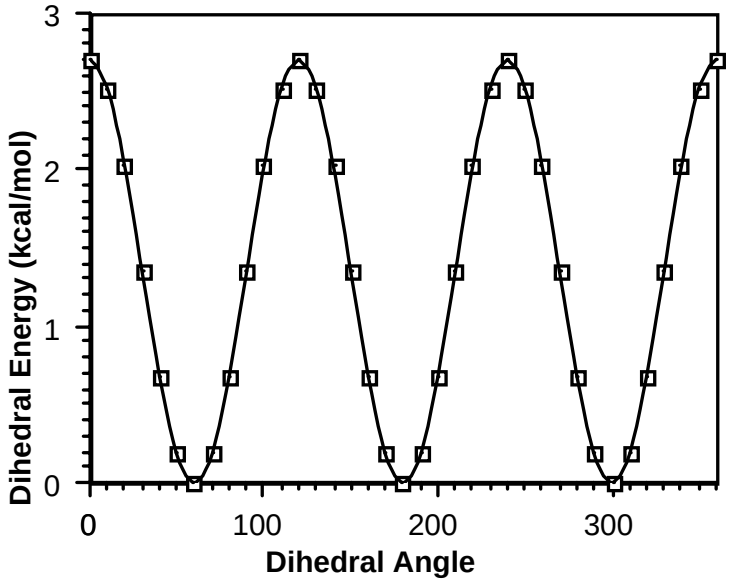
Again this system can be modeled by a spring, and the energy is given by the Hookean potential with respect to planar angle:

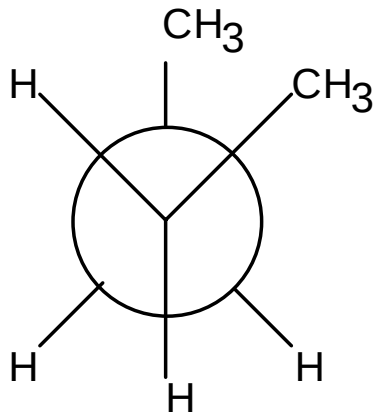
$$E_{improper} = \tfrac{1}{2}\, k_{o,ijkl}\, (\omega_{ijkl} - \omega_{o})^2$$

Bonded Interactions: Torsion

E_{tor} is the energy of torsion needed to rotate about bonds:







Butane

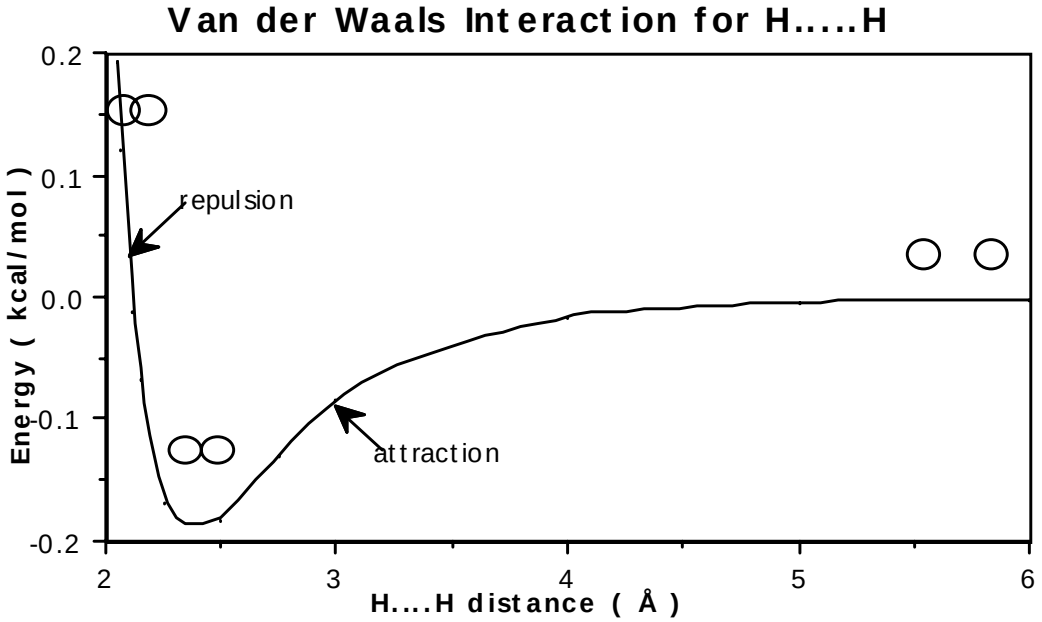
Torsional interactions are modeled by the potential:

$$E_{tor} = \tfrac{1}{2} k_{tor,1} (1 - \cos \phi) + \tfrac{1}{2} k_{tor,2} (1 - \cos 2\phi) + \tfrac{1}{2} k_{tor,3} (1 - \cos 3\phi)$$

asymmetry (butane) 2-fold groups e.g. COO- standard tetrahedral torsions

Non-Bonded Interactions: van der Waals

E_{vdW} is the steric exclusion and long-range attraction energy (QM origins):



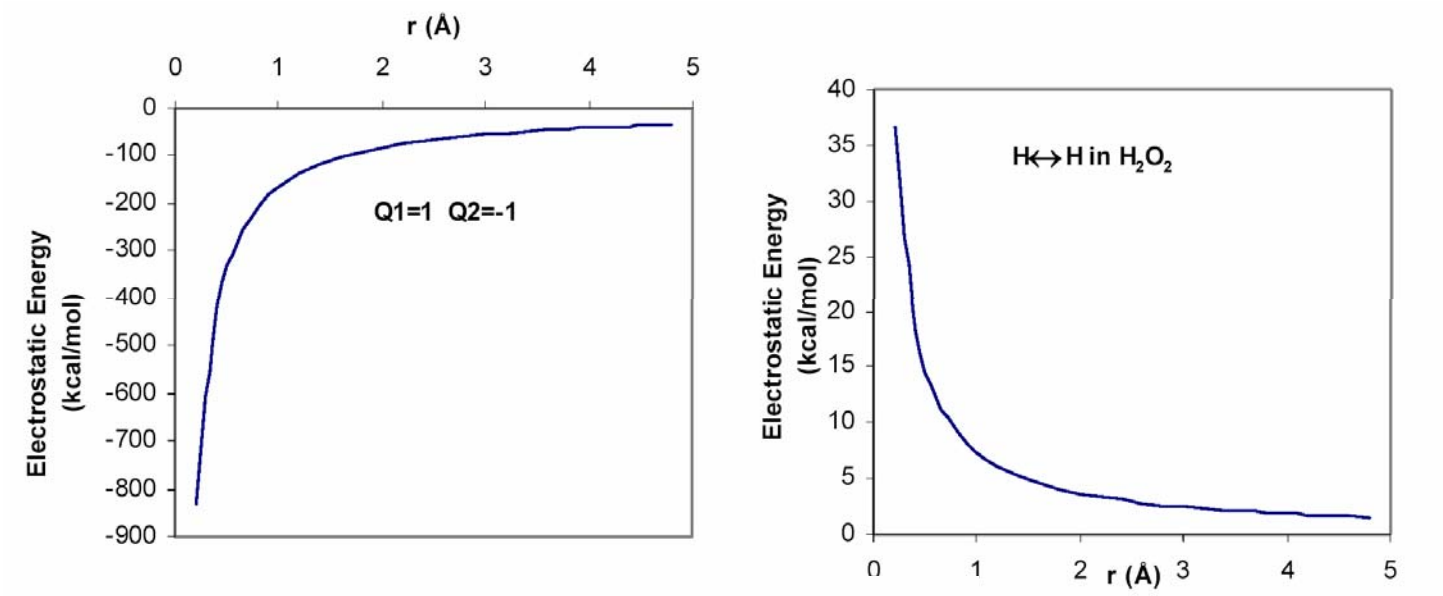
Two frequently used formulas:

$$E_{VDW}(R) = \frac{A}{R^{12}} - \frac{B}{R^6}$$

$$E_{VDW}(R) = 4\varepsilon\left(\left(\frac{\sigma}{R}\right)^{12} - \left(\frac{\sigma}{R}\right)^{6}\right)$$

Non-Bonded Interactions: Coulomb

E_{qq} is the Coulomb potential function for electrostatic interactions of charges:



Formula:

$$E_{qq,ij} = \frac{k\ Q_i\ Q_j}{4\pi\varepsilon\ r_{ij}}$$

The Q_i and Q_j are the partial atomic charges for atoms i and j separated by a distance r_{ij}. ε is the relative dielectric constant. For gas phase calculations ε is normally set to 1. Larger values of ε are used to approximate the dielectric effect of intervening solute (ε~60-80) or solvent atoms in solution. k is a units conversion constant; for kcal/mol, k=2086.4.

Newton's Law

$E_{\text{steric energy}} = E_{\text{str}} + E_{\text{bend}} + E_{\text{improper}} + E_{\text{tor}} + E_{\text{vdW}} + E_{\text{qq}}$

Potential Function $\rightarrow$ Force

$$F_i = -\frac{\partial E_i}{\partial r_i}$$

Newton's Law:

$$F_i = m_i a_i$$

Verlet's Numeric Integration Method

Taylor expansion:

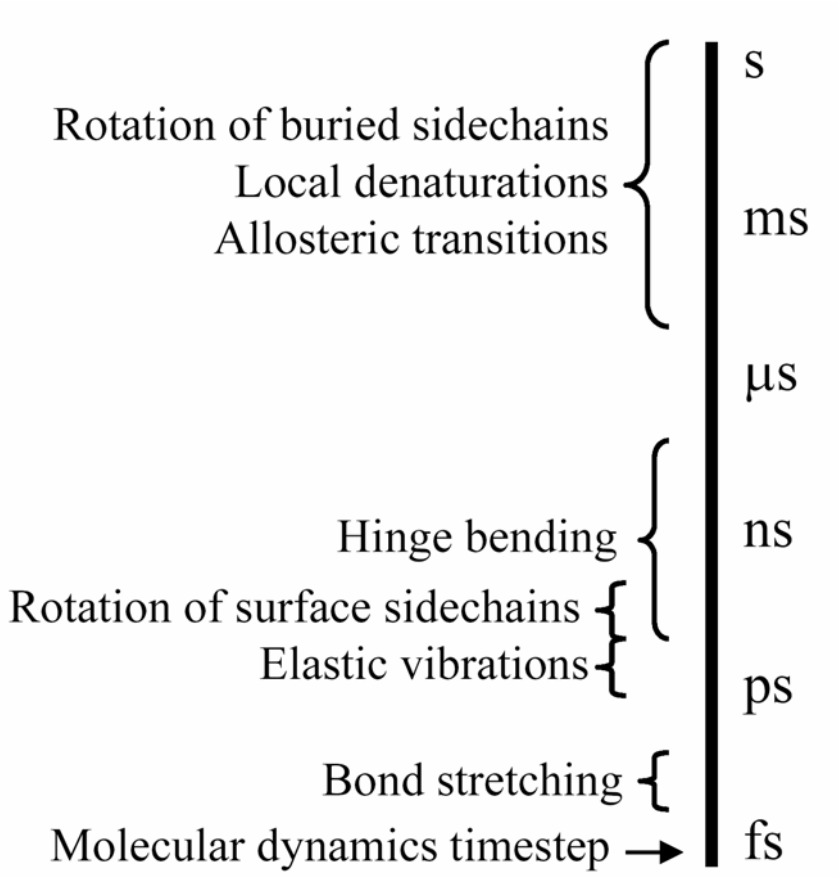
$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2$$

$$r(t-\delta t) = r(t) - v(t)\delta t + \frac{1}{2}a(t)\delta t^2$$

$$\boxed{r(t+\delta t) = 2r(t) - r(t-\delta t) + a(t)\delta t^2}$$

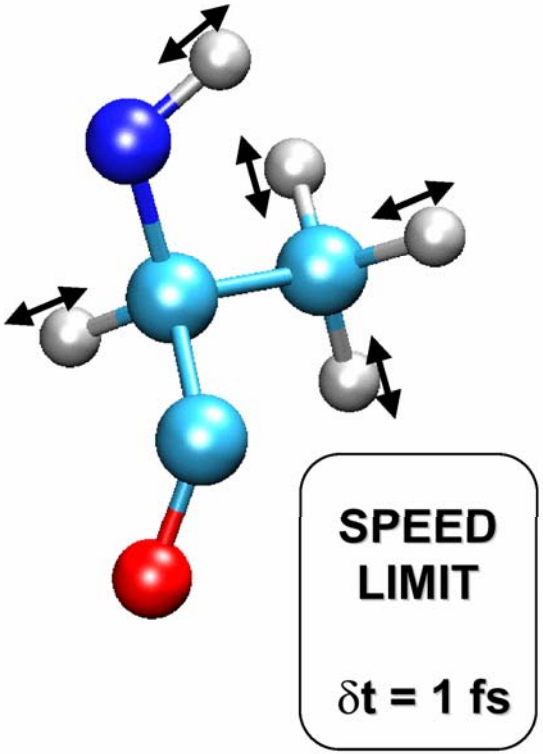
Verlet's Method

Timescale Limitations



- Protein Folding - milliseconds/seconds (10^{-3}-1s)
- Ligand Binding - micro/milliseconds (10^{-6}-10^{-3} s)
- Enzyme catalysis - micro/milliseconds (10^{-6}-10^{-3} s)
- Conformational transitions - pico/nanoseconds (10^{-12}-10^{-9} s)
- Collective vibrations - 1 picosecond (10^{-12} s)
- Bond vibrations - 1 femtosecond (10^{-15} s)

Timescale Limitations



Molecular dynamics:

Integration timestep - 1 fs, set by fastest varying force.

Accessible timescale: about 10 nanoseconds.

Cutting Corners

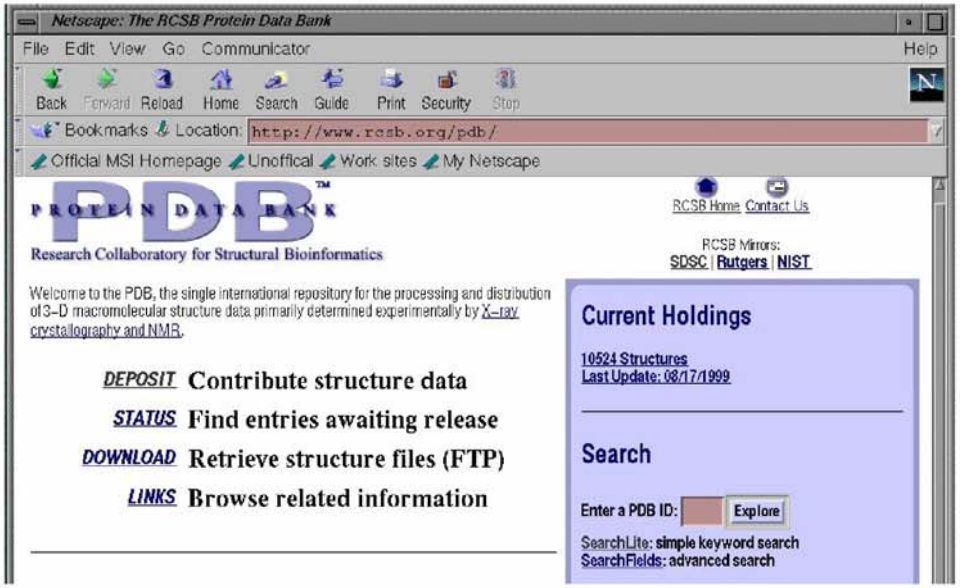
- Nonbonded interactions require order N^2 computer time!
  - Truncating at R_{cutoff} reduces this to order N R_{cutoff}^3
  - Particle mesh Ewald (PME) method adds long range electrostatics at order N log N, only minor cost compared to cutoff calculation.

PME, Schlick 9.4

- Can we extend the timestep, and do this work fewer times?
  - Bonds to hydrogen atoms, which require a 1fs timestep, can be held at their equilibrium lengths, allowing 2fs steps. **SHAKE, Schlick 12.5**
  - Long range electrostatics forces vary slowly, and may be evaluated less often, such as on every second or third step. **MTS, Schlick 13.3**

Input Files: PDB

- Simulations start with atomic structures from the Protein Data Bank, in the standard PDB file format.
- PDB files contain standard records for species, tissue, authorship, citations, sequence, secondary structure, etc.
- We only care about the atom records…
  - atom name (N, C, CA)
  - residue name (ALA, HIS)
  - residue id (integer)
  - coordinates (x, y, z)
  - occupancy (0.0 to 1.0)
  - temp. factor (a.k.a. beta)
  - segment id (6PTI)
- No hydrogen atoms!
  (We must add them ourselves.)



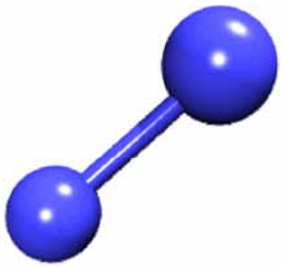
Input Files: PDB

Go to http://www.rcsb.org/pdb/ and explain file content of example PDF file, e.g. PDB entry 1ATN

Input Files: PSF

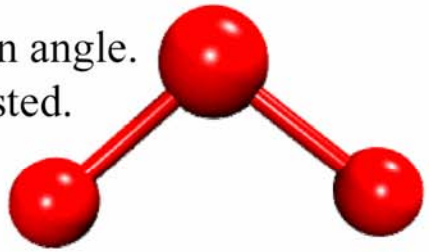
- Every atom in the simulation is listed.
- Provides all static atom-specific values:
  - atom name (N, C, CA)
  - atom type (NH1, C, CT1)
  - residue name (ALA, HIS)
  - residue id (integer)
  - segment id (6PTI)
  - atomic mass (in atomic mass units)
  - partial charge (in electronic charge units)
- What is not in the PSF file?
  - coordinates (dynamic data, initially read from PDB file)
  - velocities (dynamic data, initially from Boltzmann distribution)
  - force field parameters (non-specific, used for many molecules)

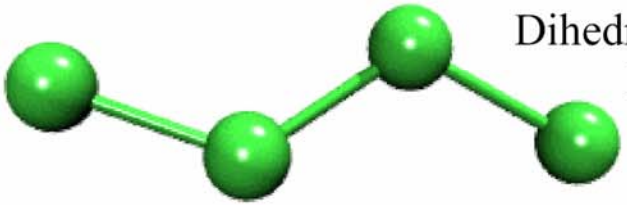
Input Files: PSF



Bonds: Every pair of covalently bonded atoms is listed.

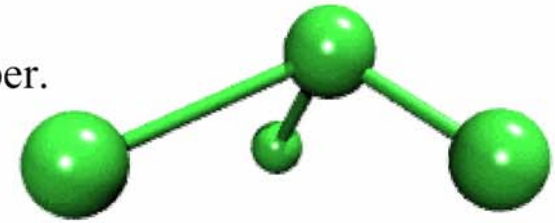
Angles: Two bonds that share a common atom form an angle.
Every such set of three atoms in the molecule is listed.





Dihedrals: Two angles that share a common bond form a dihedral.
Every such set of four atoms in the molecule is listed.

Impropers: Any *planar* group of four atoms forms an improper.
Every such set of four atoms in the molecule is listed.



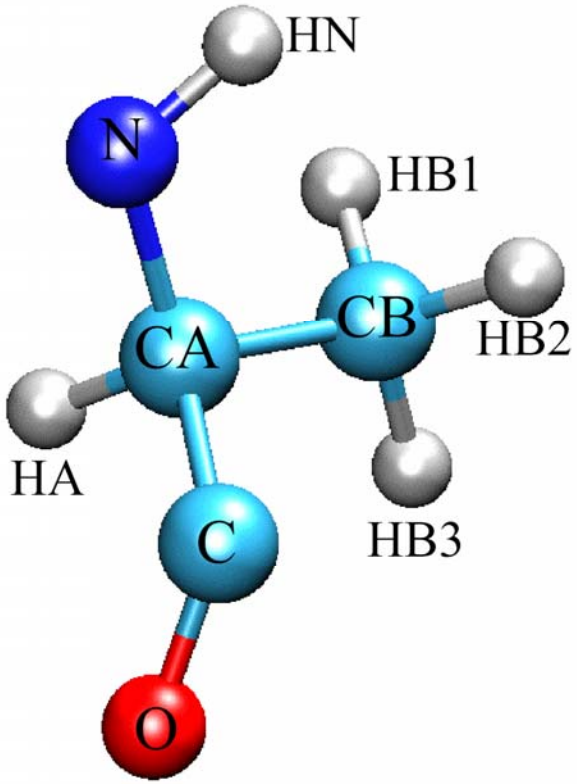
Input Files: PSF

PSF example: download
http://biomachina.org/courses/modeling/download/example.psf

Input Files: Topology Files

blueprints for building a PSF file

- For every type of residue known:
  - atom name, type, mass, and charge
  - bonds within the residue
  - bonds to other residues
  - any planar impropers (rare)
- Additional "patches" for:
  - terminating protein segments
  - joining protein segments
  - modifying protonation states
  - adding disulphide bonds
  - deoxygenating nucleic acids



Input Files: Topology Files

blueprints for building a PSF file

Example: download
http://biomachina.org/courses/modeling/download/topallh22x.pro

Input Files: Parameter Files

defining the MM energy terms

- Equilibrium value and spring constant for
  - every pair of atom types that can form and bond
  - every triple of atom types that can form an angle
  - every quad of atom types that can form a dihedral or improper (many wildcard cases)
- vdW radius and well depth for every atom type
  - actually need these for every pair of atoms types!
  - pair radius calculated from arithmetic mean
  - pair well depth calculated from geometric mean
- Closely tied to matching topology file!

Input Files: Parameter Files

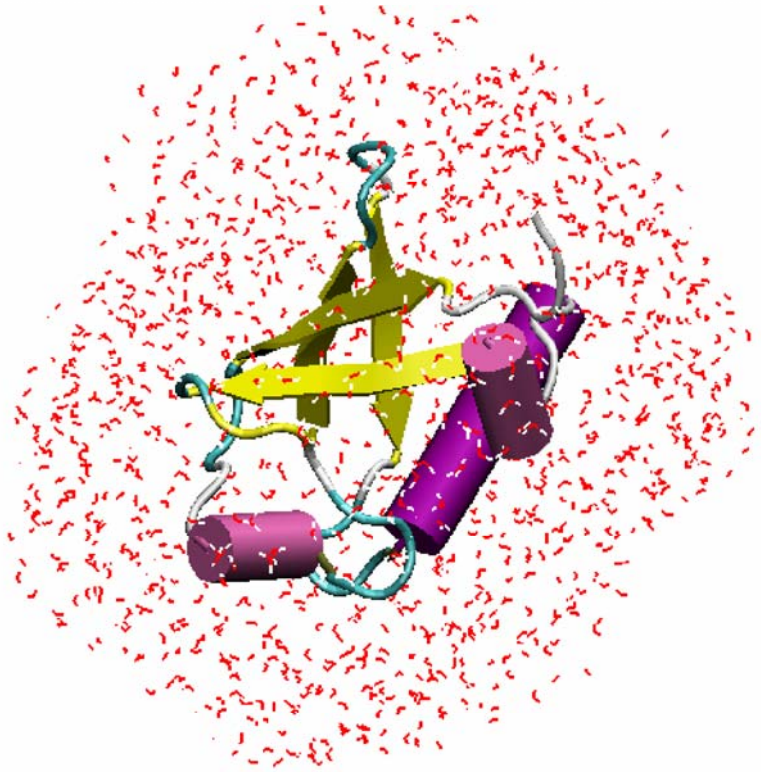
defining the MM energy terms

Example: download
http://biomachina.org/courses/modeling/download/parallh22x.pro

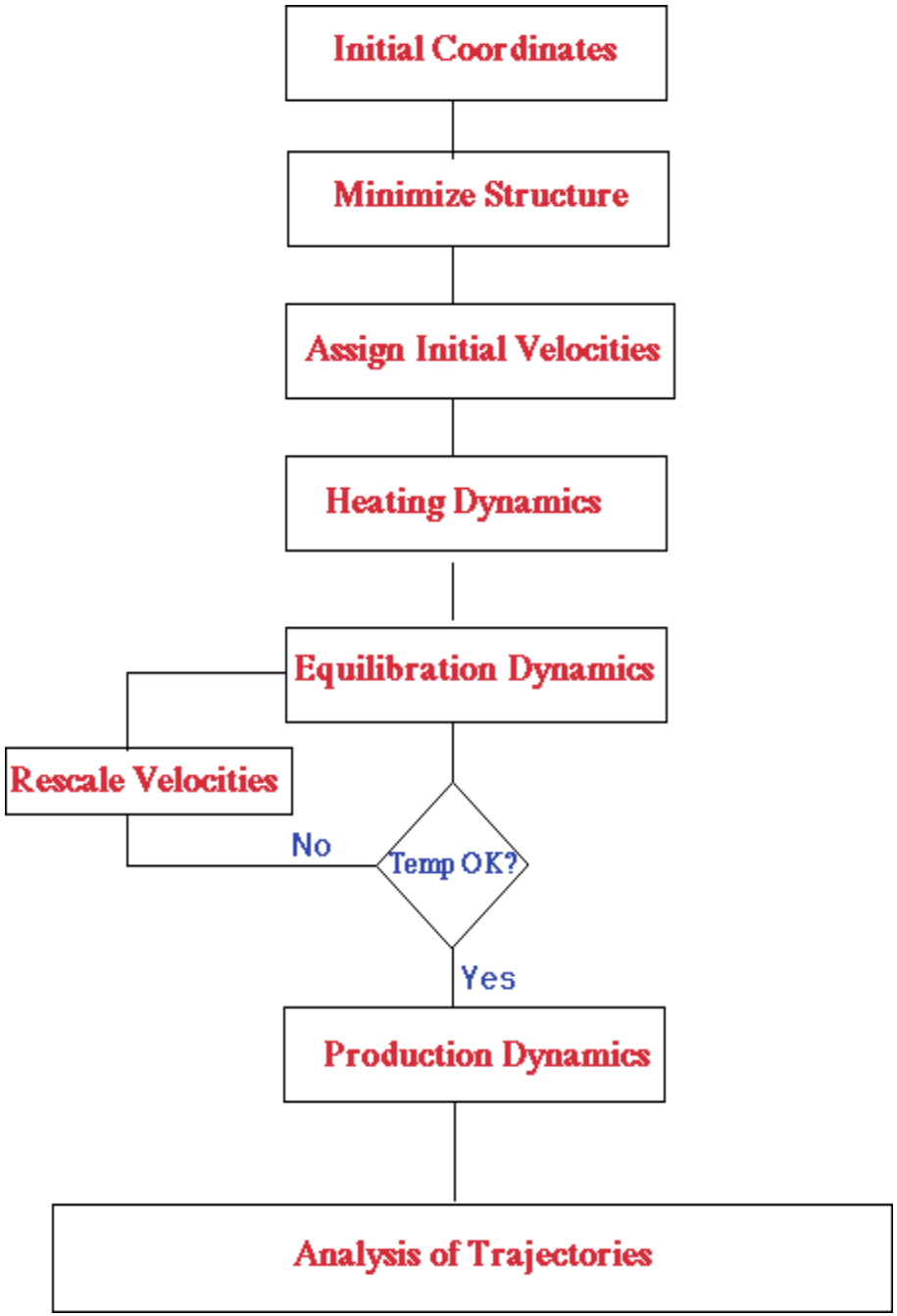
Setting up a MD Simulation

example: ubiquitin

- Obtain file from PDB (1ubq);
- Add missing hydrogen atoms;
- Determine protonation state of HIS residues;
- Add a water box;
- Trim the water box down to a sphere.



Production Run Protocol

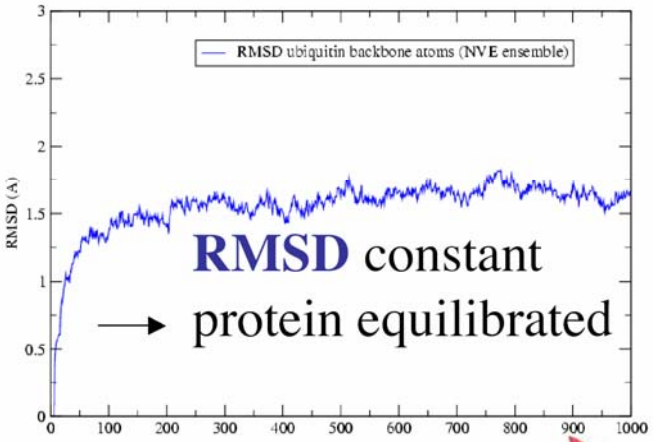


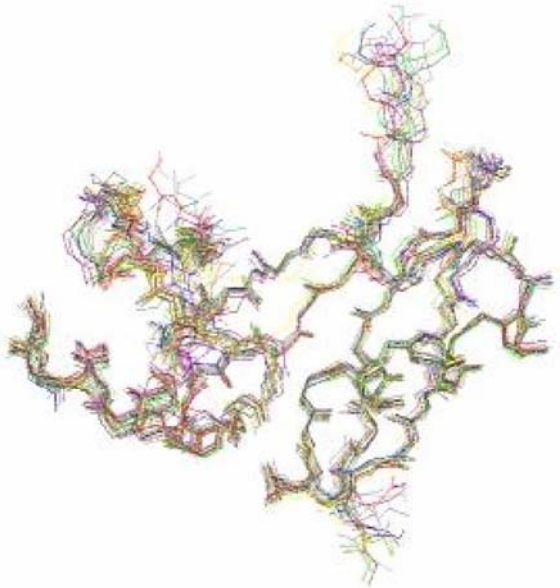
What can be learnt from MD?

equilibrium properties

Root Mean Squared Deviation: measure for equilibration and protein flexibility

$$RMSD_\alpha = \sqrt{\frac{\sum_{j=1}^{N_t}\sum_{\alpha=1}^{N_\alpha}(\vec{r}_\alpha(t_j)-\langle\vec{r}_\alpha\rangle)^2}{N_\alpha}}$$

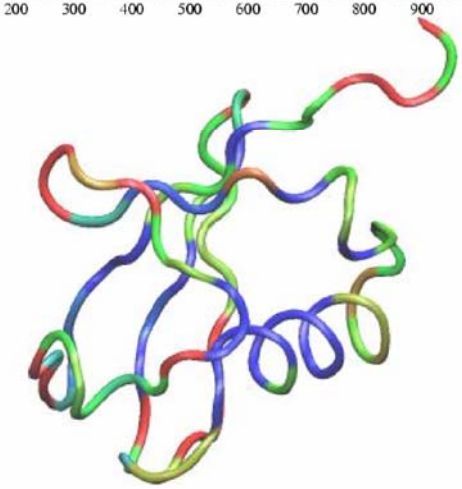




NMR structures

aligned together to see flexibility

Protein sequence exhibits characteristic permanent flexibility!

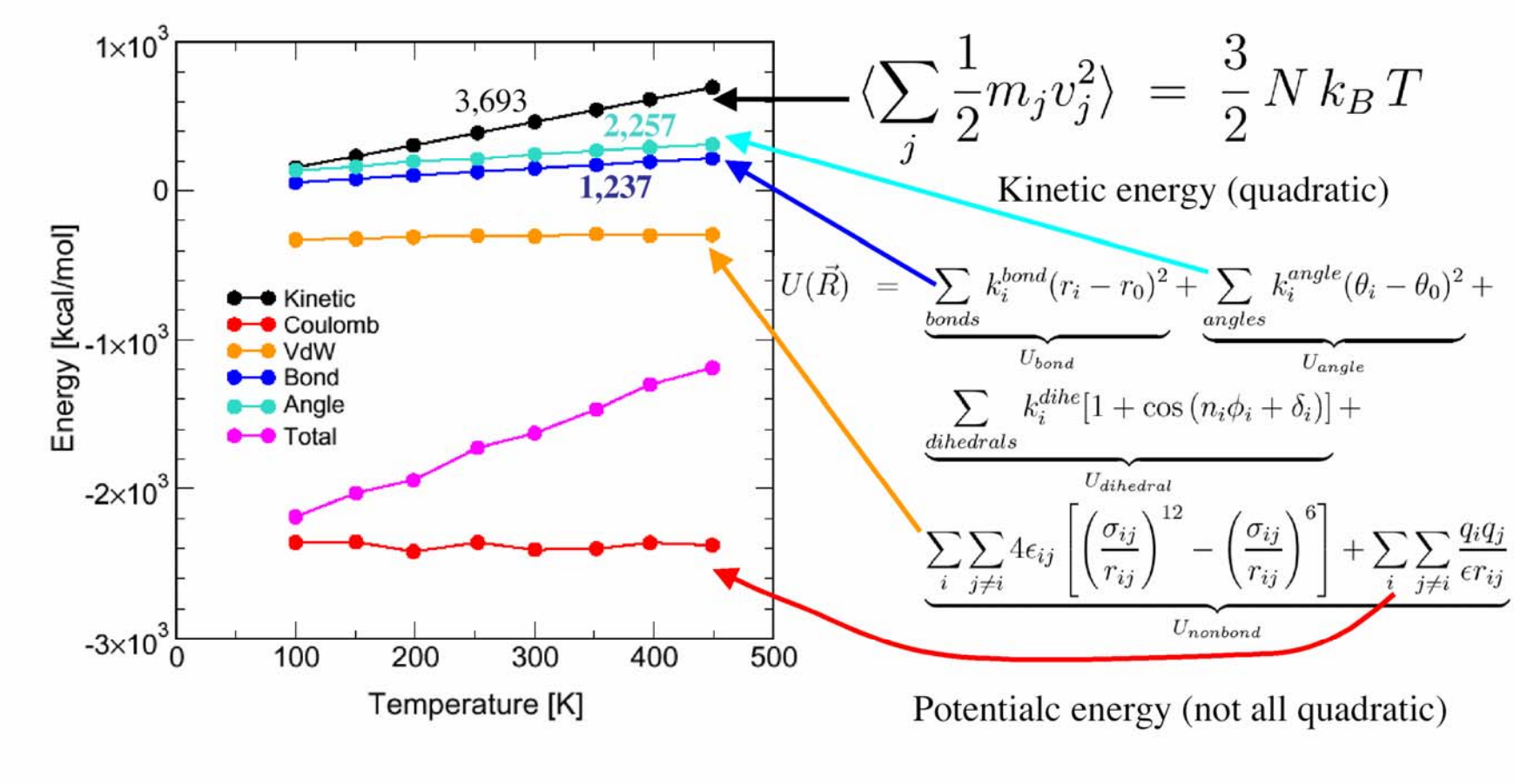


MD simulation

The color represents mobility of the protein through simulation (red = more flexible)

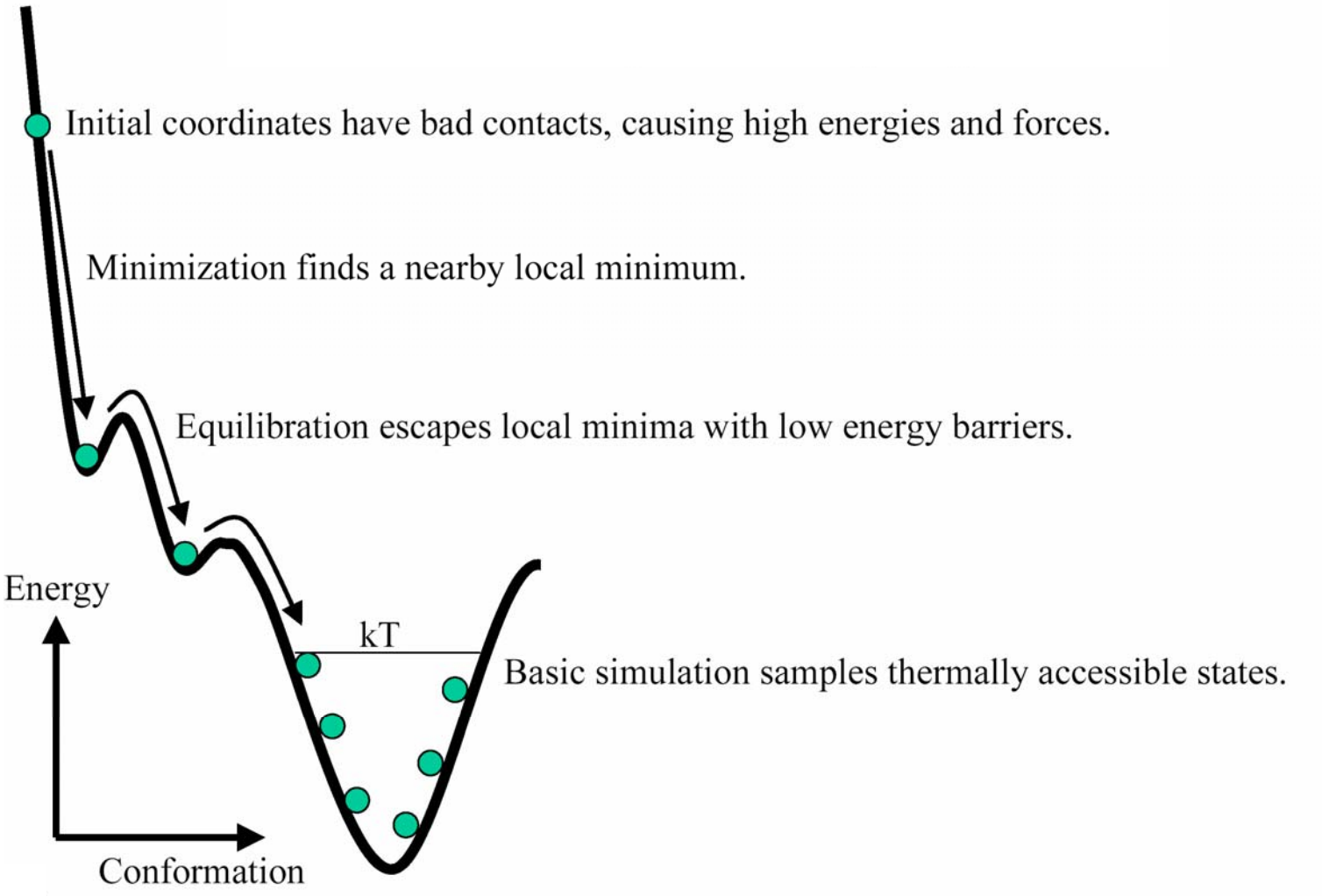
What can be learnt from MD?

energies: kinetic and potential



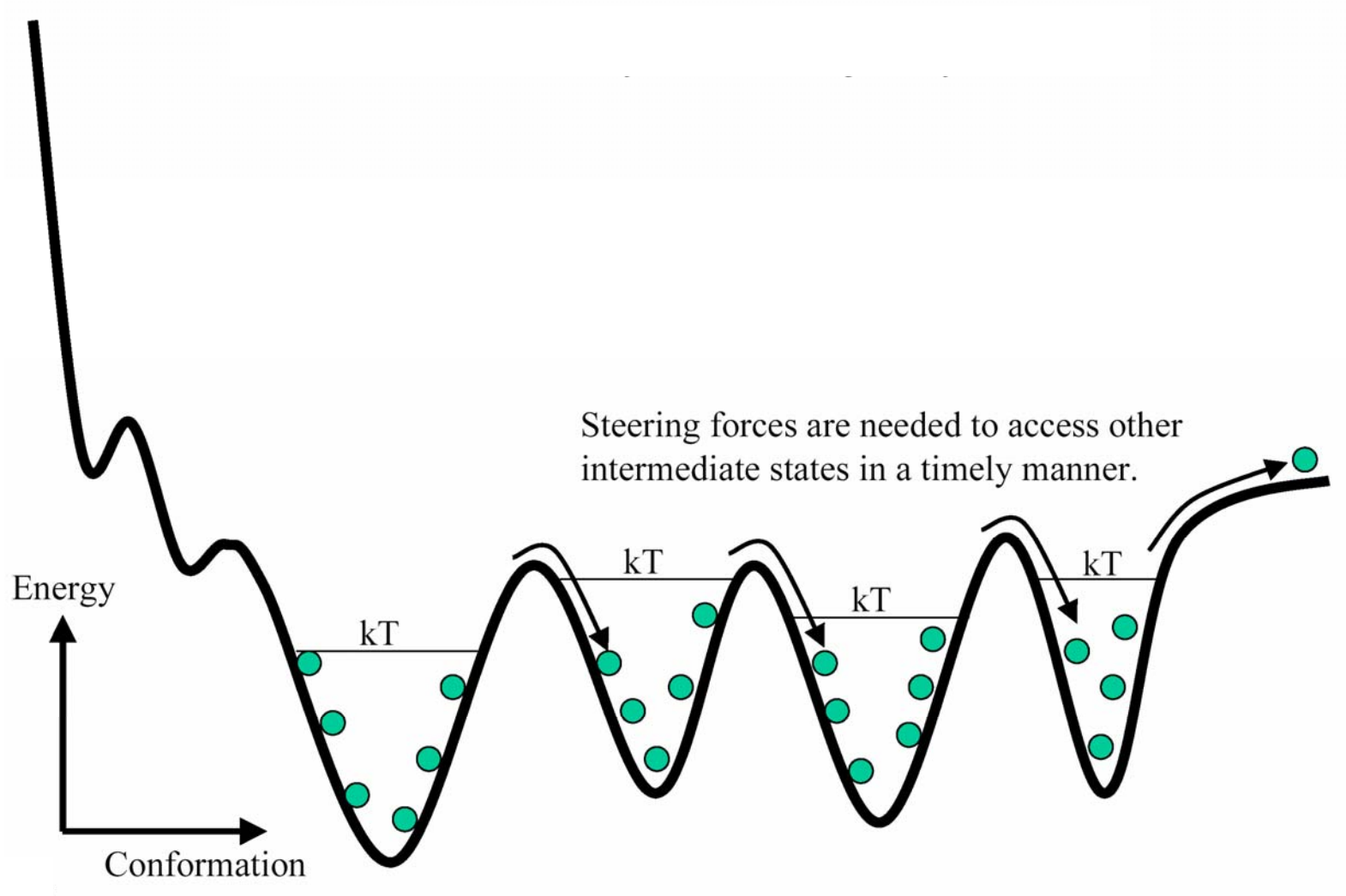
What can be learnt from MD?

exploring conformations



What can be learnt from MD?

exploring conformations



What can be learnt from MD?

exploring conformations

Actin SMD movies, see URL

http://www.biomachina.org/research/projects/actin

Go To:

2. Actin Phosphate Release – The Movie

Further Reading

WWW:

http://cmm.info.nih.gov/intro_simulation

Textbooks:

Schlick, Chapters 8, 9, 12, 13

Acknowledgements/Credits

Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign

Thomas W. Shattuck, Department of Chemistry, Colby College, Waterville, Maine 04901

Jeremy Smith, Biocomputing Unit, U Heidelberg, Germany