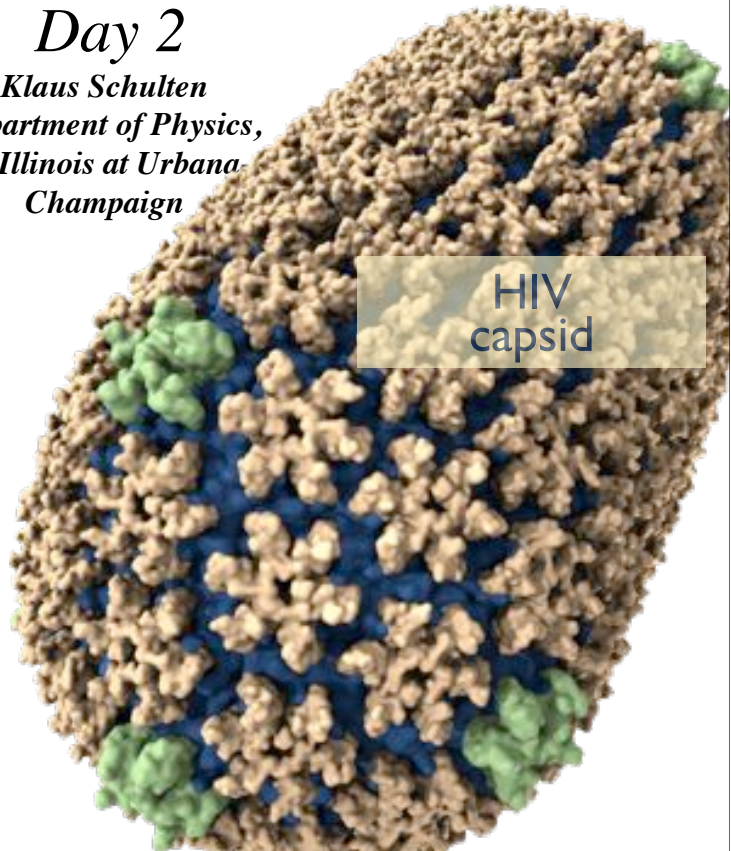


Introduction to Molecular Dynamics

Day 2

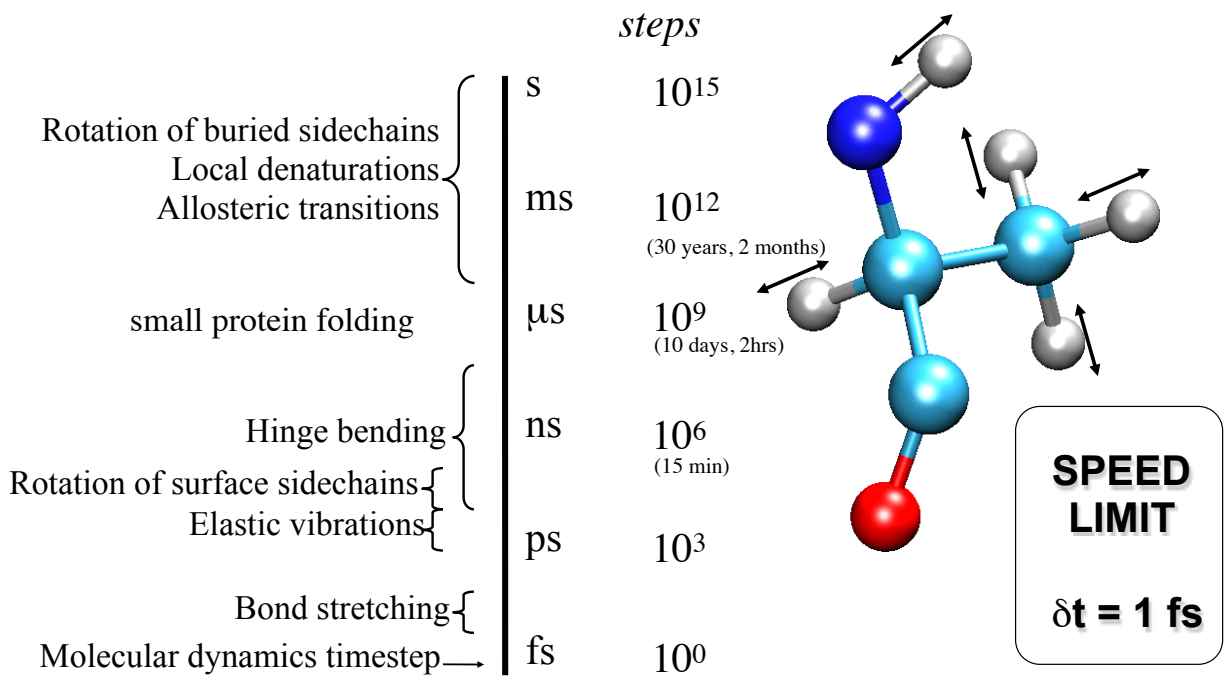
Klaus Schulten
 Department of Physics,
 U. Illinois at Urbana-
 Champaign

Discoveries
 Through the
 Computational
 Microscope



But long is still a problem!

biomolecular timescale and timestep limits

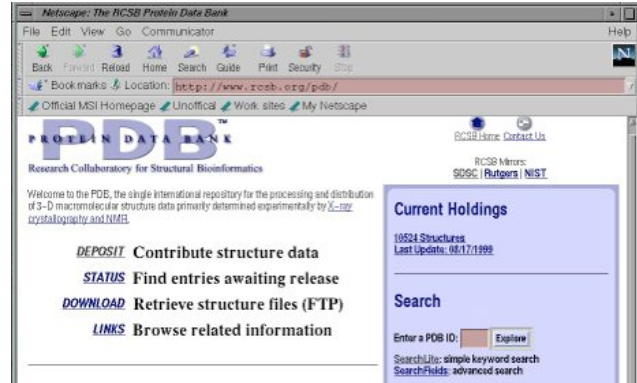


(NSF center, Shaw Res.)

PDB Files

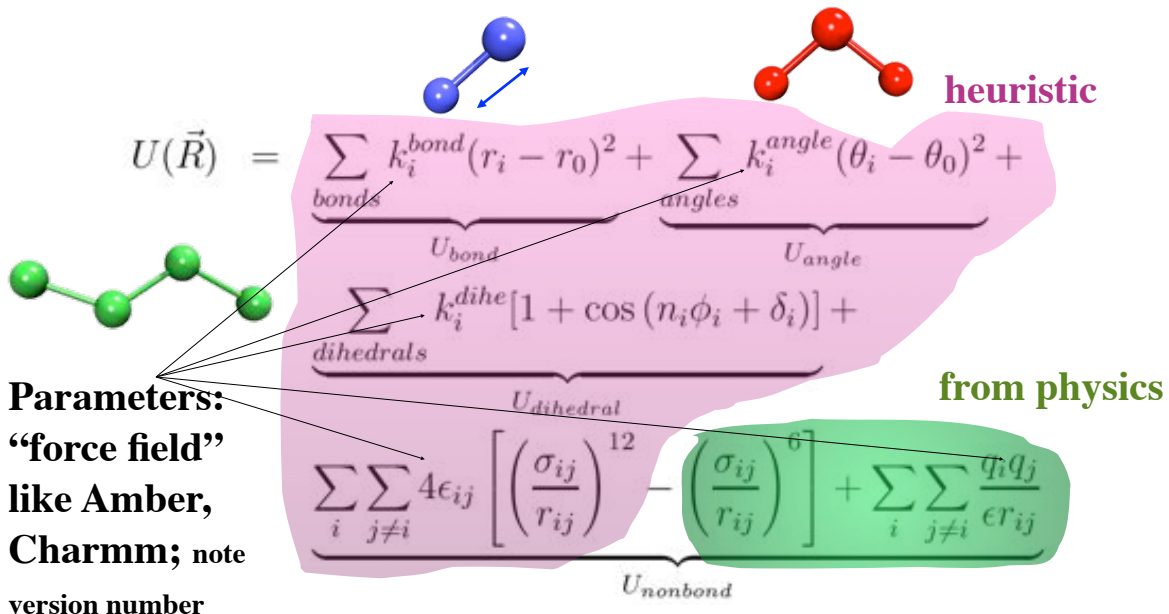
gives one the structure and starting position

- 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.)



Potential Energy Function of Biopolymer

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



Preparing Your System for MD

Solvation

Biological activity is the result of interactions between molecules and occurs at the interfaces between molecules (protein-protein, protein-DNA, protein-solvent, DNA-solvent, etc).

*mitochondrial
bc1 complex*

Why model solvation?

- many biological processes occur in aqueous solution
- solvation effects play a crucial role in determining molecular conformation, electronic properties, binding energies, etc

How to model solvation?

- explicit treatment: solvent molecules are added to the molecular system
- implicit treatment: solvent is modeled as a continuum dielectric or so-called implicit force field



Preparing Your System for MD

Solvation

Biological activity is the result of interactions between molecules and occurs at the interfaces between molecules (protein-protein, protein-DNA, protein-solvent, DNA-solvent, etc).

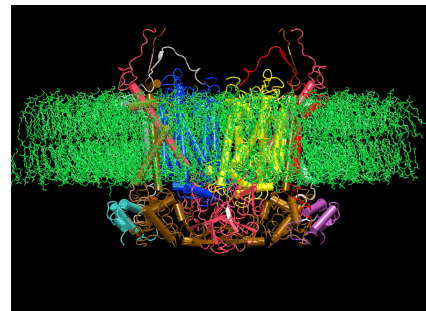
*mitochondrial
bc1 complex*

Why model solvation?

- many biological processes occur in aqueous solution
- solvation effects play a crucial role in determining molecular conformation, electronic properties, binding energies, etc

How to model solvation?

- explicit treatment: solvent molecules are added to the molecular system
- implicit treatment: solvent is modeled as a continuum dielectric



Preparing Your System for MD

Solvation

Biological activity is the result of interactions between molecules and occurs at the interfaces between molecules (protein-protein, protein-DNA, protein-solvent, DNA-solvent, etc).

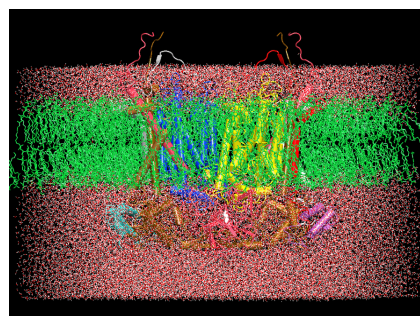
*mitochondrial
bc1 complex*

Why model solvation?

- many biological processes occur in aqueous solution
- solvation effects play a crucial role in determining molecular conformation, electronic properties, binding energies, etc

How to model solvation?

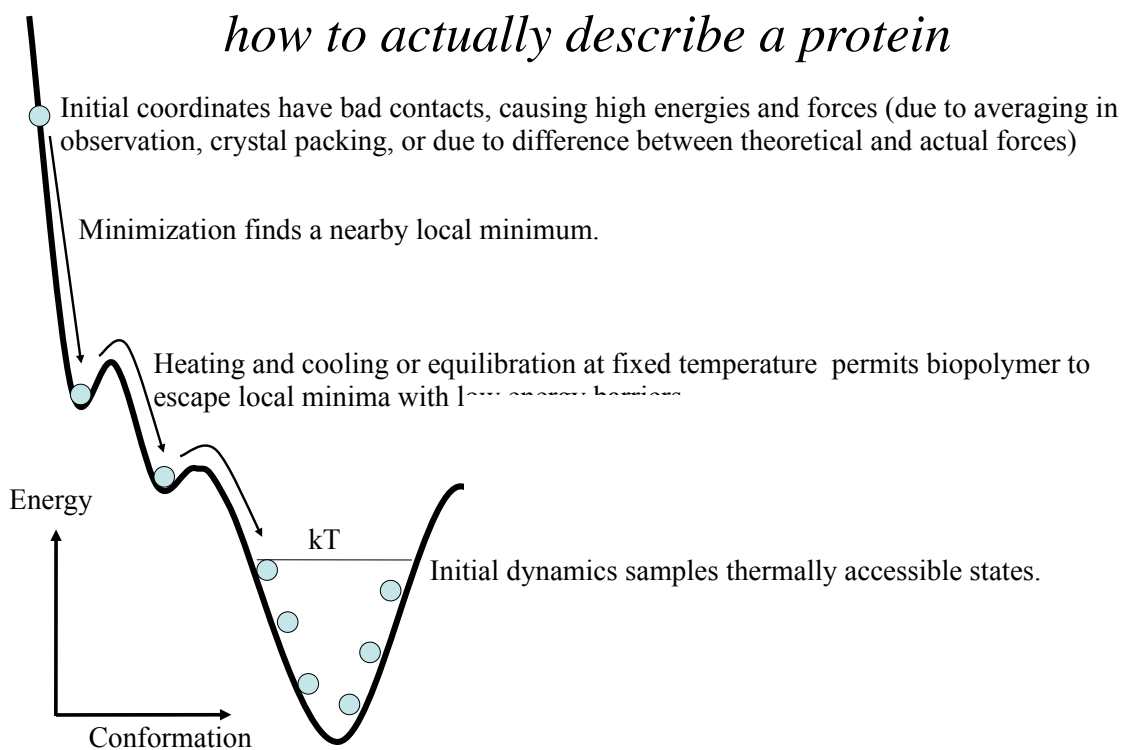
- explicit treatment: solvent molecules are added to the molecular system
- implicit treatment: solvent is modeled as a continuum dielectric



**(Usually periodic!
Avoids surface effects)**

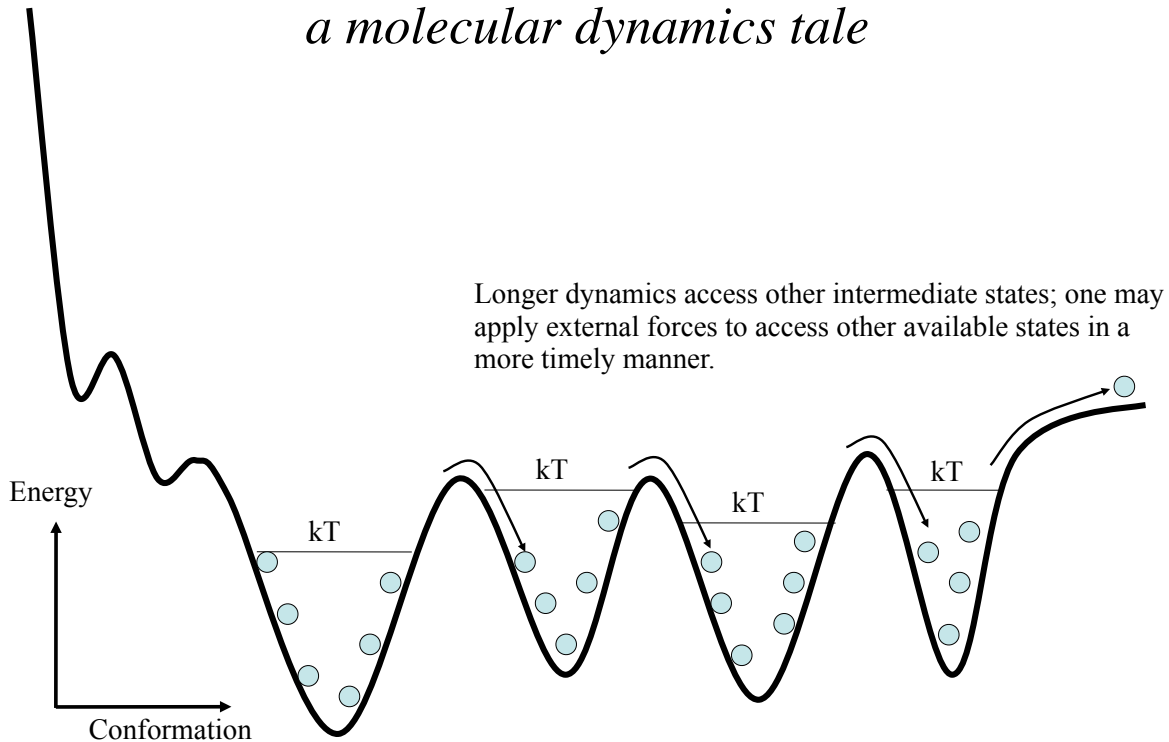
From the Mountains to the Valleys

how to actually describe a protein



From the Mountains to the Valleys

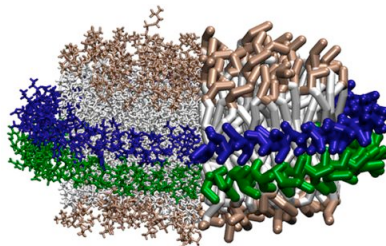
a molecular dynamics tale



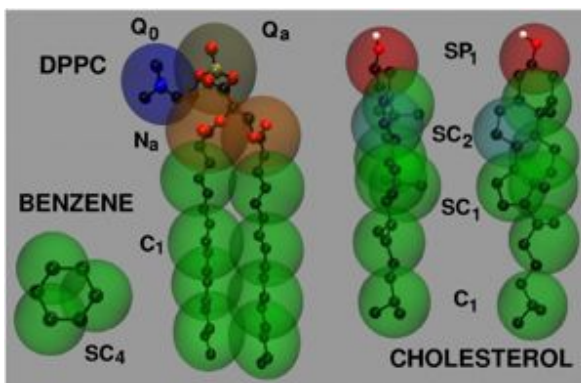
Cutting Corners

cutoffs, PME, rigid bonds, and multiple timesteps

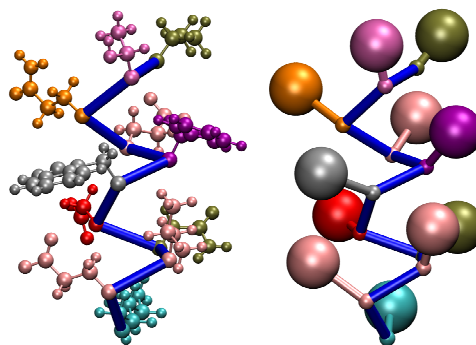
- Nonbonded interactions require order N^2 computer time!
 - Truncating at R_{cutoff} reduces this to order $N R_{\text{cutoff}}^3$
 - Particle mesh Ewald (PME) method adds long range electrostatics at order $N \log N$, only minor cost compared to cutoff calculation.
- Can we extend the timestep, and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



Residue-Based Coarse-Grained Model



- Protein model uses two CG beads per residue
- One CG bead per side chain another for backbone



All-atom peptide

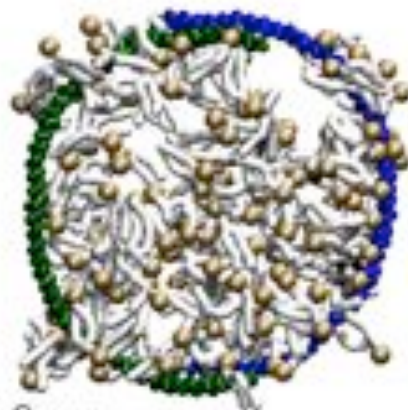
CG peptide

- Lipid model: MARTINI
- Level of coarse-graining: ~4 heavy atoms per CG bead
- Interactions parameterized based on experimental data and thermodynamic properties of small molecules

Peter L. Freddolino, Anton Arkhipov, Amy Y. Shih, Ying Yin, Zhongzhou Chen, and Klaus Schulten. **Application of residue-based and shape-based coarse graining to biomolecular simulations.** In Gregory A. Voth, editor, *Coarse-Graining of Condensed Phase and Biomolecular Systems*, chapter 20, pp. 299-315. Chapman and Hall/CRC Press, Taylor and Francis Group, 2008.

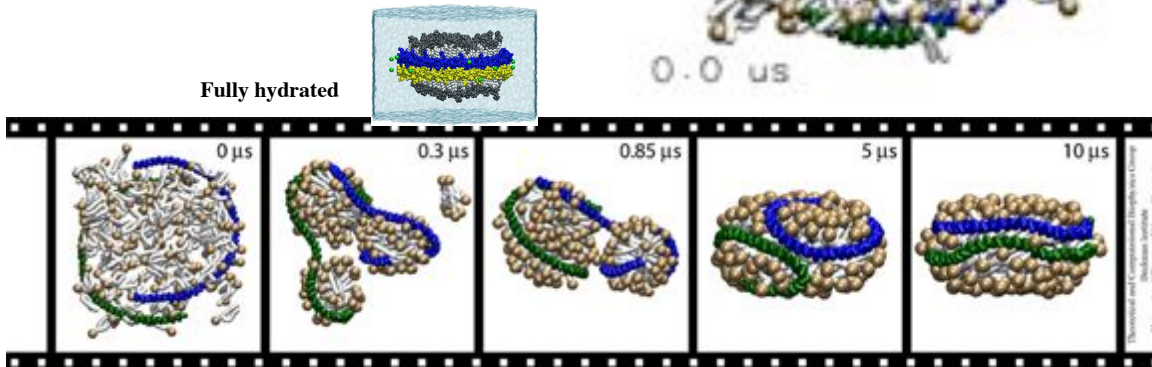
Nanodisc Assembly CG MD Simulation

- 10 μs simulation
- Assembly proceeds in two steps:
 - Aggregation of proteins and lipids driven by the hydrophobic effect
 - Optimization of the protein structure driven by increasingly specific protein-protein interactions
- Formation of the generally accepted double-belt model for discoidal HDL



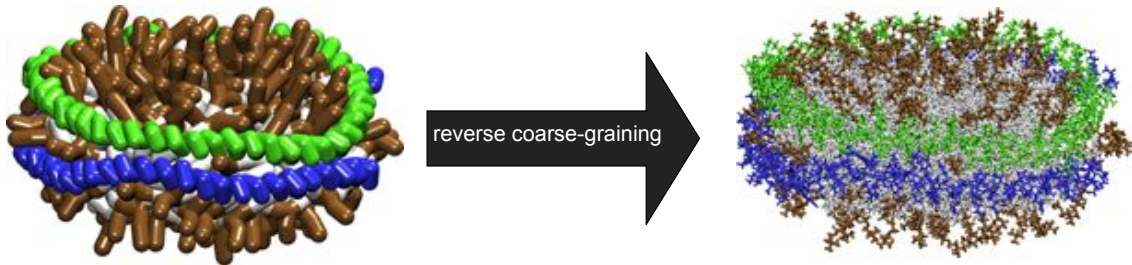
0.0 μs

Fully hydrated



Validation of Simulations

reverse coarse-graining and small-angle X-ray scattering

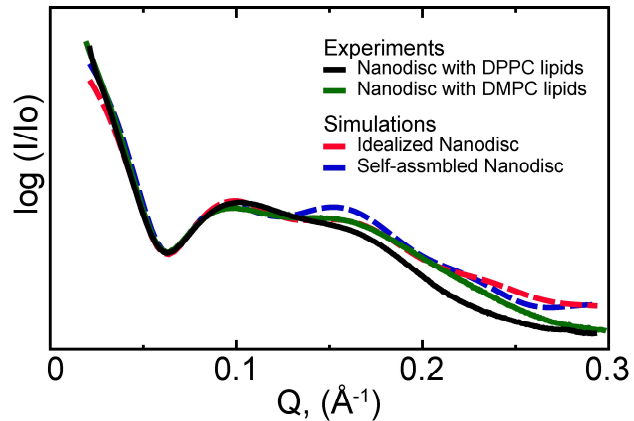


Reverse coarse-graining:

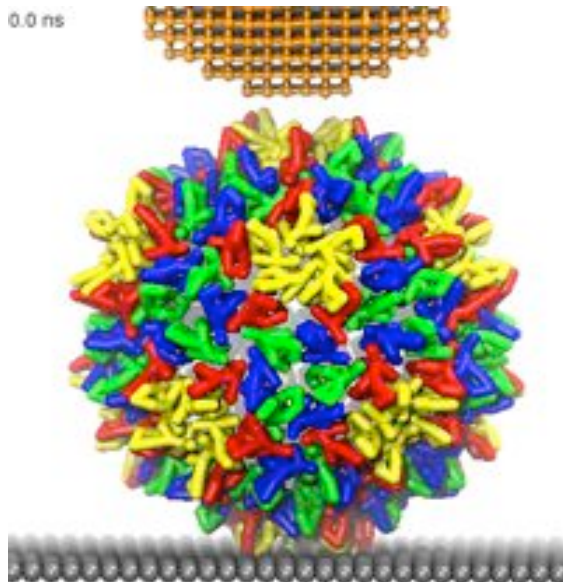
1. Map center of mass of the group of atoms represented by a single CG bead to that beads location
2. MD minimization, simulated annealing with restraints, and equilibration to get all-atom structure

Small-angle X-ray scattering:

Calculated from reverse coarse-grained all-atom model and compared with experimental measurements



Shape-Based Coarse-Grained (CG) model



Anton Arkhipov, Wouter H. Roos, Gijs J. L. Wuite, and Klaus Schulten.
Elucidating the mechanism behind irreversible deformation of viral capsids. *Biophysical Journal*, 97, 2009. In press.

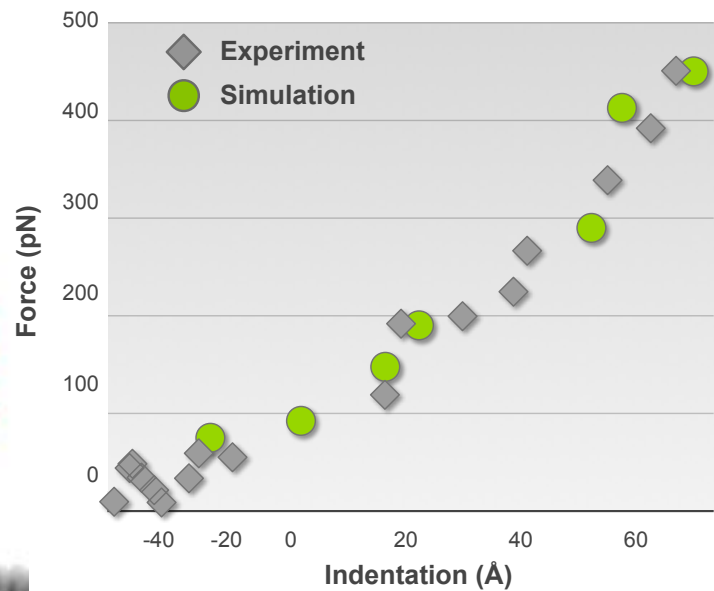
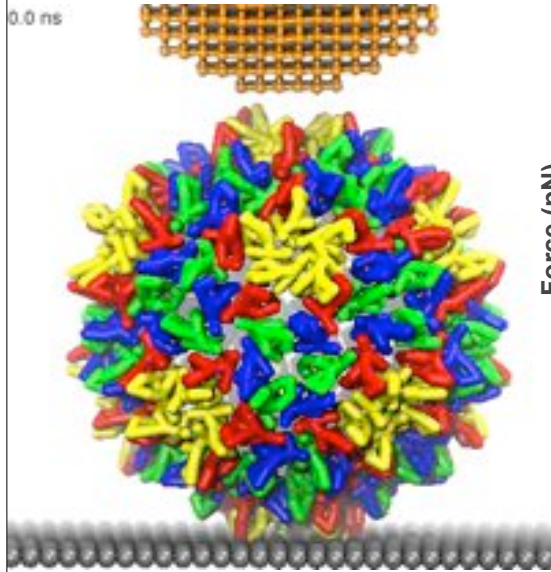
- Fully automatic
- Number of CG beads is chosen by a user (we used ~200 atoms per CG bead)

Peter L. Freddolino, Anton Arkhipov, Amy Y. Shih, Ying Yin, Zhongzhou Chen, and Klaus Schulten. **Application of residue-based and shape-based coarse graining to biomolecular simulations.** In Gregory A. Voith, editor, *Coarse-Graining of Condensed Phase and Biomolecular Systems*, chapter 20, pp. 299-315. Chapman and Hall/CRC Press, Taylor and Francis Group, 2008.

Virus Capsid Mechanics

Atomic Force Microscope

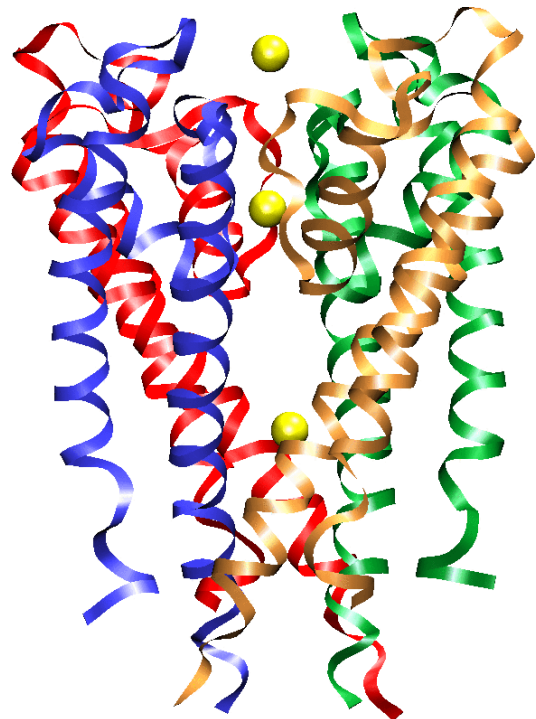
— Hepatitis B Virus —



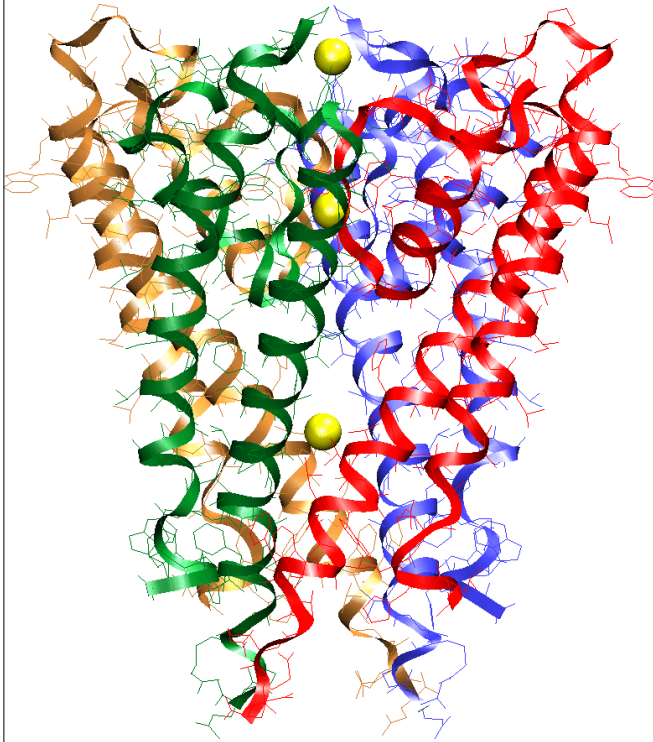
Example: MD Simulations of the K⁺ Channel Protein

Ion channels are membrane - spanning proteins that form a pathway for the flux of inorganic ions across cell membranes.

Potassium channels are a particularly interesting class of ion channels, managing to distinguish with impressive fidelity between K⁺ and Na⁺ ions while maintaining a very high throughput of K⁺ ions when gated.

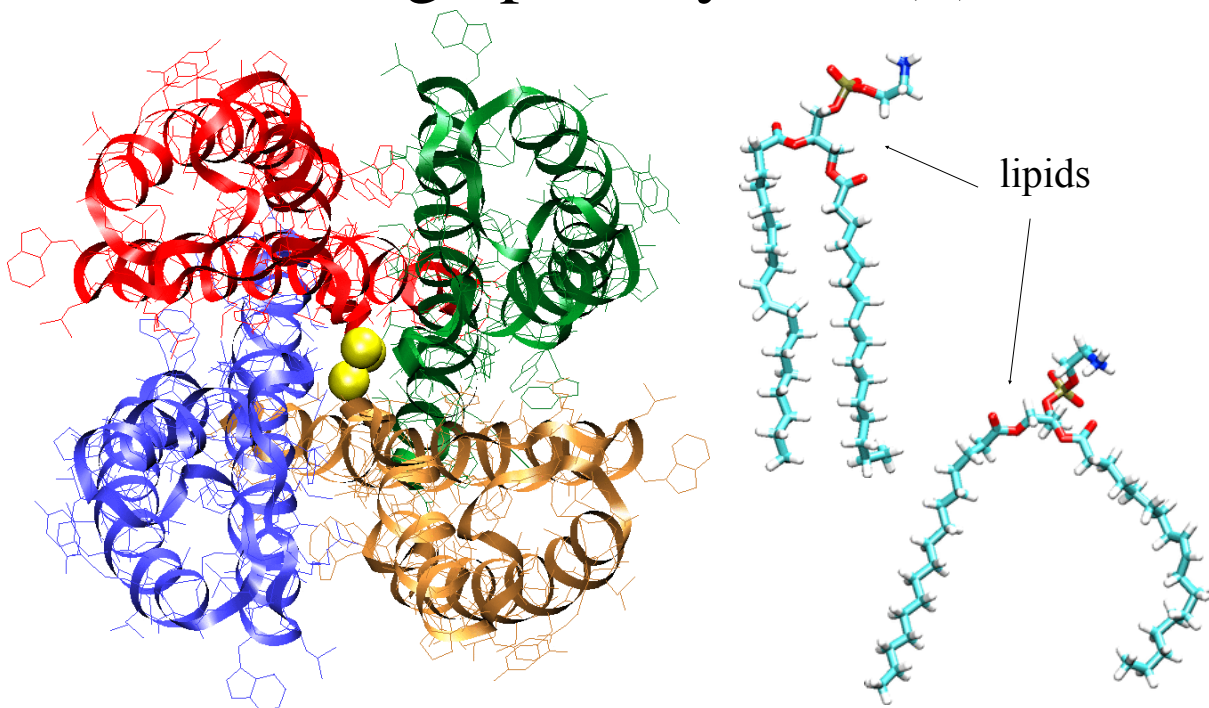


Setting up the system (1)



- retrieve the PDB (coordinates) file from the Protein Data Bank
- add hydrogen atoms using PSFGEN
- use psf and parameter files to set up the structure; needs better than available in Charmm to describe well the ions
- minimize the protein structure using NAMD2

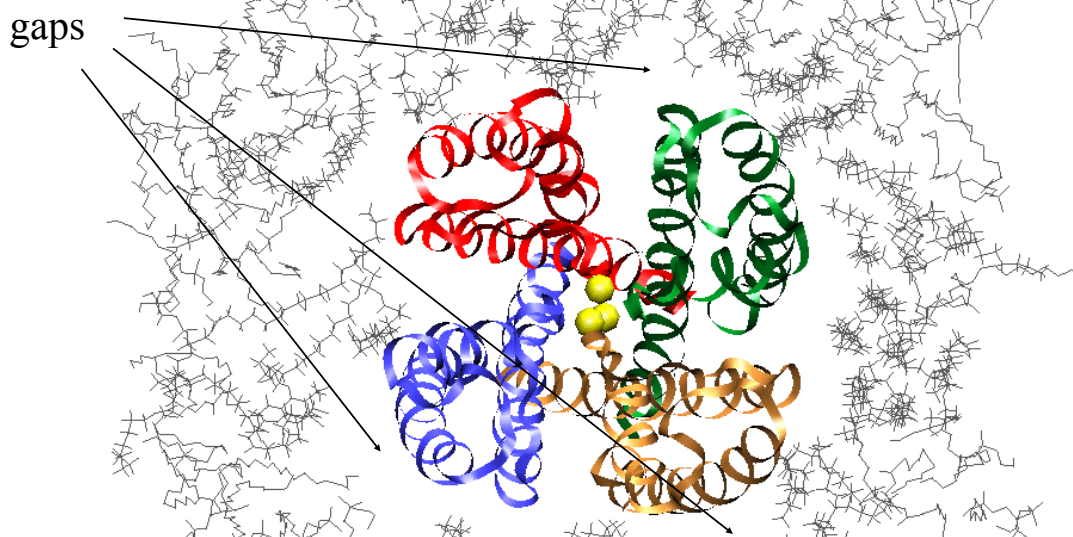
Setting up the system (2)



Simulate the protein in its natural environment: solvated lipid bilayer

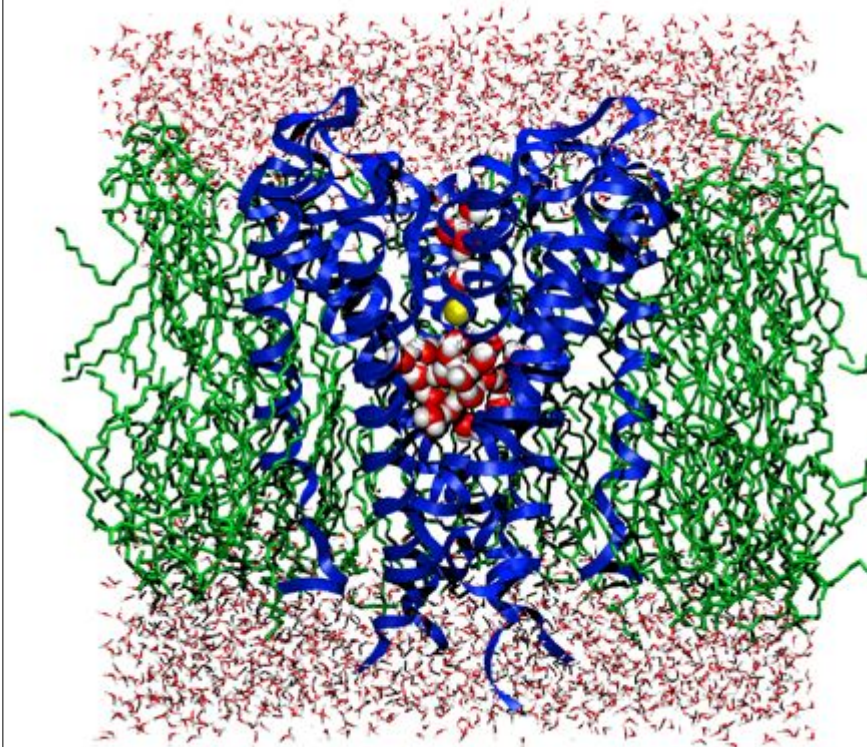
Setting up the system (3)

Inserting the protein in the lipid bilayer



Automatic insertion into the lipid bilayer leads to big gaps between the protein and the membrane => long equilibration time required to fill the gaps.
Solution: manually adjust the position of lipids around the protein. Employ constant (lateral and normal) pressure control.

The system



solvent

Kcsa channel protein (in blue) embedded in a (3:1) POPE/POPG lipid bilayer. Water molecules inside the channel are shown in vdW representation.

solvent

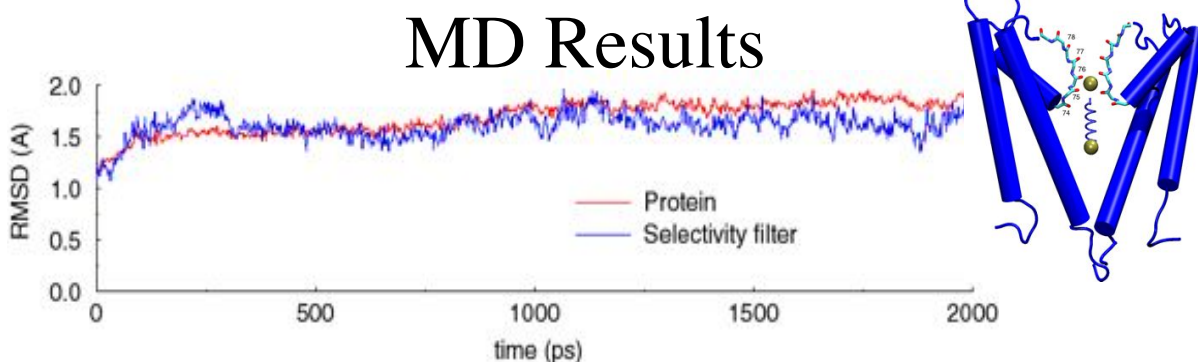
Simulating the system: Free MD

Summary of simulations:

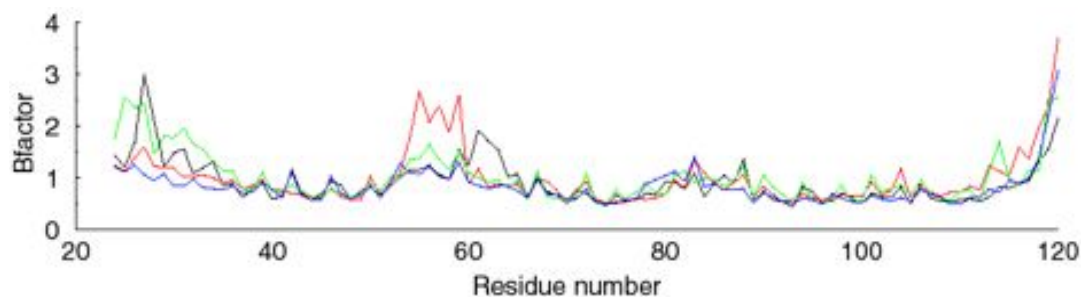
- protein/membrane system contains 38,112 atoms, including 5117 water molecules, 100 POPE and 34 POPG lipids, plus K^+ counterions
- CHARMM26 forcefield
- periodic boundary conditions, PME electrostatics
- 1 ns equilibration at 310K, NpT
- 2 ns dynamics, NpT

Program: NAMD2

Platform: Cray T3E (Pittsburgh Supercomputer Center) or local computer cluster; choose ~1000 atoms per processor.

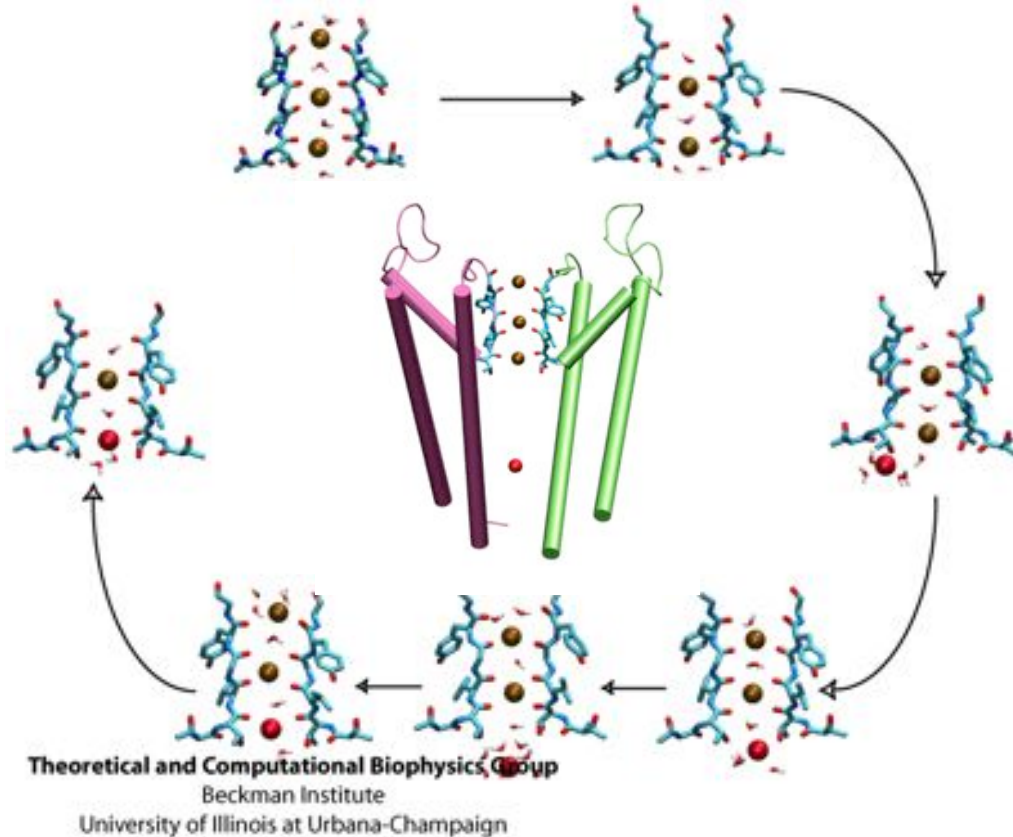


RMS deviations for the KcsA protein and its selectivity filter indicate that the protein is stable during the simulation with the selectivity filter the most stable part of the system.



Temperature factors for individual residues in the four monomers of the KcsA channel protein indicate that the most flexible parts of the protein are the N and C terminal ends, residues 52-60 and residues 84-90. Residues 74-80 in the selectivity filter have low temperature factors and are very stable during the simulation.

Simulation of Ion Conduction (here for Kv1.2)

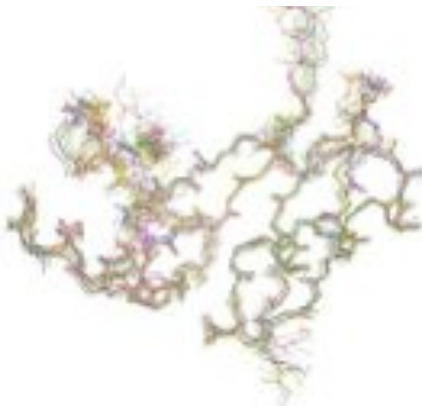
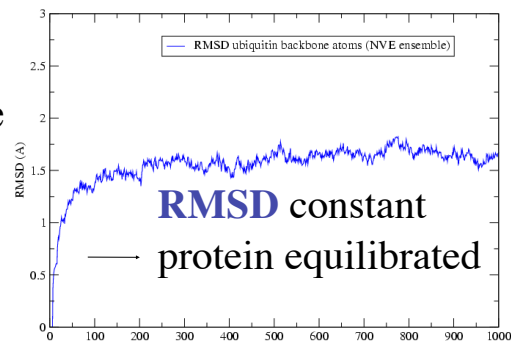


Equilibrium Properties of Proteins

Ubiquitin

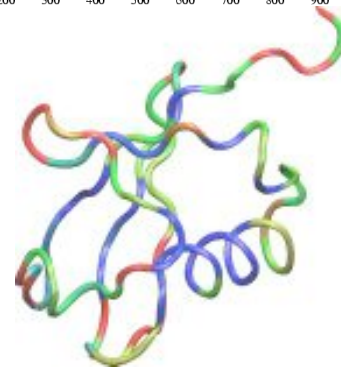
Root Mean Squared Deviation: measure for equilibration and protein flexibility

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



NMR structures
aligned together to see flexibility

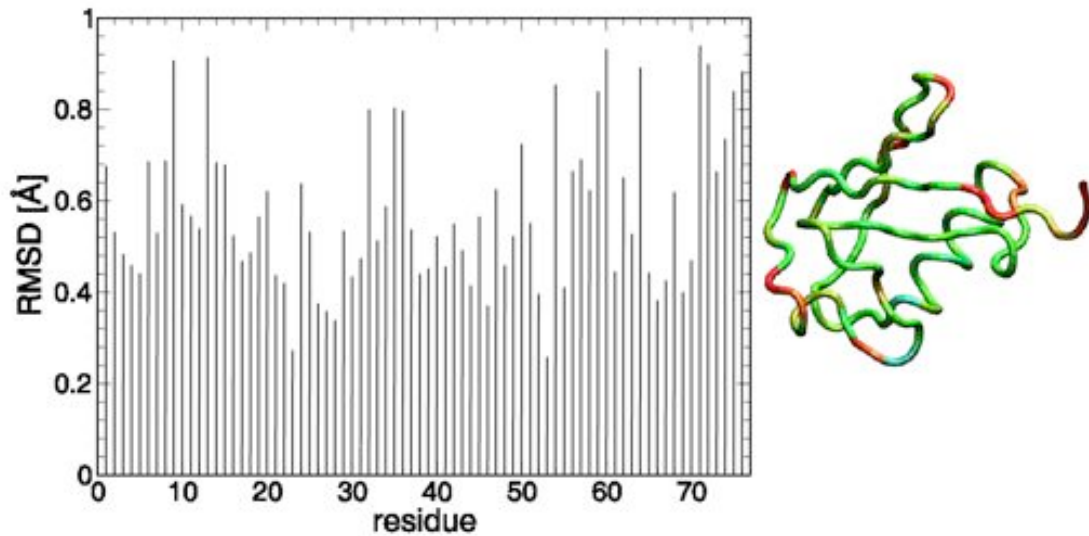
**Protein sequence
exhibits
characteristic
permanent
flexibility!**



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

Thermal Motion of Ubiquitin from MD

RMSD values per residue



Thermal Motion of Ubiquitin from MD

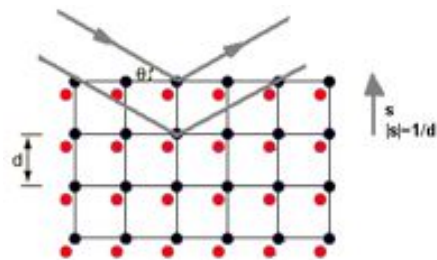
Temperature Dependence of Crystal Diffraction (Debye-Waller factor)

Bragg's law

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

structure factor

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



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:

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

Spatial average for harmonic oscillator: $\langle (\vec{s} \cdot \vec{u}_j)^2 \rangle = \frac{1}{3} s^2 \langle u_j^2 \rangle$

One can carry out the expansion further and show

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

Using for the thermal amplitude of the harmonic oscillator

$$\frac{1}{2} m \omega^2 u_j^2 = \frac{3}{2} k_B T$$

one obtains

Debye-Waller factor

$$\langle f_j \exp[-i\vec{s} \cdot \vec{r}_j] \rangle = f_j \overbrace{\exp[-s^2 k_B T / 2m\omega^2]}^{\text{Debye-Waller factor}} \exp[-i\vec{s} \cdot \vec{x}_j]$$

Equilibrium Properties of Proteins

Energies: kinetic and potential



temperature
dependence

$$\begin{aligned}
 & \left\langle \sum_j \frac{1}{2} m_j v_j^2 \right\rangle \\
 & \text{Kinetic energy (quadratic)} \\
 U(\vec{R}) = & \underbrace{\sum_{\text{bonds}} k_i^{\text{bond}} (r_i - r_0)^2}_{U_{\text{bond}}} + \underbrace{\sum_{\text{angles}} k_i^{\text{angle}} (\theta_i - \theta_0)^2}_{U_{\text{angle}}} + \\
 & \underbrace{\sum_{\text{dihedrals}} k_i^{\text{dihe}} [1 + \cos(n_i \phi_i + \delta_i)]}_{U_{\text{dihedral}}} + \\
 & \underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]}_{U_{\text{nonbond}}} + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}} \\
 & \text{Potential energy (not all quadratic)}
 \end{aligned}$$

Equilibrium Properties of Proteins

Energies: kinetic and potential



temperature dependence

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

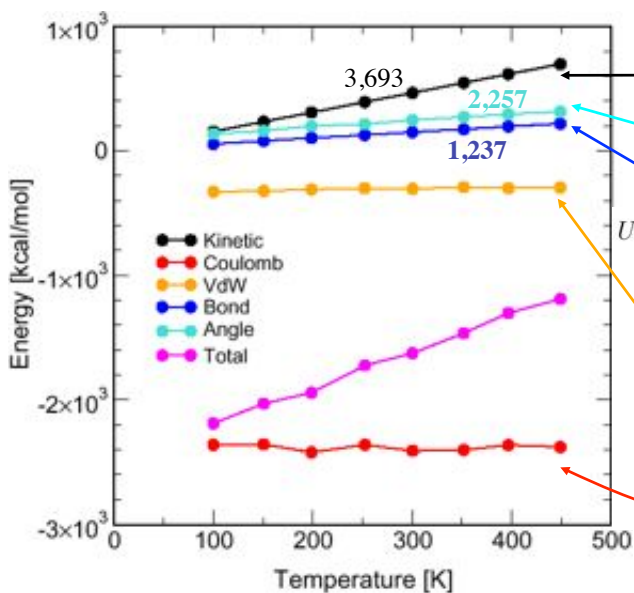
Kinetic energy (quadratic)

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

Potential energy (not all quadratic)

Equilibrium Properties of Proteins

Energies: kinetic and potential



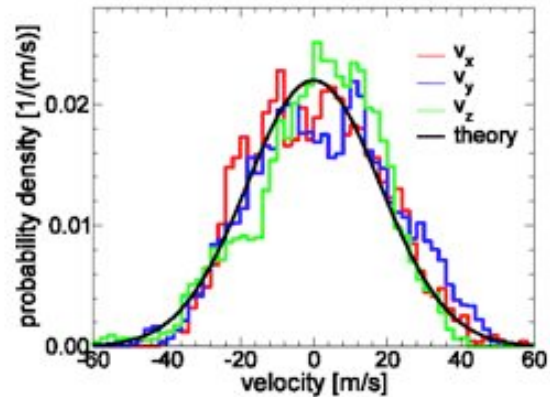
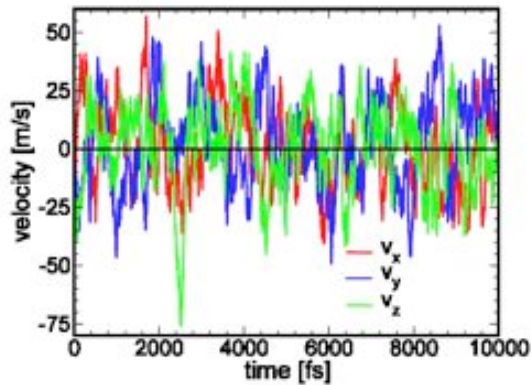
$$\left\langle \sum_j \frac{1}{2} m_j v_j^2 \right\rangle = \frac{3}{2} N k_B T$$

Kinetic energy (quadratic)

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

Potential energy (not all quadratic)

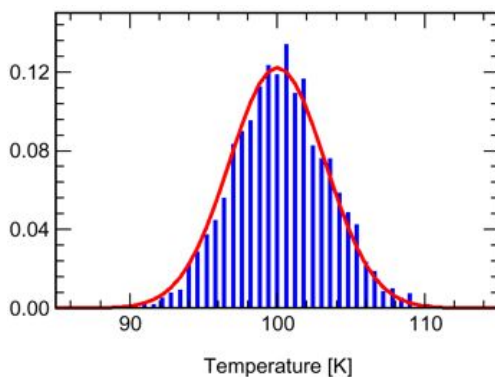
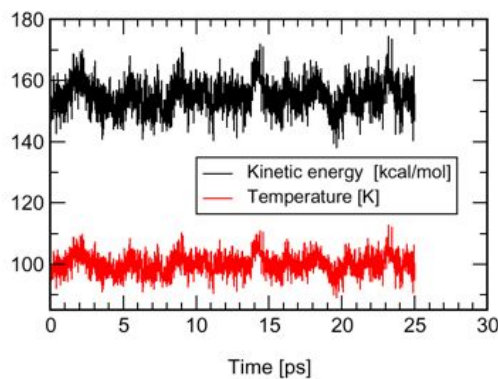
Maxwell Distribution of Atomic Velocities



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

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

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



Definition of Temperature

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

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

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

Temperatur Fluctuations

Maxwell distribution

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

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

$$dP(\epsilon_n) = (\pi T_0 \epsilon_n)^{-1/2} \exp(-\epsilon_n/k_B T_0) d\epsilon_n \quad (8)$$

One can derive (temperature T_0 in units k_B)

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

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

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

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

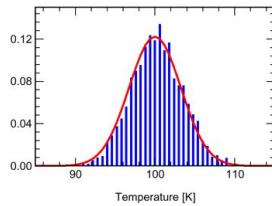
$$P(E_{kin}) = c \exp\left(\frac{-(E_{kin} - \langle E_{kin} \rangle)^2}{2 \left(\frac{3Nk_B^2 T_0^2}{2}\right)}\right) \quad (12)$$

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

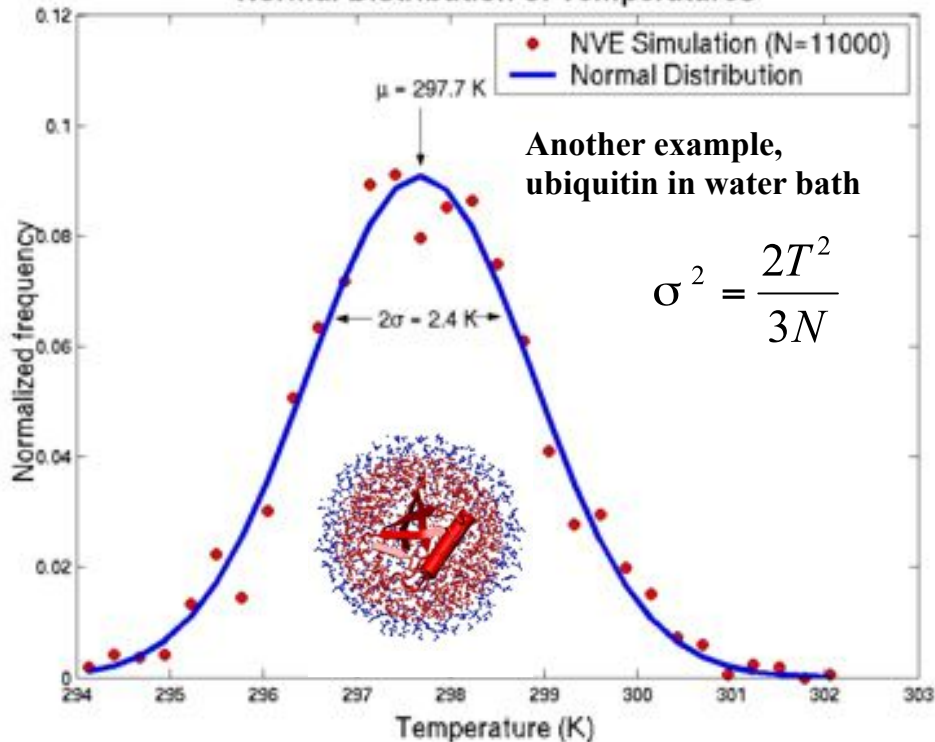
$$P(\Delta T) = c \exp[-(\Delta T)^2/2\sigma^2], \quad \sigma^2 = 2T^2/3N \quad (13)$$

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

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

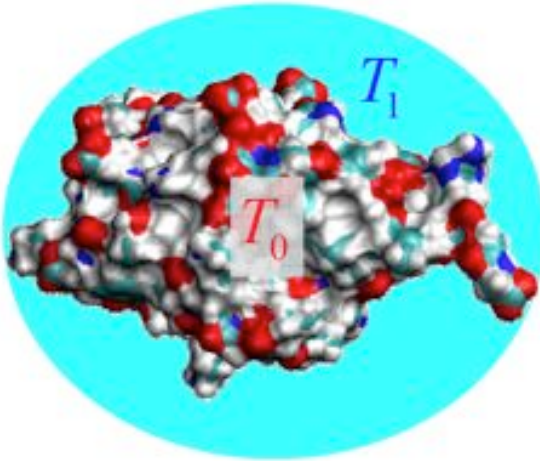


Normal Distribution of Temperatures



Simulated Cooling of Ubiquitin

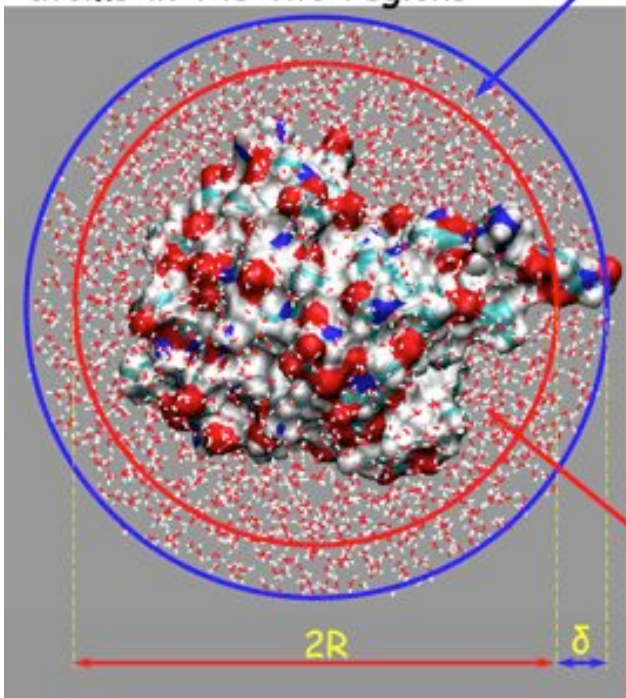
- Proteins function in a narrow (physiological) temperature range. What happens to them when the temperature of their surrounding changes significantly (temperature gradient) ?
- Can the heating/cooling process of a protein be simulated by molecular dynamics ? If yes, then how?



- What can we learn from the simulated cooling/heating of a protein ?

How to simulate cooling ?

Heat transfer through mechanical coupling between atoms in the two regions



coolant layer of atoms

motion of atoms is subject to stochastic Langevin dynamics

$$m \ddot{\mathbf{r}} = \mathbf{F}_{FF} + \mathbf{F}_H + \mathbf{F}_f + \mathbf{F}_L$$

\mathbf{F}_{FF} → force field

\mathbf{F}_H → harmonic restrain

\mathbf{F}_f → friction

\mathbf{F}_L → Langevin force

atoms in the inner region follow Newtonian dynamics

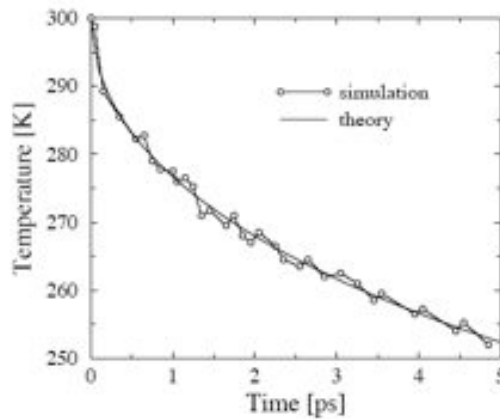
$$m \ddot{\mathbf{r}} = \mathbf{F}_{FF}$$

Simulated Cooling - Result

t	$\langle T_{sim} \rangle$	t	$\langle T_{sim} \rangle$	t	$\langle T_{sim} \rangle$	t	$\langle T_{sim} \rangle$
0.05	298.75	1.05	276.00	1.95	267.00	3.25	261.00
0.15	289.25	1.15	276.50	2.05	268.50	3.45	258.50
0.35	285.50	1.25	275.25	2.25	266.50	3.55	259.50
0.55	282.25	1.35	271.00	2.35	264.50	3.95	256.50
0.65	282.75	1.45	271.75	2.55	263.50	4.05	257.25
0.75	279.00	1.65	269.50	2.65	264.50	4.45	254.00
0.85	277.75	1.75	271.00	2.85	262.00	4.55	255.25
1.00	277.50	1.85	268.00	3.05	262.50	4.85	252.00

Result from simulation

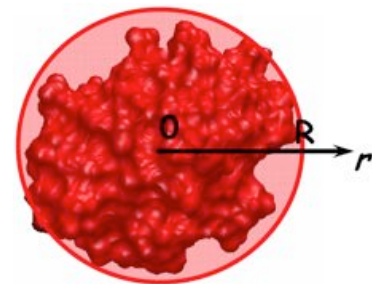
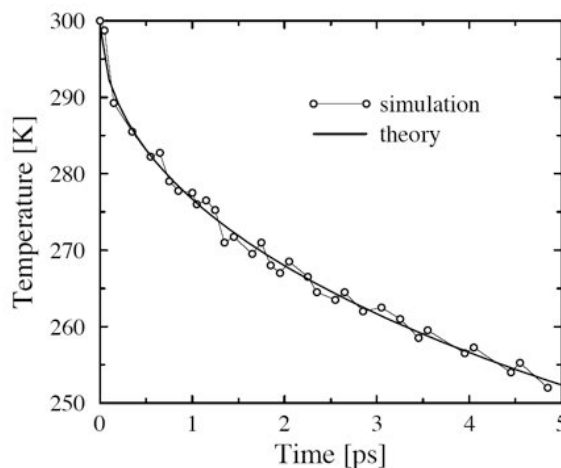
Table 1: Mean temperature $\langle T_{sim} \rangle$ [K] of the protein as a function of time t [ps].



Solution of the Heat Equation

Temperature averaged over volume

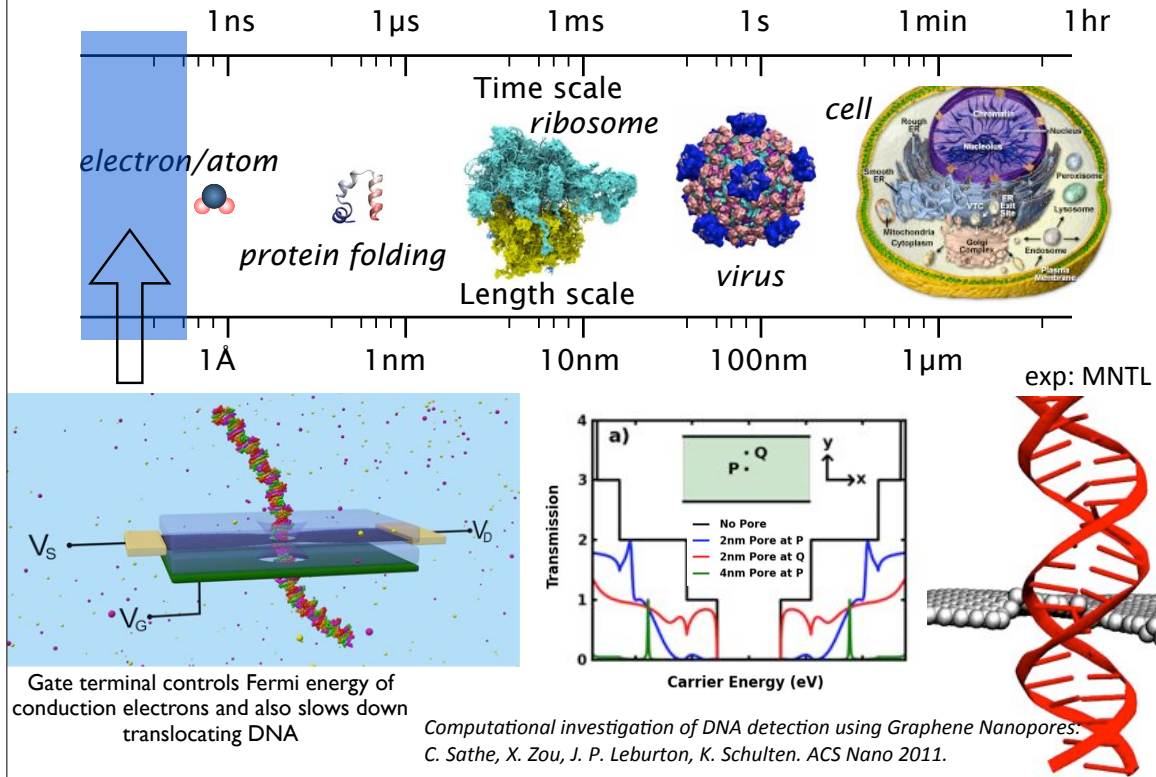
$$\begin{aligned}
 \langle T \rangle(t) &= \left(\frac{4\pi R^3}{3} \right)^{-1} \int d^3\mathbf{r} T(\mathbf{r}, t) = \frac{3}{R^3} \int_0^R r^2 dr T(r, t) \\
 &= T_{bath} + \sum_{n=1}^{\infty} a_n \exp \left[- \left(\frac{n\pi}{R} \right)^2 D t \right] \frac{3}{R^3} \int_0^R r dr \sin \left(\frac{n\pi r}{R} \right) \\
 &= T_{bath} + 6 \frac{\Delta T}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \exp \left[- \left(\frac{n\pi}{R} \right)^2 D t \right]
 \end{aligned}$$



$$D \approx 0.38 \times 10^{-3} \text{cm}^2 \text{s}^{-1}$$

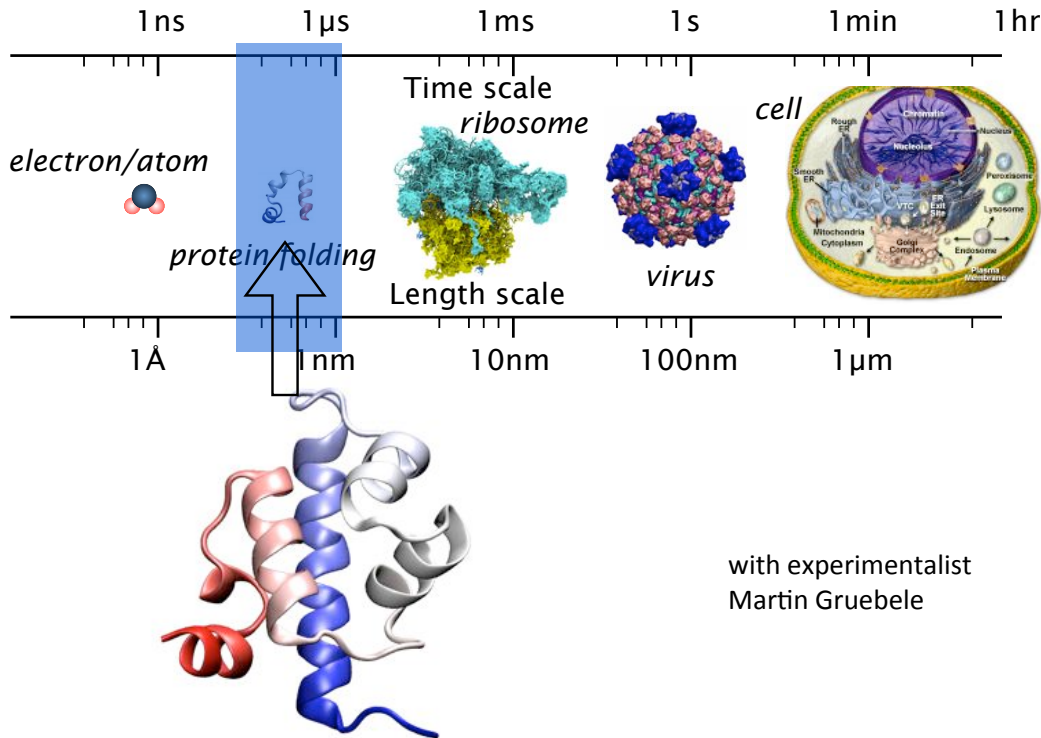
water $1.4 \times 10^{-3} \text{cm}^2 \text{s}^{-1}$

Increasing Biological Realism in Theory and Computation



Aksimentiev: *Advanced Functional Materials* 21:1040-1050 (2011); *ACS Nano*, 5: 9345–9353 (2011); *Nano Letters* 12: 1038–1044 (2012).
Schulten: *ACS Nano* 5: 8842–8851 (2011); *Nucleic Acids Res.* 39: 8740–8751 (2011); *ACS Nano* 6: 8847–8856 (2012); *Scientific Reports* 3: 1389 - 1397 (2013).

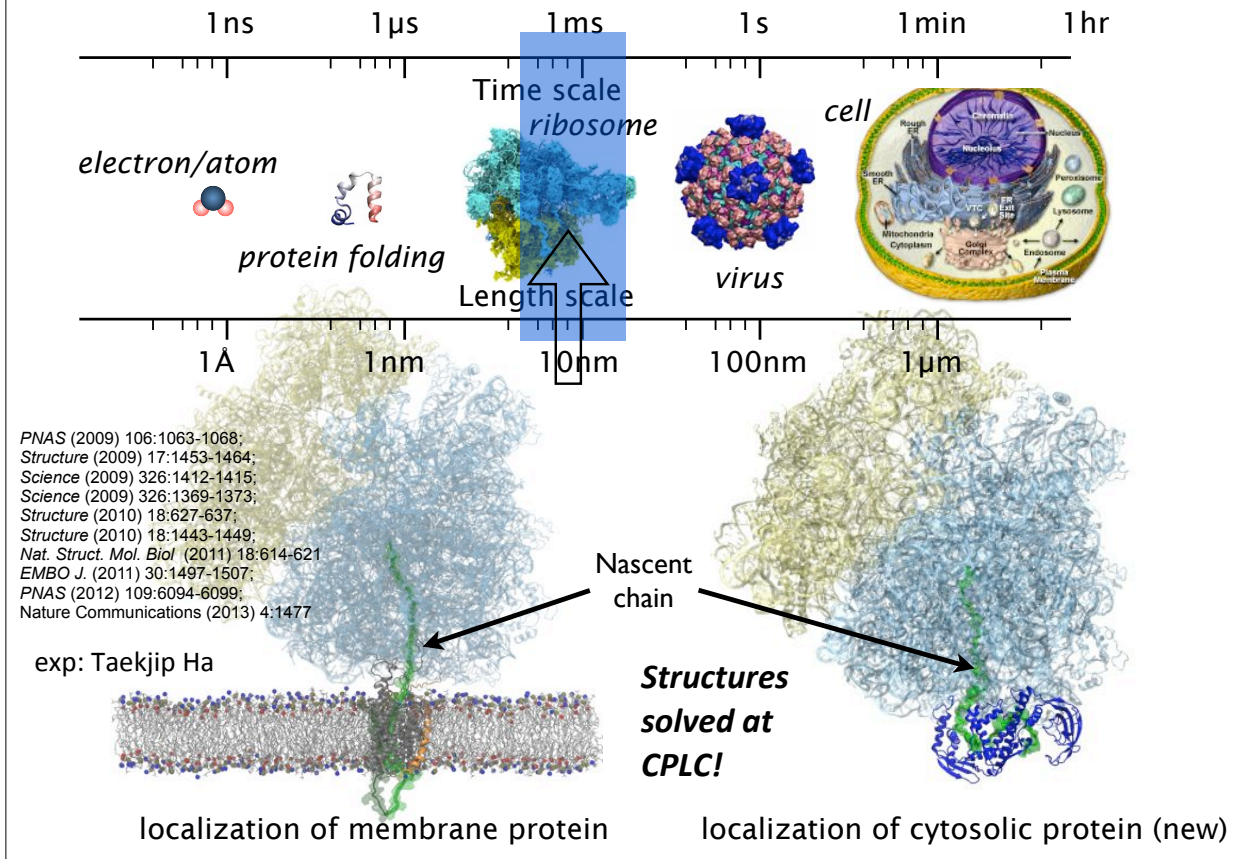
Increasing Biological Realism in Theory and Computation



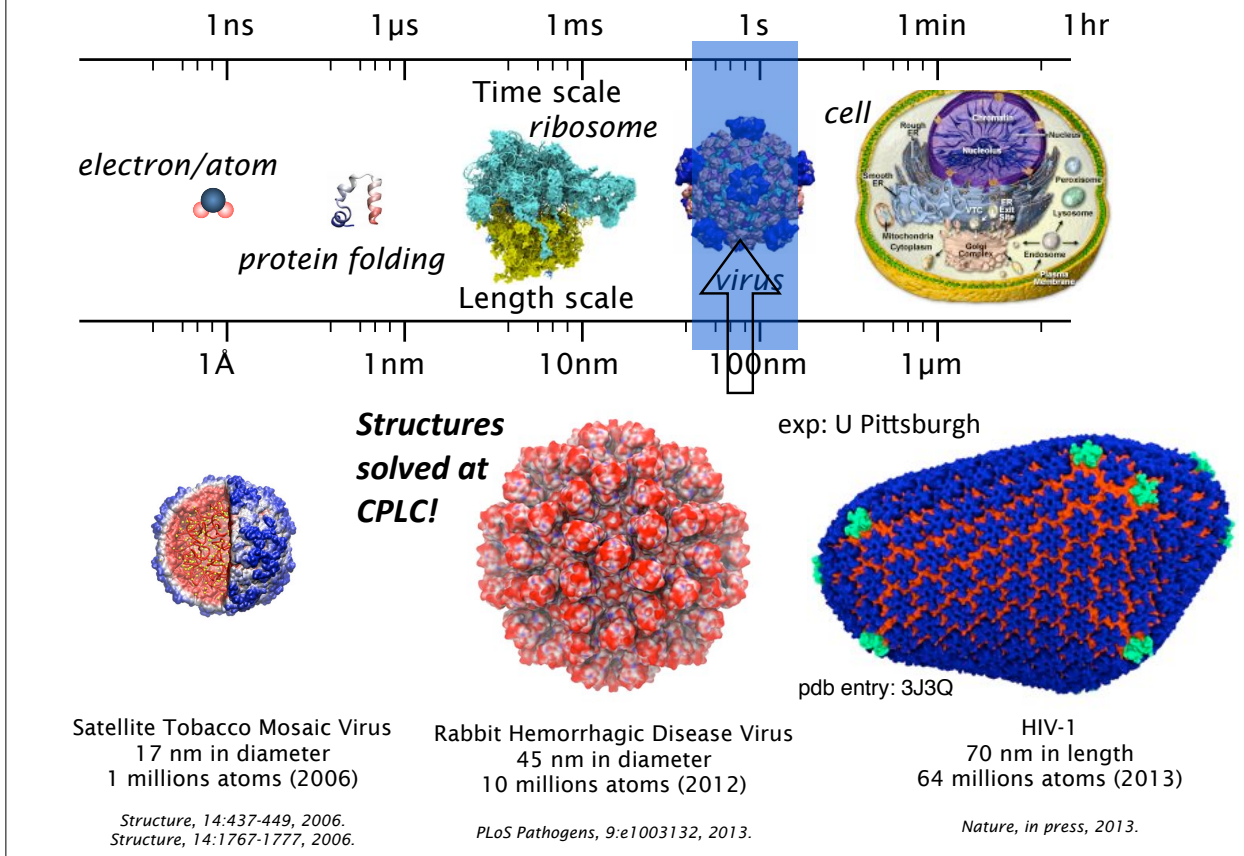
lambda repressor, 80 amino acids, 100 μ s

Proceedings of the National Academy of Sciences, USA, In press, 2013.

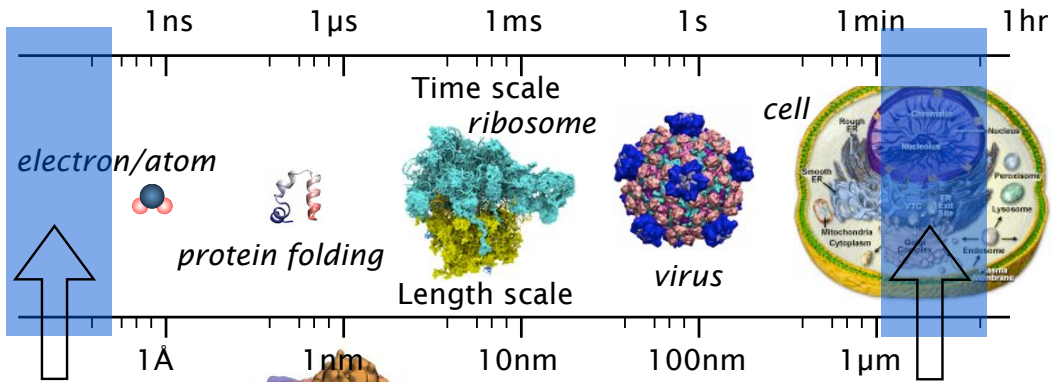
Increasing Biological Realism in Theory and Computation



Increasing Biological Realism in Theory and Computation



Increasing Biological Realism in Theory and Computation



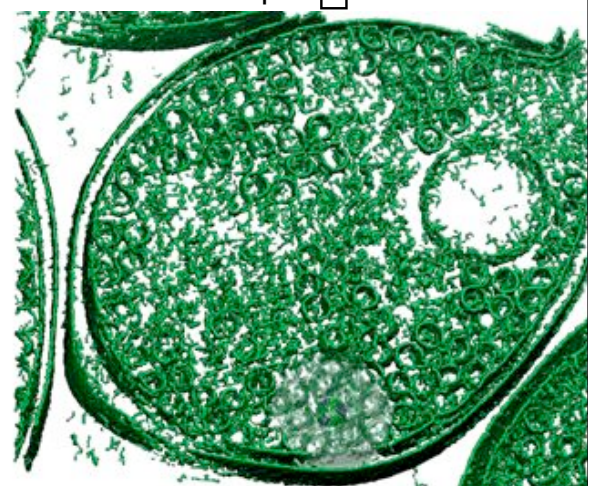
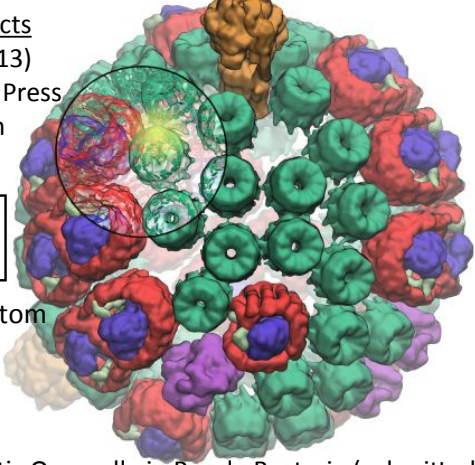
Quantum Effects
in Biology (2013)

Cambridge U. Press
+ 10 articles in
2011-2013

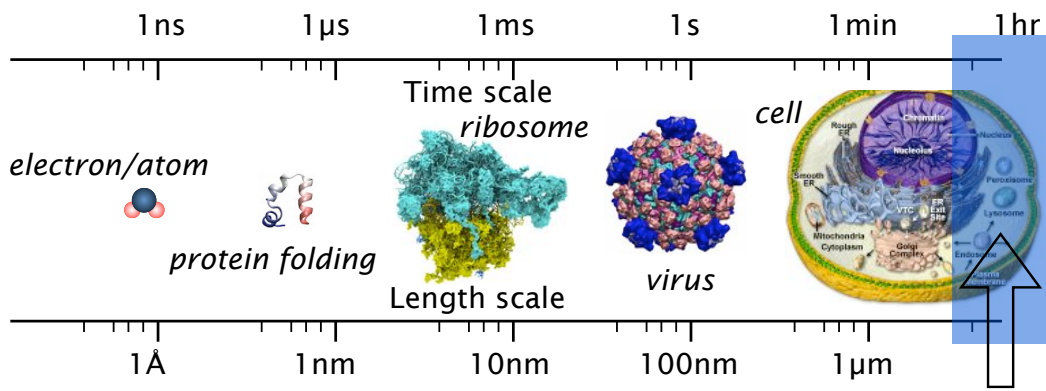
exp:
several labs

100 million atom
simulation
Blue Waters
(Oct. 2012)

Photosynthetic Organelle in Purple Bacteria (submitted)



Increasing Biological Realism in Theory and Computation



whole cell
simulation

Luthey-Schulten,
presents later

Theoretical and Computational Biophysics Group Developers



- focus on systems biology
- theoretical biophysics
- develops renewable energy
- focus on quantum biology
- computational biophysics
- guides bionanotechnology