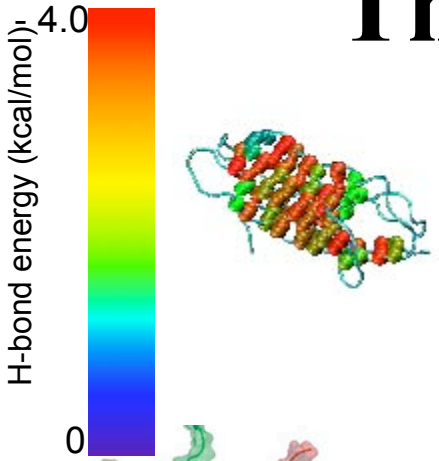
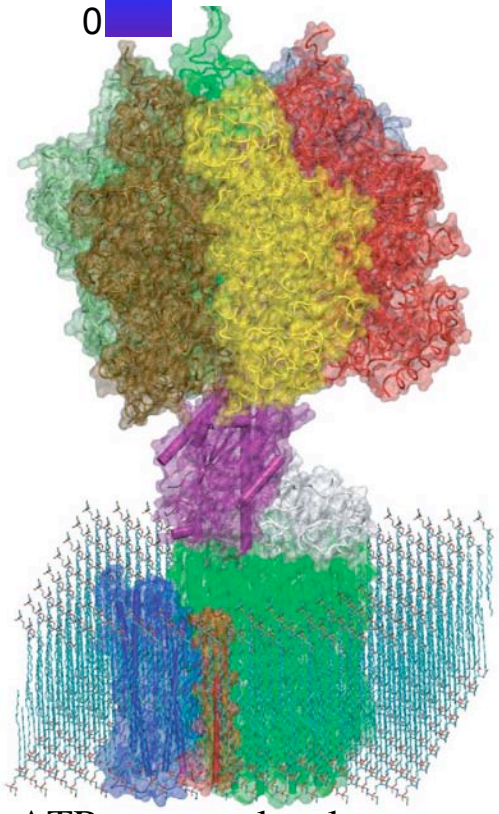
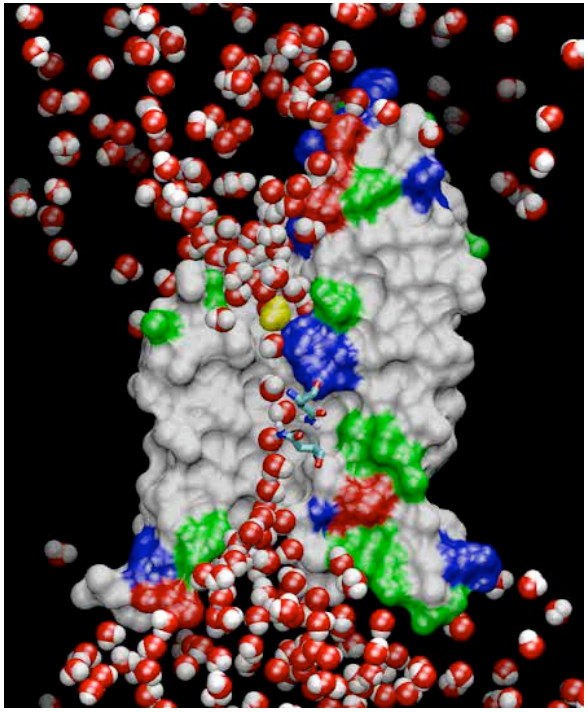
The Molecular Dynamics Method



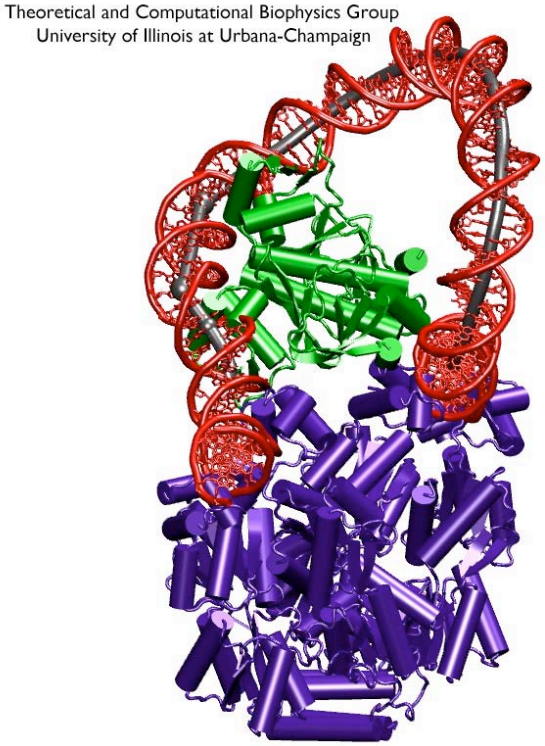
Fibronectin III_1, a mechanical protein that glues cells together in wound healing and in preventing tumor metastasis



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



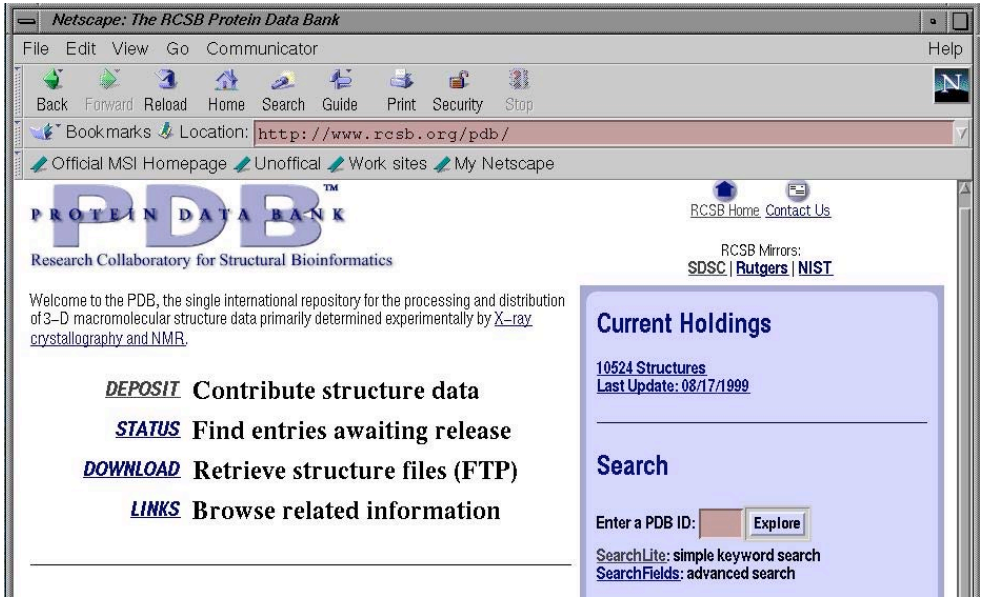
A ternary complex of DNA, lac repressor, and CAP controlling gene expression

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.)



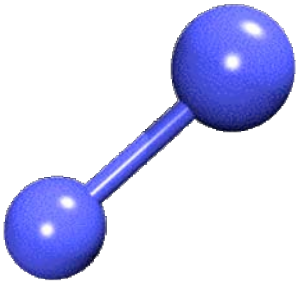
PSF Files

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)

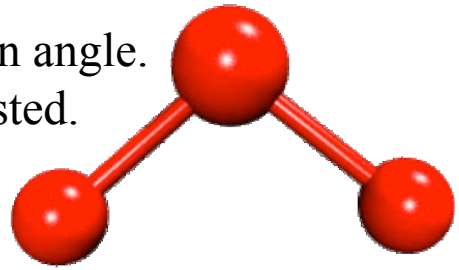
PSF Files

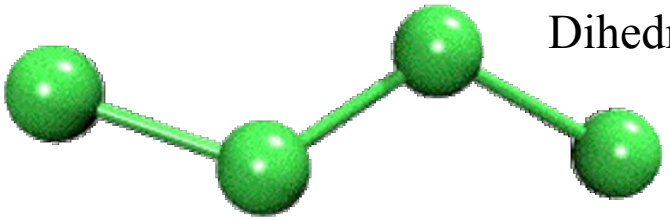
molecular structure (bonds, angles, etc.)



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.





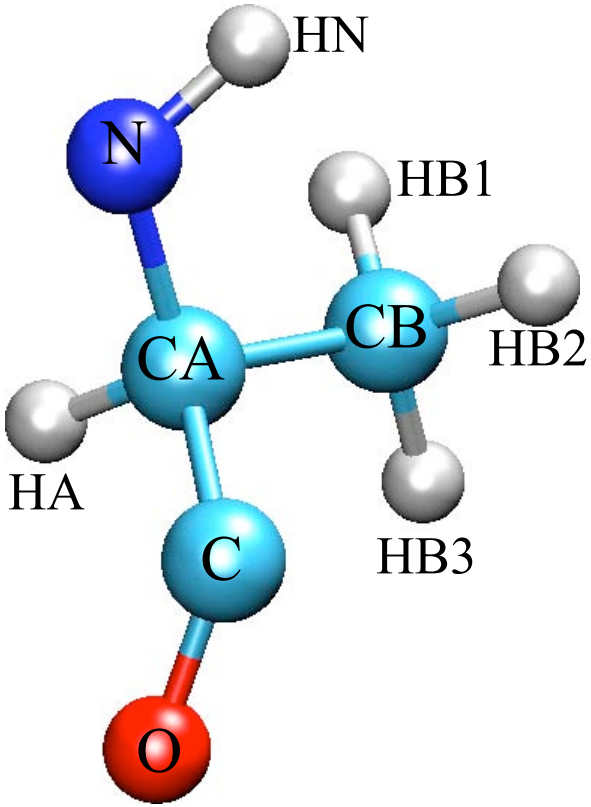
Dihedrals: Two angles that share a common bond form a dihedral.
Every such set of four atoms in the molecule is listed.

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

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



CHARMM Potential Function

form without substance

- Simple, fixed algebraic form for every type of interaction.
- 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 tomorrow afternoon

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

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

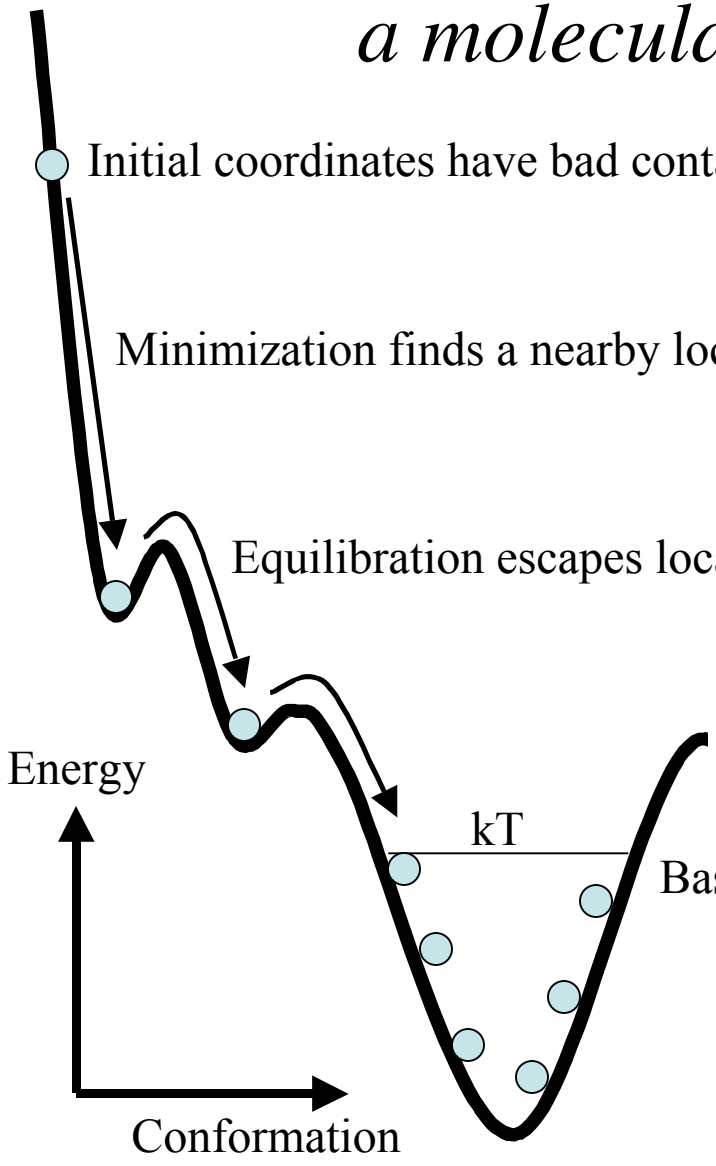
come on, feel the noise

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

$$\dot{\vec{v}} = \vec{F}(\vec{r})/m - \gamma\vec{v} + \vec{F}_{\text{random}}(t)$$

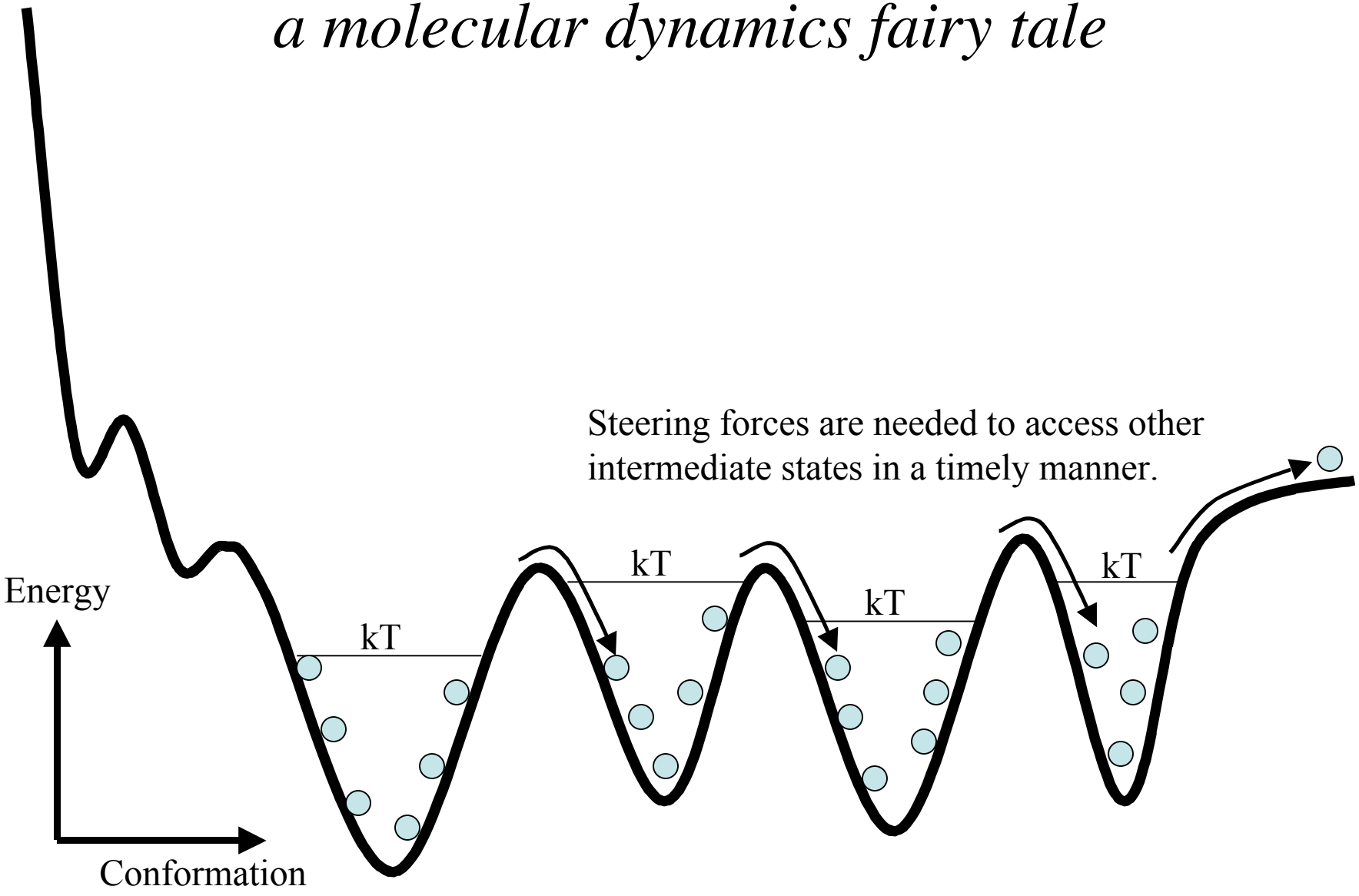
From the Mountains to the Valleys

a molecular dynamics fairy tale



From the Mountains to the Valleys

a molecular dynamics fairy tale



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.

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

+

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

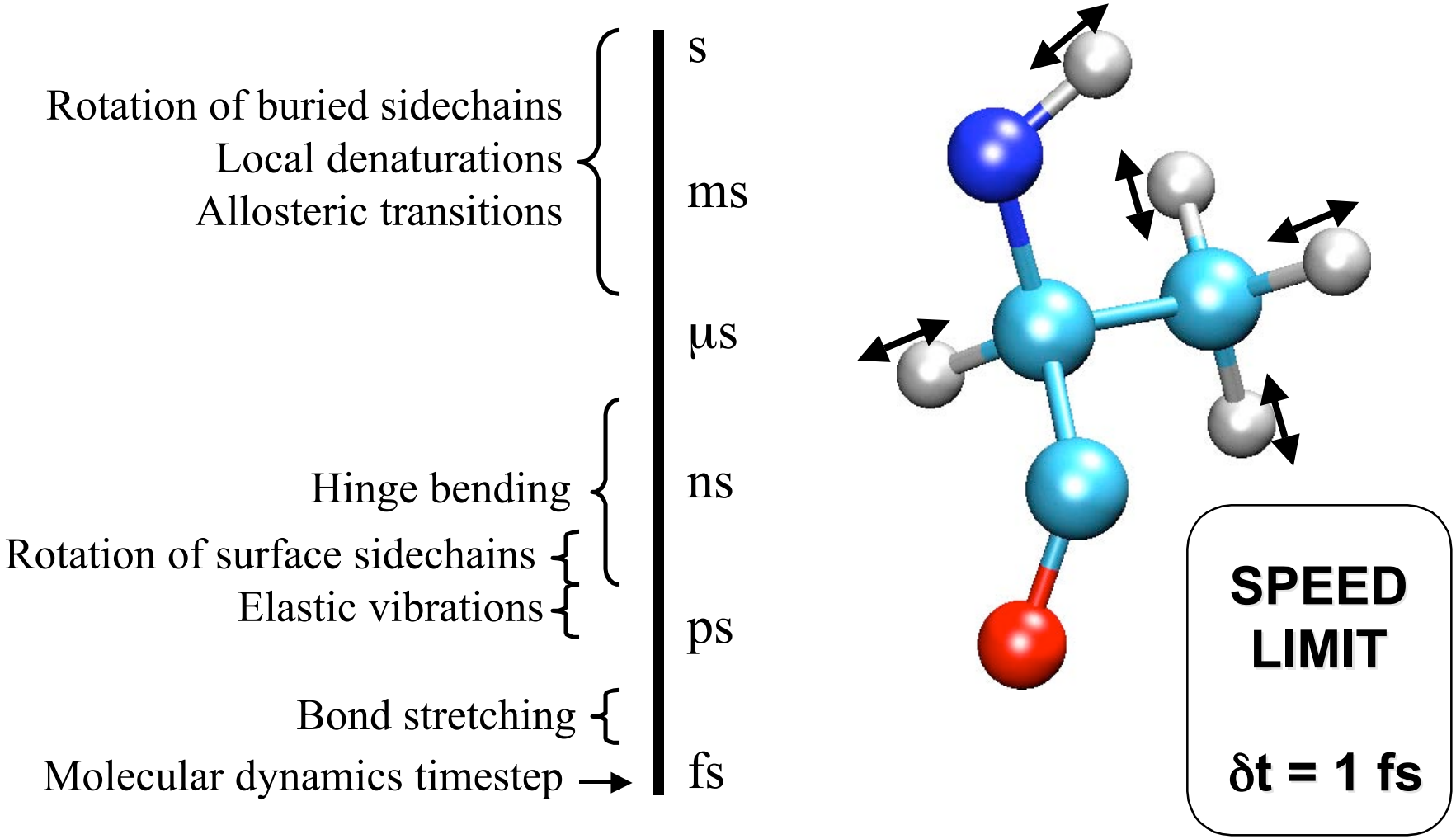
$\cup$

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

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

Hurry Up and Wait

biomolecular timescales and timestep limits



Cutting Corners

cutoffs, PME, rigid bonds, and multiple timesteps

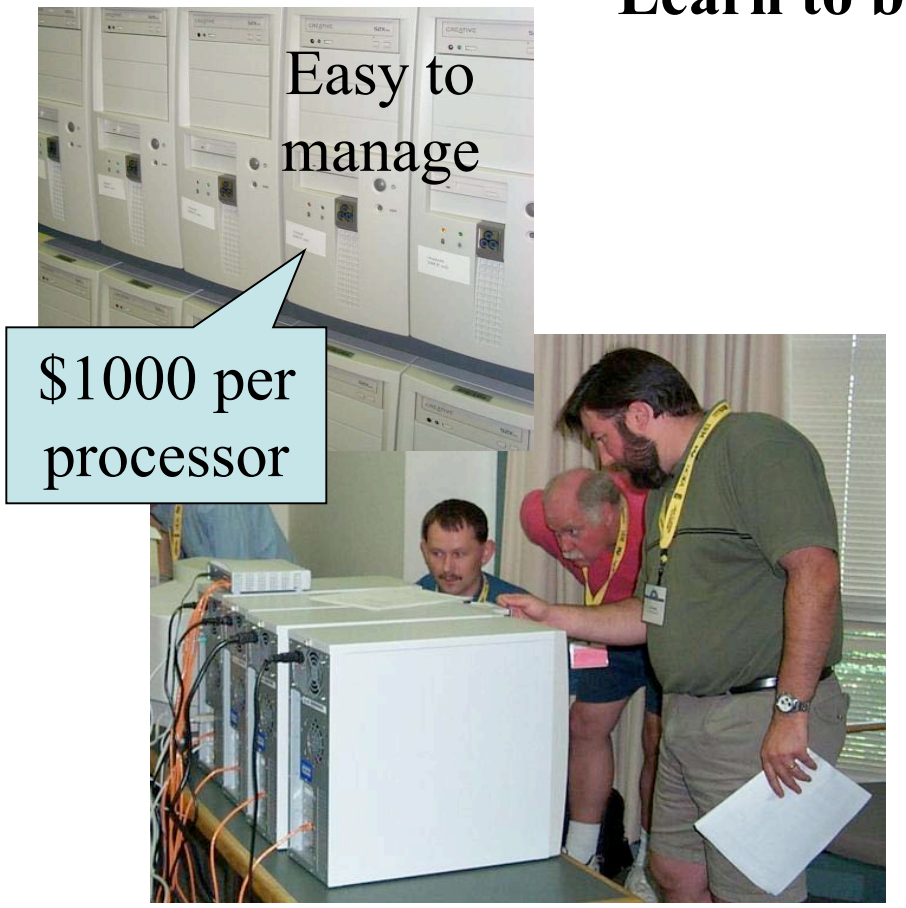
- Nonbonded interactions require order N^2 computer time!
  - Truncating at R_{cutoff} reduces this to order $N\ R_{cutoff}^3$
  - Particle mesh Ewald (PME) method adds long range electrostatics at order N log N, only minor cost compared to cutoff calculation.

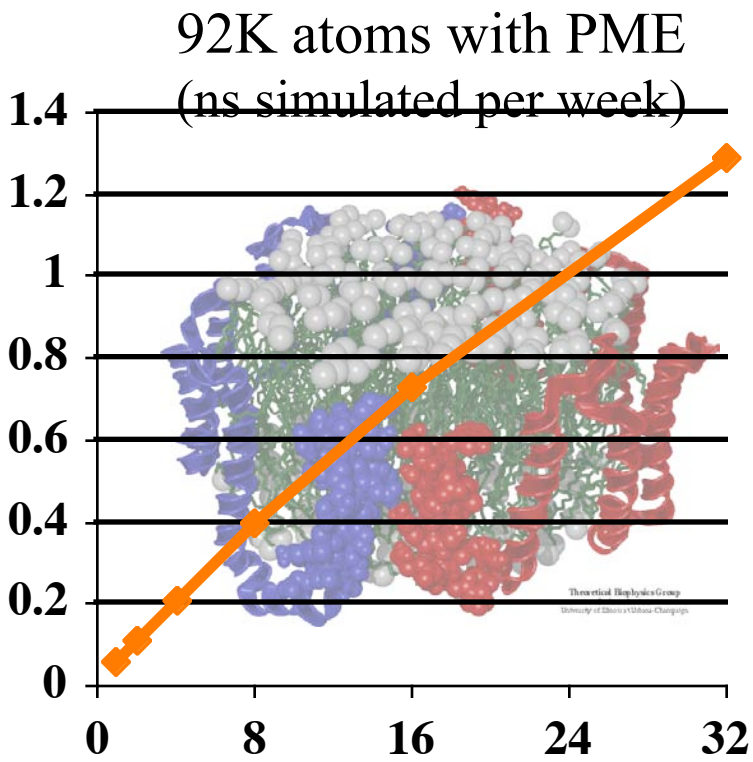
- Can we extend the timestep, and do this work fewer times?
  - Bonds to hydrogen atoms, which require a 1fs timestep, can be held at their equilibrium lengths, allowing 2fs steps.
  - Long range electrostatics forces vary slowly, and may be evaluated less often, such as on every second or third step.

Linux Clusters 101

parallel computing on a professor's salary

Learn to build your own Linux cluster!



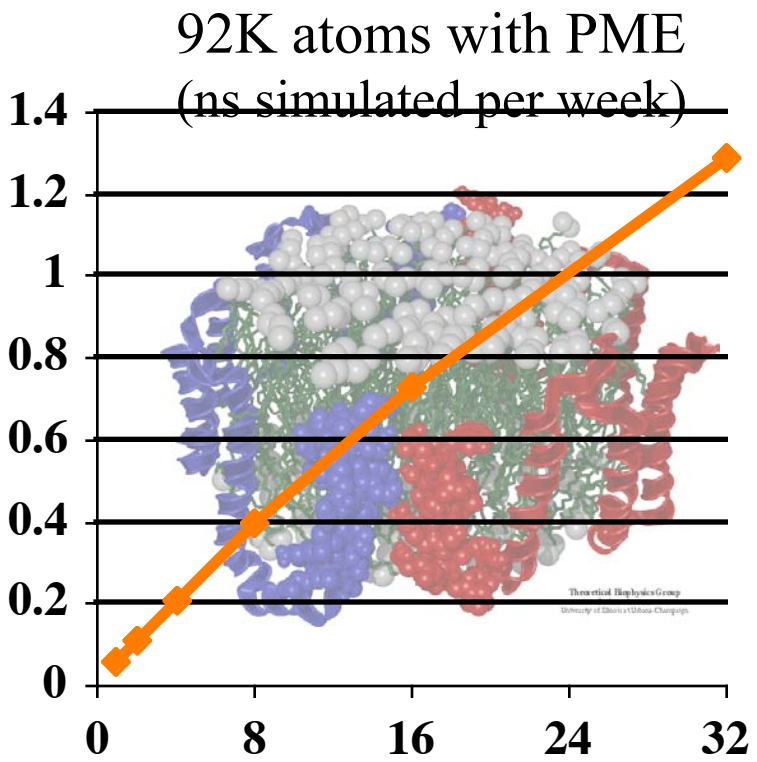


Linux Clusters 101

parallel computing on a professor's salary

Learn to build your own Linux cluster!

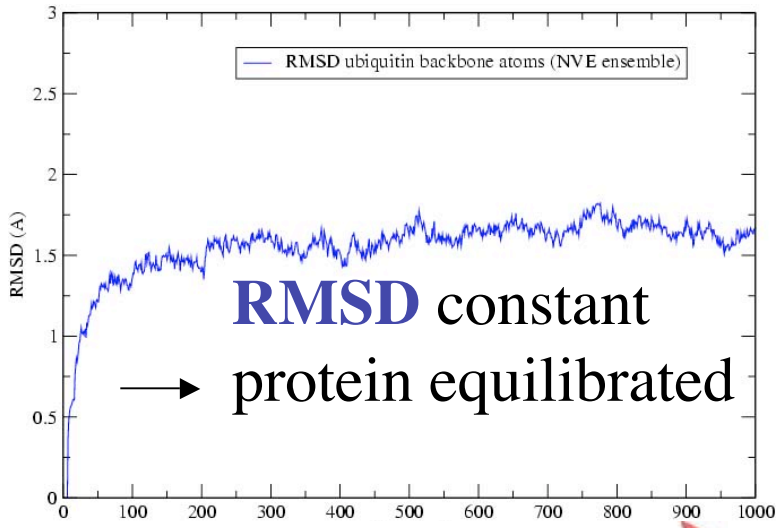
$1000 per processor

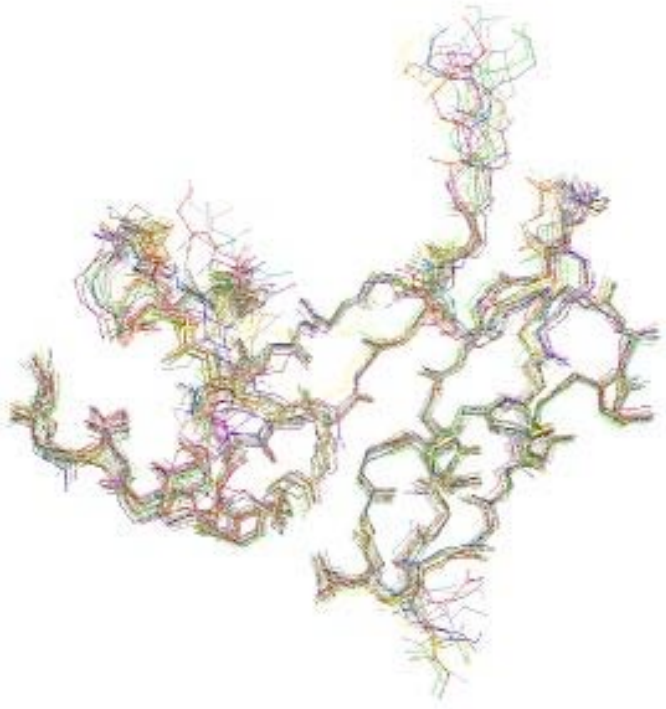


Equilibrium Properties of Proteins

Root Mean Squared Deviation: measure for equilibration and protein flexibility

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

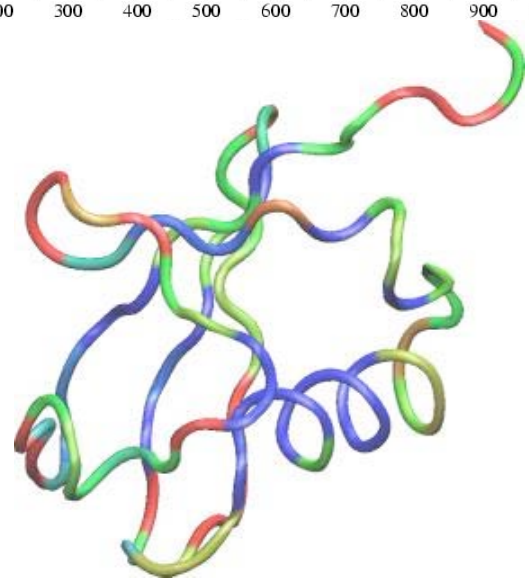




NMR structures

aligned together to see flexibility

Protein sequence exhibits characteristic permanent flexibility!



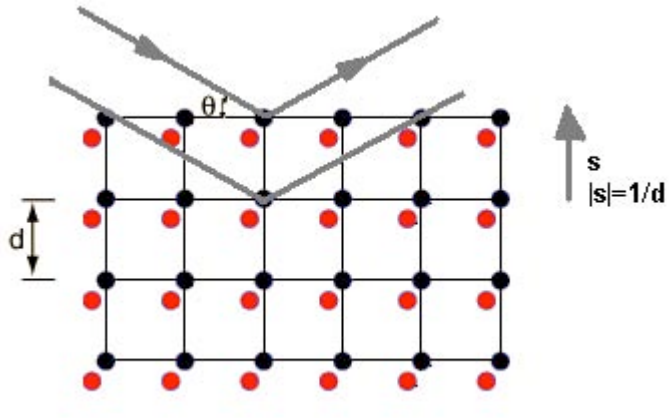
MD simulation

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

Temperature Factor

$2d \sin\theta = \lambda$ Bragg's Law

$F = \Sigma f_j \exp(i2\pi \mathbf{r}_j.\mathbf{s})$
structure factor



More than one atom in the unit cell

$\mathbf{r}_j = \mathbf{x}_j + \mathbf{u}_j$ ($\mathbf{x}_j$:equilibrium position, $\mathbf{u}_j$: vibrations $T > 0$)

$[F]_T = < \Sigma f_j \exp(2\pi i\, \mathbf{s}.(\mathbf{u}_j + \mathbf{x}_j)) >$

averaged over time and unit cells

$I \sim \{ [F]_T \}^2$

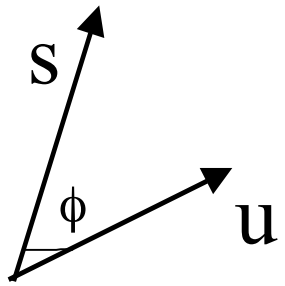
The Atomic B Values in Protein Crystallography

$$< \exp (2\pi i \, \mathbf{s}.\mathbf{u}_j)) > \approx 1 + 2\pi i \, \mathbf{s}. <\mathbf{u}_j> - 2\pi^2 <(\mathbf{s}.\mathbf{u}_j)^2> + \ldots.$$

For small vibrations and symmetric potentials (like harmonic oscillator) $<u>=0$

$$[F]_T = \Sigma \, f_j \exp (i2\pi \, \mathbf{x}_j.\mathbf{s}) \exp (-2\pi^2 <(\mathbf{u}_j.\mathbf{s})^2>)$$

$$<(\mathbf{u}_j.\mathbf{s})^2 > = \mathbf{s}^2 <\mathbf{u}_{j\perp}{}^2 > = (2 \, \mathrm{Sin} \, \theta / \lambda)^2 <\mathbf{u}_{j\perp}{}^2 >$$



$$[F]_T = \Sigma \, f_j \exp (i2\pi \, \mathbf{x}_j.\mathbf{s}) \exp (-B_j{}^2 \, \mathrm{Sin}^2 \, \theta / \lambda^2)$$

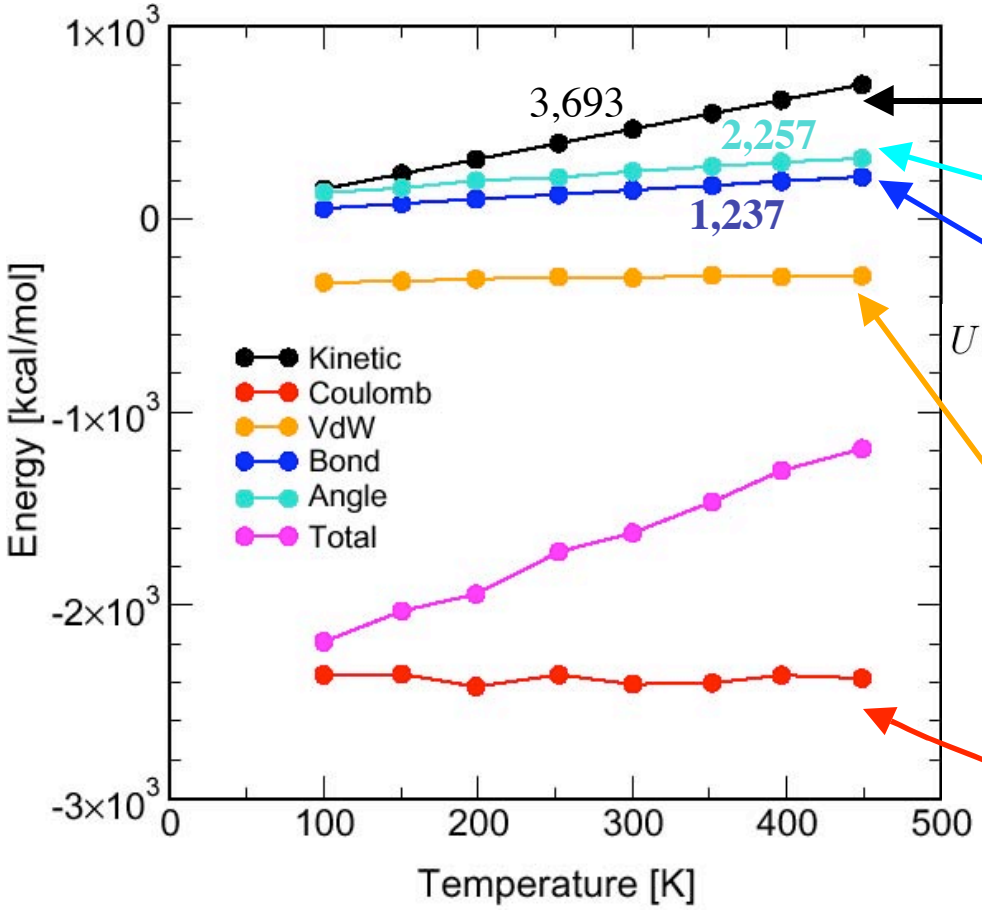
$$B_j = -8 \, \pi^2 < \mathbf{u}_j{}^2 \, \mathrm{Cos}^2 \, \phi > \quad \text{Temperature Factor}$$

Isotropic harmonic potential m, ω:

$$B_j = -8/3 \, \pi^2 < \mathbf{u}^2 > = -16/3 \, \pi^2 \, k_B T / m \, \omega^2$$

Equilibrium Properties of Proteins

Energies: kinetic and potential



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

Kinetic energy (quadratic)

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

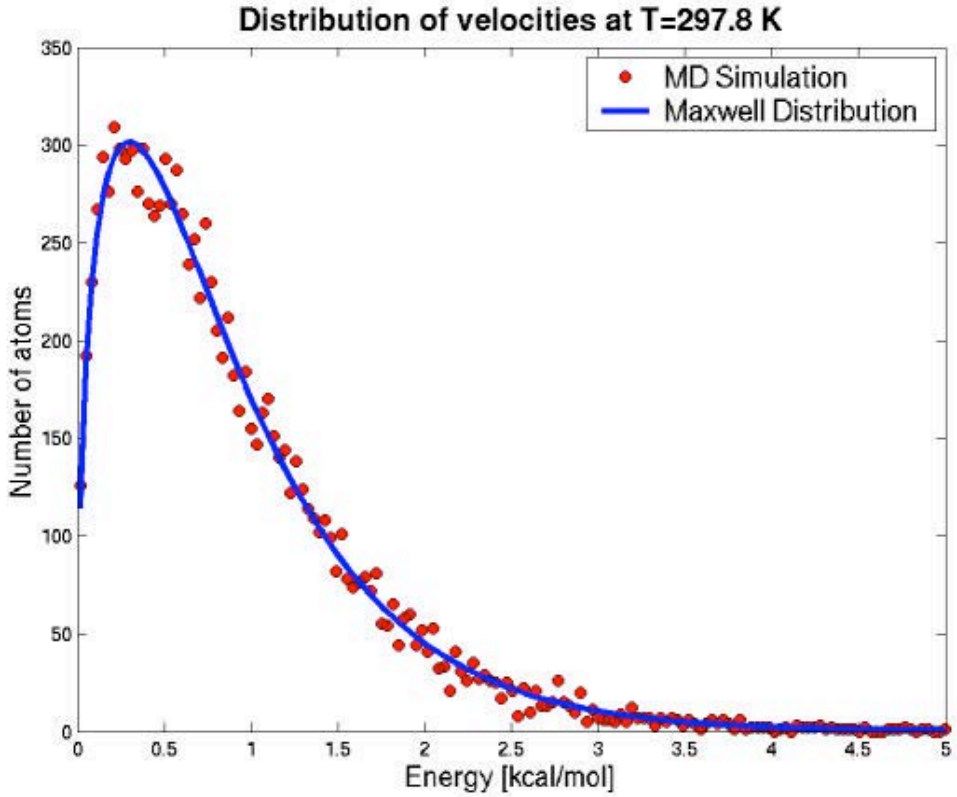
$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i \phi_i + \delta_i)]}_{U_{dihedral}} +$$

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

Potentialc energy (not all quadratic)

Maxwell-Boltzmann distribution

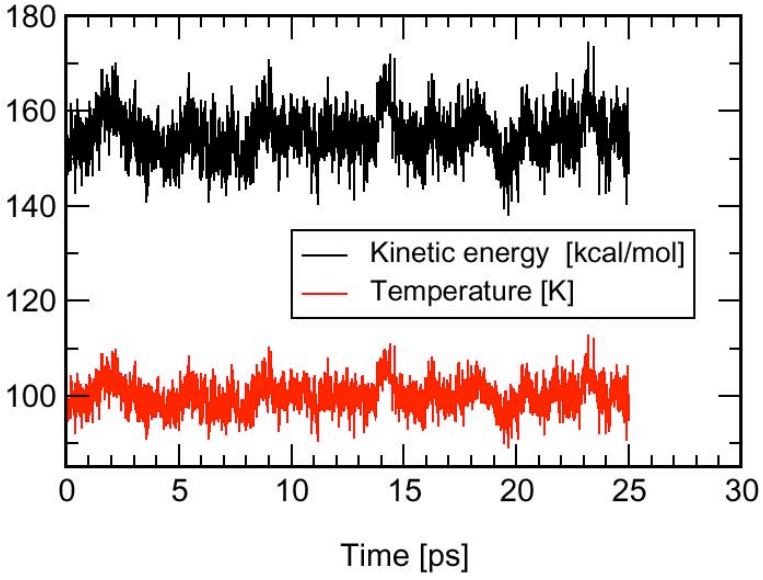
An NVT (temperature held constant) ensemble simulation at T=300K was terminated and continued as an NVE (energy constant) ensemble simulation. After an equilibration phase, the distribution of velocities from all atoms was determined (red) and fitted to the Maxwell velocity distribution (blue); the best fit corresponded to a temperature T=297.8 K.

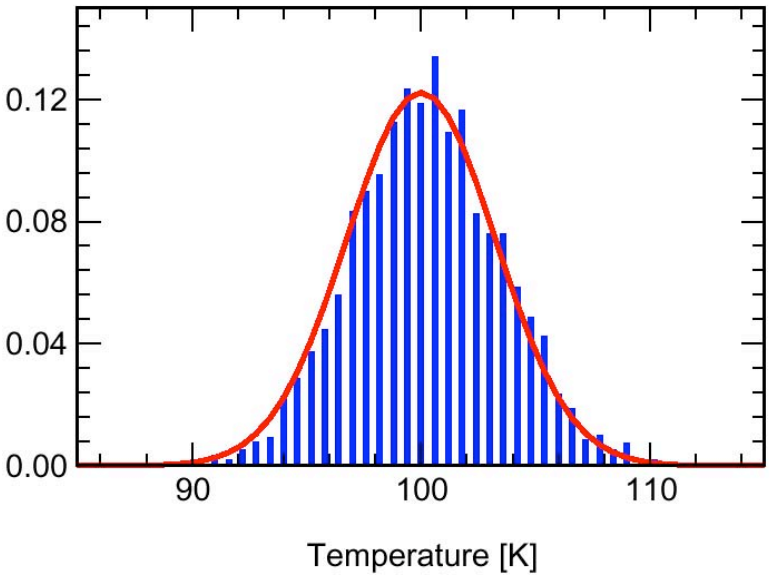


Maxwell distribution for kinetic energies

$$f(\epsilon_k) = \frac{2}{\sqrt{\pi}} \frac{1}{(k_B T)^{\frac{3}{2}}} \sqrt{\epsilon_k} \exp\left(-\frac{\epsilon_k}{k_B T}\right)$$

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





Definition of Temperature

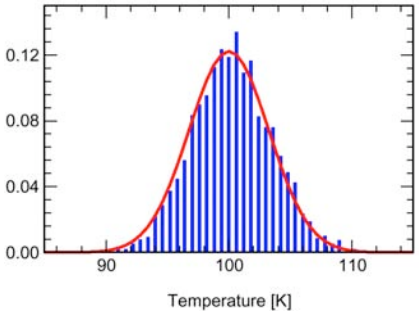
$$\langle \sum_j \frac{1}{2} m_j v_j^2 \rangle = \frac{3}{2} N k_B T$$

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

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

Temperatur Fluctuations

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



Maxwell distribution

$$dP(v_n) = c\,\exp(-m\,v_n^2/2k_BT)\,dv_n \tag{7}$$

Individual kinetic energy $\epsilon_n = mv_n^2/2$

$$dP(\epsilon_n) = (\pi T_0\epsilon_n)^{-1/2}\exp(-\epsilon_n/k_BT_0)\,d\epsilon_n \tag{8}$$

One can derive

$$\langle\epsilon_n\rangle = T_0/2 \tag{9}$$

$$\langle\epsilon_n^2\rangle = 3\,T_0^2/4 \tag{10}$$

$$\langle\epsilon_n^2\rangle - \langle\epsilon_n\rangle^2 = T_0^2/2 \tag{11}$$

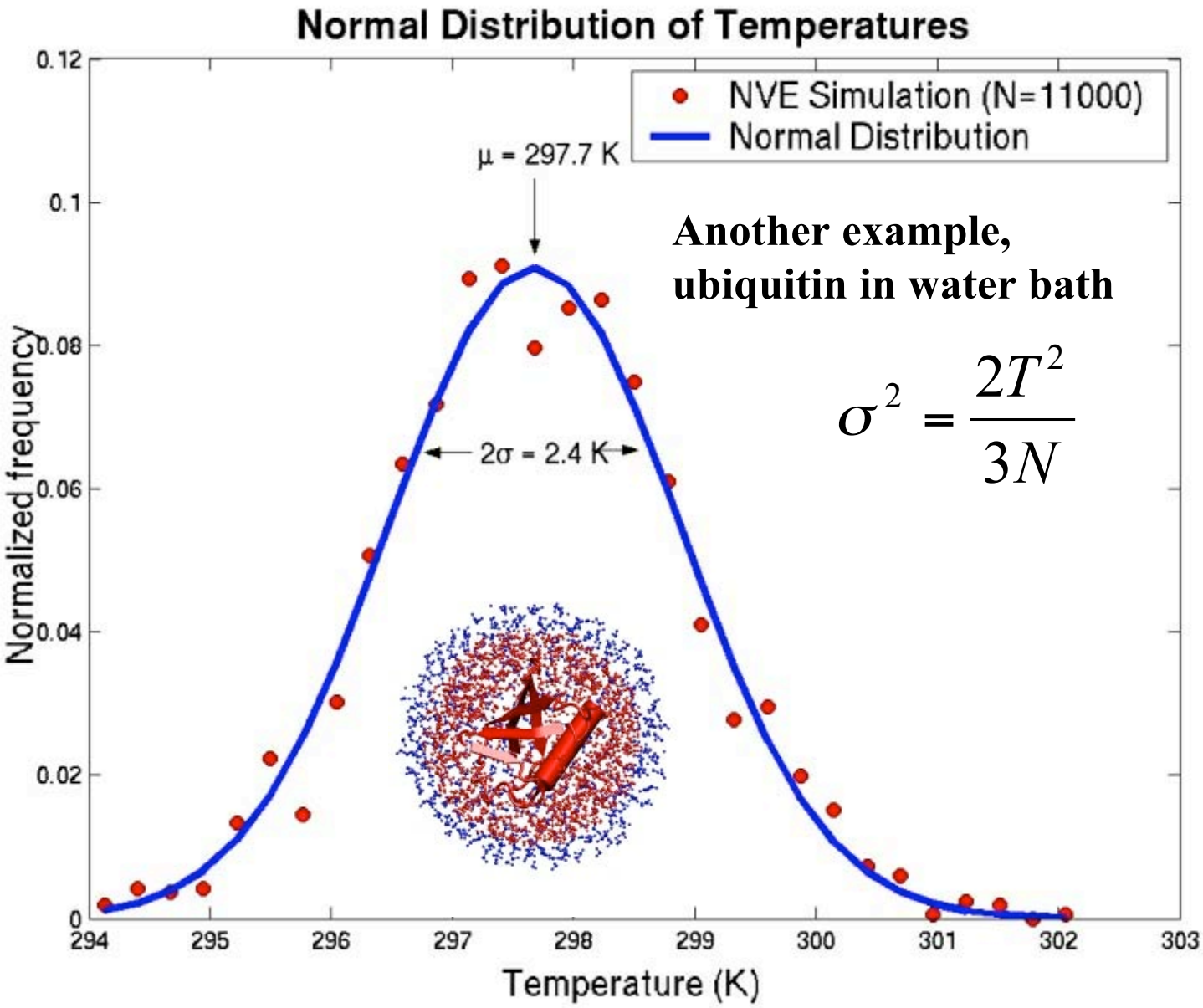
The distribution of the total kinetic energy $E_{kin} = \sum_j \frac{1}{2}m_jv_j^2$, according to the central limit theorem, is approximately Gaussian

$$P(E_{kin}) = c\,\exp\left(\frac{-(E_{kin} - \langle E_{kin}\rangle)^2}{2\left(\frac{3Nk_B^2T_0^2}{2}\right)}\right) \tag{12}$$

The distribution function for the temperature ($T = 2E_{kin}/3k_B$) fluctuations $\Delta T = T - T_0$ is then

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

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



Another example, ubiquitin in water bath

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

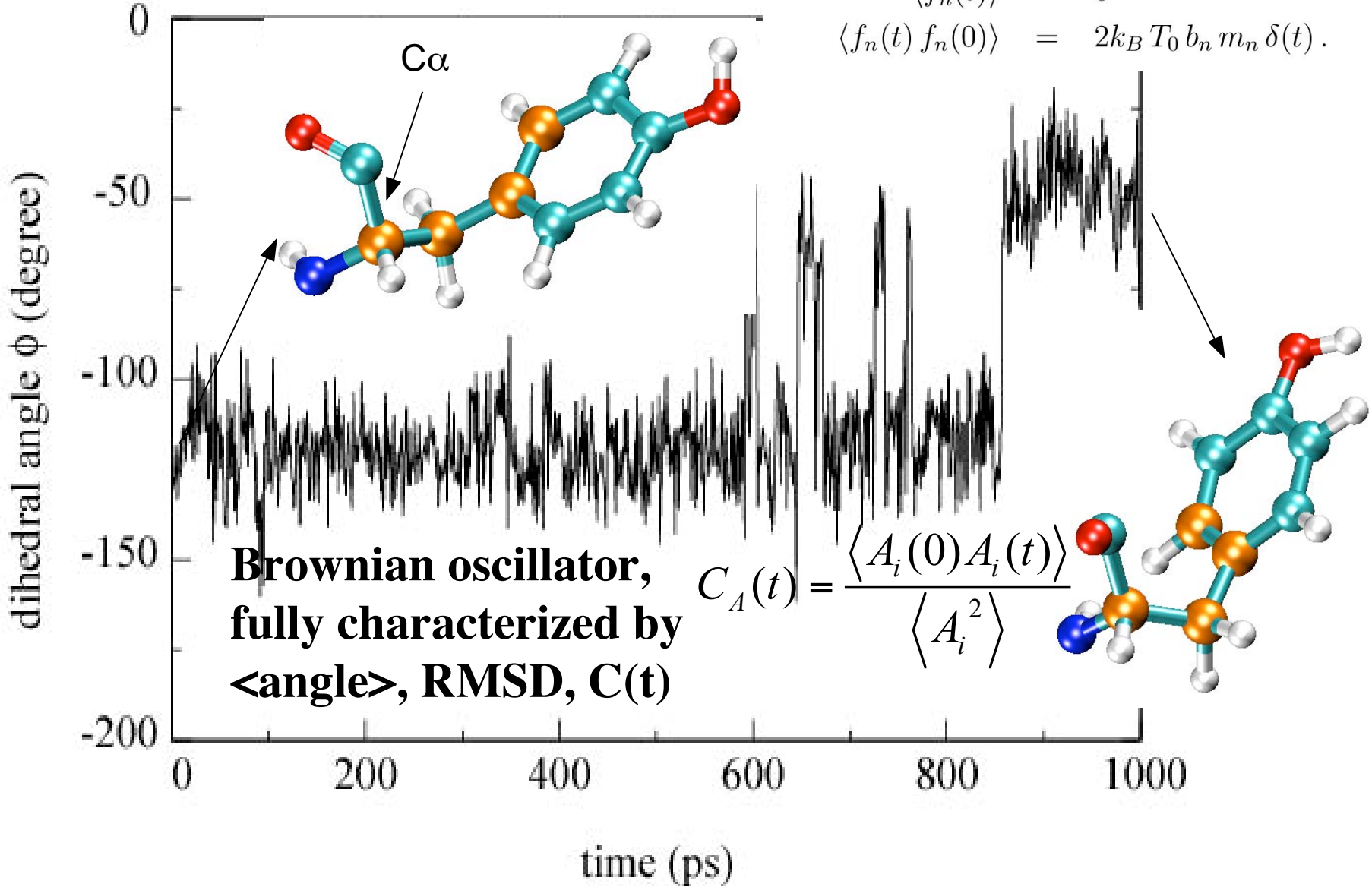
Show BPTI trajectory

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

Dihedral Angle

Langevin dynamics in strong friction limit

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



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