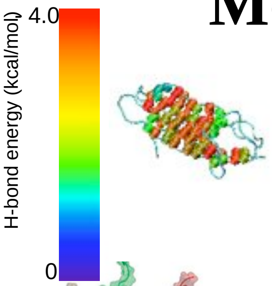
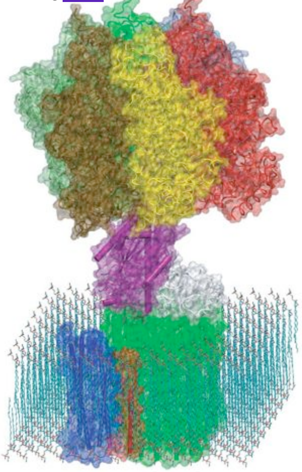
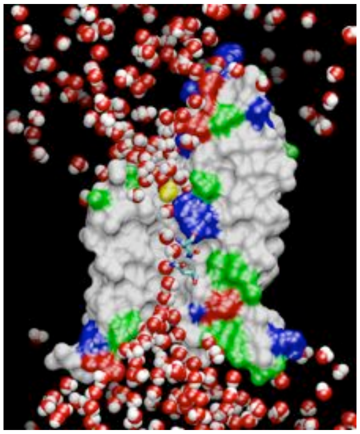
Molecular Dynamics of Proteins



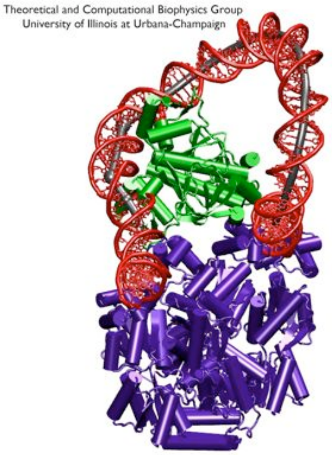
Fibronectin III_1, a mechanical protein that glues cells together in wound healing and in preventing tumor metastasis



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



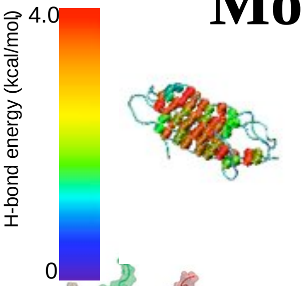
AQP filtering a bath tub of the body's water a day



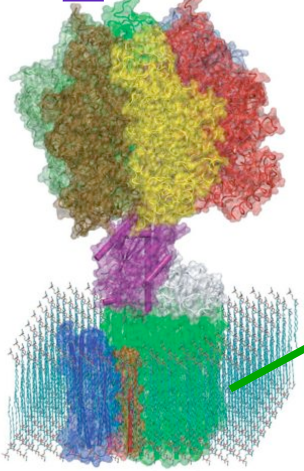
A ternary complex of DNA, lac repressor, and CAP controlling gene expression

---

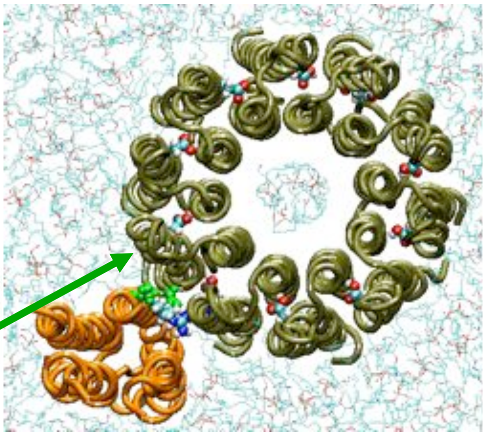
Molecular Dynamics of Proteins

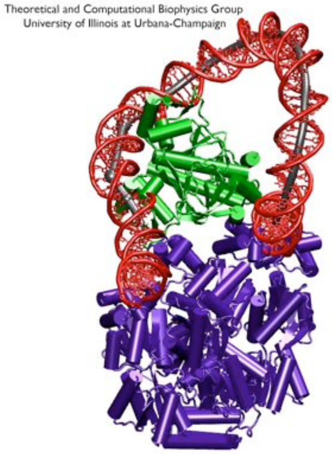


Fibronectin III_1, a mechanical protein that glues cells together in wound healing and in preventing tumor metastasis



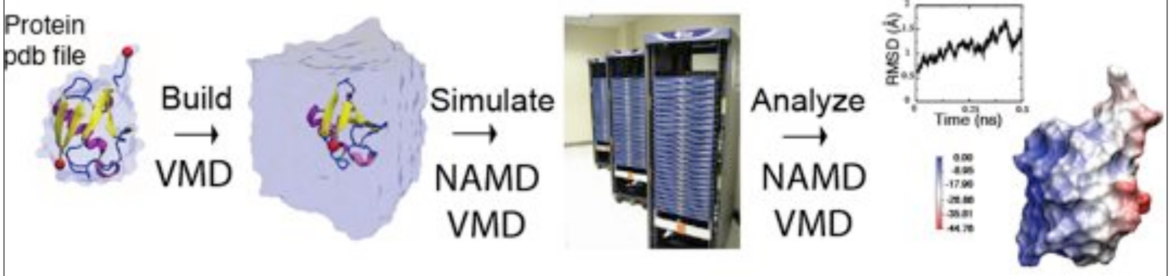
ATPase, a molecular motor that synthesizes the body's weight of ATP a day





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

The Molecular Dynamics Simulation Process



Classical Dynamics

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 2^{nd}-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

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.

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

+

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

“Verlet algorithm”

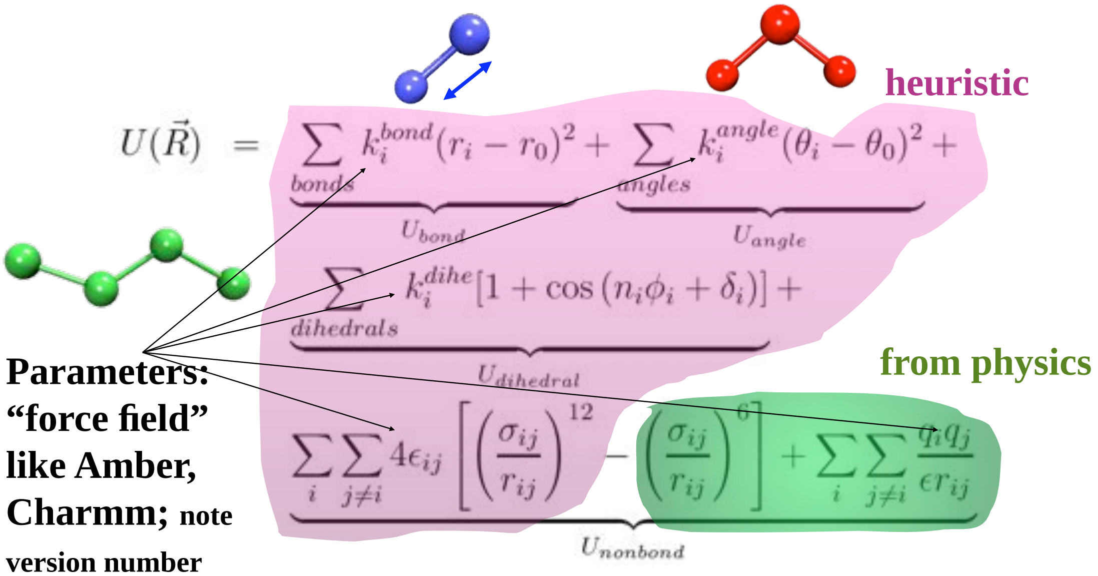
⬇

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

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

Potential Energy Function of Biopolymer

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



heuristic

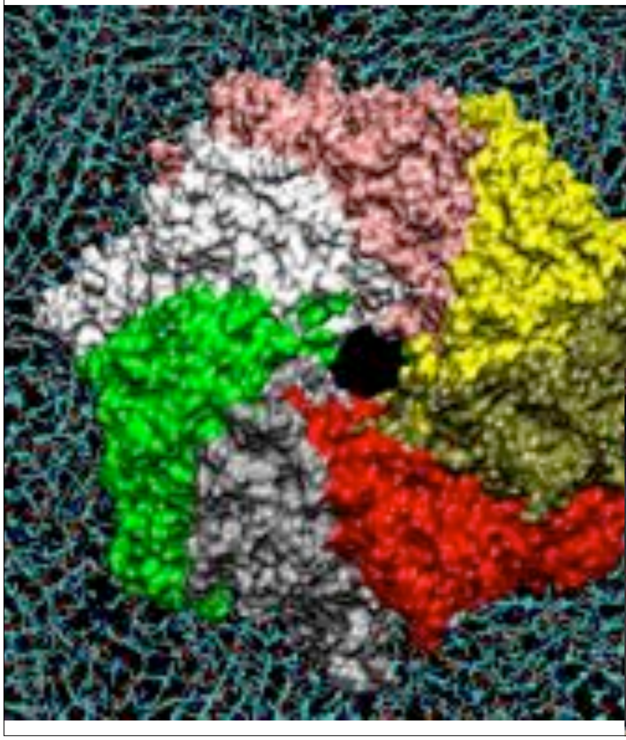
$$U(\vec{R}) = \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} + \underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} + \underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}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

from physics

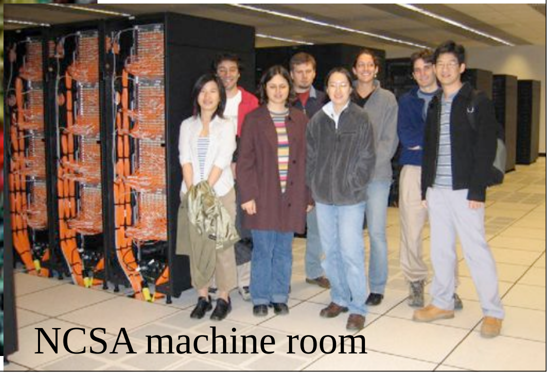
Parameters: "force field" like Amber, Charmm; note version number

---

Large is no problem. But …



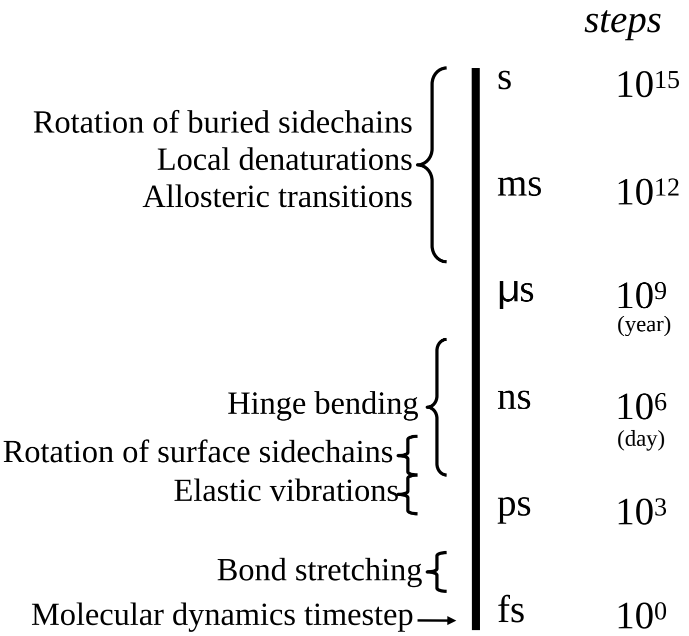
Molecular dynamics simulation of alpha-hemolysin with about 300,000 atoms

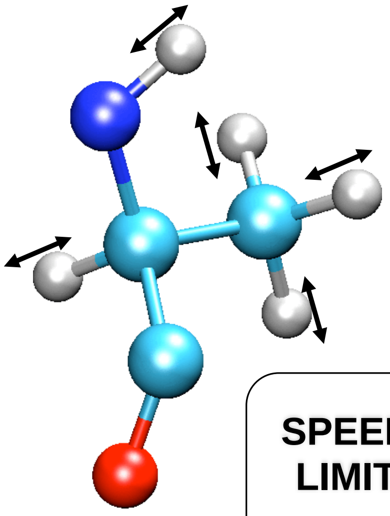


NCSA machine room

But long is!

biomolecular timescale and timestep limits





SPEED LIMIT

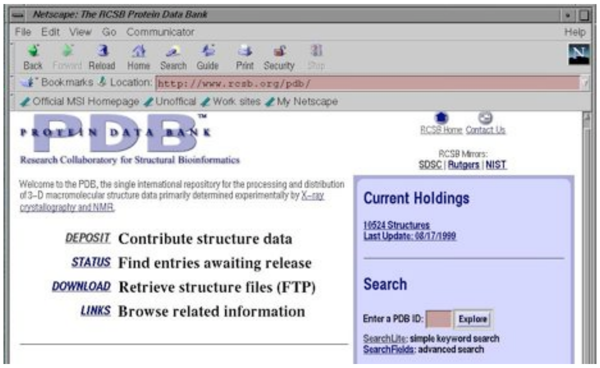
δt = 1 fs

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.)



Research Collaboratory for Structural Bioinformatics

Welcome to the PDB, the single international repository for the processing and distribution of 3-D macromolecular structure data primarily determined experimentally by X-ray crystallography and NMR.

DEPOSIT Contribute structure data

STATUS Find entries awaiting release

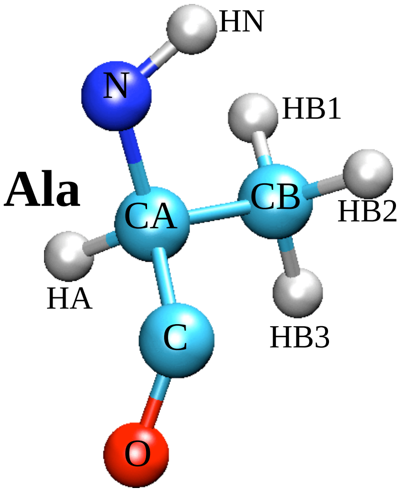
DOWNLOAD Retrieve structure files (FTP)

LINKS Browse related information

10524 Structures
Last Update: 08/17/1999

SearchLite: simple keyword search
SearchFields: advanced search

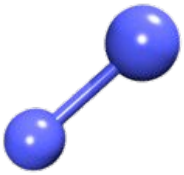
PSF Files



- 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)

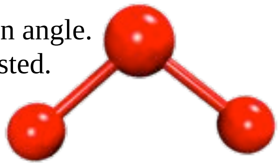
PSF Files

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.

Preparing Your System for MD

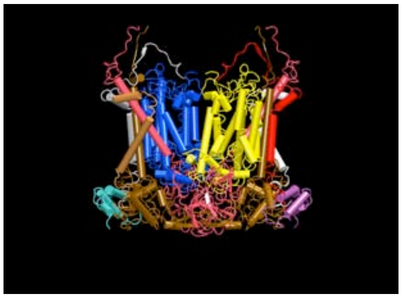
Solvation

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).

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

mitochondrial bc1 complex



Preparing Your System for MD

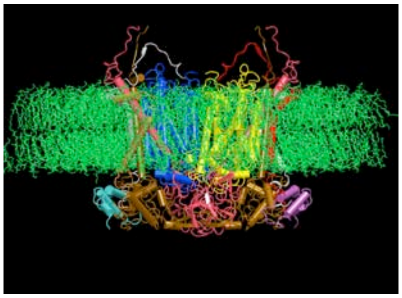
Solvation

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).

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

mitochondrial bc1 complex



Preparing Your System for MD

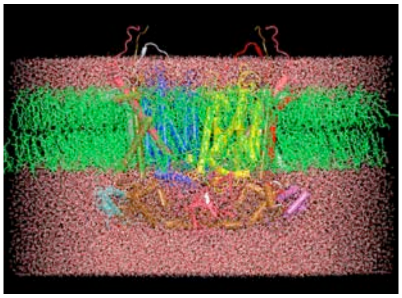
Solvation

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).

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

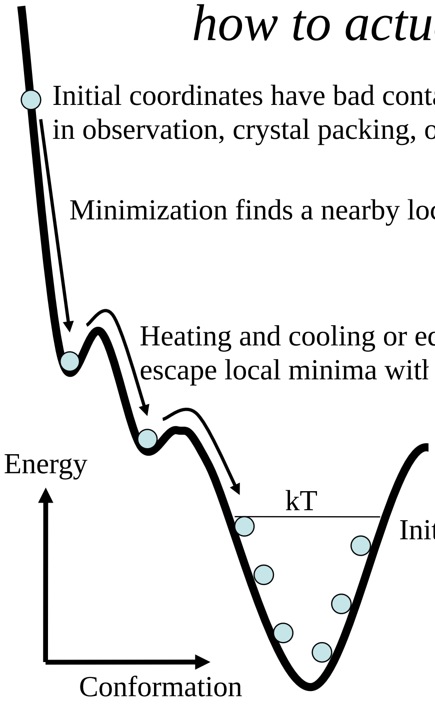
mitochondrial bc1 complex



(Usually periodic! Avoids surface effects)

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)

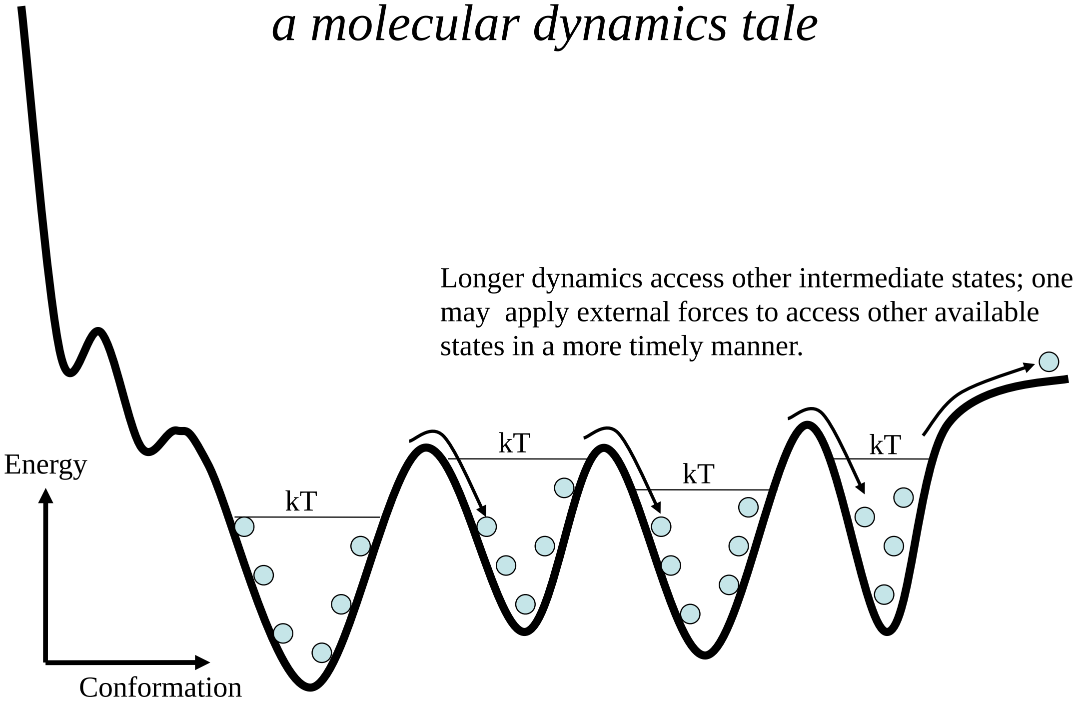
Minimization finds a nearby local minimum.

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

Initial dynamics samples thermally accessible states.

From the Mountains to the Valleys

a molecular dynamics tale



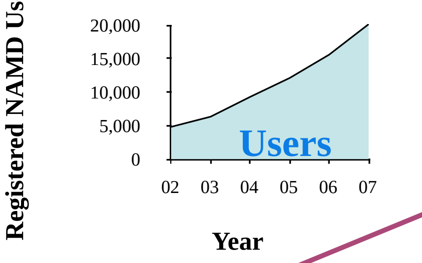
Longer dynamics access other intermediate states; one may apply external forces to access other available states in a more timely manner.

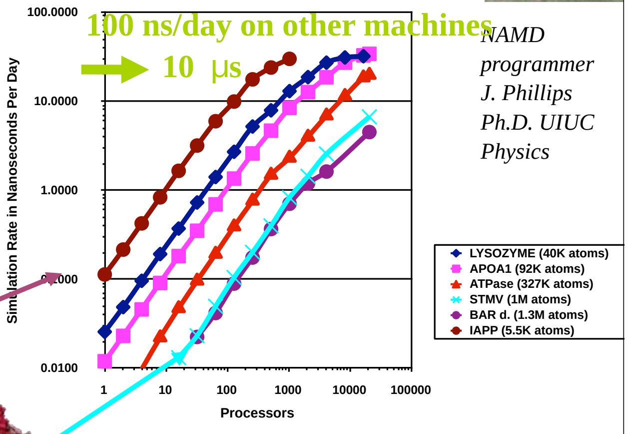
NAMD: The Program we will Use

NAMD programmer J. Phillips Ph.D. UIUC Physics

Development cost 1990 - 2007: $20 million

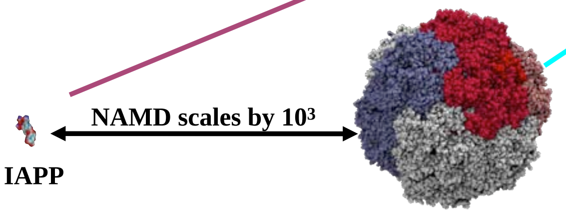
NAMD Registrants
19,995 Registrants (3336 NIH)
4,111 Repeat Users
NAMD 2.6 released Aug 2006
4181 NAMD 2.6 users (742 NIH)





100 ns/day on other machines

10 μs



STMV

"We haven't found a hard limit on scaling up the number of processors."

-- Philip Blood and Greg Voth, Univ Utah

Commenting on NAMD performance for the PSC XT3 Cray

National Center for Research Resources

Molecular Dynamics Ensembles

Constant energy, constant number of particles (NE)

Constant energy, constant volume (NVE)

Constant temperature, constant volume (NVT)

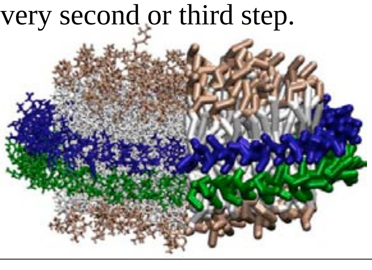
Constant temperature, constant pressure (NPT)

Choose the ensemble that best fits your system and start the simulations, but use NE to check on accuracy of the simulation.

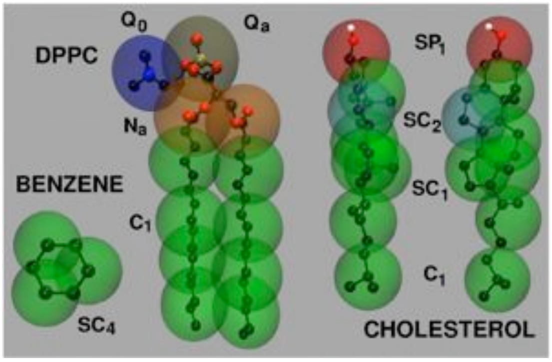
Cutting Corners

cutoffs, PME, rigid bonds, and multiple timesteps

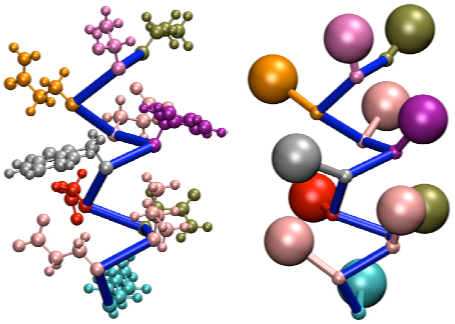
- Nonbonded interactions require order N^2 computer time!
  - Truncating at R_{cutoff} reduces this to order $N R_{cutoff}^3$
  - Particle mesh Ewald (PME) method adds long range electrostatics at order N log N, only minor cost compared to cutoff calculation.
- Can we extend the timestep, and do this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



Residue-Based Coarse-Grained Model



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



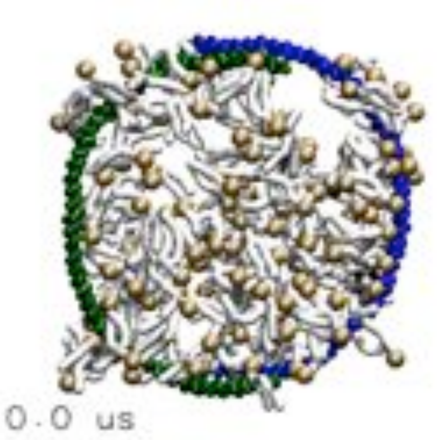
All-atom peptide CG peptide

- 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

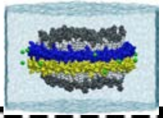
Peter L. Freddolino, Anton Arkhipov, Amy Y. Shih, Ying Yin, Zhongzhou Chen, and Klaus Schulten. **Application of residue-based and shape-based coarse graining to biomolecular simulations.** In Gregory A. Voth, editor, *Coarse-Graining of Condensed Phase and Biomolecular Systems*, chapter 20, pp. 299-315.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

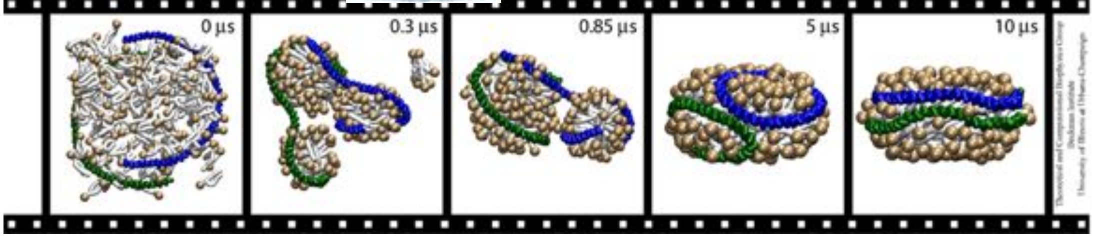
Nanodisc Assembly CG MD Simulation

- 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



Fully hydrated

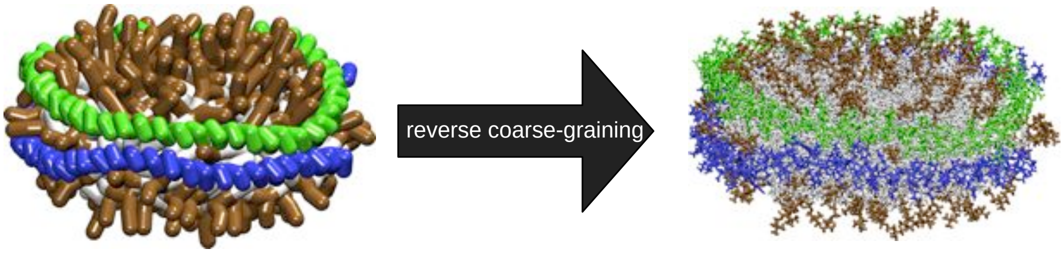




A. Shih, A. Arkhipov, P. Freddolino, and K. Schulten. *J. Phys. Chem. B*, 110:3674–3684, 2006; A. Shih, P. Freddolino, A. Arkhipov, and K. Schulten. *J. Struct. Biol.*, 157:579–592,2007; A. Shih, A. Arkhipov, P. Freddolino, S. Sligar, and K. Schulten. *Journal of Physical Chemistry B*, 111: 11095 - 11104, 2007; A. Shih, P. Freddolino, S. Sligar, and K. Schulten. *Nano Letters*, 7:1692-1696, 2007.

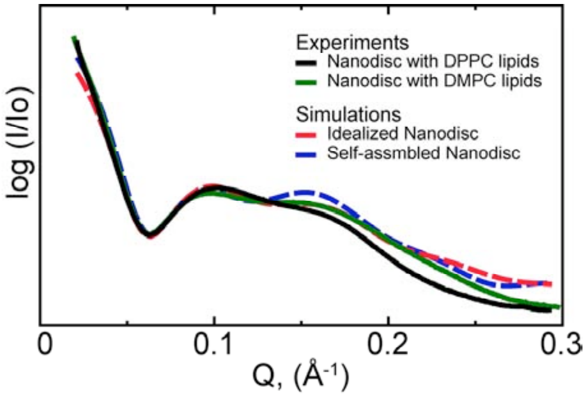
Validation of Simulations

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

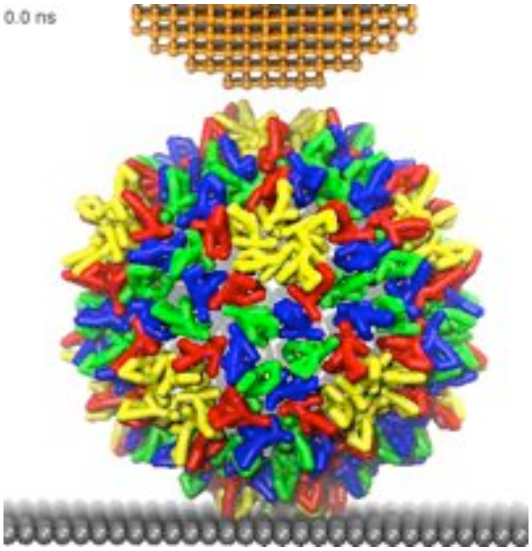


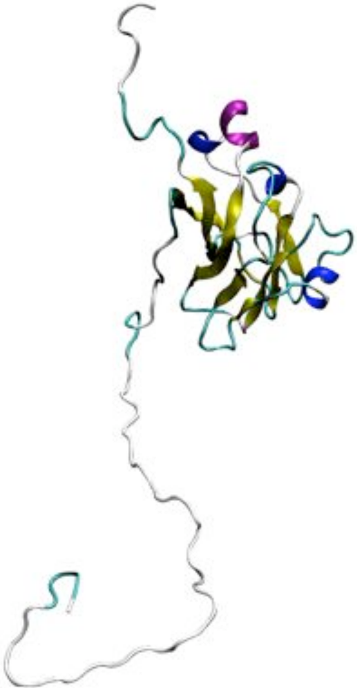
Reverse coarse-graining:
1. Map center of mass of the group of atoms represented by a single CG bead to that beads location
2. MD minimization, simulated annealing with restraints, and equilibration to get all-atom structure

Small-angle X-ray scattering:
Calculated from reverse coarse-grained all-atom model and compared with experimental measurements



Shape-Based Coarse-Grained (CG) model

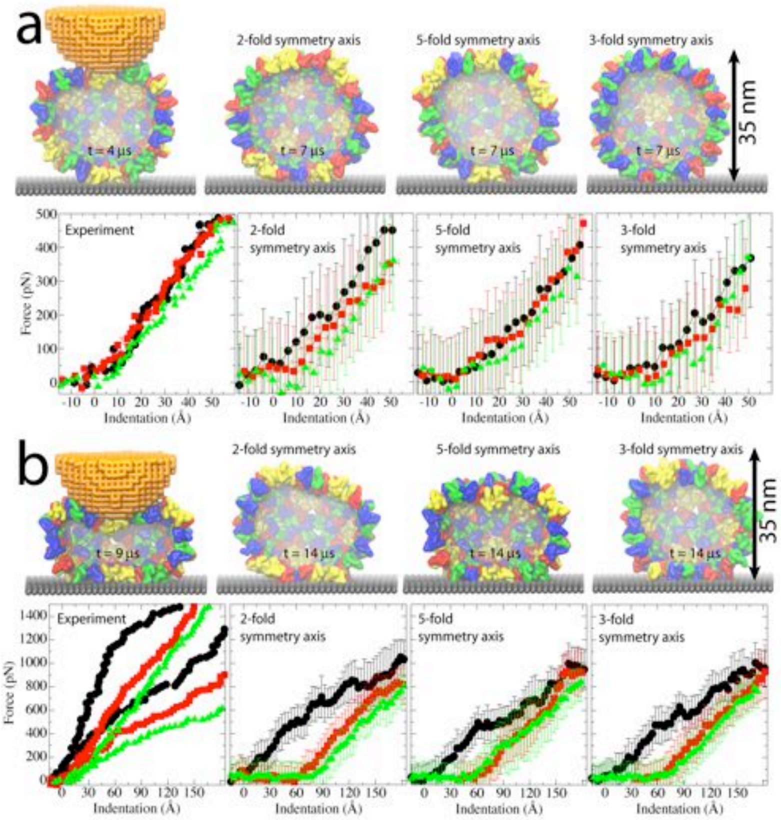




- **Fully automatic**
- **Number of CG beads is chosen by a user (we used ~200 atoms per CG bead)**

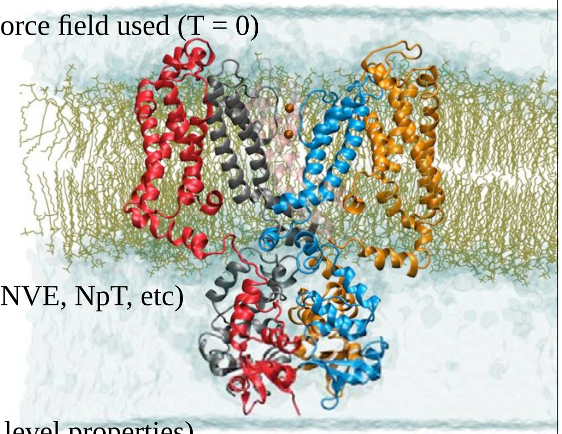
Peter L. Freddolino, Anton Arkhipov, Amy Y. Shih, Ying Yin, Zhongzhou Chen, and Klaus Schulten. **Application of residue-based and shape-based coarse graining to biomolecular simulations.** In Gregory A. Voth, editor, *Coarse-Graining of Condensed Phase and Biomolecular Systems*, chapter 20, pp. 299-315.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

Reversible and irreversible indentations



Steps in a Typical MD Simulation

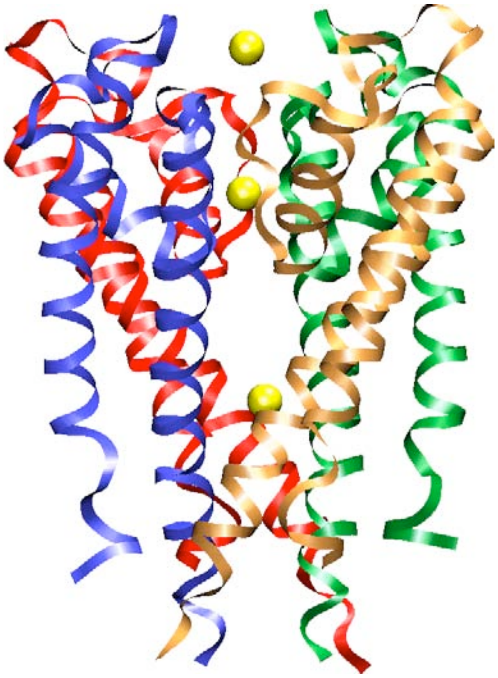
- 1. Prepare molecule
  - Read in pdb and psf file
  - Usually requires setting up the system, e.g., solvation
  - Many tools available in VMD
- 2. Minimization
  - Reconcile observed structure with force field used (T = 0)
- 3. Heating
  - Raise temperature of the system
- 4. Equilibration
  - Ensure system is stable
- 5. Dynamics
  - Simulate under desired conditions (NVE, NpT, etc)
  - Collect your data
- 6. Analysis
  - Evaluate observables (macroscopic level properties)
  - Or relate to single molecule experiments



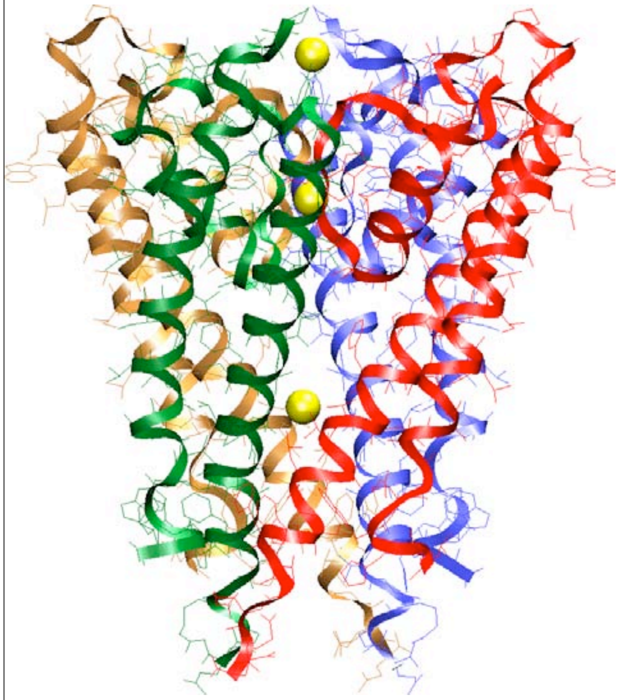
Example: MD Simulations of the K^+ Channel Protein

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.

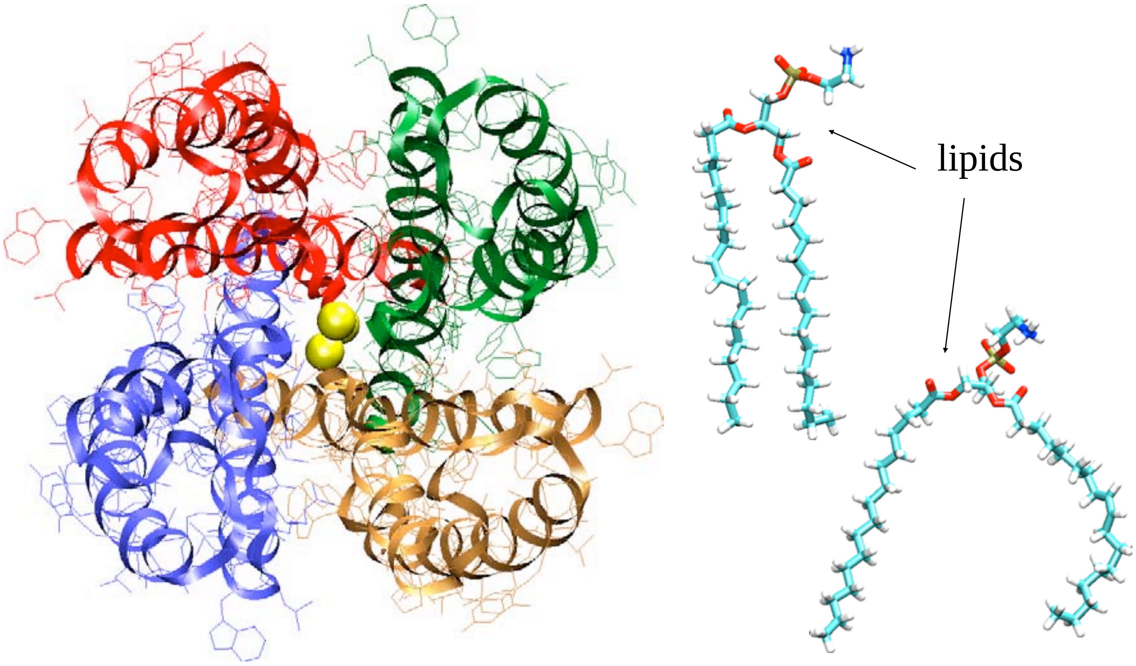


Setting up the system (1)



- 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
- minimize the protein structure using NAMD2

Setting up the system (2)

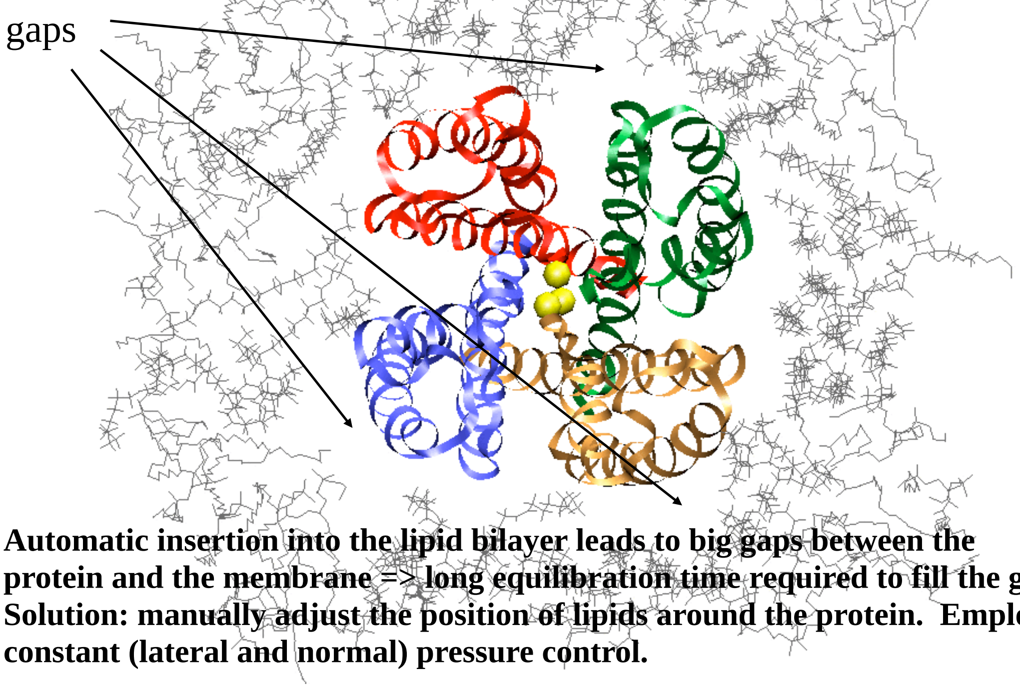


Simulate the protein in its natural environment: solvated lipid bilayer

---

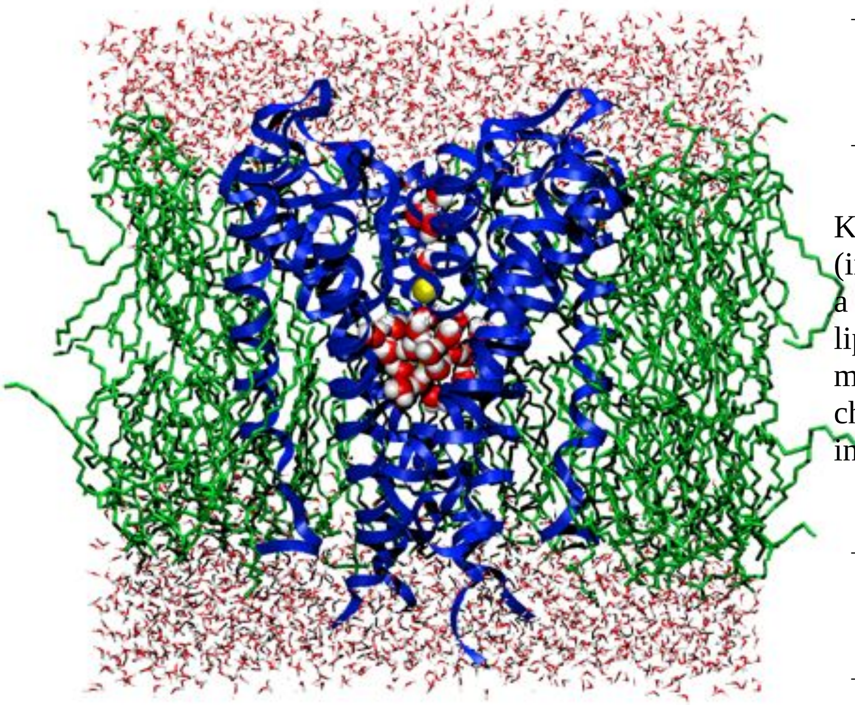
Setting up the system (3)

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

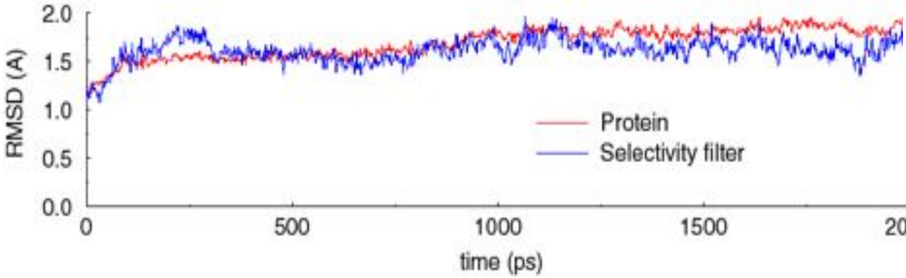
Summary of simulations:

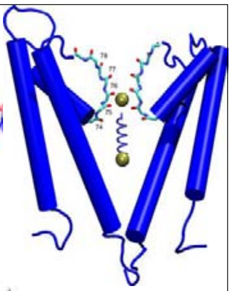
- protein/membrane system contains 38,112 atoms, including 5117 water molecules, 100 POPE and 34 POPG lipids, plus K^+ counterions
- CHARMM26 forcefield
- periodic boundary conditions, PME electrostatics
- 1 ns equilibration at 310K, NpT
- 2 ns dynamics, NpT

Program: NAMD2

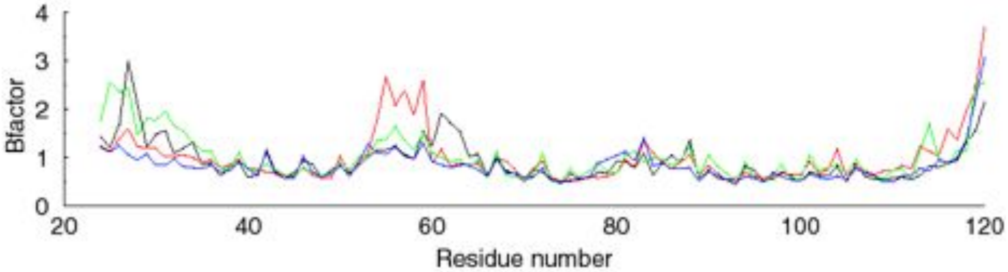
Platform: Cray T3E (Pittsburgh Supercomputer Center) or local computer cluster; choose ~1000 atoms per processor.

MD Results



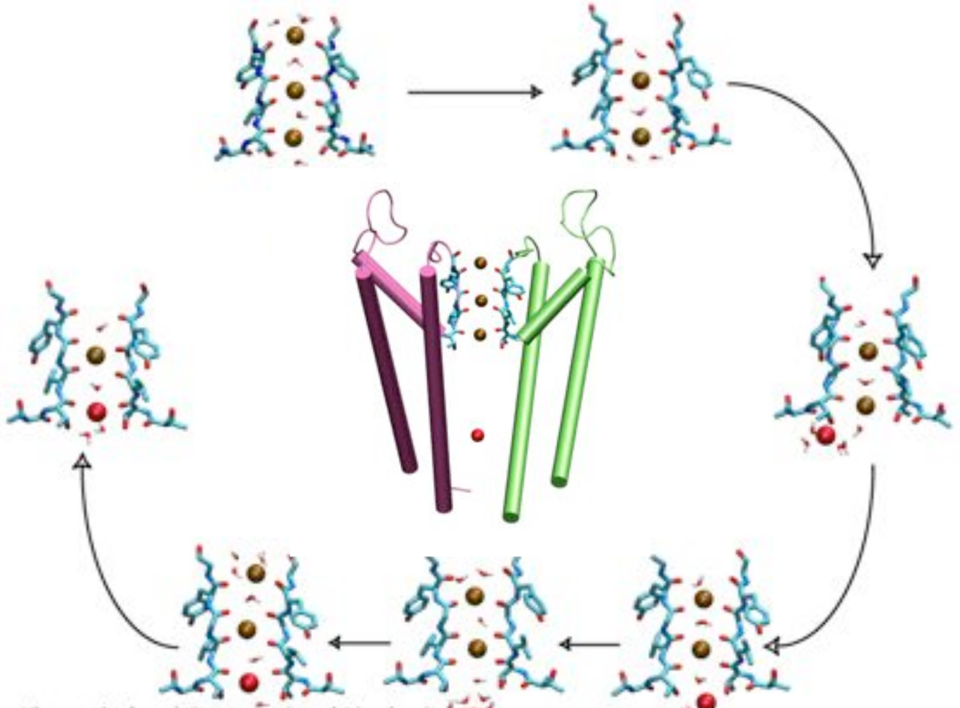


RMS deviations for the KcsA protein and its selectivity filer indicate that the protein is stable during the simulation with the selectivity filter the most stable part of the system.



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)



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

Theoretical and Computational Biophysics Group Developers

Funding: NIH, NSF

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