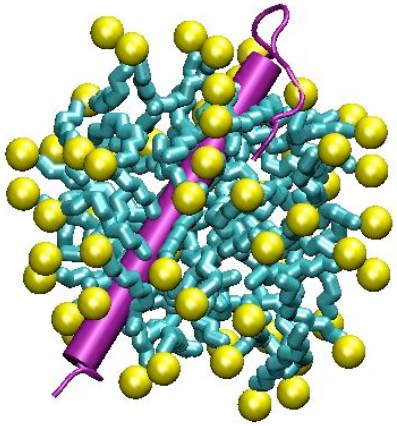
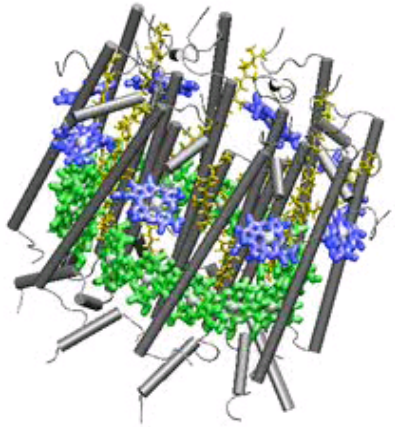
What can be learnt from MD

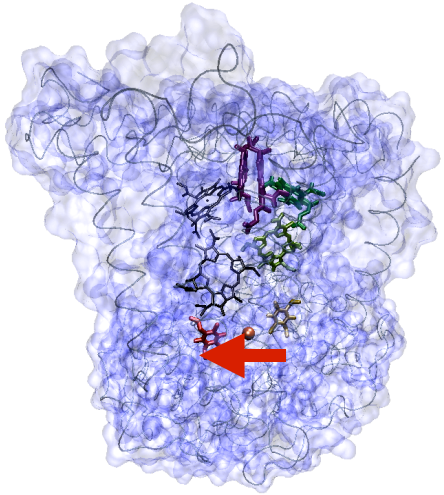
Titin

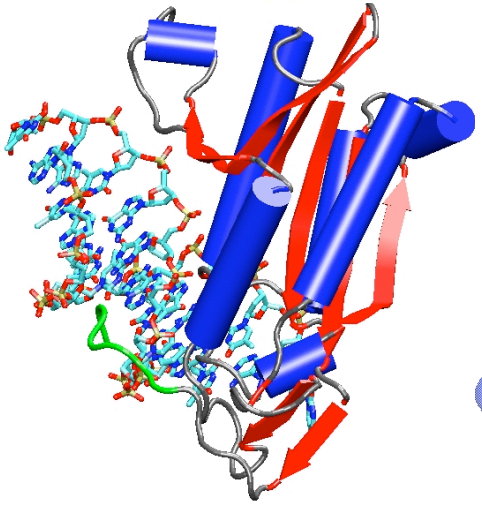
Micelle Light harvesting complex Reaction Center

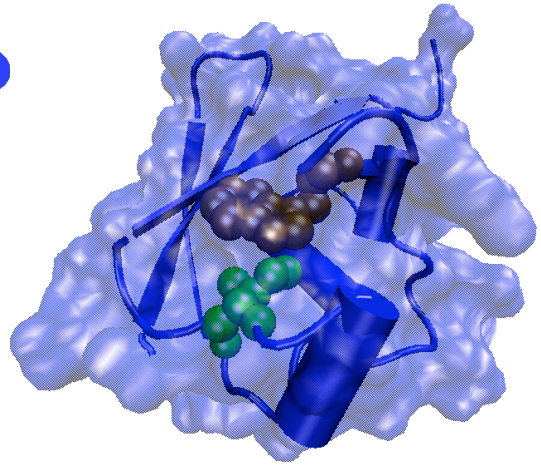
Restriction enzyme Ubiquitin Biotin - Avidin

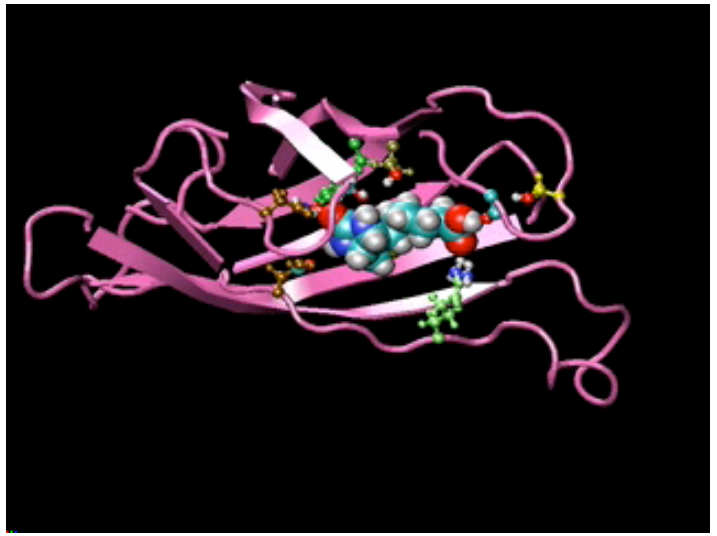










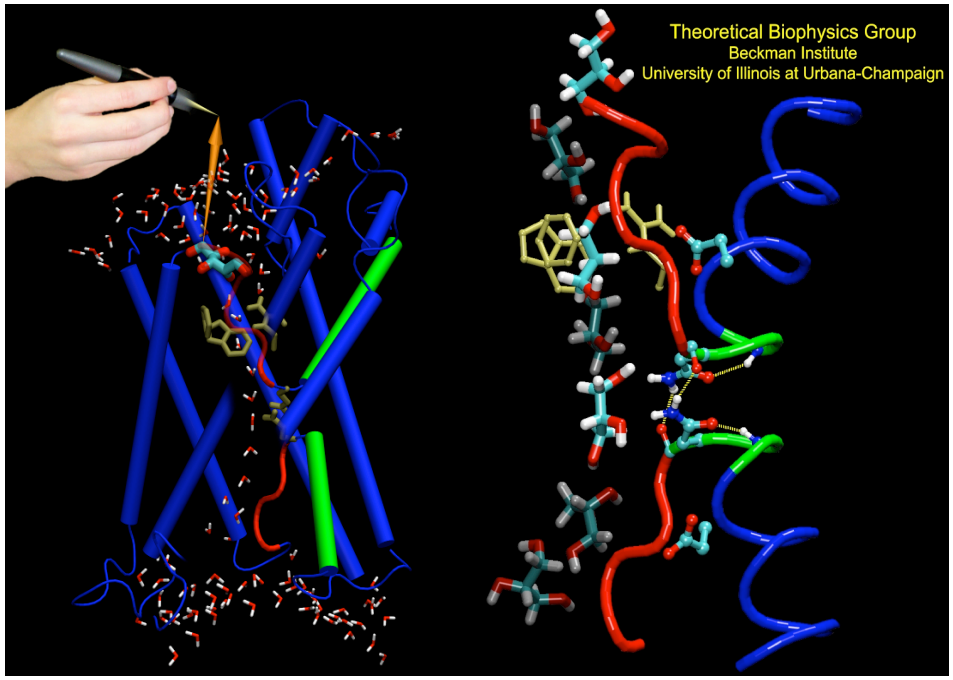


Putting Your Hands on a Protein with Interactive Molecular Dynamics

- Interactive molecular dynamics is a grid-based computational methodology that promises to revolutionize molecular biomedicine.

"If I could just get my hands on that protein!" Single molecule manipulation techniques like atomic force microscopy have brought researchers closer to this frequently expressed wish, but these techniques do not "see" the atomic level detail needed to relate mechanism to protein architectures. Only computational modeling can do this.

Aquaporins are channel proteins abundantly present in all life forms, for example, bacteria, plants, and in the kidneys, the eyes, and the brain of humans. These proteins conduct water and small molecules, but no ions, across the cell walls. Their defective forms are known to cause diseases, *e.g.*, diabetes insipidus, or cataracts. Crystallographers from the University of California at San Francisco (R. Stroud and coworkers) that discovered the structure of one type of aquaporins, aquaglyceroporins, have teamed up with the UIUC researchers to determine how these channels achieve their very high water throughput in humans as well as capture and transport key food molecules into the cells of bacteria. The team, combining 106,000 atom simulations using NAMD on NCSA and PSC machines and crystallography, explained how the cproteins achieve their impressive filtering function. In a paper just accepted for publication in Biophysics Journal, they report the first use of **interactive molecular dynamics** to study filtering by aquaporins. Linking advanced molecular graphics with ongoing molecular dynamics simulations on NCSA's machines, and utilizing a haptic device to connect forces from a user's hand with forces in the simulation, researchers can interact with "live" proteins.



This figure shows on the right snapshots of the passage of glycerol through the aquaporin channel and on the left a researcher pulling glycerol "by hand" through the channel with a haptic device.

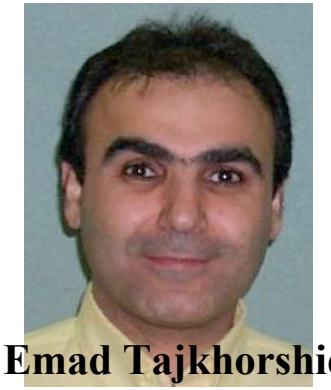
John Stone, Justin Gullingsrud, Paul Grayson, and Klaus Schulten. **A system for interactive molecular dynamics simulation.** In John F. Hughes and Carlo H. Séquin, editors, *2001 ACM Symposium on Interactive 3D Graphics*, pages 191-194, New York, 2001. ACM SIGGRAPH.

Paul Grayson, Emad Tajkhorshid, and Klaus Schulten. **Mechanisms of selectivity in channels and enzymes studied with interactive molecular dynamics.** *Biophysical Journal*, 2003. In press.

Spectral Tuning in Archaeal Retinal Proteins

Hayashi et al. Structural determinants of spectral tuning in retinal proteins – bacteriorhodopsin *vs* sensory rhodopsin II. *J. Phys. Chem. B*, 105:10124–10131, 2001.

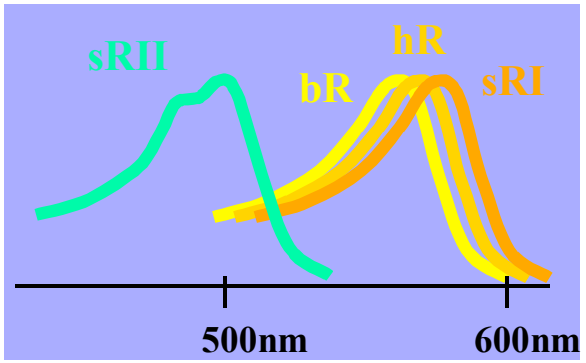
Shigehiko Hayashi

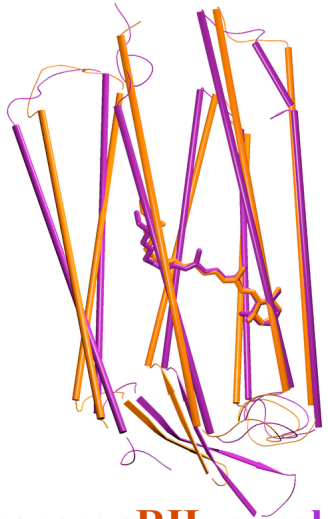


Emad Tajkhorshid

Sensory rhodopsin I (sRI):
attractant (repellent) to orange (near UV) light
Sensory rhodopsin II (sRII):
repellent to blue-green light

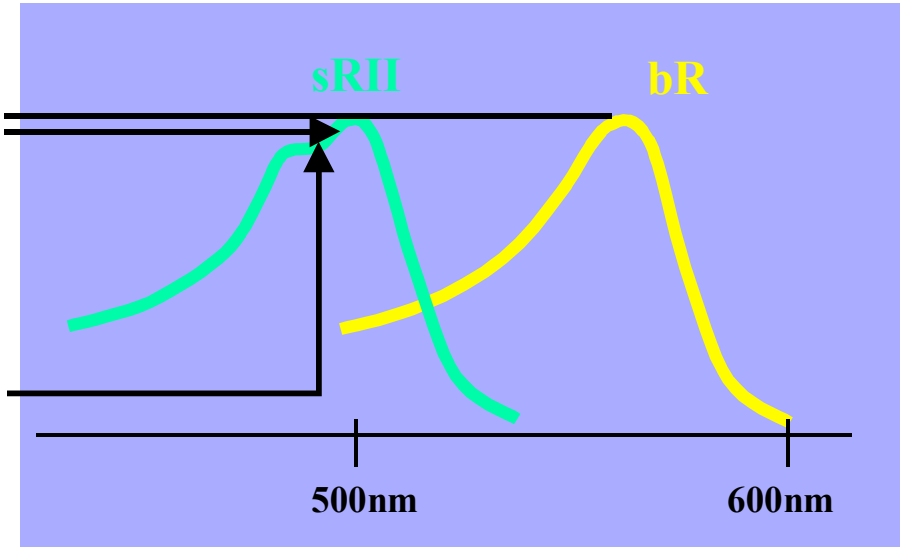
phototaxis





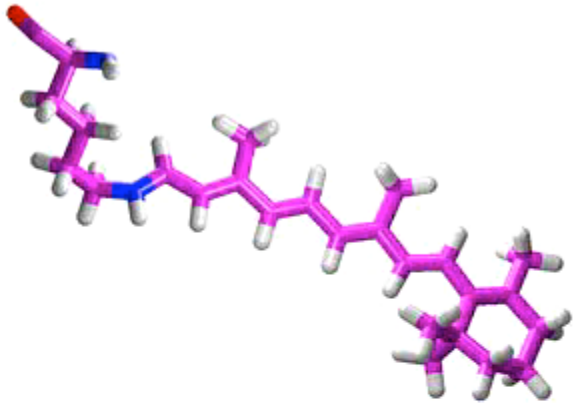
orange: sRII, purple: bR

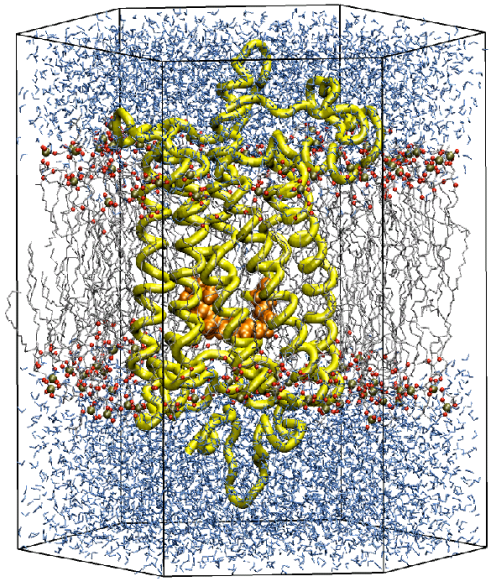
- **Absorption maximum of sRII (vs. bR) is blue-shifted (70 nm) despite close homology**
- **sRII spectrum exhibits a prominent sub-band.**

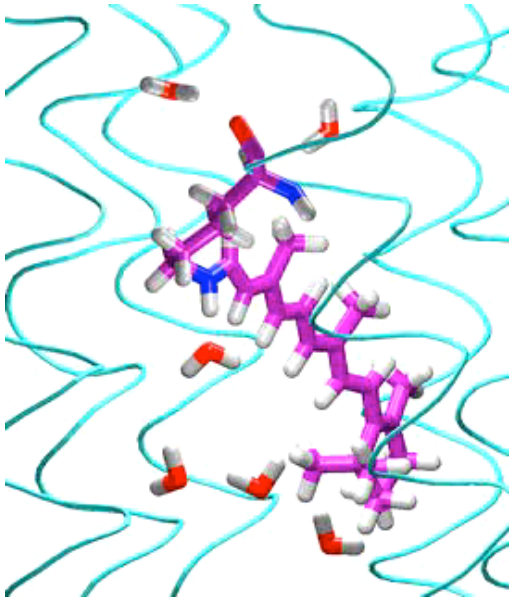


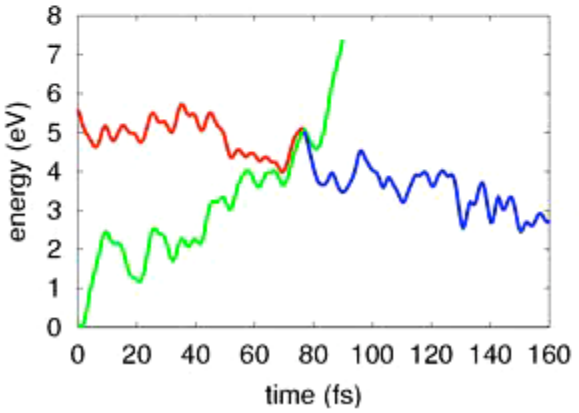
On-the-fly Ab Initio QM/MM Excited State MD Simulation

- Multielectronic state (S_0 and S_1) dynamics of photoisomerization of retinal in bR
- An analogue of retinal (three double bonds, 20 atoms)
- State-average CASSCF(6,6)/DZV/AMBER94
- 11 trajectories starting from initial configurations generated by classical MD.





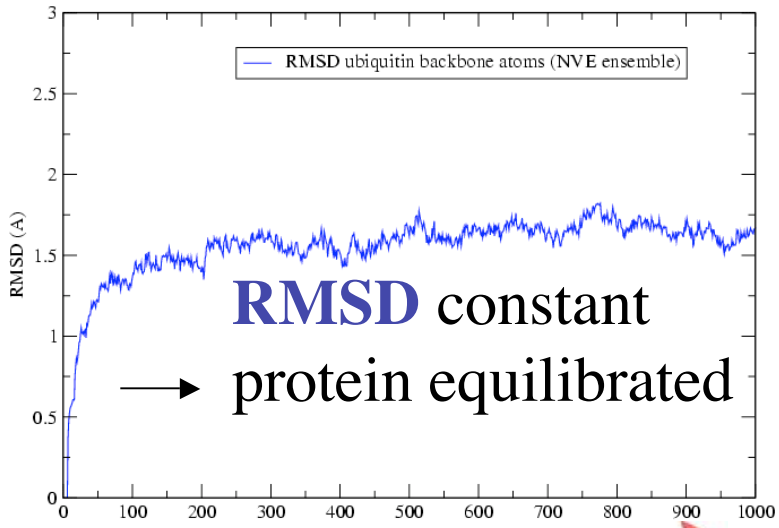


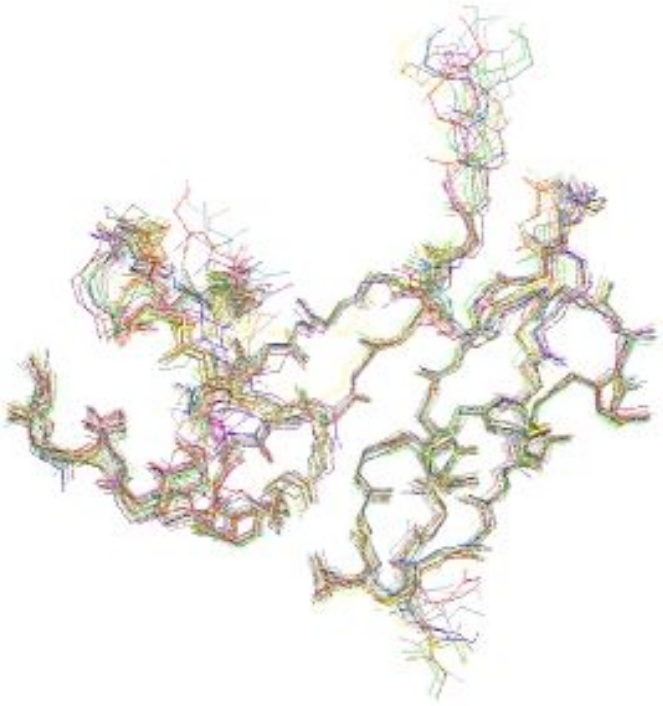


Equilibrium Properties of Proteins

Root Mean Squared Deviation: measure for equilibration and protein flexibility

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

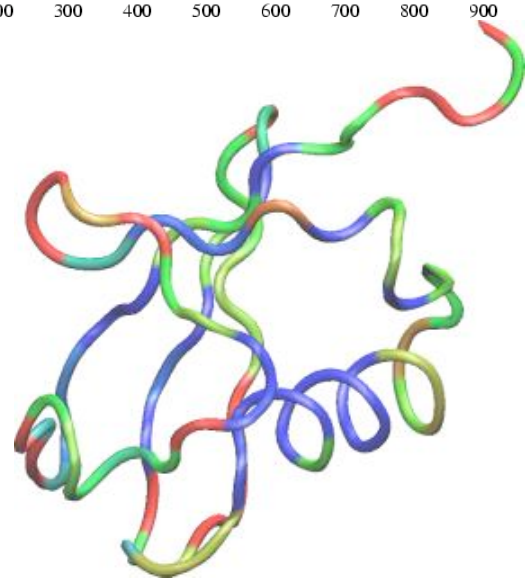




NMR structures

aligned together to see flexibility

Protein sequence exhibits characteristic permanent flexibility!



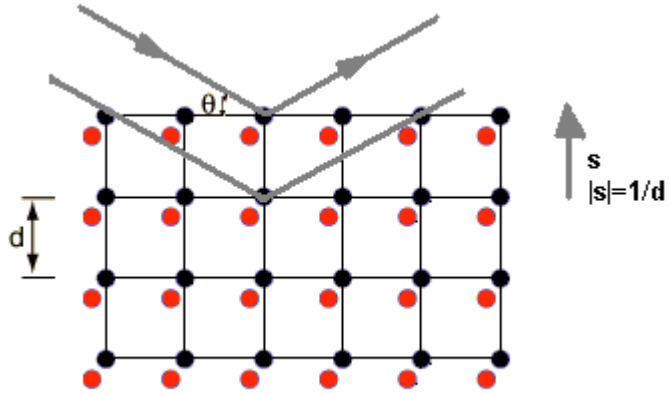
MD simulation

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

Temperature Factor

2d Sin □ = □ Bragg's Law

F=□ f_j exp(i2π $\mathbf{r_j}$.**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)

$[F]_T = <$ □ f_j exp (2πi **s.**($\mathbf{u_j} + \mathbf{x_j}$)) >

averaged over time and unit cells

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

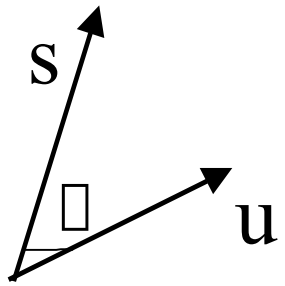
The Atomic B Values in Protein Crystallography

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

For small vibrations and symmetric potentials (like harmonic oscillator) $\langle u \rangle = 0$

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

$$\langle (\mathbf{u}_j.\mathbf{s})^2 \rangle = \mathbf{s}^2 \langle \mathbf{u}_{j\perp}^2 \rangle = (2 \sin \theta / \lambda)^2 \langle \mathbf{u}_{j\perp}^2 \rangle$$



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

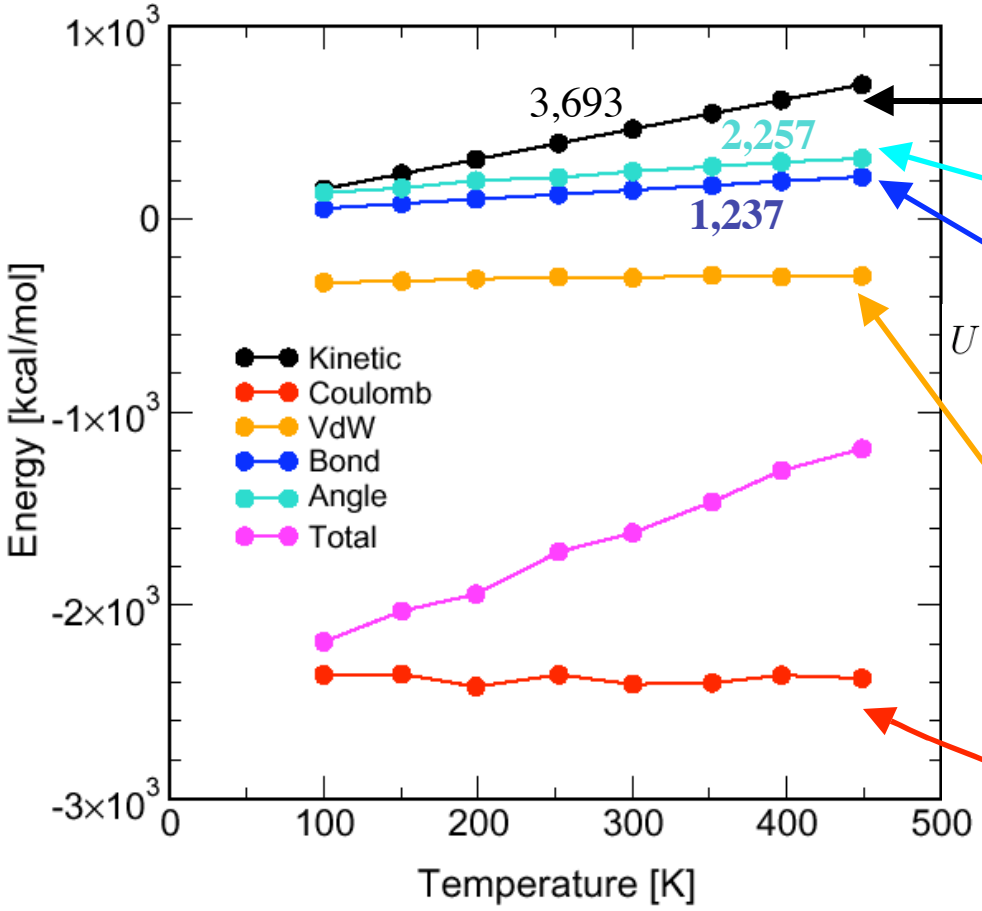
$B_j = -8\pi^2 < \mathbf{u}_j^2 \cos^2 \theta >$ Temperature Factor

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)

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

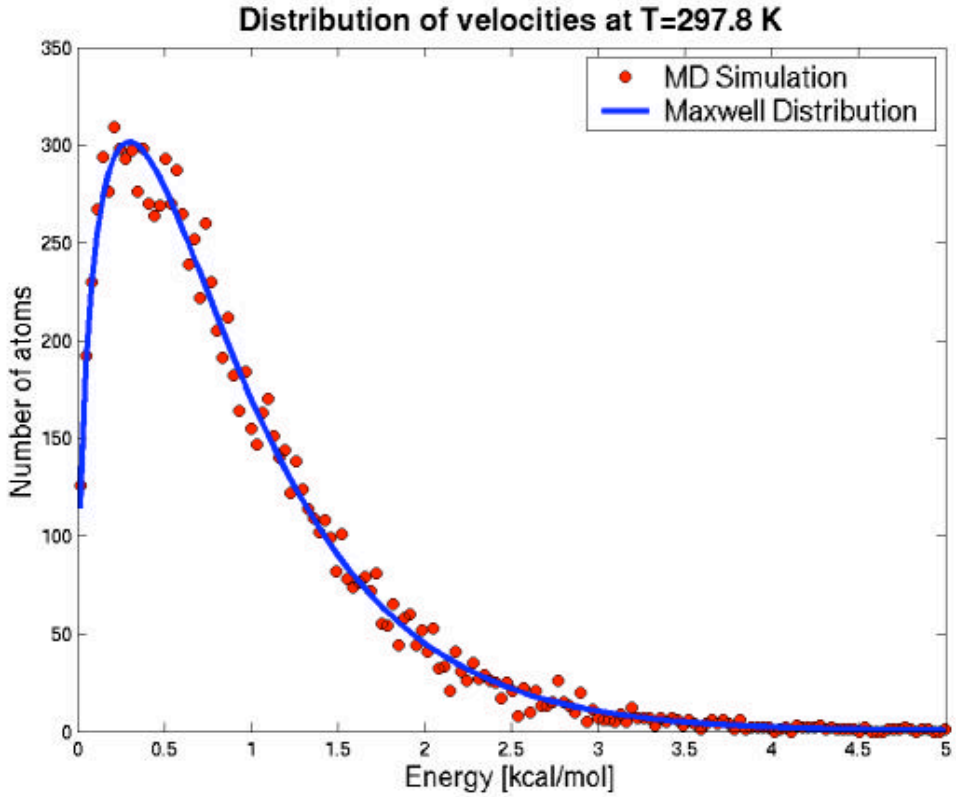
$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i \phi_i + \delta_i)]}_{U_{dihedral}} +$$

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

Potentialc energy (not all quadratic)

Maxwell-Boltzmann distribution

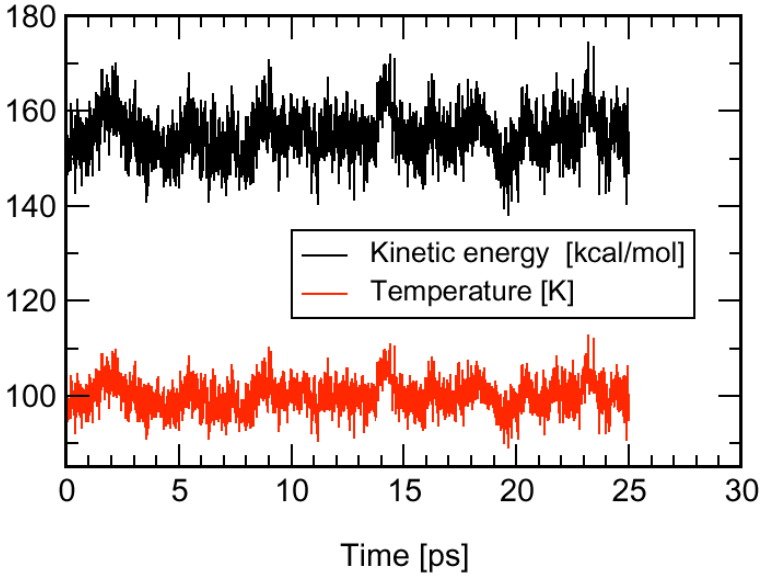
An NVT ensemble simulation at T=300K was terminated and continued as an NVE ensemble simulation. After an equilibration phase, the distribution of velocities from all atoms was determined (red) and fitted to the Maxwell velocity distribution (blue); the best fit corresponded to a temperature T=297.8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

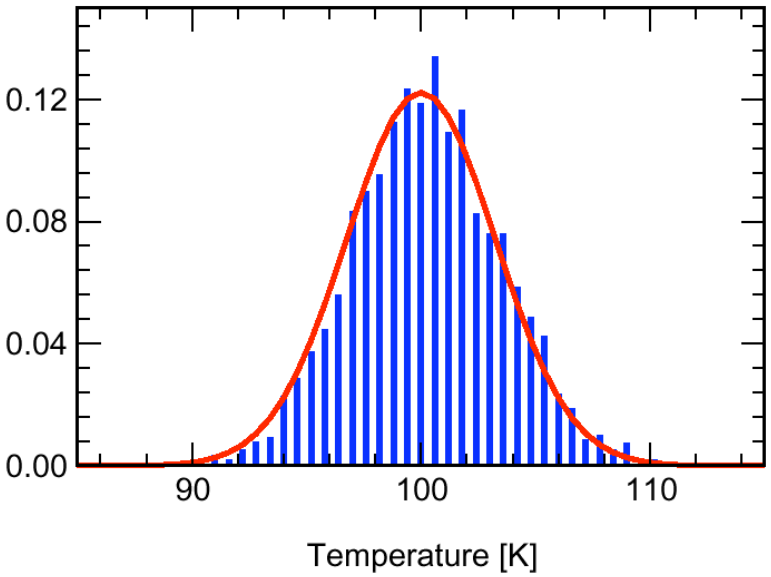


Maxwell distribution for kinetic energies

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

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





Definition of Temperature

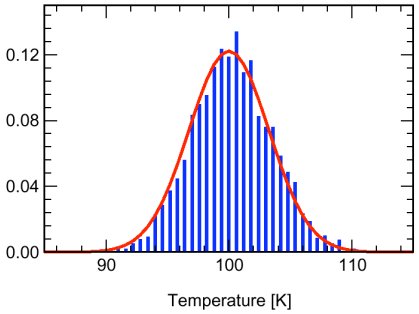
$$\langle \sum_j \frac{1}{2} m_j v_j^2 \rangle = \frac{3}{2} N k_B T$$

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

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

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

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

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

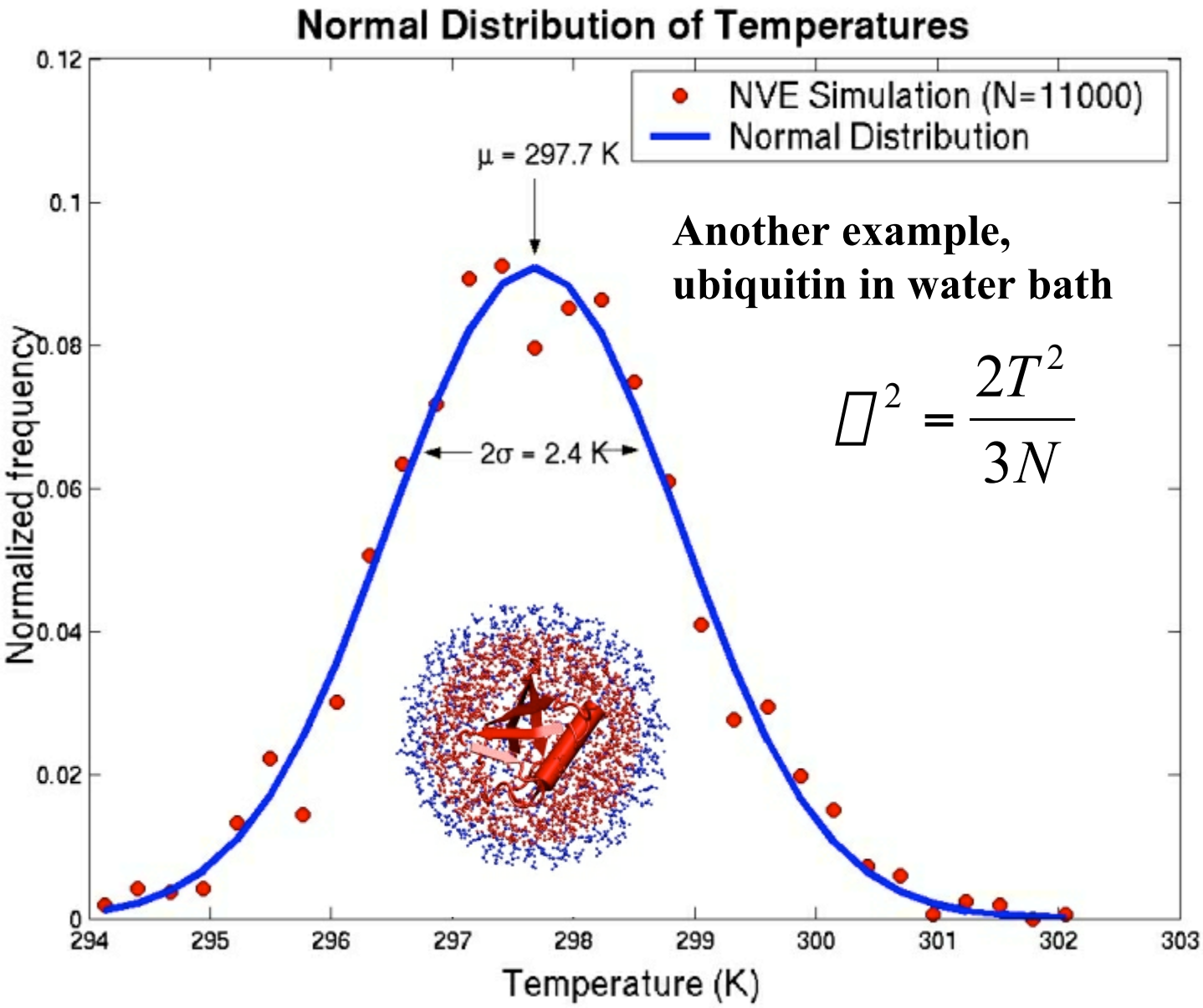
The distribution of the total kinetic energy $E_{kin} \;=\; \sum_j \frac{1}{2}m_jv_j^2$, according to the central limit theorem, is approximately Gaussian

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

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

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

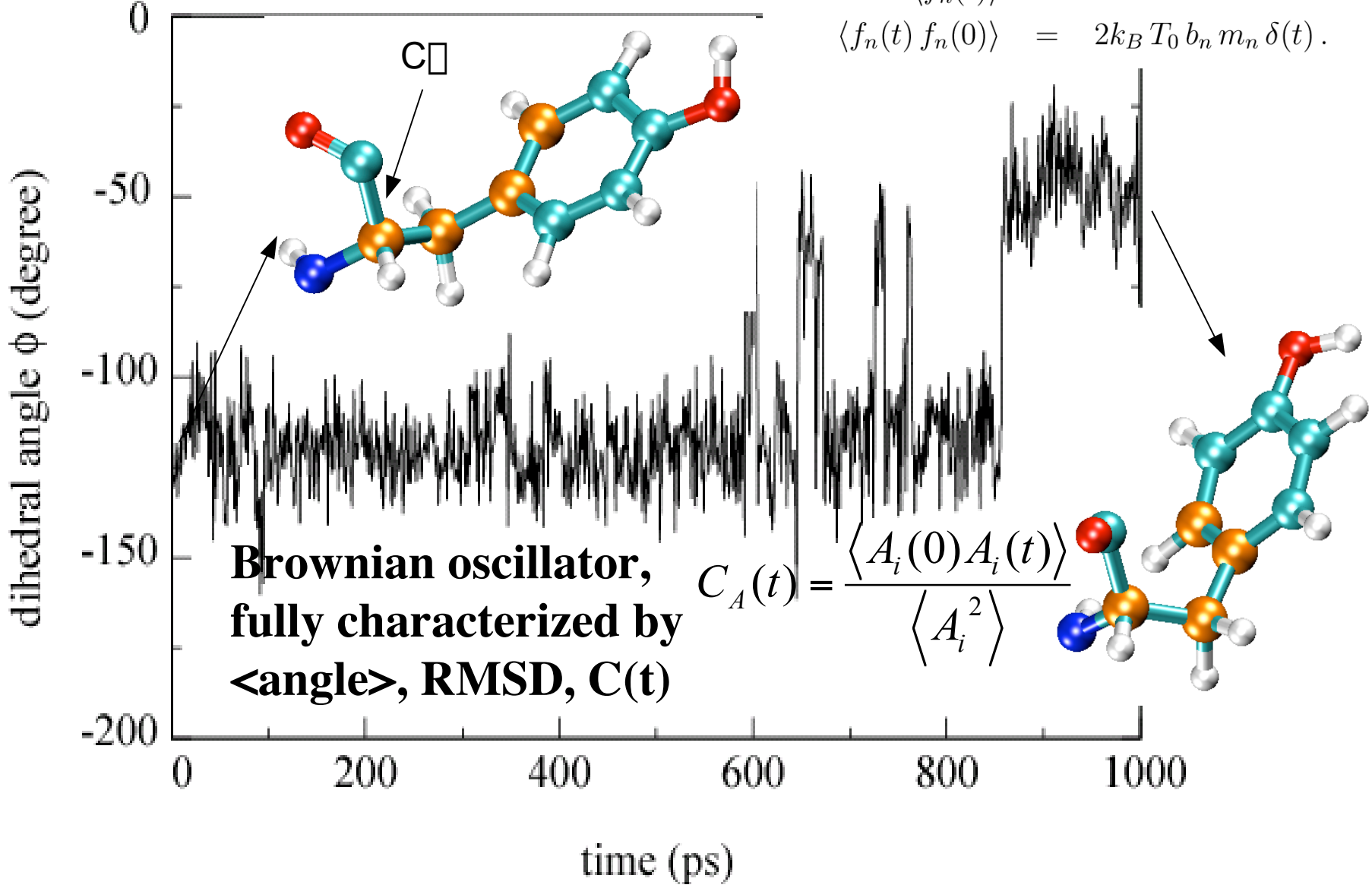
Show BPTI trajectory

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

Dihedral Angle

Langevin dynamics in strong friction limit

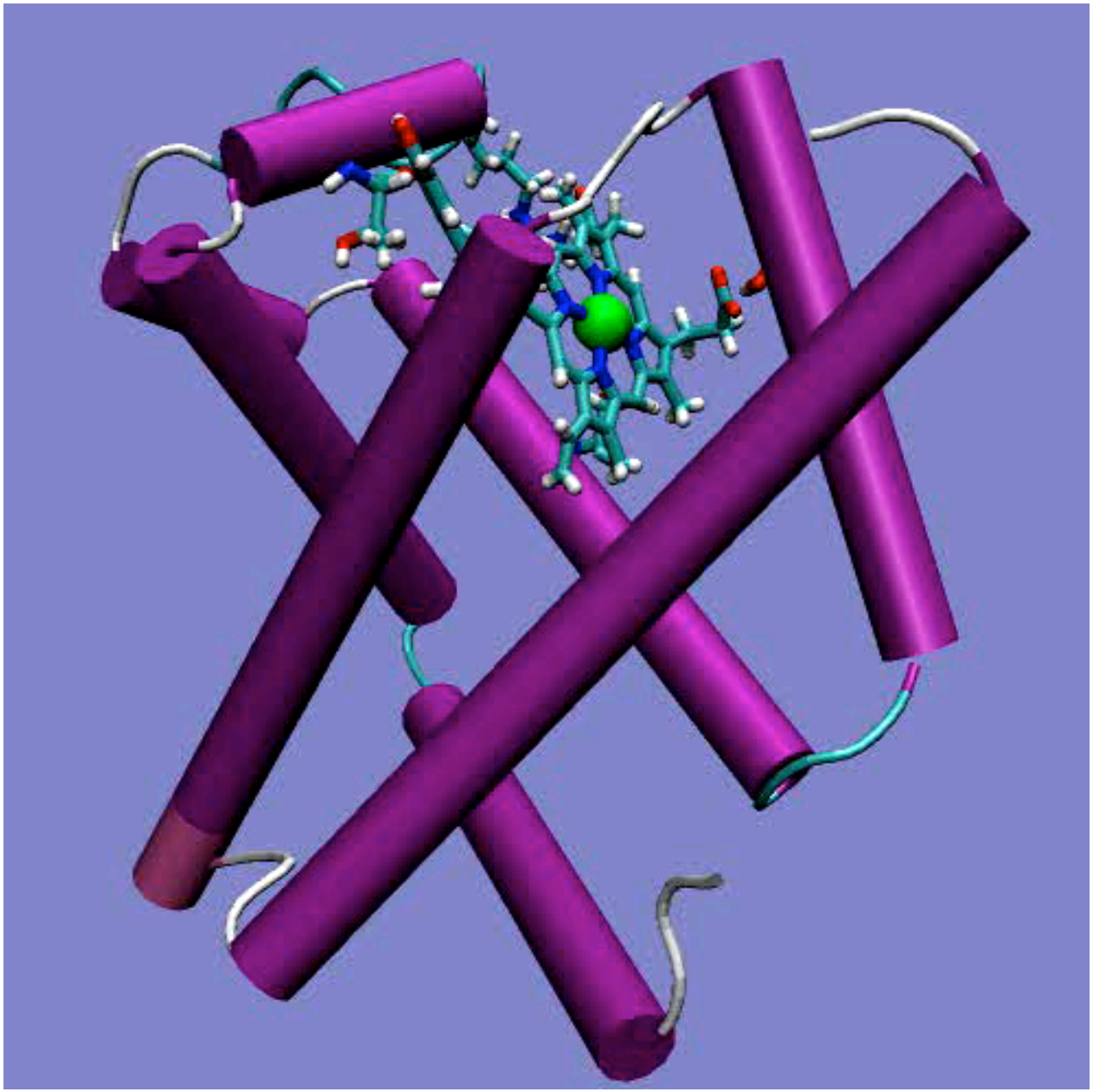
$$\langle f_n(t) \rangle = 0$$
$$\langle f_n(t)\, f_n(0) \rangle = 2 k_B T_0\, b_n\, m_n\, \delta(t)\,.$$



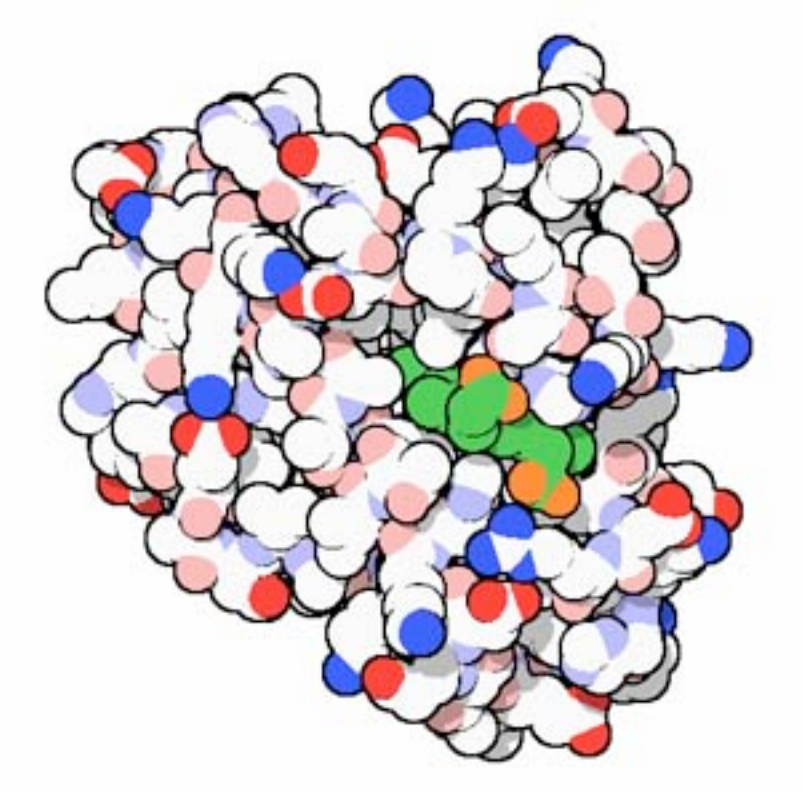
Brownian oscillator, fully characterized by <angle>, RMSD, C(t)

$$C_A(t) = \frac{\langle A_i(0) A_i(t) \rangle}{\langle A_i^2 \rangle}$$

Myoglobin Dynamics



Myoglobin

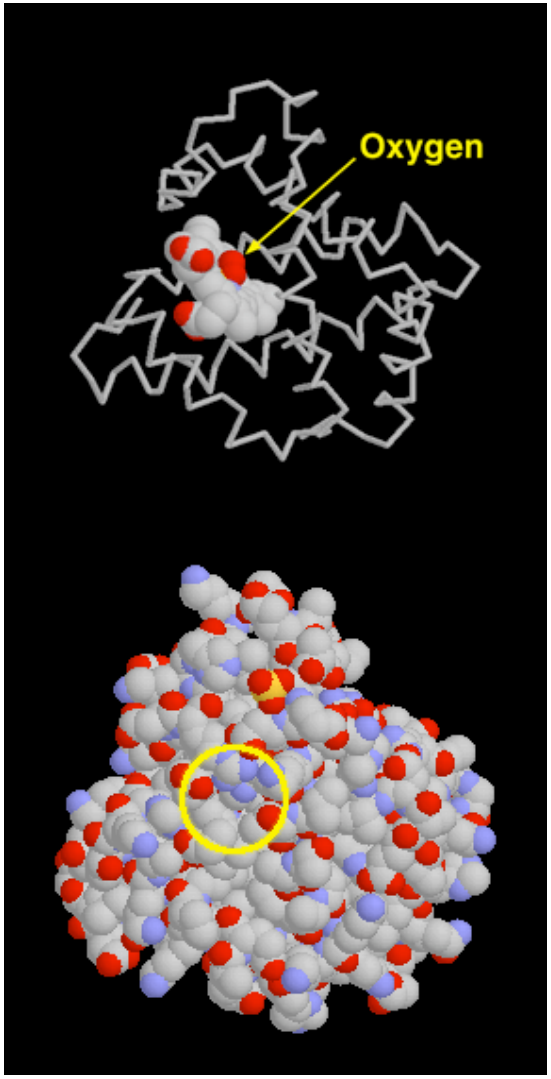


Myoglobin is a small, bright red protein. It is very common in muscle cells, and gives meat much of its red color. Its job is to store oxygen, for use when muscles are hard at work. If you look at John Kendrew's PDB file, you will notice that the myoglobin that he used was taken from sperm whale muscles. As you can imagine, marine whales and dolphins have a great need for myoglobin, so that they can store extra oxygen for use in their deep dives undersea.

PDB Molecule of the Month: Myoglobin

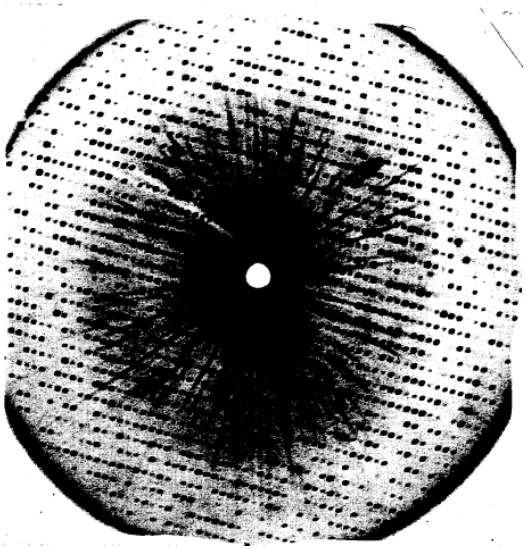
Oxygen Bound to Myoglobin

This structure of myoglobin, with the accession code **1mbo**, shows the location of oxygen. The iron atom at the center of the heme group holds the oxygen molecule tightly. Compare the two pictures. The first shows only a set of thin tubes to represent the protein chain, and the oxygen is easily seen. But when all of the atoms in the protein are shown in the second picture, the oxygen disappears, buried inside the protein.So how does the oxygen get in and out, if it is totally surrounded by protein? In reality, myoglobin (and all other proteins) are constantly in motion, performing small flexing and breathing motions. Temporary openings constantly appear and disappear, allowing oxygen in and out. The structure in the PDB is merely one snapshot of the protein, caught when it is in a tightly-closed form. Looking at the static structure held in the PDB, we must imagine the dynamic structure that actually exists in nature.The two pictures above were created with RASMOL. You can create similar pictures by accessing the PDB file **1mbo**, and then clicking on "View Structure." Try switching between the two types of pictures shown above, to prove to yourself that the oxygen is buried in this structure!



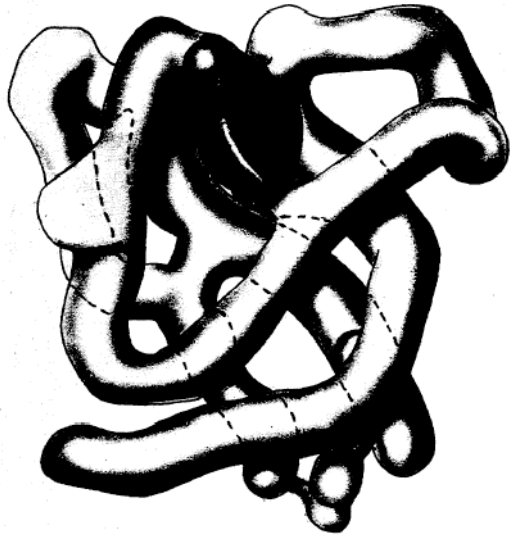
PDB Molecule of the Month: Myoglobin

Myogobin, the first protein with known structure



Diffraction pattern observed

John Cowdery Kendrew
Nobel Prize in Chemistry
Jointly with Max Perutz

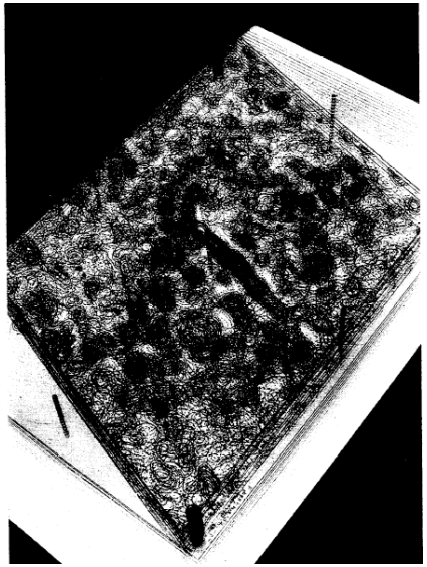


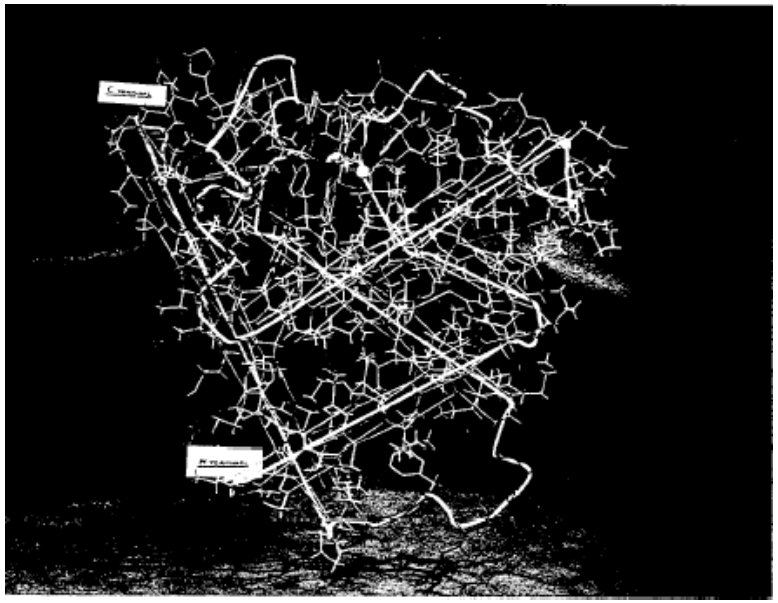
Struture model at 6 A resolution

Higher resolution Model:

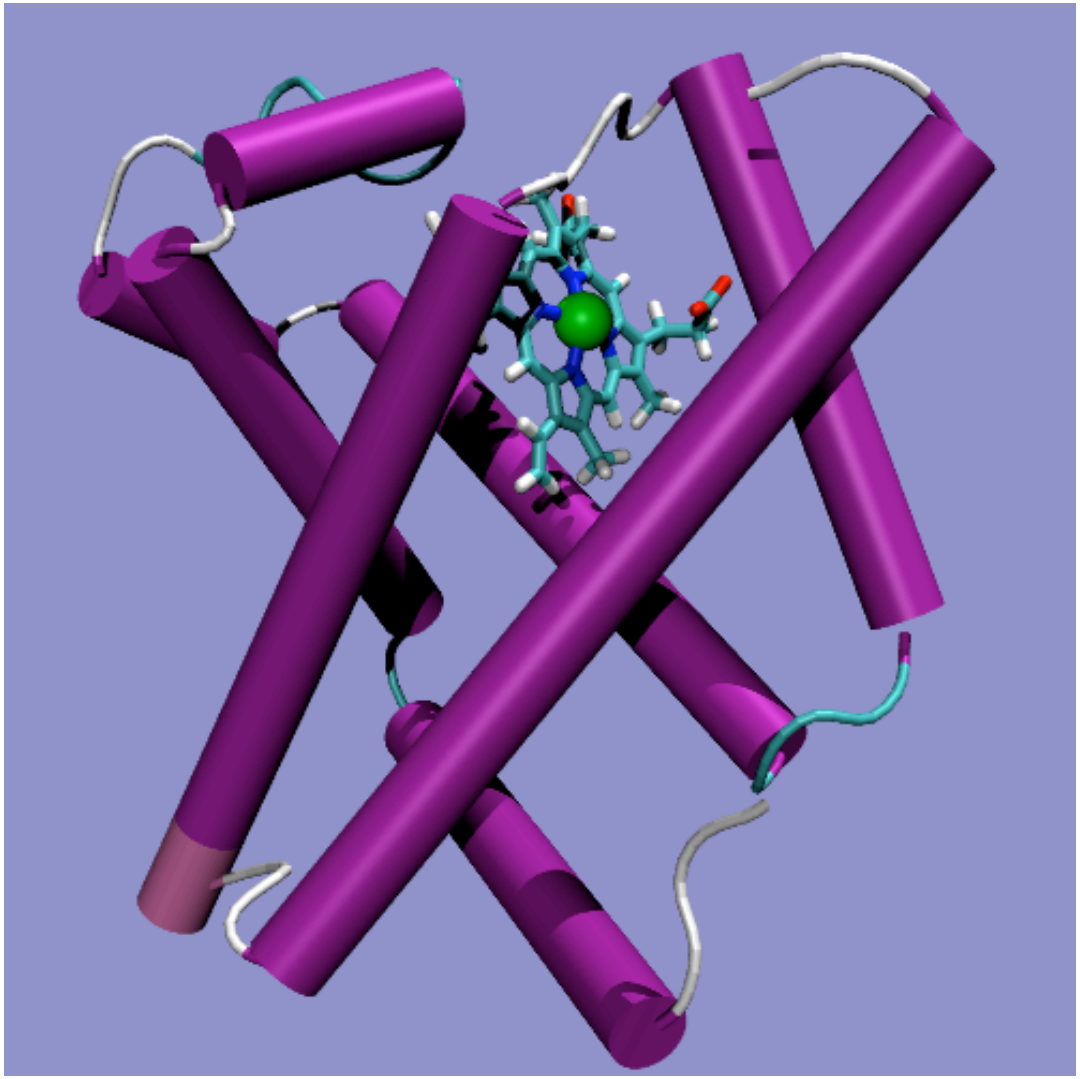
1) Construct electron density map

2) Build model





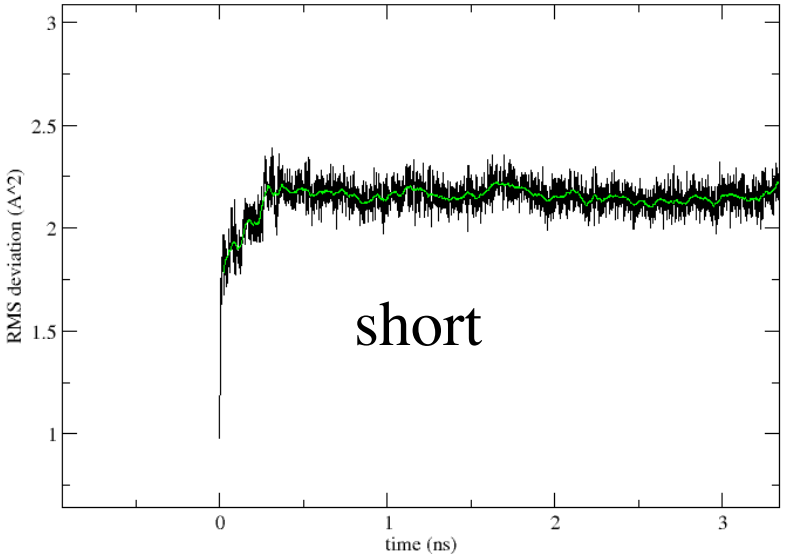
Myoglobin with heme group

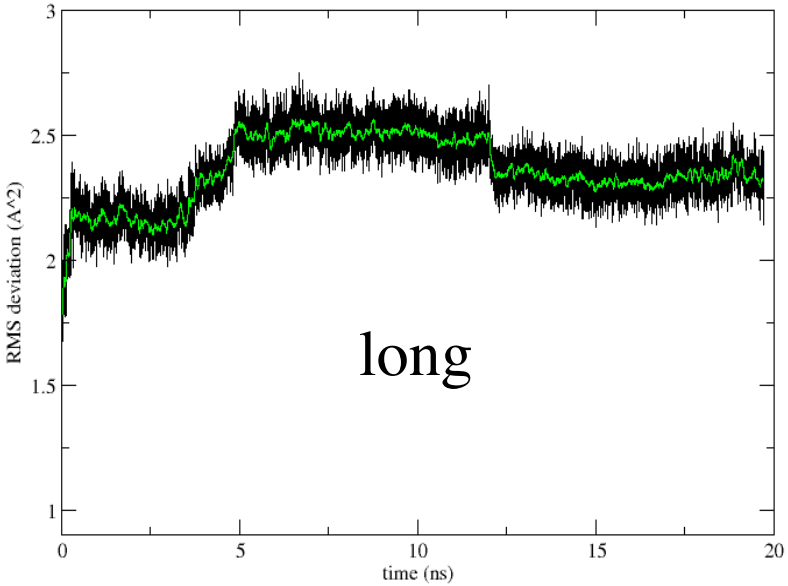


- Myoglobin from PDB structure 1A6M
- X-ray crystal structure at 1.00 A resolution.
- Steps seen in RMSD are due primarily to tilting of the helix to the upper right of the heme in the picture…

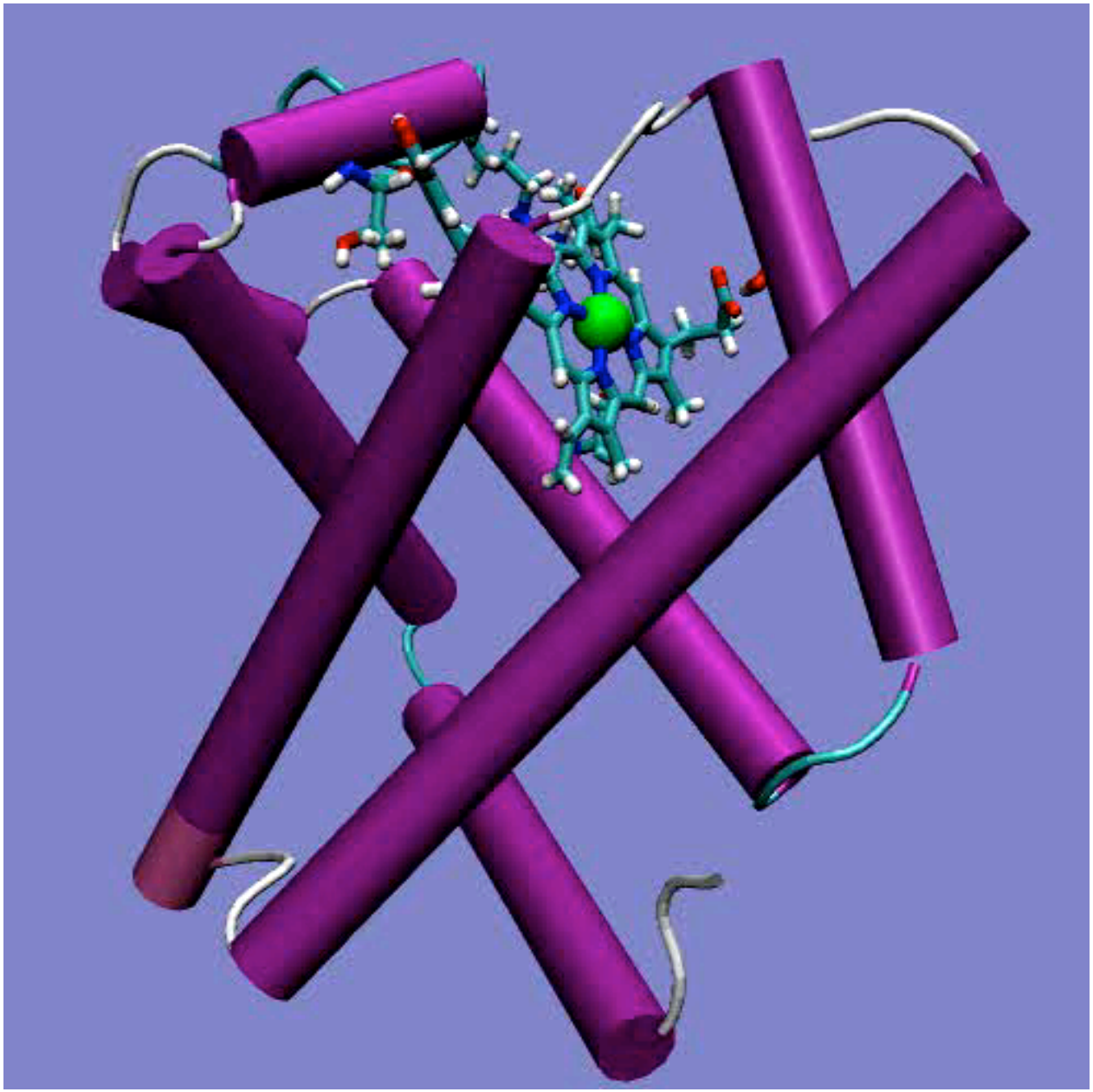
Setup and Equilibration

- Remove oxygen liganded to the Fe
- Minimize 1000 steps, fixing the Ca atoms.
- Heat for 5 ps with Langevin dynamics at 300 K, fixed Ca atoms.
- Simulate in NVT ensemble for 19 ns, saving coordinates every ps.



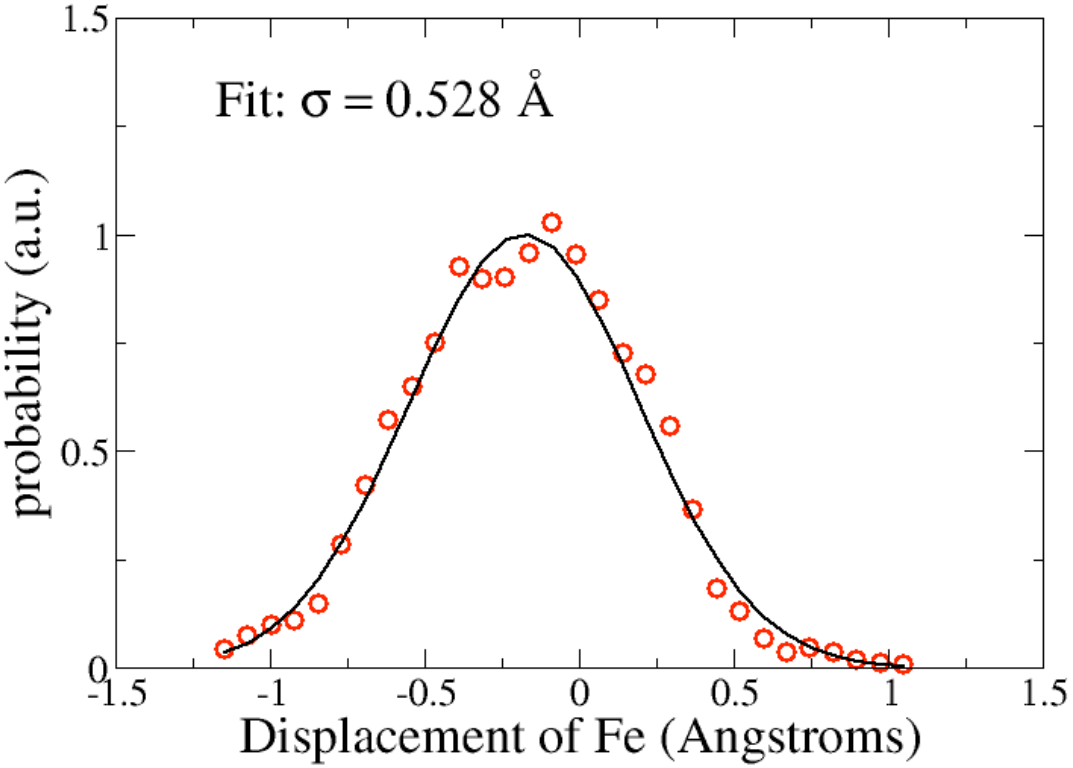


Myoglobin Dynamics



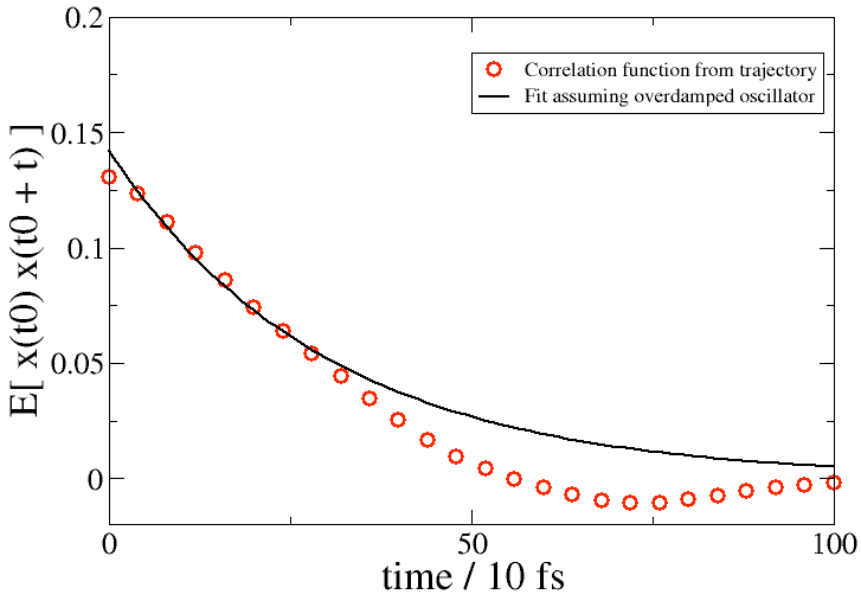
Obtain “f” from position distribution

- Best fit “by eye” is □=0.528 Å.
- However: standard deviation gives □=0.36 (f=kT/ □^2 = 319 pN/A); this is what we use below.



Obtain diffusion coefficient from position autocorrelation function

- Once we have the restoring force, the diffusion coefficient can be obtained from the position autocorrelation function:



$$h(t-t_o) = \int_{-\infty}^{\infty} dx \int_{\infty}^{\infty} dx_o \, x x_o \, p(x,t|x_o,t_o) p_o(x_o)$$
$$= (kT/2f) e^{-(D\,f/kT)(t-t_o)}$$

D f/kT = .0321

D = 0.0042 A^2/10 fs

= 0.42 A^2/ps.

Compare: water = 0.24 A^2/ps

Position autocorrelation: underdamped case

The Langevin equation governing underdamped motion is

$$\ddot{x} + b\dot{x} + \omega^2 x = \eta(t)$$

The position correlation function is given by

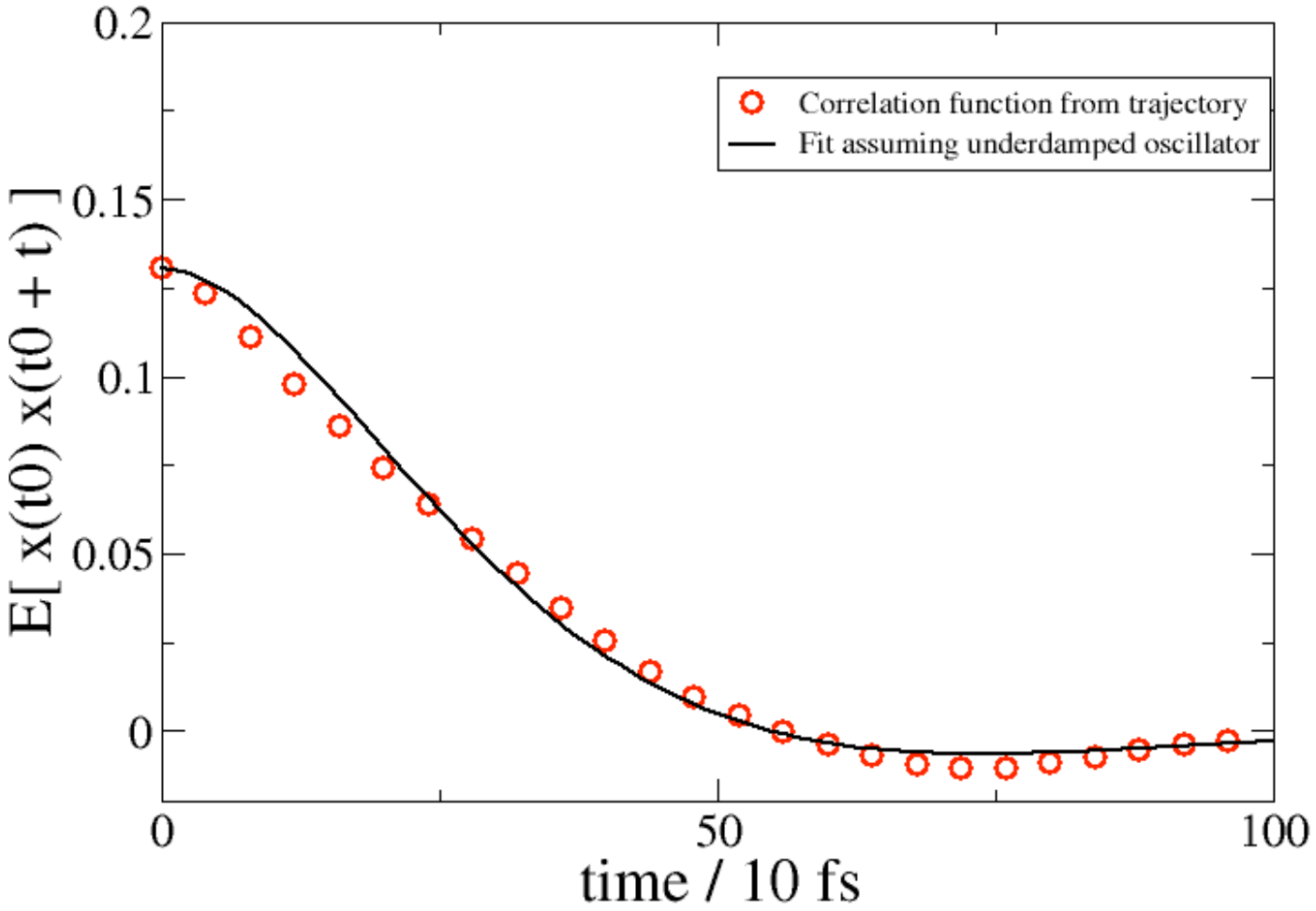
$$\langle x(0)^2 \rangle e^{-bt/2} \left[\cos(\Omega t) + \frac{b}{2\Omega} \sin(\Omega t) \right]$$

Using

$$\begin{aligned} m &= f/\omega^2 \quad \text{(from F = ma)} \\ \gamma &= mb \\ D &= k_B T/\gamma = \frac{k_B T}{f}\frac{\omega^2}{b} \end{aligned}$$

we can solve for D.

Diffusion coefficient from underdamped oscillator



Fitting parameters: □=0.0426; b=0.0811; □^2=
□^2+b^2/4 = 34.59/ps^2. D = 0.556 A^2/ps.

Mossbauer line shape function

The lineshape $I(\omega)$ we are trying to calculate is given by

$$I(\omega) = \frac{\sigma_0 \Gamma}{4} \int_{-\infty}^{\infty} dt\, e^{-i\omega t - \frac{1}{2}\Gamma|t|} G(\mathbf{k}, t) \qquad (1)$$

where $G(\mathbf{k}, t)$ is given by

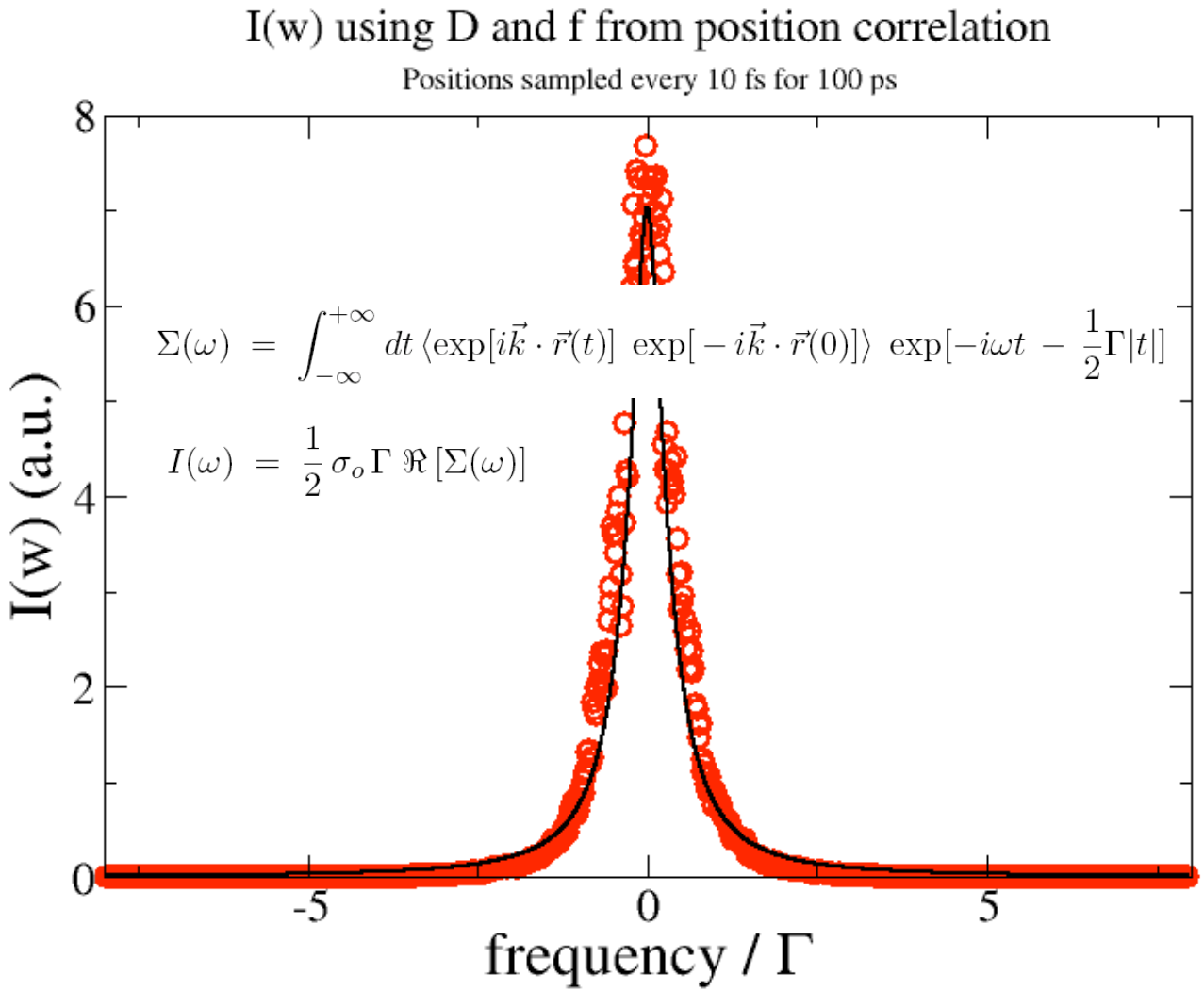
$$G(\mathbf{k}, t) = \int d\mathbf{r} \int d\mathbf{r}_o e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}_o)} p(\mathbf{r}, t|\mathbf{r}_o, 0) p_0(\mathbf{r}_o) \qquad (2)$$

Notice that $G(\mathbf{k}, t)$ is just the autocorrelation function of $\exp(-i\mathbf{k}\cdot\mathbf{r})$.

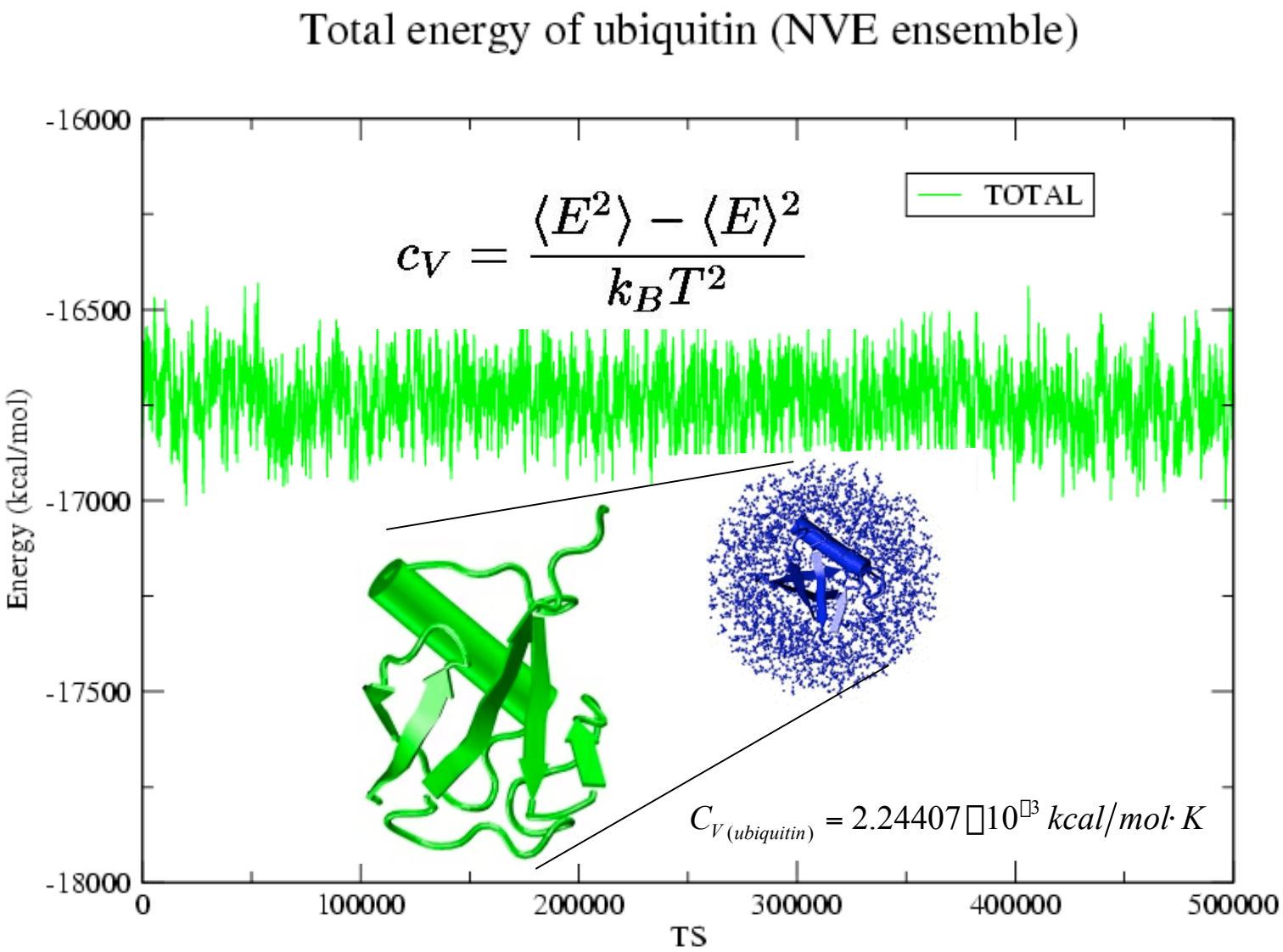
Mossbauer line shape function

- To relate the lineshape to a simulation we must define **r** in terms of atomic coordinates.
- Define **k** = k**n**, where **n** is the normal to the plane of the heme, let **r** be the Fe coordinate. We must then compute **n**(t) . **r**(t) for all frames in our simulation.

Moessbauer Line Shape Function - Sampled and Matched to Analytical Formula



Specific Heat of a Protein



Now the dancing ubiquitin