The Molecular Dynamics Method

H-bond energy (kcal/mol) 4.0 — 0
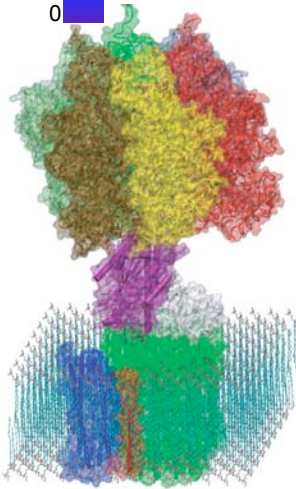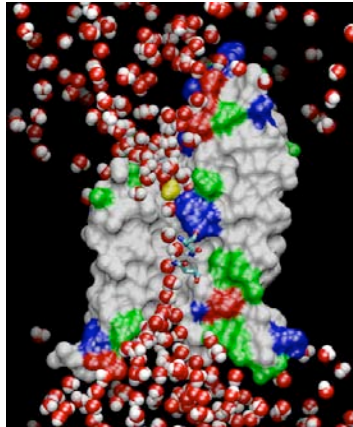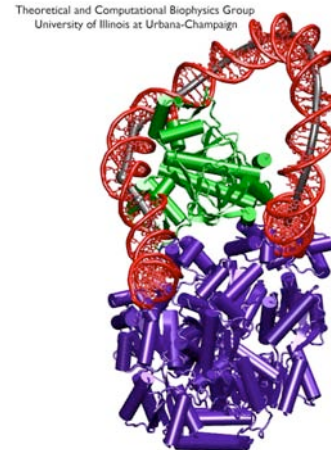
Fibronectin III_1, a mechanical protein that glues cells together in wound healing and in preventing tumor metastasis

Theoretical and Computational Biophysics Group
University of Illinois at Urbana-Champaign

ATPase, a molecular motor that synthesizes the body's weight of ATP a day

AQP filtering a bath tub of the body's water a day

A ternary complex of DNA, lac repressor, and CAP controlling gene expression

Classical Dynamics
F=ma at 300K

Energy function: $U(\vec{r}_1, \vec{r}_2, \cdots \vec{r}_N) = U(\vec{R})$

used to determine the force on each atom:

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i = -\vec{\nabla} U(\vec{R})$$

yields a set of 3N coupled 2nd-order differential equations that can be propagated forward (or backward) in time.

Initial coordinates obtained from crystal structure, velocities taken at random from Boltzmann distribution.

Maintain appropriate temperature by adjusting velocities.

Langevin Dynamics
come on, feel the noise

Langevin dynamics deals with each atom separately, balancing a small friction term with Gaussian noise to control temperature:

$$m\ddot{\vec{r}} = \vec{F}(\vec{r}) - \gamma m \dot{\vec{r}} + \vec{R}(t)$$

$$\langle \vec{R}(t) \cdot \vec{R}(t') \rangle = 6 k_B T \gamma \, \delta(t - t')$$

Classical Dynamics
discretization in time for computing

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i = -\vec{\nabla} U(\vec{R})$$

Use positions and accelerations at time t and the positions from time t-δt to calculate new positions at time t+δt.

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

$$+$$

$$\mathbf{r}(t - \delta t) \approx \mathbf{r}(t) - \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2$$
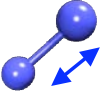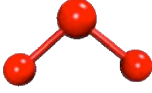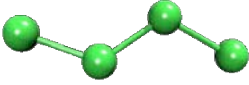
$$\cup \qquad\qquad -\vec{\nabla} U(\vec{R})\big/ m_i$$

$$\boxed{\mathbf{r}(t + \delta t) \approx 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + \mathbf{a}(t)\delta t^2}$$
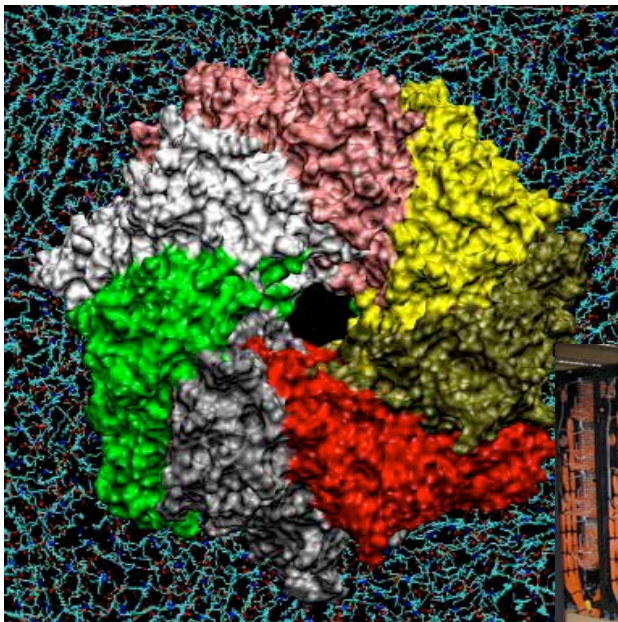
Potential Energy Function of Biopolymer

- Simple, fixed algebraic form for every type of interaction.
- Variable parameters depend on types of atoms involved.

$$U(\vec{R}) = \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} +$$

$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} +$$

$$\underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$
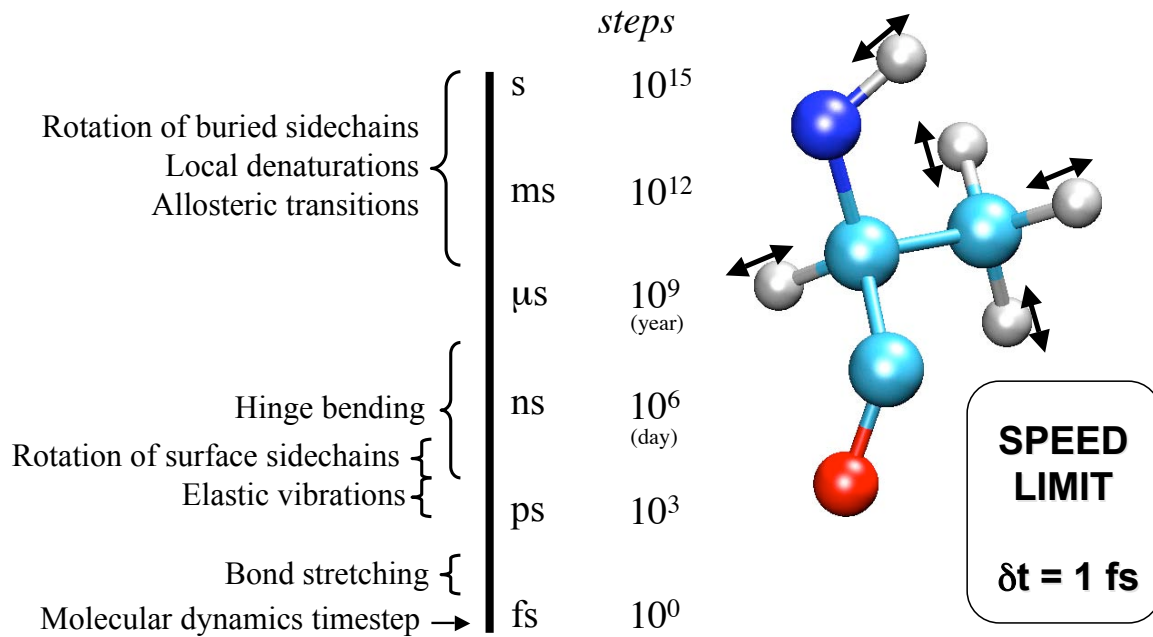
Large is no problem. But …

Molecular dynamics simulation of alpha-hemolysin with about 300,000 atoms

NCSA machine room

But long is!
## *biomolecular timescale and timestep limits*

steps

Rotation of buried sidechains	s	10^{15}
Local denaturations		
Allosteric transitions	ms	10^{12}
	μs	10^9 (year)
Hinge bending	ns	10^6 (day)
Rotation of surface sidechains		
Elastic vibrations	ps	10^3
Bond stretching		
Molecular dynamics timestep →	fs	10^0

SPEED
LIMIT

$\delta t = 1$ fs

PDB Files
## *a little information*

- Simulations start with a crystal structure 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.)
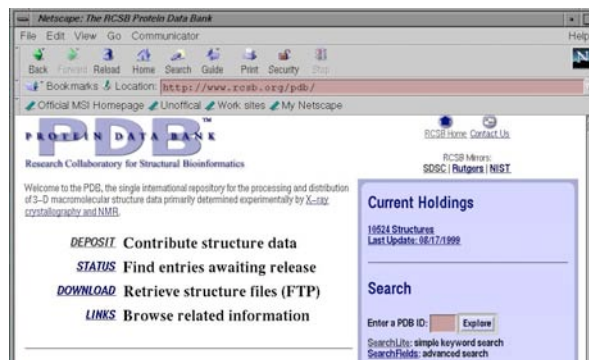
PSF Files
atomic properties (mass, charge, type)

- 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)
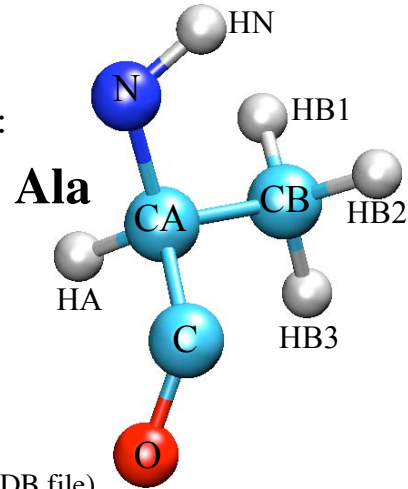
PSF Files
molecular structure (bonds, angles, etc.)
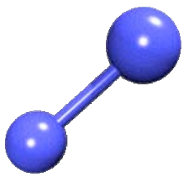
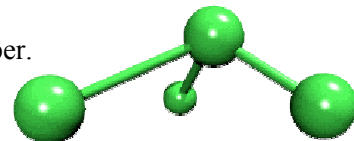Bonds: Every pair of covalently bonded atoms is listed.

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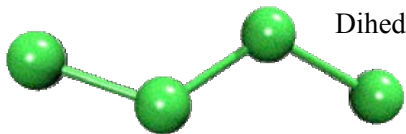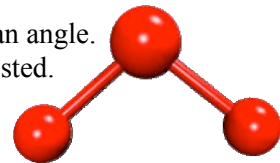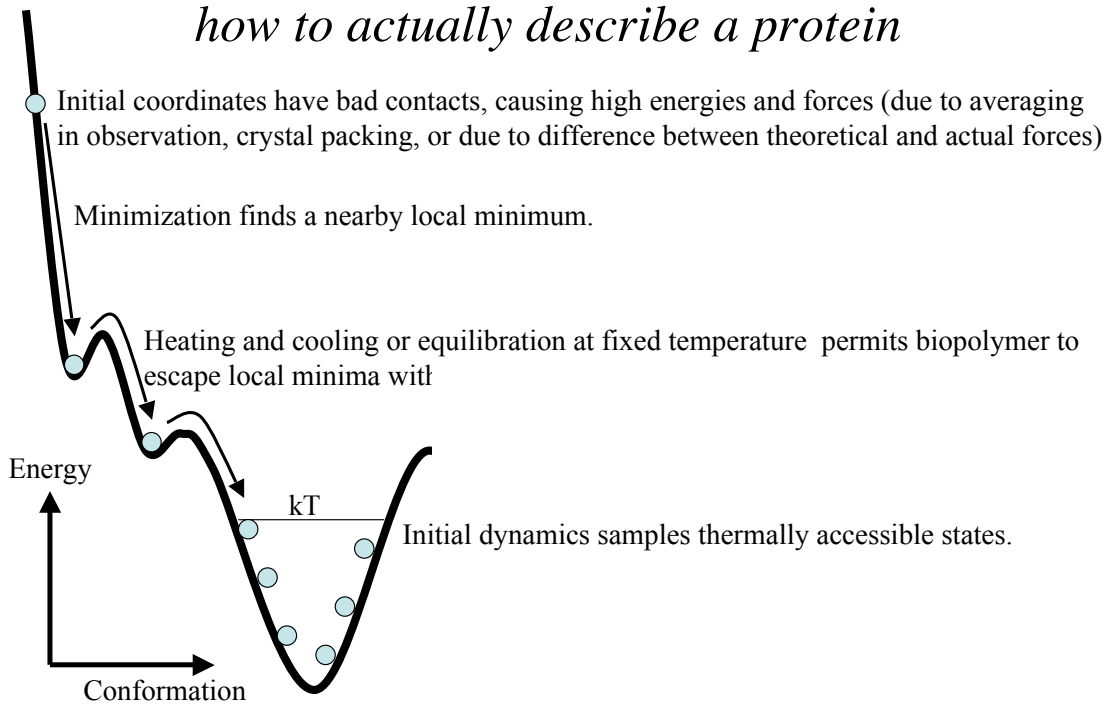
Dihedrals: Two angles that share a common bond form a dihedral. Every such set of four atoms in the molecule is listed.

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

From the Mountains to the Valleys
how to actually describe a protein

Initial coordinates have bad contacts, causing high energies and forces (due to averaging in observation, crystal packing, or due to difference between theoretical and actual forces)

Minimization finds a nearby local minimum.

Heating and cooling or equilibration at fixed temperature permits biopolymer to escape local minima with

Energy

kT

Initial dynamics samples thermally accessible states.

Conformation

From the Mountains to the Valleys
a molecular dynamics tale

Longer dynamics access other intermediate states; one may apply external forces to access other available states in a more timely manner.

Energy

kT

kT

kT

kT

kT

Conformation

NAMD: The Program we will Use

NAMD programmer J. Phillips Ph.D. UIUC Physics

Simulation of large biomolecular systems

2002 Gordon Bell Award for parallel scalability.

Runs at NSF centers, on clusters, and on desktop.

Available for **FREE** as precompiled binaries; includes source code.
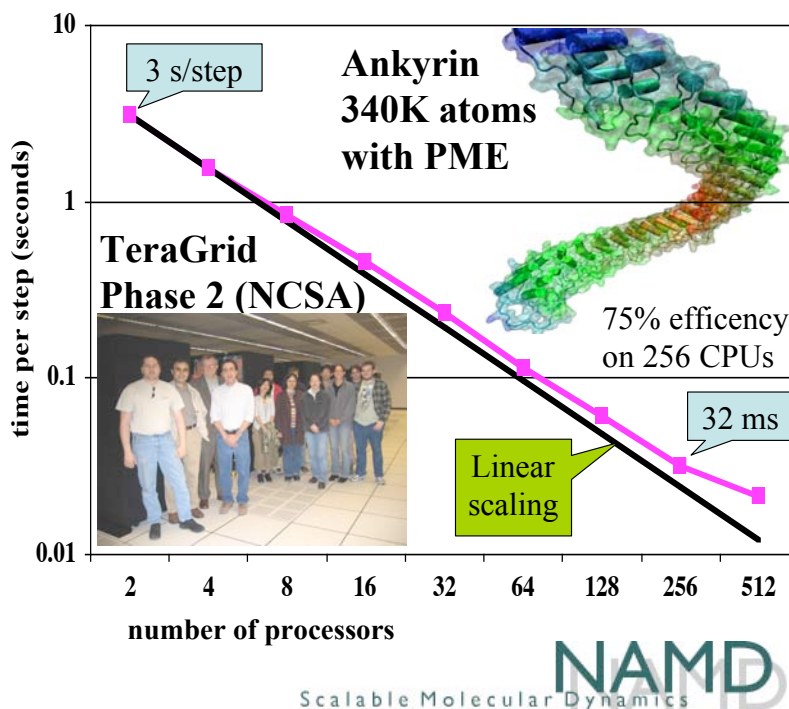
10,000 registered users.

CHARMM Potential Function

- Simple, fixed algebraic form for every type of interaction.
- Form stems from compromise between accuracy and speed.
- Variable parameters depend on types of atoms involved.

$$
\begin{aligned}
U(\vec{R}) =\ & \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} + \\
& \underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} + \\
& \underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}
\end{aligned}
$$

Parameter Files
biomolecular paint by numbers

- 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
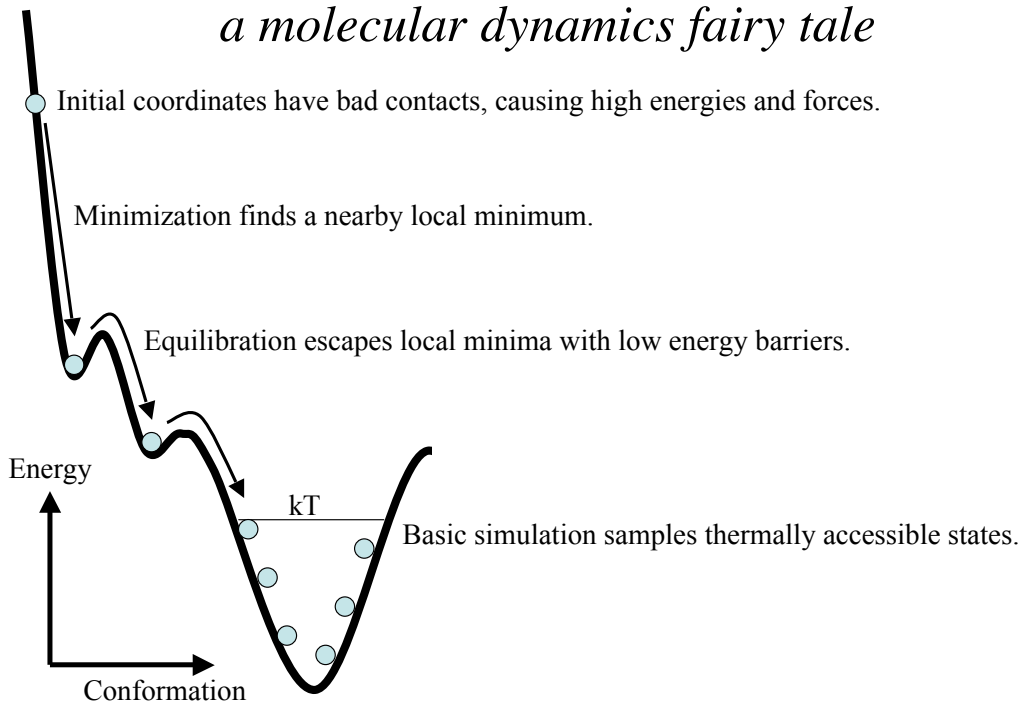- Closely tied to matching topology file!

Molecular Dynamics Method

- PDB, PSF, topology, and parameter files
- Molecular dynamics
  - …in an ideal world
    - …and in our world
      - …with computers
        - …using NAMD
- Preparing a protein using VMD
- You prepare a protein using VMD
  - …and simulate it using NAMD
    - …in the hands-on Tuesday afternoon

  Don't worry, the written tutorial is very complete.
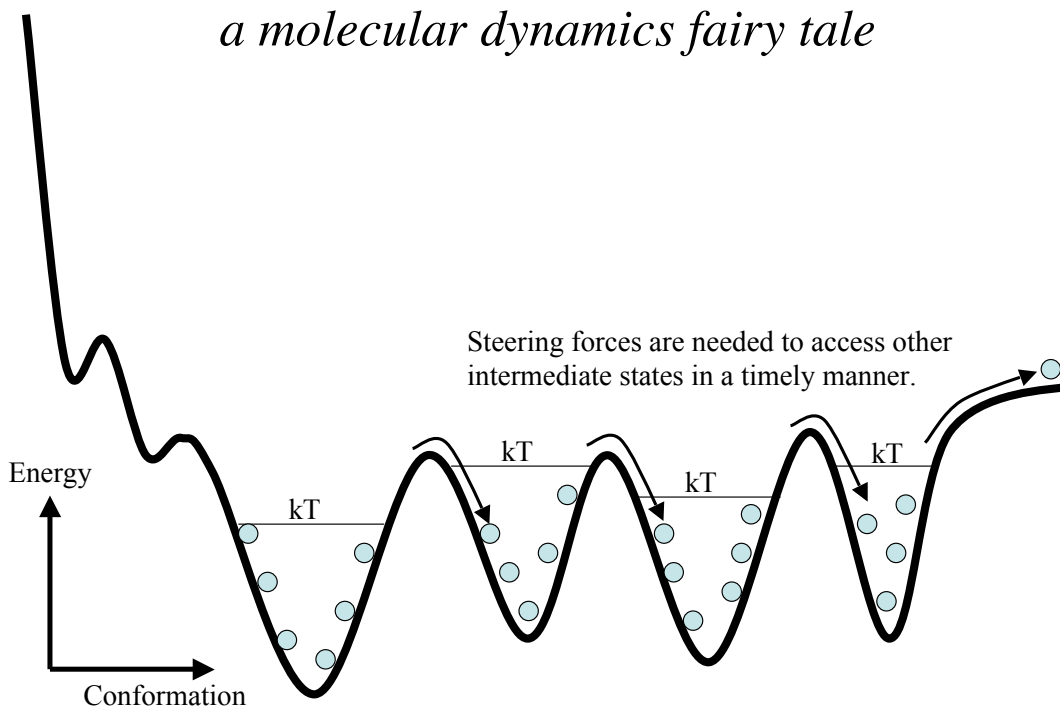 ***You will learn by doing. This talk is an overview.***

From the Mountains to the Valleys
a molecular dynamics fairy tale

Initial coordinates have bad contacts, causing high energies and forces.

Minimization finds a nearby local minimum.

Equilibration escapes local minima with low energy barriers.

Energy

kT

Basic simulation samples thermally accessible states.

Conformation

From the Mountains to the Valleys
a molecular dynamics fairy tale

Steering forces are needed to access other intermediate states in a timely manner.

Energy

kT

kT

kT

kT

kT

Conformation

Step by Step
discretization in time

Use positions and accelerations at time t and the positions
from time t-δt to calculate new positions at time t+δt.

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

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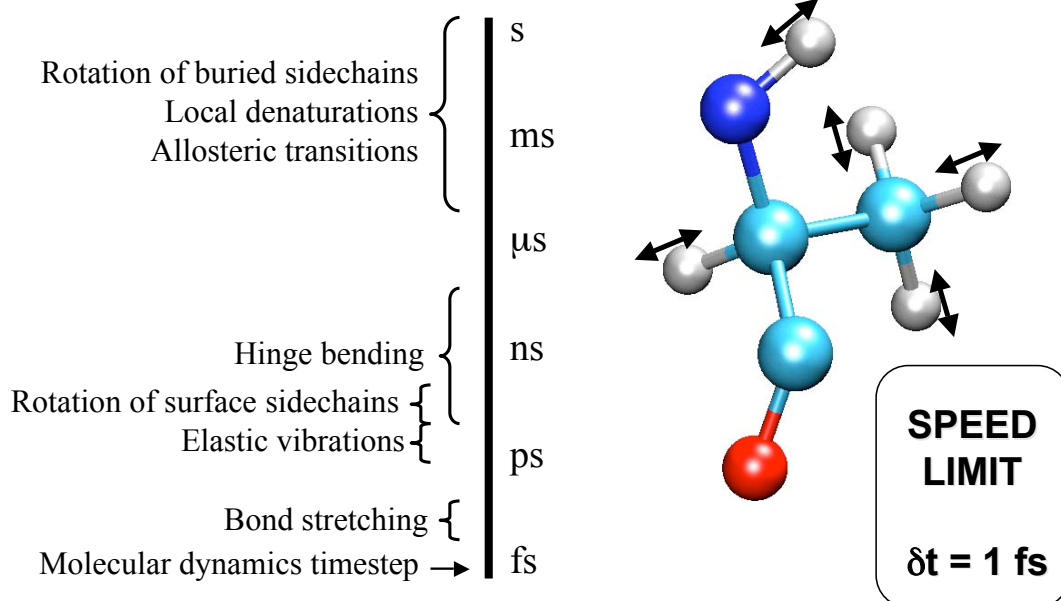
+

$$\cup \qquad\qquad -\vec{\nabla}U(\vec{R})\big/ m_i$$

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

Hurry Up and Wait
biomolecular timescale and timestep limits

Rotation of buried sidechains ⎱
Local denaturations ⎰ s
Allosteric transitions

ms

μs

Hinge bending { ns

Rotation of surface sidechains {
Elastic vibrations { ps

Bond stretching {

Molecular dynamics timestep → fs

**SPEED LIMIT**

δt = 1 fs

Cutting Corners
cutoffs, PME, rigid bonds, and multiple timesteps

- 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.

- 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.
    - Long range electrostatics forces vary slowly, and may be evaluated less often, such as on every second or third step.

Linux Clusters 101
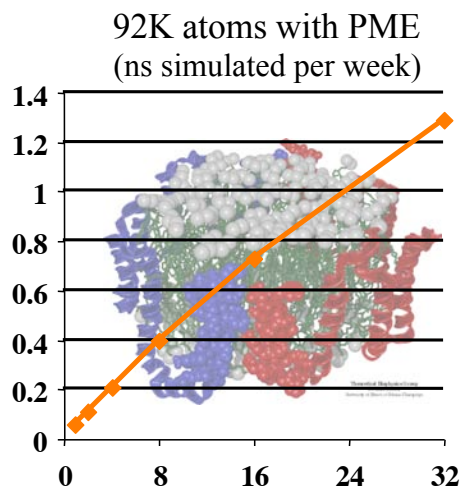## *parallel computing on a professor's salary*

Learn to build your own Linux cluster!

Easy to manage

$1000 per processor

92K atoms with PME
(ns simulated per week)

Linux Clusters 101
parallel computing on a professor's salary

Learn to build your own Linux cluster!

Easy to manage

$1000 per processor

92K atoms with PME
(ns simulated per week)
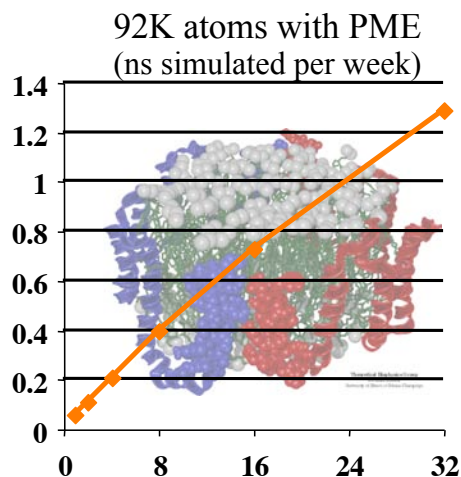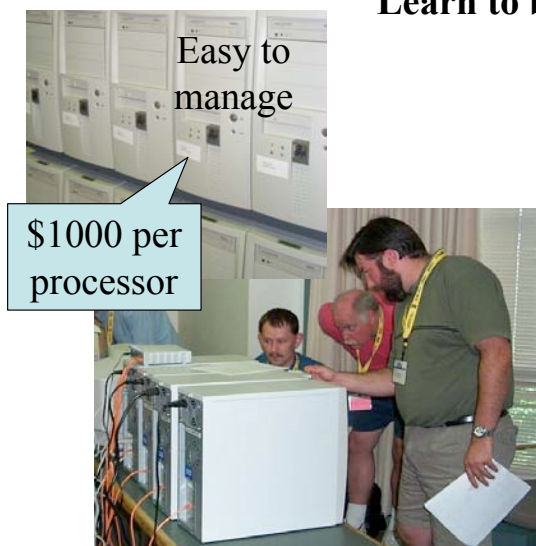
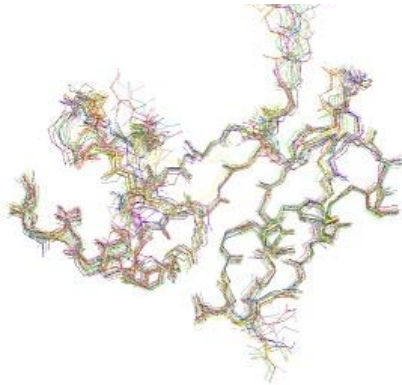# Molecular Dynamics of Proteins

Ubiquitin
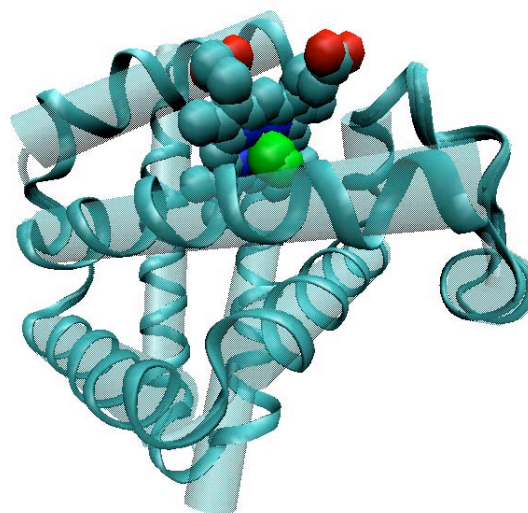
Myoglobin
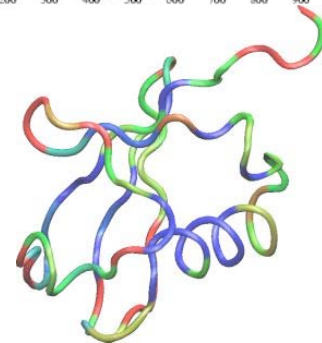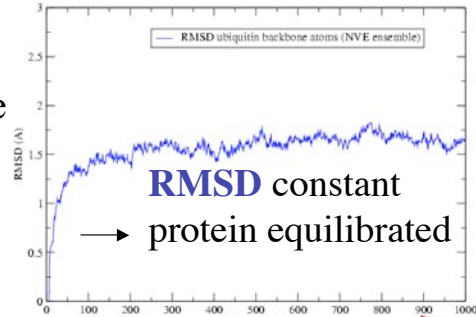
Equilibrium Properties of Proteins

Ubiquitin

Root Mean Squared Deviation: measure for equilibration and protein flexibility

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

RMSD constant
→ protein equilibrated

Protein sequence exhibits characteristic permanent flexibility!

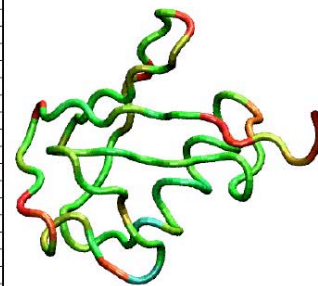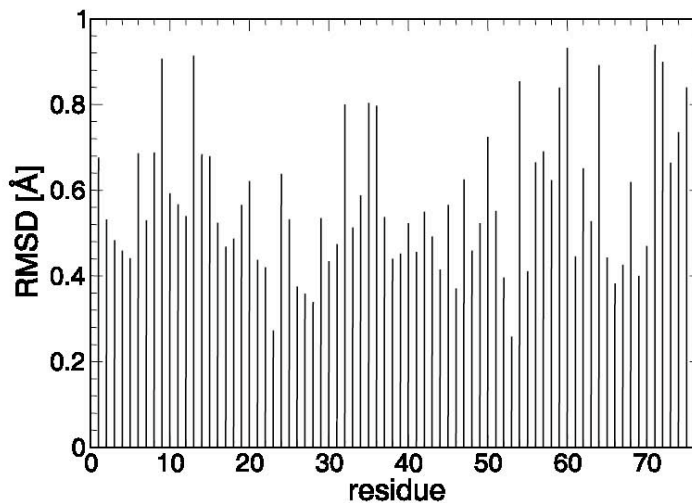NMR structures
aligned together to see flexibility

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

Thermal Motion of Ubiquitin from MD

RMSD values per residue
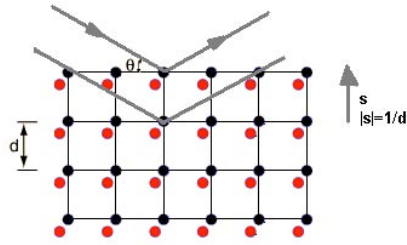
Thermal Motion of Ubiquitin from MD

Temperature Dependence of Crystal Diffraction (Debye-Waller factor)

Bragg's law

$$2\,d\,\sin\theta \;=\; \lambda$$

structure factor

$$f_j \exp\left[-i\vec{s}\cdot\vec{r}_j\right]$$

But the atom carries out thermal vibrations around equilibrium position $\vec{x}_j$

$$\vec{r}_j(t) \;=\; \vec{x}_j \;+\; \vec{u}_j(t)$$

Accordingly:

$$\langle f_j \exp[-i\vec{s}\cdot\vec{r}_j]\rangle \;=\; f_j \exp[-i\vec{s}\cdot\vec{x}_j]\,\langle\exp[-i\vec{s}\cdot\vec{u}_j]\rangle$$

Thermal Motion of Ubiquitin from MD

Temperature Dependence of Crystal Diffraction (Debye-Waller factor)

One can expand:

$$\langle\exp[-i\vec{s}\cdot\vec{u}_j]\rangle \;=\; 1 \;-\; i\underbrace{\langle\vec{s}\cdot\vec{u}_j\rangle}_{=0} \;-\;\tfrac{1}{2}\langle(\vec{s}\cdot\vec{u}_j)^2\rangle \;\;+\;\ldots$$

Spatial average:
$$\langle(\vec{s}\cdot\vec{u}_j)^2\rangle \;=\; \tfrac{1}{3}s^2\langle u_j^2\rangle$$

One can carry out the expansion further and show

$$\langle\exp[-i\vec{s}\cdot\vec{u}_j]\rangle \;=\; \exp\left[-\tfrac{1}{6}s^2\langle\langle u_j^2\rangle\rangle\right]$$

Using for the thermal amplitude of the harmonic oscillator
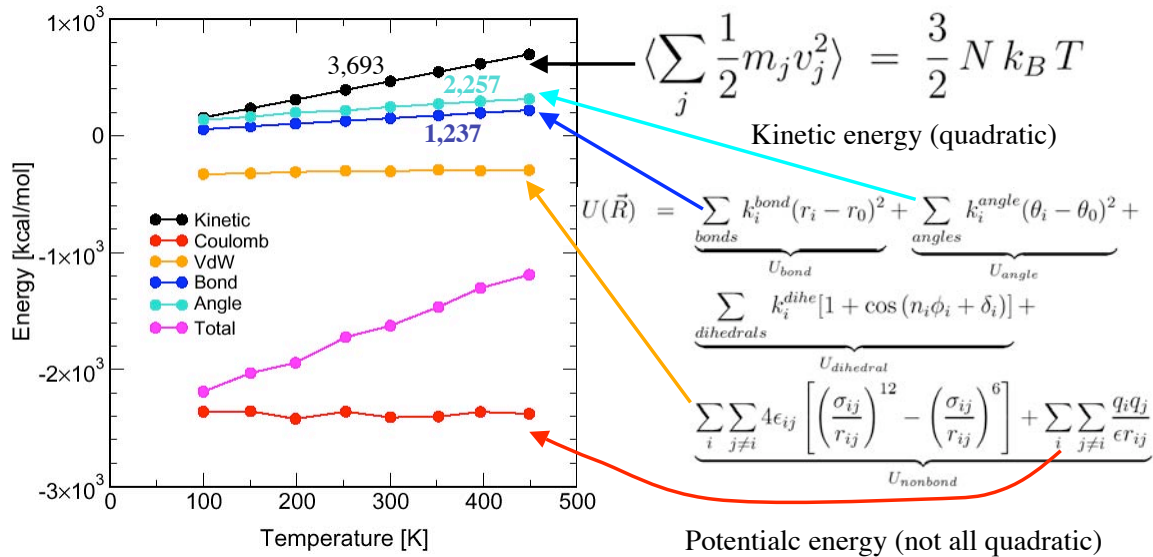
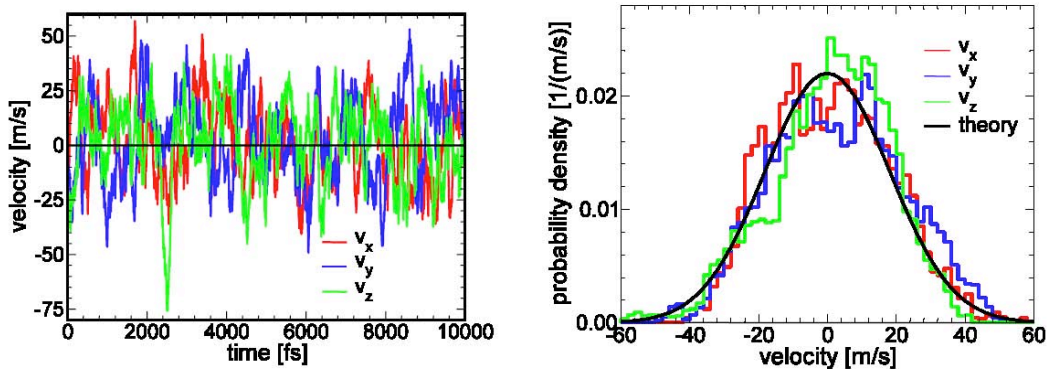$$\tfrac{1}{2}m\omega^2 u_j^2 \;=\; \tfrac{3}{2}k_B T$$

one obtains

Debye-Waller factor

$$\langle f_j \exp\left[-i\vec{s}\cdot\vec{r}_j\right]\rangle \;=\; f_j \overbrace{\exp[-s^2 k_B T/2m\omega^2]}\exp[-i\vec{s}\cdot\vec{x}_j]$$

Equilibrium Properties of Proteins

Energies: kinetic and potential

$$\left\langle \sum_j \frac{1}{2} m_j v_j^2 \right\rangle = \frac{3}{2} N k_B T$$

Kinetic energy (quadratic)

$$U(\vec{R}) = \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} +$$

$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i \phi_i + \delta_i)]}_{U_{dihedral}} +$$

$$\underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$

Potentialc energy (not all quadratic)

Maxwell Distribution of Atomic Velocities

$$p(v_\sigma) = \sqrt{\frac{m}{2\pi k_B T}} \exp\left[-\frac{m v_\sigma^2}{2 k_B T}\right]$$

$$\sigma = x, y, z$$

Mean Kinetic Energy
Exercise in Statistics

$$\langle \tfrac{1}{2}mv^2 \rangle = \int_{-\infty}^{\infty} dv \left(\tfrac{1}{2}mv^2\right) p(v)$$

$$= \sqrt{\frac{m}{2\pi k_B T}} \int_{-\infty}^{\infty} dv \left(\tfrac{1}{2}mv^2\right) \exp\left[-\frac{mv^2}{2k_B T}\right]$$

$$= k_B T \sqrt{\tfrac{1}{\pi}} \int_{-\infty}^{\infty} \sqrt{\frac{m}{2k_B T}} \, dv \left(\frac{mv^2}{2k_B T}\right) \exp\left[-\frac{mv^2}{2k_B T}\right]$$

$$= k_B T \sqrt{\tfrac{1}{\pi}} \int_{-\infty}^{\infty} dy \, y^2 \exp{-y^2}]$$

Use formula below: $\qquad \langle \tfrac{1}{2}mv^2 \rangle = \tfrac{1}{2}k_B T$

$$\int_0^{\infty} dy \, y^m \exp[-y^2] = \tfrac{1}{2}\Gamma\left(\tfrac{m+1}{2}\right)$$
$$\Gamma(x+1) = x\,\Gamma(x), \quad \Gamma\left(\tfrac{1}{2}\right) = \sqrt{\tfrac{1}{2}}$$

Maxwell Kinetic EnergyDistribution
Second Exercise in Statistics

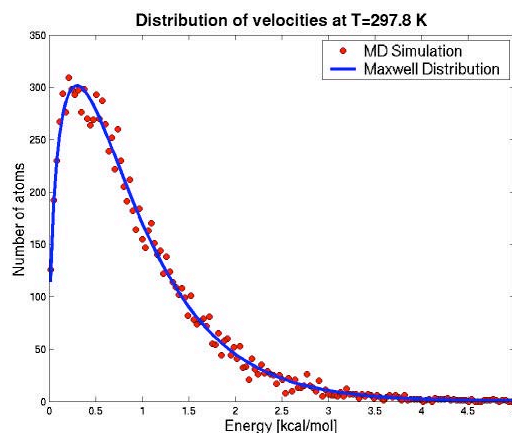One-dimensional kinetic energy: $\quad \epsilon_k = \tfrac{1}{2}mv_\sigma^2$

$$\tilde{p}(\epsilon_k) = p(v_\sigma)\frac{dv_\sigma}{d\epsilon_k} \quad \longrightarrow \quad \tilde{p}(\epsilon_k) = \sqrt{1/\pi k_B T}\sqrt{1/\epsilon_k}\exp[-\epsilon_k/k_B T]$$
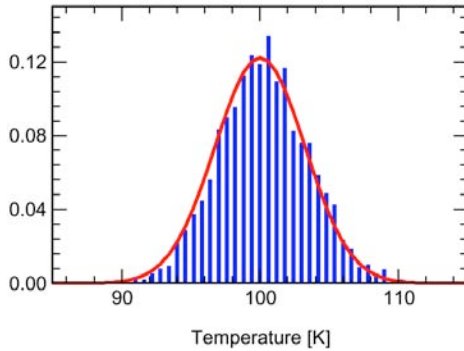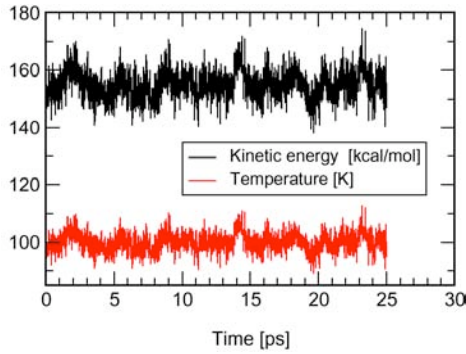
(factor 2 from restriction of integration to positive values)

For the total kinetic energy
(in three dimensions)
holds then

$$\tilde{p}(\epsilon_k) =$$

$$\frac{4}{\sqrt{\pi}}\left(k_B T\right)^{-\frac{3}{2}}\sqrt{\epsilon_k}\exp[-\epsilon_k/k_B T]$$

Distribution of velocities at T=297.8 K

Analysis of E_{kin}, T (free dynamics)

Definition of Temperature

$$\langle \sum_j \frac{1}{2} m_j v_j^2 \rangle = \frac{3}{2} N k_B T$$

$$T = \frac{2}{3N k_B} \langle \sum_j \frac{1}{2} m_j v_j^2 \rangle$$

The atomic velocities of a protein establish a thermometer, but is it accurate?

The atomic velocity thermometer is inaccurate due to the finite size of a protein!

Temperatur Fluctuations

Maxwell distribution

$$dP(v_n) = c \exp(-m\, v_n^2/2k_B T)\, dv_n \qquad (7)$$

Individual kinetic energy $\epsilon_n = m v_n^2/2$

$$dP(\epsilon_n) = (\pi T_0 \epsilon_n)^{-1/2} \exp(-\epsilon_n/k_B T_0)\, d\epsilon_n \qquad (8)$$
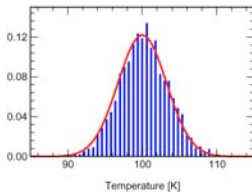
One can derive

$$\langle \epsilon_n \rangle = T_0/2 \qquad (9)$$

$$\langle \epsilon_n^2 \rangle = 3 T_0^2/4 \qquad (10)$$

$$\langle \epsilon_n^2 \rangle - \langle \epsilon_n \rangle^2 = T_0^2/2 \qquad (11)$$

The distribution of the total kinetic energy $E_{kin} = \sum_j \frac{1}{2} m_j v_j^2$, according to the central limit theorem, is approximately Gaussian

$$P(E_{kin}) = c \exp\left(\frac{-(E_{kin} - \langle E_{kin} \rangle)^2}{2\left(\frac{3Nk_B^2 T_0^2}{2}\right)} \right) \qquad (12)$$
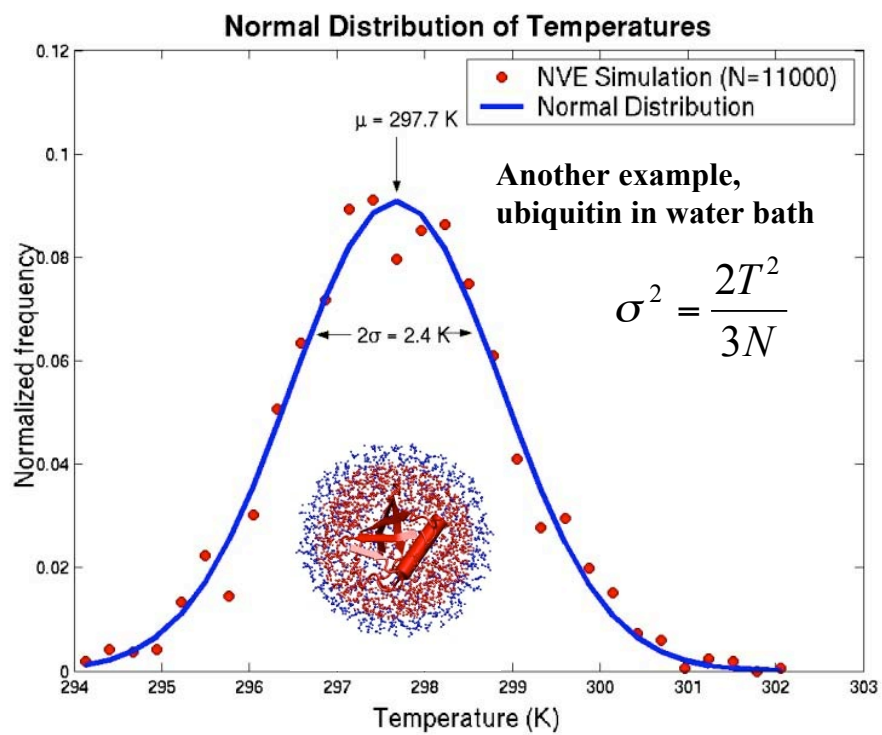
The distribution function for the temperature ($T = 2E_{kin}/3k_B$) fluctuations $\Delta T = T - T_0$ is then

$$P(\Delta T) = c \exp[-(\Delta T)^2/2\sigma^2], \qquad \sigma^2 = 2T^2/3N \qquad (13)$$

For $T_0 = 100$K and $N = 557$, this gives $\sigma = 3.6$.

Normal Distribution of Temperatures

Another example, ubiquitin in water bath

$$\sigma^2 = \frac{2T^2}{3N}$$

Show BPTI trajectory

Dihedral Angle

Langevin dynamics in strong friction limit

$$m_n \frac{d^2 x_n}{dt^2}(t) = \nabla_{x_n} V - m_n b_n \frac{dx_n}{dt}(t) + f_n(t)$$

$$\langle f_n(t) \rangle = 0$$
$$\langle f_n(t) f_n(0) \rangle = 2 k_B T_0 b_n m_n \delta(t).$$

Cα

Brownian oscillator, fully characterized by <angle>, RMSD, C(t)

$$C_A(t) = \frac{\langle A_i(0) A_i(t) \rangle}{\langle A_i^2 \rangle}$$
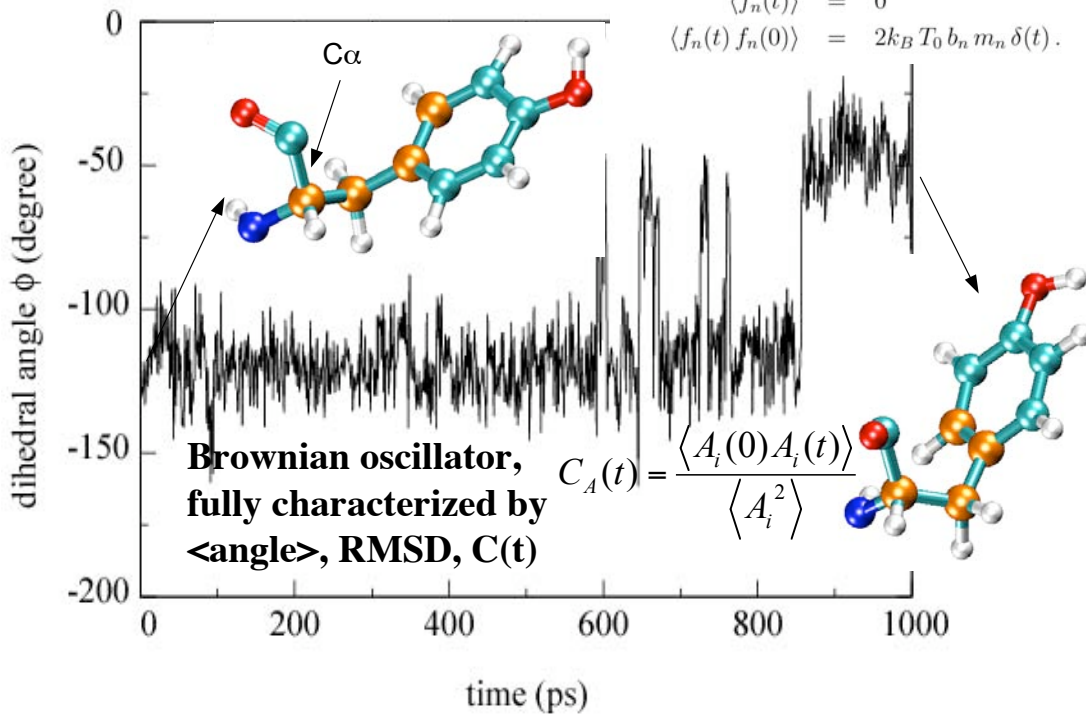
Specific Heat of a Protein

Total energy of ubiquitin (NVE ensemble)

$$c_V = \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_B T^2}$$

TOTAL

$$C_{V(ubiquitin)} = 2.24407 \times 10^{-3} \, kcal/mol \cdot K$$