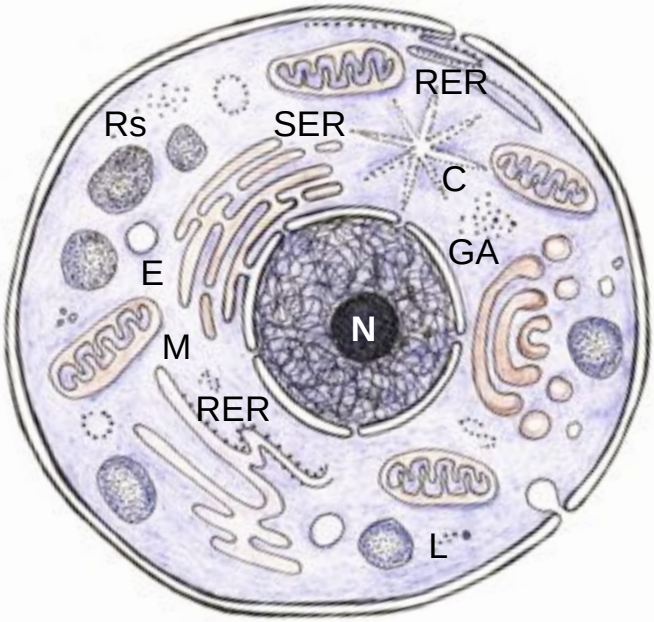
Computational microscope views at atomic resolution ...



... how living cells maintain health and battle disease

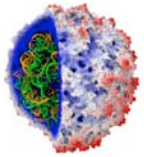
John Stone

Our Microscope is Made of...

Chemistry

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

NAMD Software



Virus

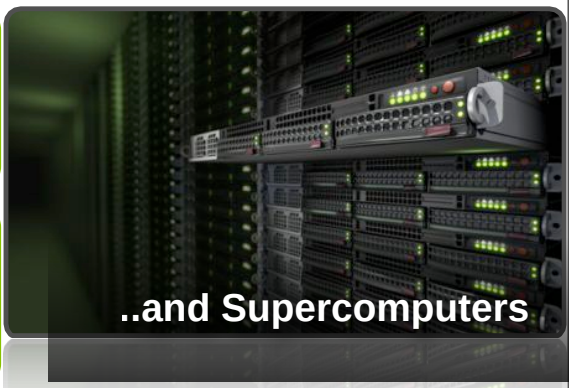
Physics

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

Math

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{\Delta t^2}{m_i}\vec{F}_i(t)$$

(repeat ***one billion times*** = microsecond)



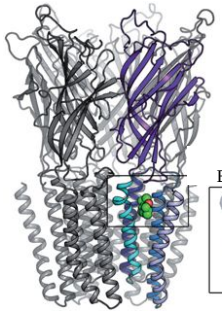
NAMD impact is broad and deep

- Comprehensive, industrial-quality software
  - Integrated with VMD for simulation setup and analysis
  - Portable extensibility through Tcl scripts (also used in VMD)
  - Consistent user experience from laptop to supercomputer
- Large user base – 51,000 registered users
  - 9,100 (18%) are NIH-funded; many in other countries
  - 14,100 have downloaded more than one version
- Leading-edge simulations
  - "most-used software" on NICS Cray XT5 (largest NSF machine)
  - "by far the most used MD package" at TACC (2nd and 3rd largest)
  - NCSA Blue Waters early science projects and acceptance test
  - Argonne Blue Gene/Q early science project

Outside researchers choose NAMD and succeed

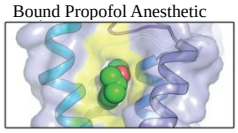
2100 external citations since 2007

Corringer, et al., *Nature*, 2011

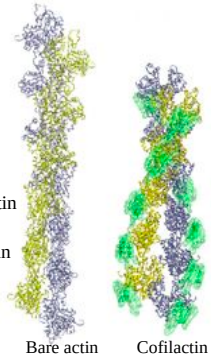


180K-atom 30 ns study of anesthetic binding to bacterial ligand-gated ion channel provided "complementary interpretations…that could not have been deduced from the static structure alone."

Bound Propofol Anesthetic



Voth, et al., *PNAS*, 2010



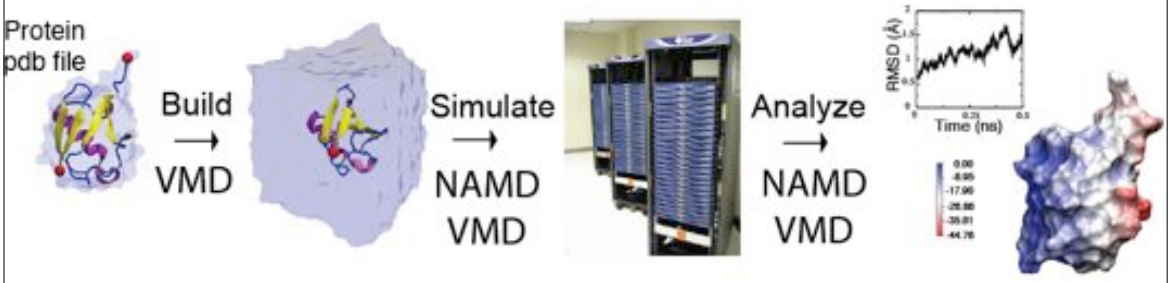
500K-atom 500 ns investigation of effect of actin depolymerization factor/cofilin on mechanical properties and conformational dynamics of actin filament.

Bare actin Cofilactin

Recent NAMD Simulations in *Nature*

- **M. Koeksal, et al.,** *Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis*. (2011)
- **C.-C. Su, et al.**, *Crystal structure of the CusBA heavy-metal efflux complex of Escherichia coli.* (2011)
- **D. Slade, et al.**, *The structure and catalytic mechanism of a poly(ADP-ribose) glycohydrolase*. (2011)
- **F. Rose, et al.,** *Mechanism of copper(II)-induced misfolding of Parkinson's disease protein*. (2011)
- **L. G. Cuello, et al.**, *Structural basis for the coupling between activation and inactivation gates in K(+) channels*. (2010)
- **S. Dang, et al.,**, *Structure of a fucose transporter in an outward-open conformation.* (2010)
- **F. Long, et al.**, *Crystal structures of the CusA efflux pump suggest methionine-mediated metal transport*. (2010)
- **R. H. P. Law, et al.**, *The structural basis for membrane binding and pore formation by lymphocyte perforin.* (2010)
- **P. Dalhaimer and T. D. Pollard**, *Molecular Dynamics Simulations of Arp2/3 Complex Activation.* (2010)
- **J. A. Tainer, et al.**, *Recognition of the Ring-Opened State of Proliferating Cell Nuclear Antigen by Replication Factor C Promotes Eukaryotic Clamp-Loading*. (2010)
- **D. Krepkiy, et al.,**, *Structure and hydration of membranes embedded with voltage-sensing domains*. (2009)
- **N. Yeung, et al.,**, *Rational design of a structural and functional nitric oxide reductase*. (2009)
- **Z. Xia, et al.**, *Recognition Mechanism of siRNA by Viral p19 Suppressor of RNA Silencing: A Molecular Dynamics Study*. (2009)

The Molecular Dynamics Simulation Process



For textbooks see:

M.P. Allen and D.J. Tildesley. ***Computer Simulation of Liquids.***Oxford University Press, New York, 1987.
D. Frenkel and B. Smit. ***Understanding Molecular Simulations. From Algorithms to Applications.*** Academic Press, San Diego, California, 1996.
A. R. Leach. ***Molecular Modelling. Principles and Applications.***Addison Wesley Longman, Essex, England, 1996.
More at http://www.biomath.nyu.edu/index/course/99/textbooks.html

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 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.

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.

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

+

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

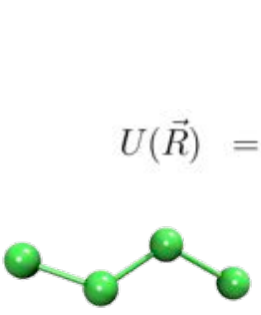
"Verlet algorithm"

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

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

Potential Energy Function of Biopolymer

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



heuristic

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

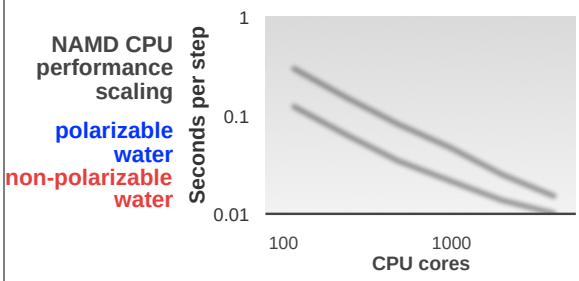
$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} +$$

from physics

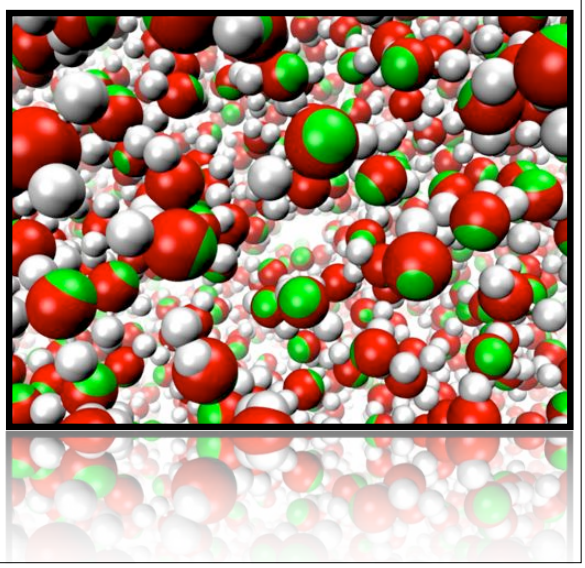
$$\underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$

Improving the Force Field

- Atomic polarizability increases computation by 2x…
- …but, the additional computations are perfectly suited to the GPU!
- For now, NAMD calculates atomic polarizability on CPUs only...soon we will also use GPUs



Atomic polarizability of water, highly accurately simulated through additional particles (shown in green)



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.

Langevin Dynamics

for temperature control

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

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

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

Langevin Dynamics

for pressure control

Underlying Langevin-Hoover barostat equation for all atoms:
Equations solved numerically in NAMD

$$\frac{d^2V(t)}{dt^2} \;=\; \frac{1}{W_{bs}}\,[P(t) \,-\, P_{\rm target}] \,-\, \frac{1}{\tau_{\rm bs}}\frac{dV(t)}{dt} \;+\; R_{\rm bs}(t)$$

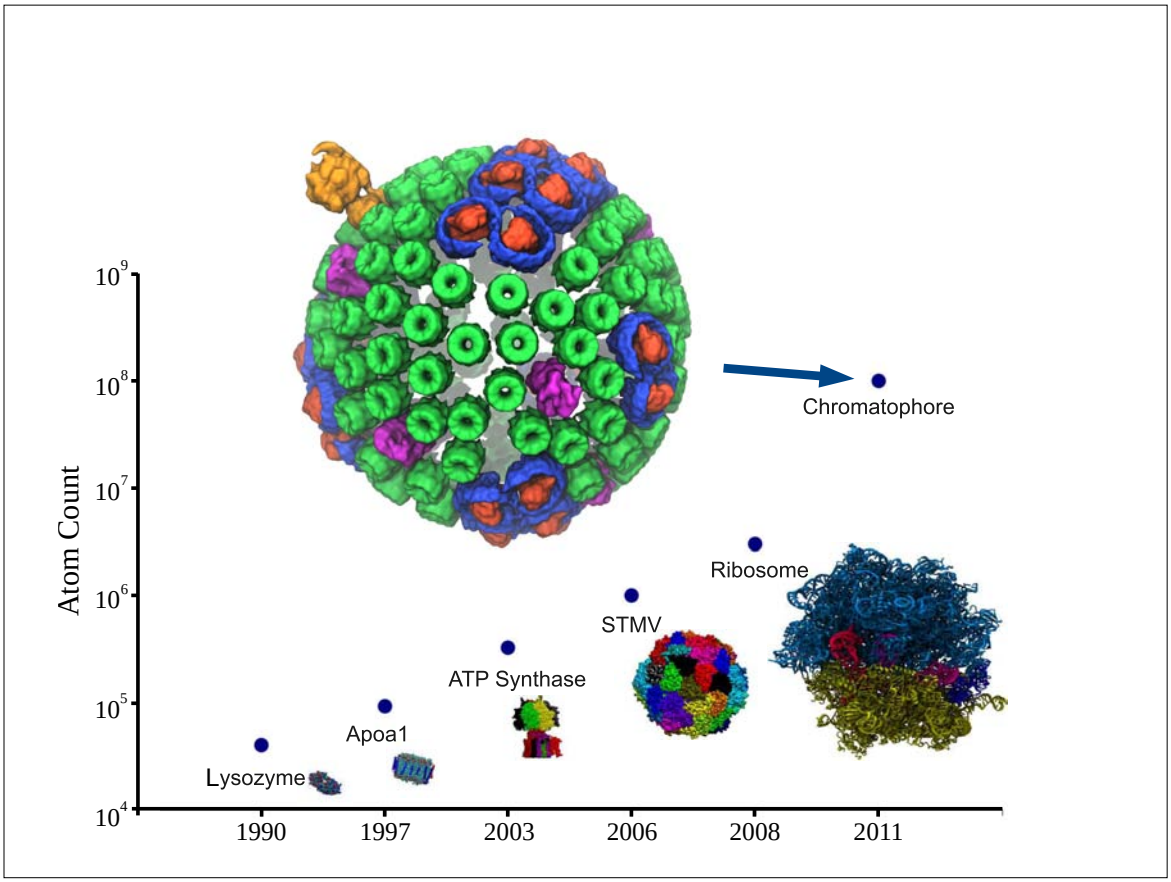
$$P = \rho k_B T + \frac{1}{Vd}\sum_{i<j}\langle r_{ij}\frac{dU_{tot}(r_{ij})}{dr_{ij}}\rangle \qquad d \;=\; \text{dimension}$$

$$\langle R_{\rm bs}(t)\, R_{\rm bs}(t')\rangle \;=\; \frac{2\,k_{\rm B}\,T_{\rm target}\,\delta(t \,-\, t')}{W_{\rm bs}\,\tau_{\rm bs}} \qquad W_{\rm bs} \;=\; d\,N_{\rm atoms}\,k_{\rm B}T_{\rm target}\,\tau^2_{period}$$

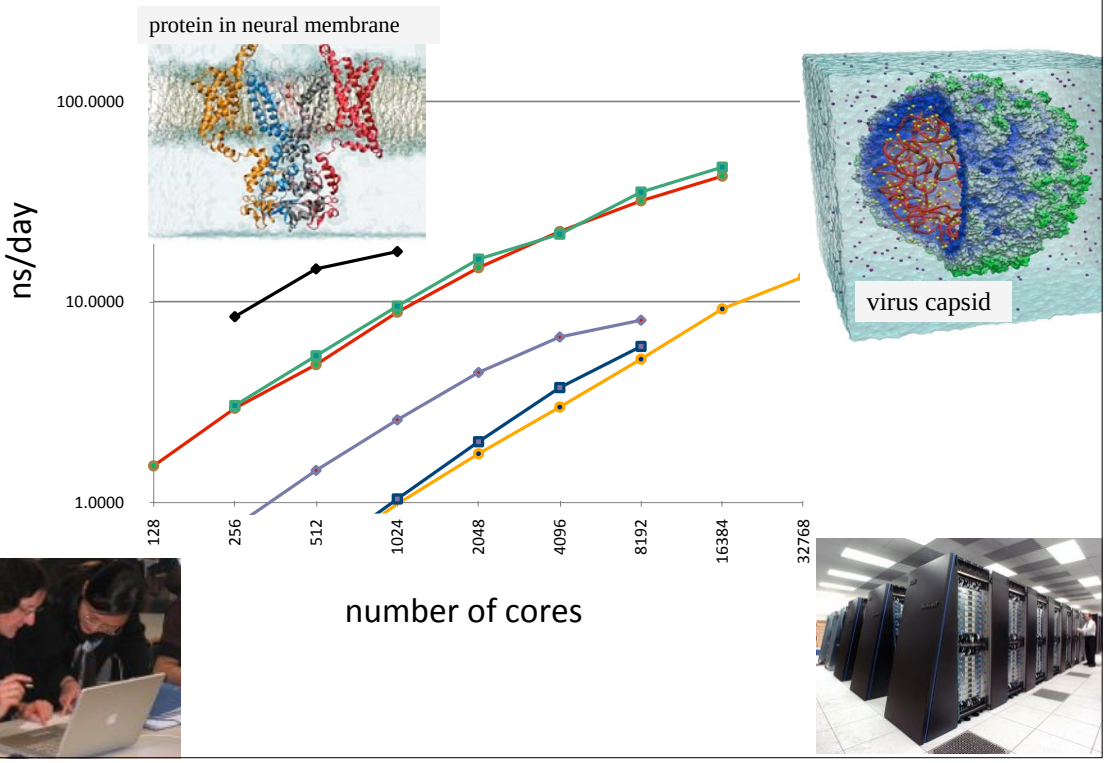
$$\dot{\mathbf{r}}_i = \mathbf{v}_i + s\mathbf{r}_i \qquad\qquad \dot{\mathbf{v}}_i = \mathbf{F}_i \,/\, m_i - s\mathbf{v}_i$$

$$\dot{V} = dVs \qquad \dot{s} = dV\,(P - P_{\rm target})\,/\,W - s\,/\,\tau_{\rm bs} \quad + R(t)$$

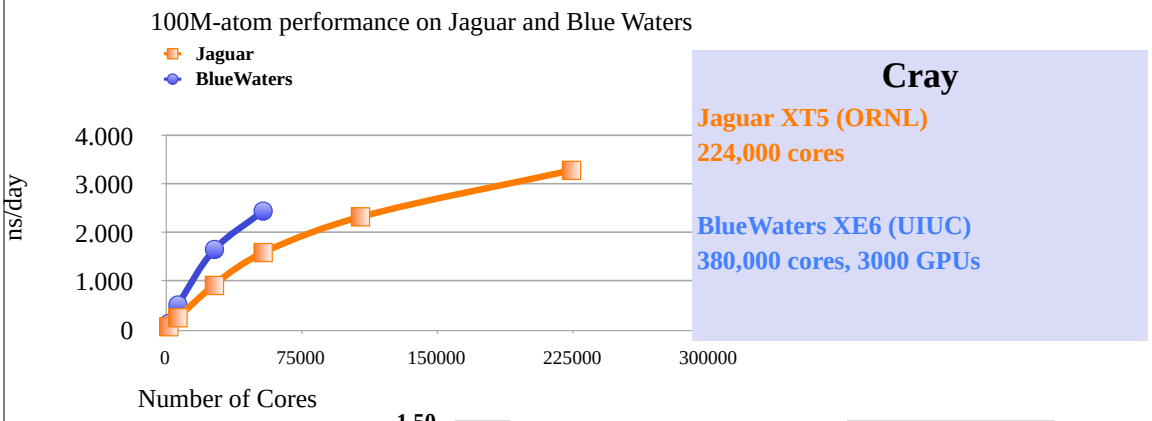
d - dimension



NAMD Scalability



Scaling of NAMD on VERY Large Machines



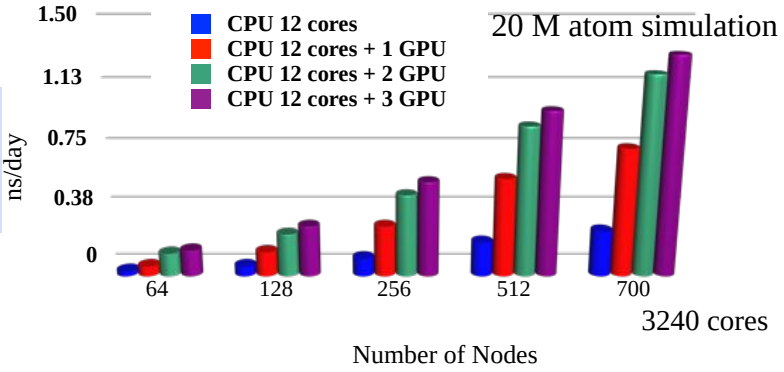
Cray

Jaguar XT5 (ORNL)
224,000 cores

BlueWaters XE6 (UIUC)
380,000 cores, 3000 GPUs

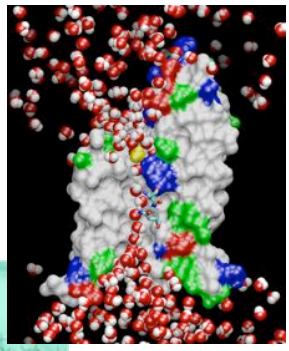
Tsubame

Tokyo Institute of Tech.
4224 GPUs



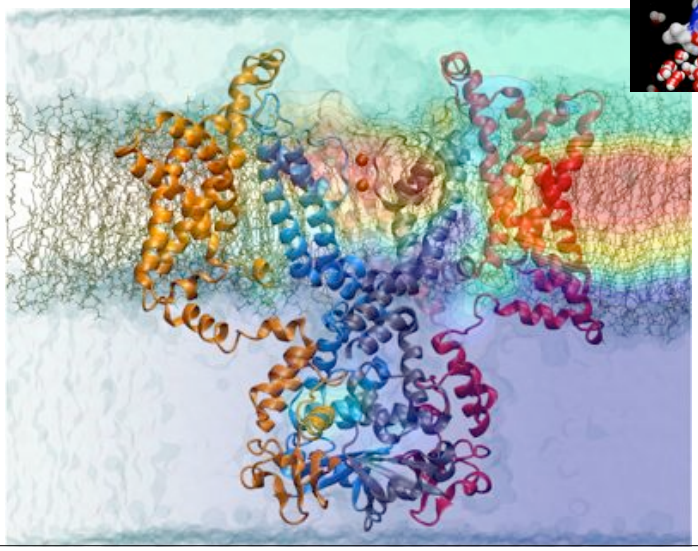
From 10,000 to 100,000 Atom MD in 2000

100k atom MD reached in 2000
- then a factor 10 increase in computation;
- **needed to describe membrane processes;**
- was achieved through cluster computing;
- produced good quality results for aquaporin;
- **is now standard.**



E.Tajkhorshid, P. Nollert, M. Jensen, L. Miercke, J. O'Connell, and K. Schulten. *Science*, **296:**525-530, 2002.

100,000 atoms, 12 ns



F. Khalili-Araghi, V. Jogini, V. Yarov-Yarovoy, E. Tajkhorshid, B. Roux, and K. Schulten. **Calculation of the gating charge for the Kv1.2 voltage-activated potassium channel.** *Biophysical Journal*, **98:**2189-2198, 2010.

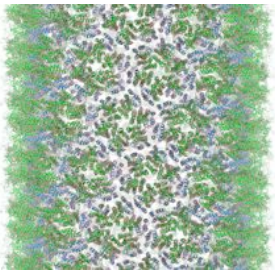
350,000 atoms, 0.5 µs

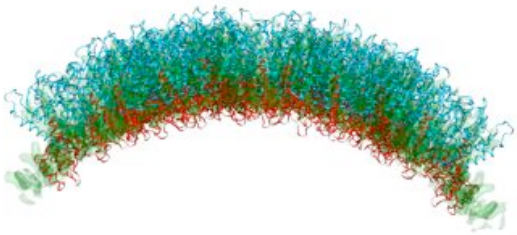
From 100,000 to 13,000,000 Atom MD Now

Recent Structural Assignment of HIV Capsid

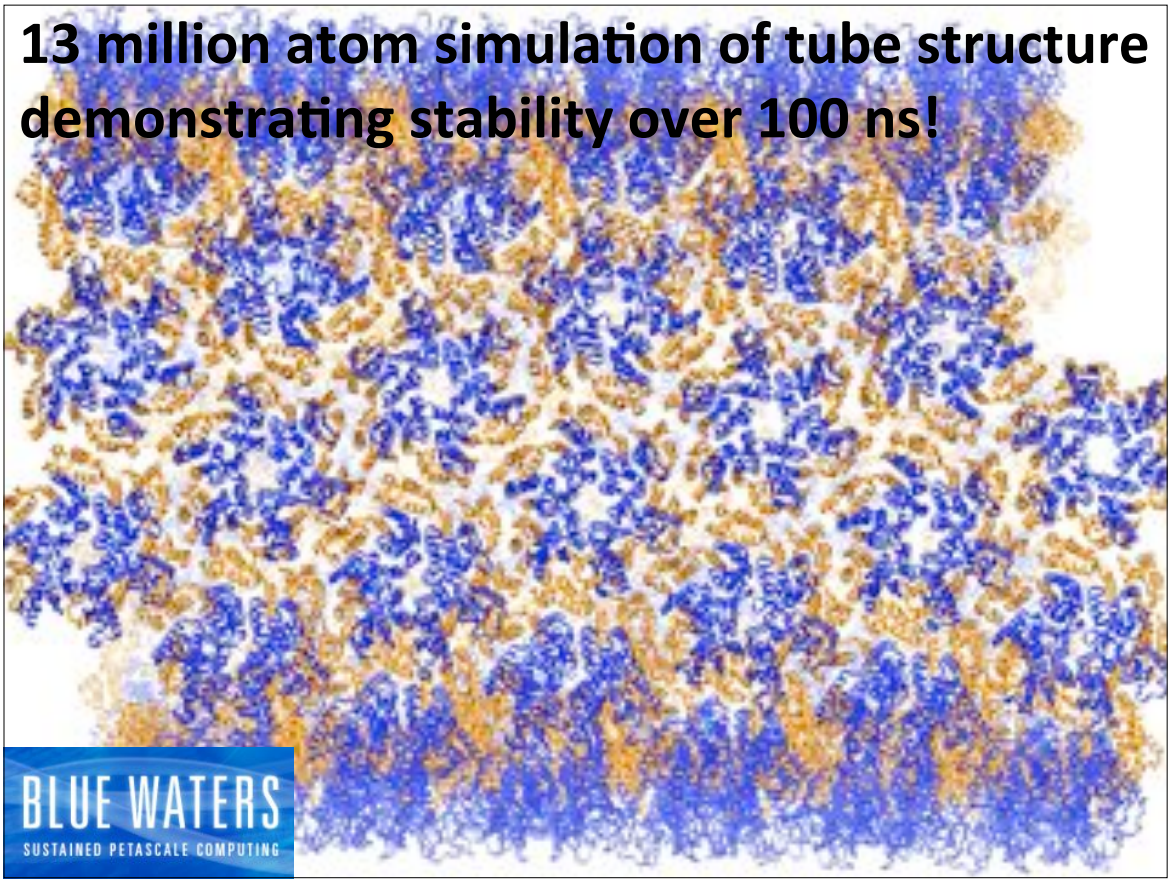
Cryo-EM density

Atomic Model

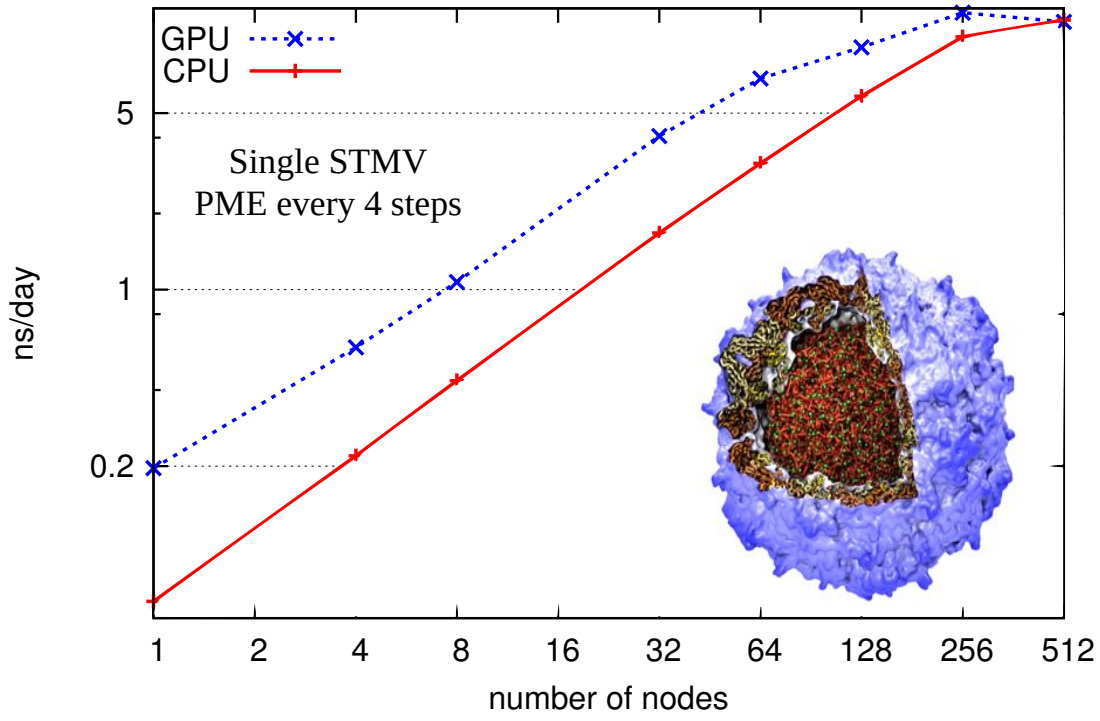




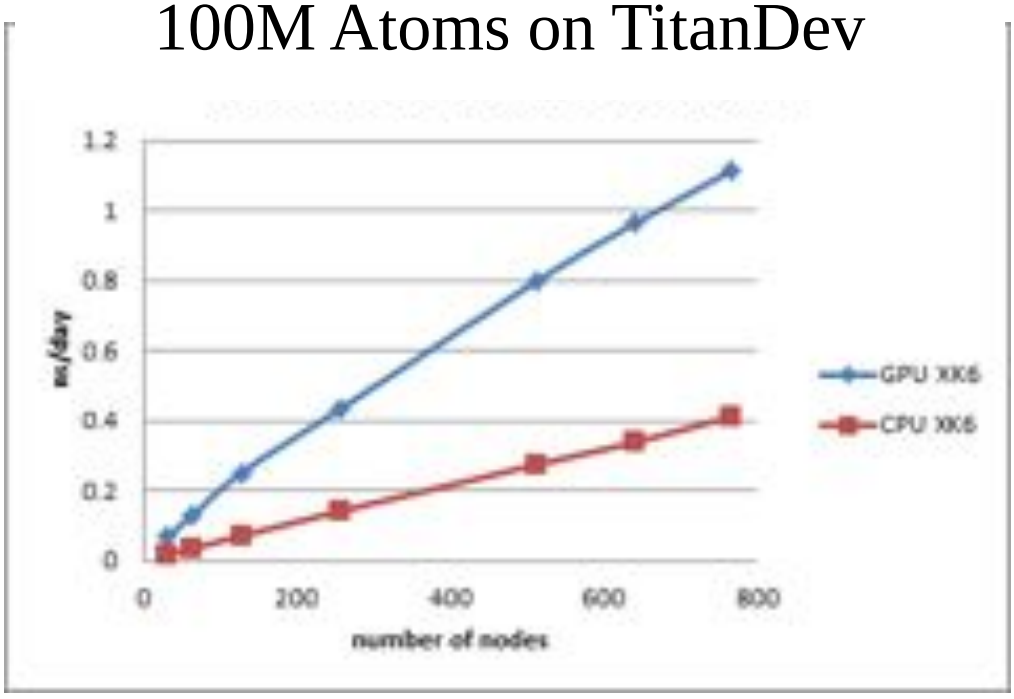
13 million atom simulation of tube structure demonstrating stability over 100 ns!

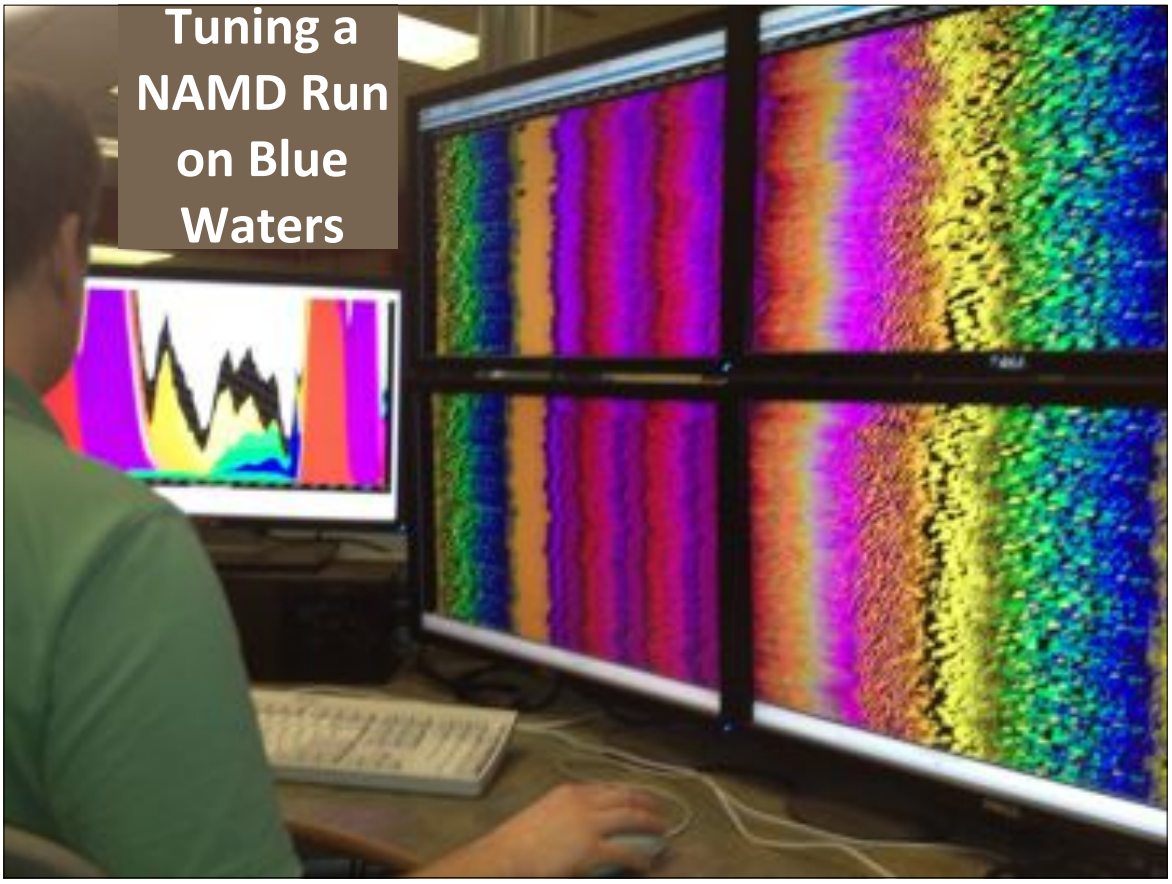


1M Atom Virus on TitanDev GPU

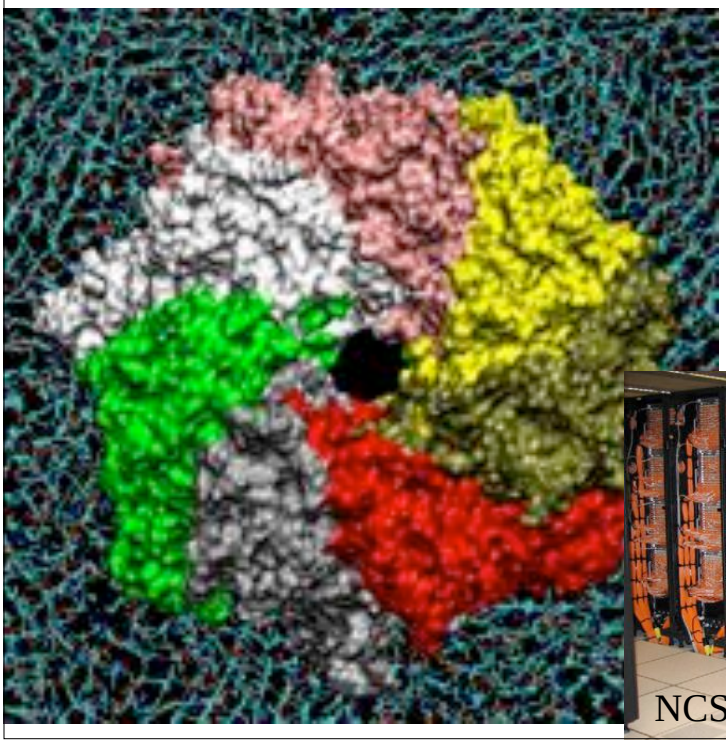


100M Atoms on TitanDev





Large is no problem. But …

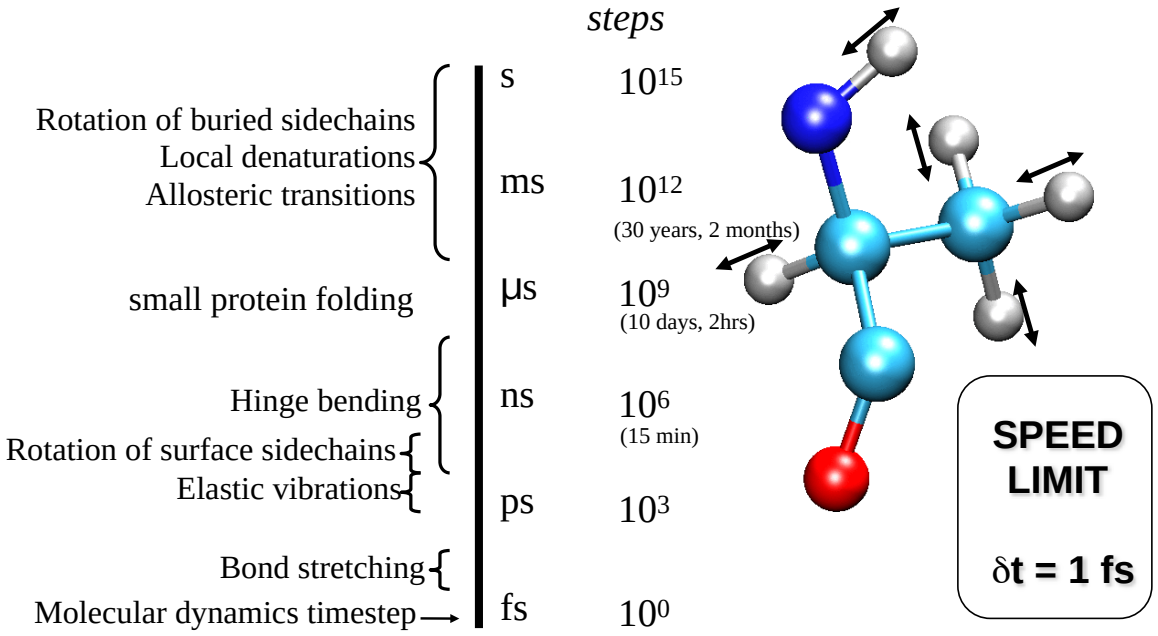


Molecular dynamics simulation of alpha-hemolysin with about 300,000 atoms; 1 million atom simulations are routine today, 20 million atom simulations are possible.

NCSA machine room

But long is still a problem!

biomolecular timescale and timestep limits



(NSF center, Shaw Res.)

Distinguishing Strengths of NAMD

Why do biomedical researchers rely on our software?

- Same user interface from laptop to world's fastest computers
- Scalable to large molecular systems on largest supercomputers
- Scalable for challenging cases, e.g. implicit solvent
- Broad, adaptable suite of free energy methods
- Consistent usability with flexible scripting interfaces

- Project-driven integration of best-in-class, broadly applicable methods
- Centralized professional development effort for dependable quality
- New capabilities co-developed with VMD, e.g. MDFF

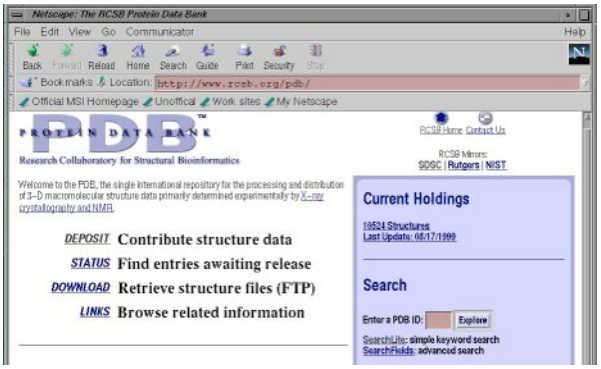
- Innovation in parallel algorithms and programming interfaces
- Early adoption of new technologies (GPUs 2007, GPU clusters 2008)
- Computer science expertise to deal with future complex hardware
- Partnership with US computing industry and supercomputer centers

PDB Files

gives one the structure and starting position

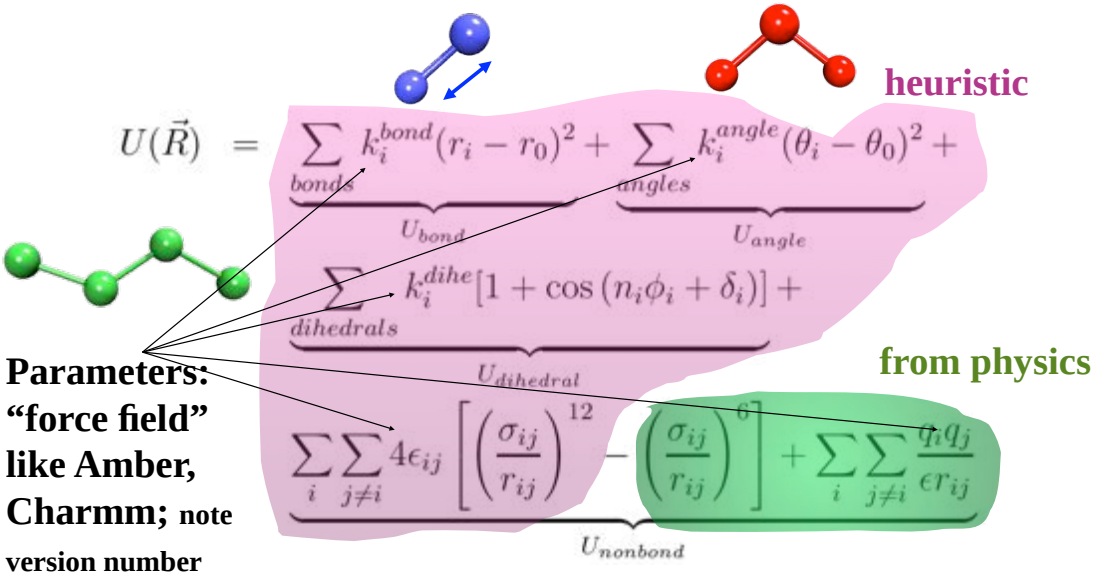
- Simulations start with a crystal structure from the Protein Data Bank, in the standard PDB file format.
- PDB files contain standard records for species, tissue, authorship, citations, sequence, secondary structure, etc.
- We only care about the atom records…
  - atom name (N, C, CA)
  - residue name (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.)



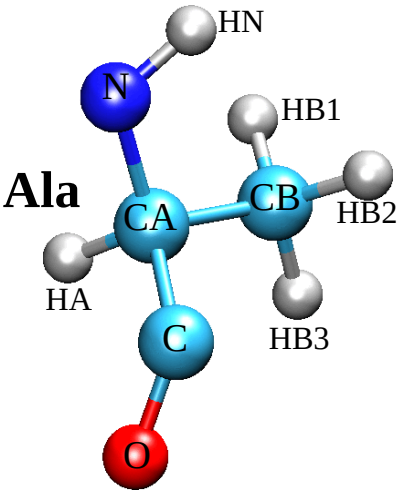
Potential Energy Function of Biopolymer

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



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

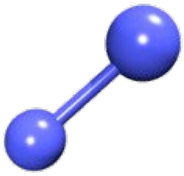
PSF Files



- 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)

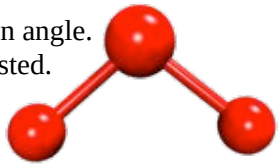
PSF Files

molecular structure (bonds, angles, etc.)



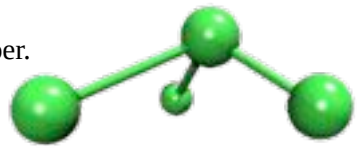
Bonds: Every pair of covalently bonded atoms is listed.

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.



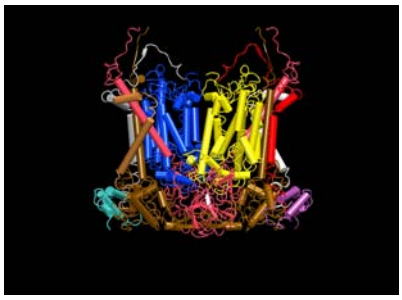
Preparing Your System for MD
Solvation

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



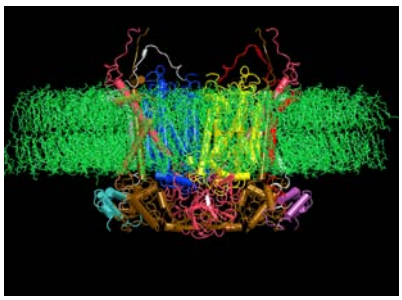
Preparing Your System for MD
Solvation

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

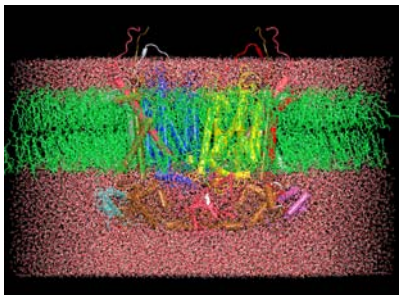
Solvation

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

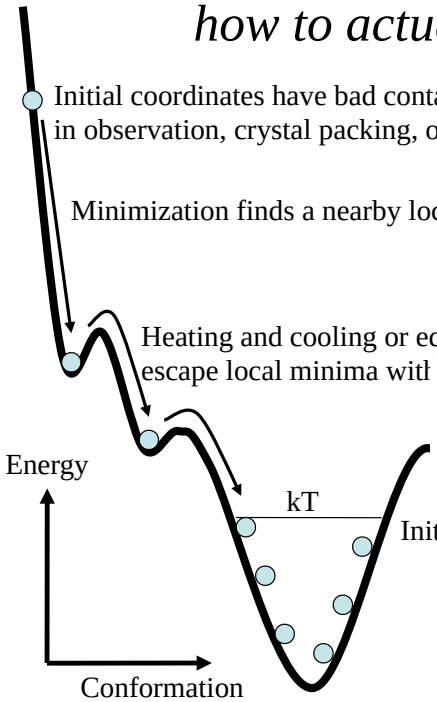
mitochondrial bc1 complex



(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)

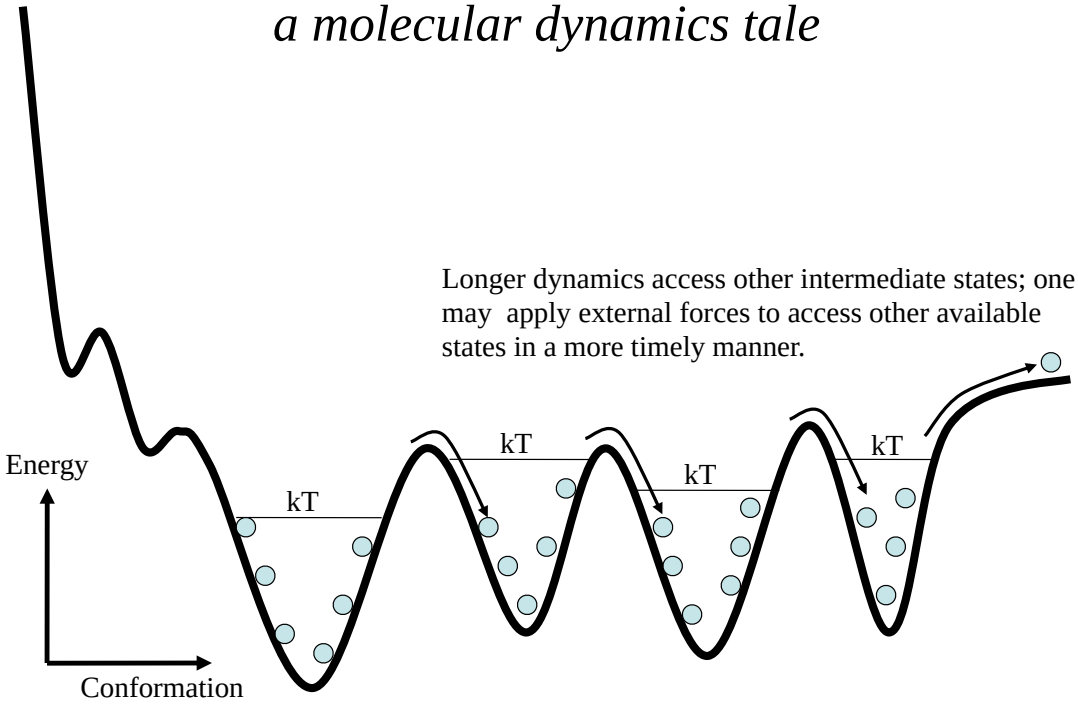
Minimization finds a nearby local minimum.

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

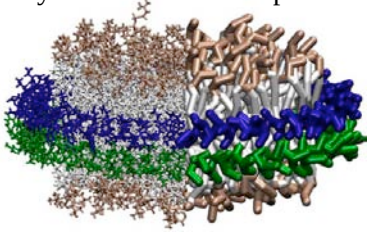
a molecular dynamics tale



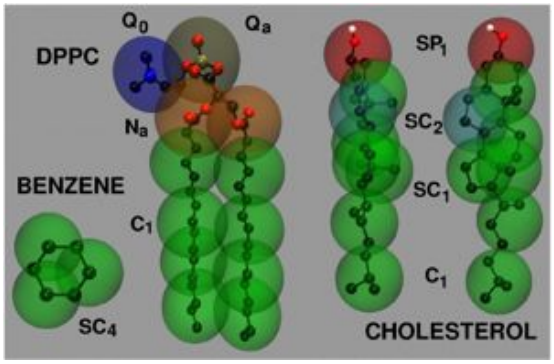
Cutting Corners

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

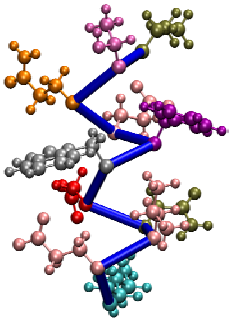


Residue-Based Coarse-Grained Model

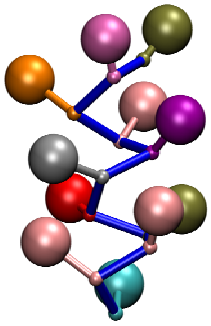


- Lipid model: MARTINI
- Level of coarse-graining: ~4 heavy atoms per CG bead
- Interactions parameterized based on experimental data and thermodynamic properties of small molecules

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



All-atom peptide

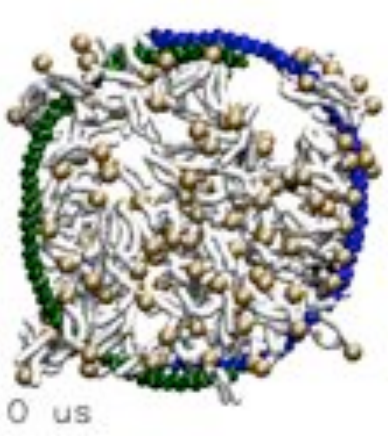


CG peptide

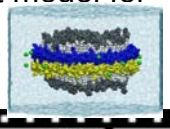
Peter L. Freddolino, Anton Arkhipov, Amy Y. Shih, Ying Yin, Zhongzhou Chen, and Klaus Schulten. **Application of residue-based and shape-based coarse graining to biomolecular simulations.** In Gregory A. Voth, editor, *Coarse-Graining of Condensed Phase and Biomolecular Systems*, chapter 20, pp. 299-315.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

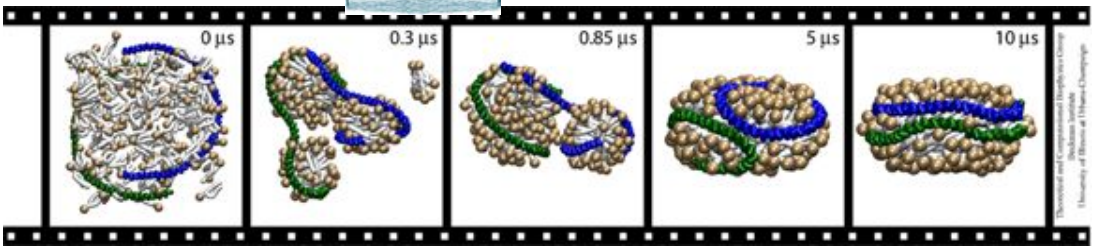
Nanodisc Assembly CG MD Simulation

- 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

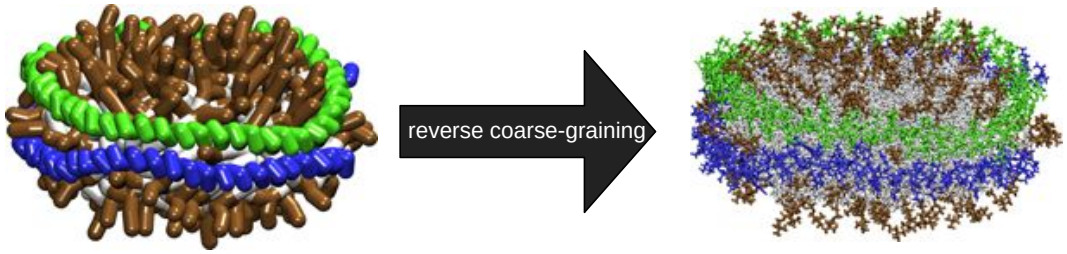




A. Shih, A. Arkhipov, P. Freddolino, and K. Schulten. *J. Phys. Chem. B*, 110:3674–3684, 2006; A. Shih, P. Freddolino, A. Arkhipov, and K. Schulten. *J. Struct. Biol.*, 157:579–592,2007; A. Shih, A. Arkhipov, P. Freddolino, S. Sligar, and K. Schulten. *Journal of Physical Chemistry B*, 111: 11095 - 11104,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

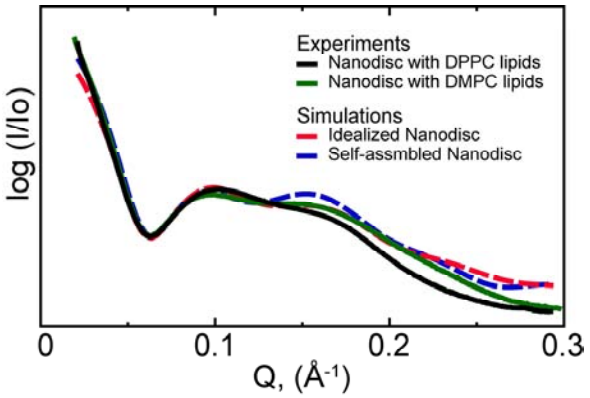
Validation of Simulations

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

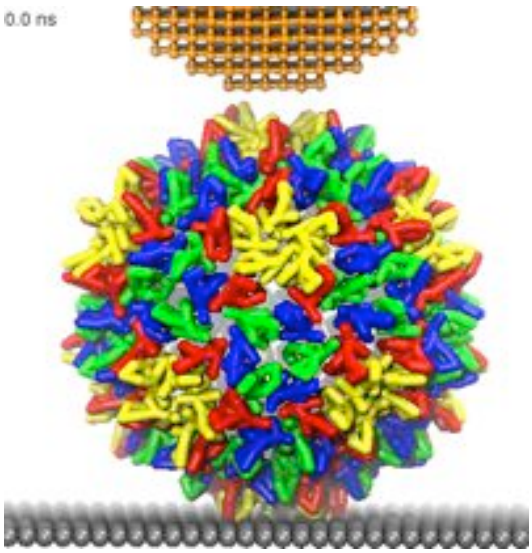


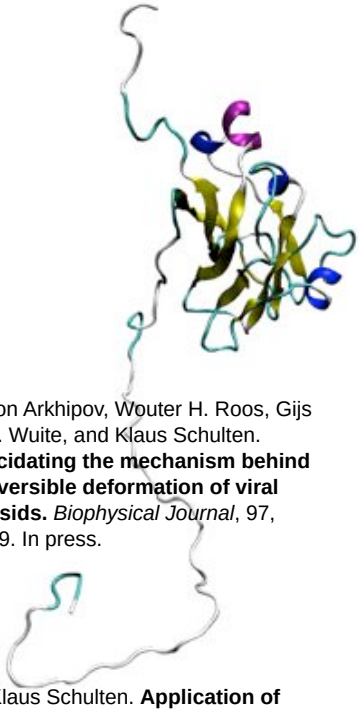
Reverse coarse-graining:
1. Map center of mass of the group of atoms represented by a single CG bead to that beads location
2. MD minimization, simulated annealing with restraints, and equilibration to get all-atom structure

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



Shape-Based Coarse-Grained (CG) model





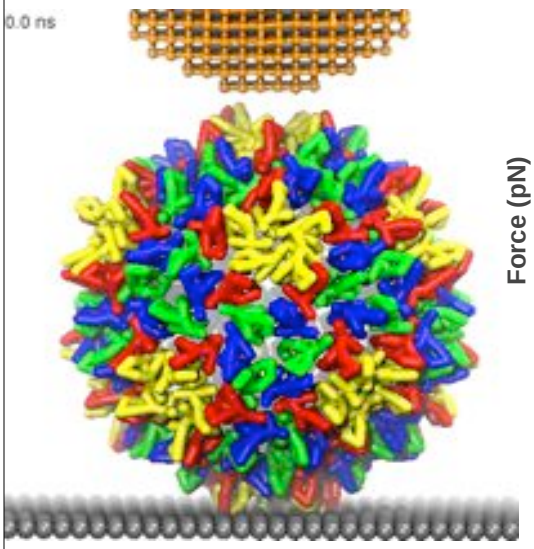
Anton Arkhipov, Wouter H. Roos, Gijs J. L. Wuite, and Klaus Schulten. **Elucidating the mechanism behind irreversible deformation of viral capsids.** *Biophysical Journal*, 97, 2009. In press.

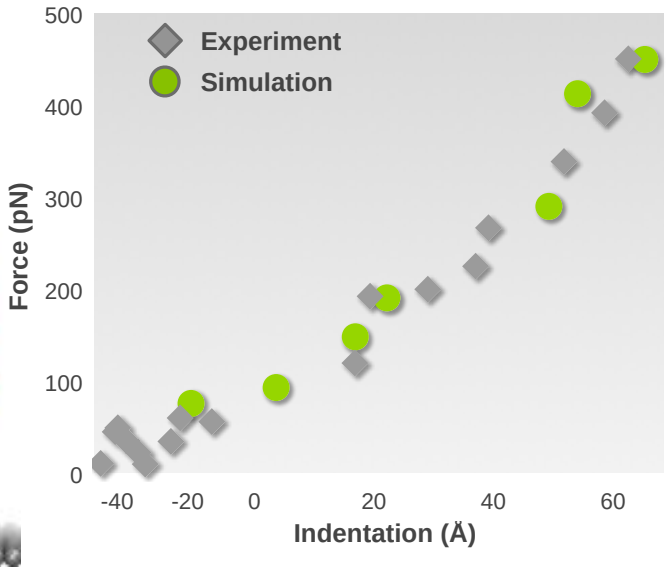
- **Fully automatic**
- **Number of CG beads is chosen by a user (we used ~200 atoms per CG bead)**

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. 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.

Virus Capsid Mechanics
Atomic Force Microscope

— Hepatitis B Virus —

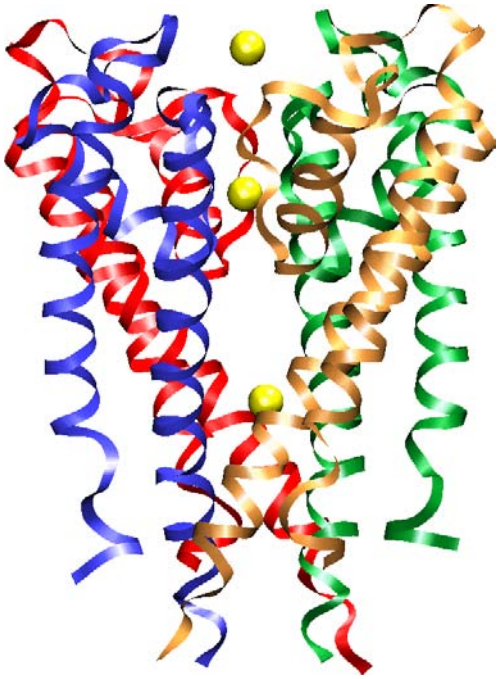




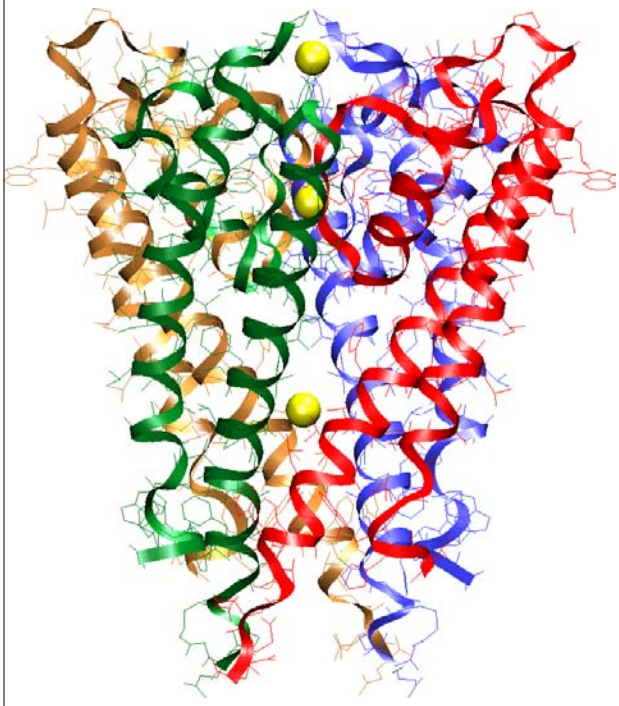
Example: MD Simulations of the K^+ Channel Protein

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.

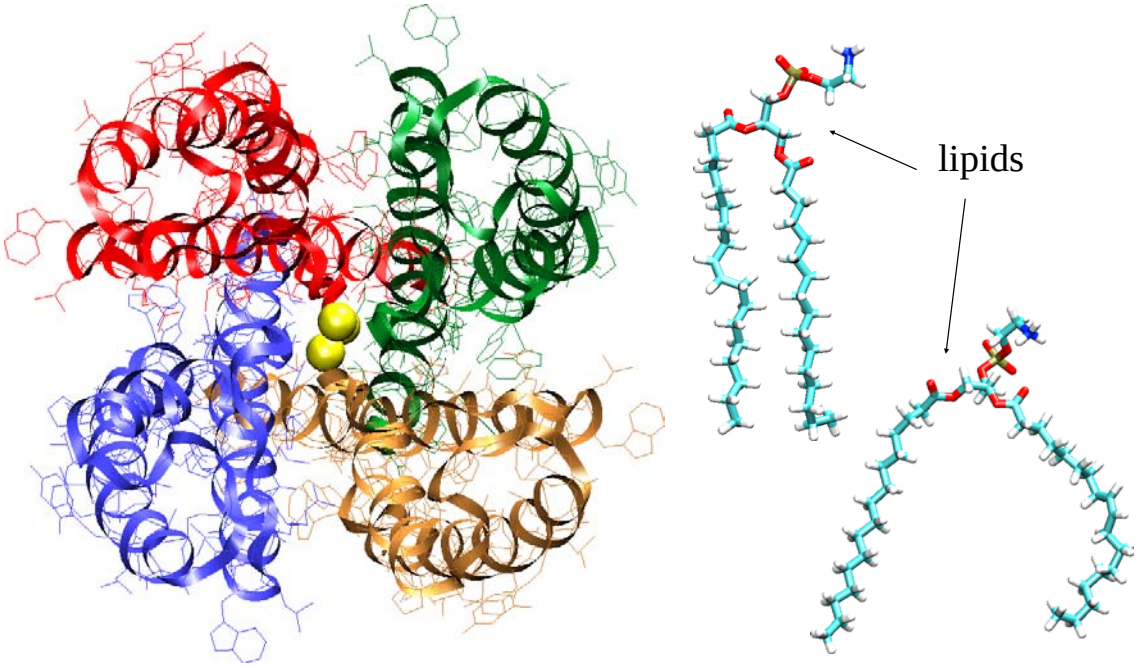


Setting up the system (1)



- 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

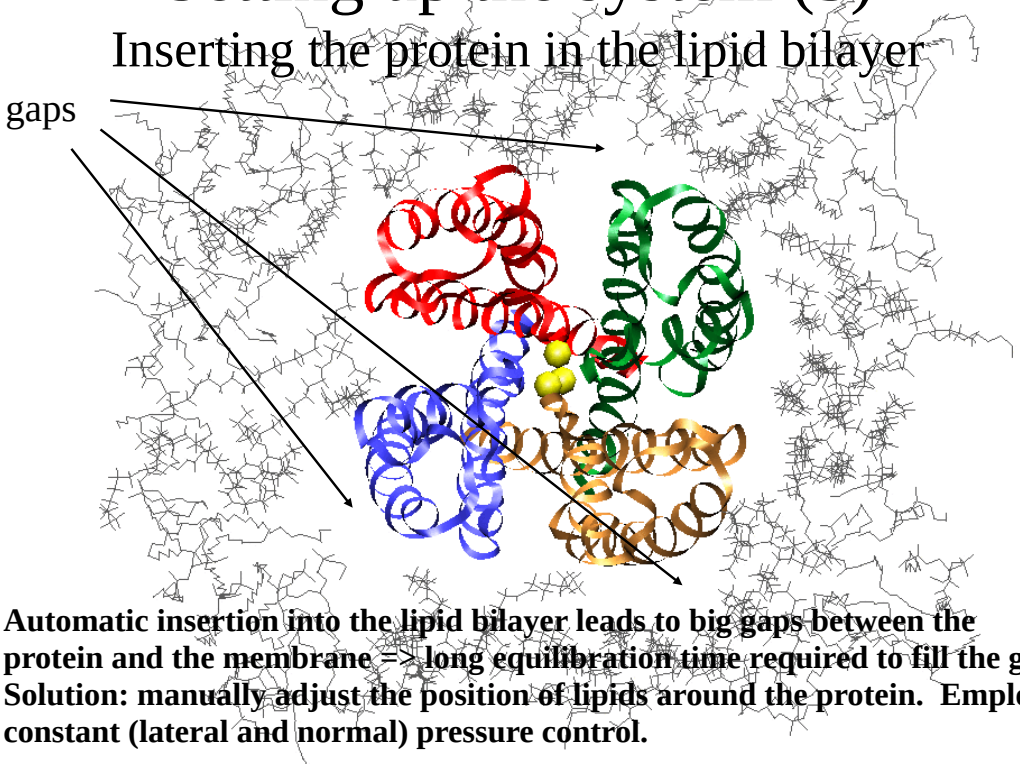
Setting up the system (2)



Simulate the protein in its natural environment: solvated lipid bilayer

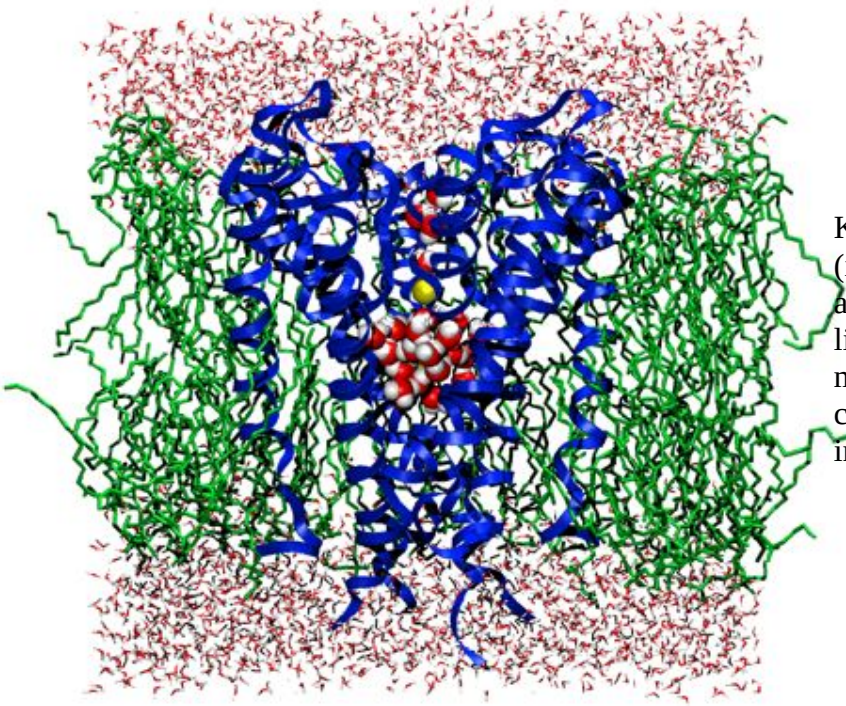
Setting up the system (3)

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

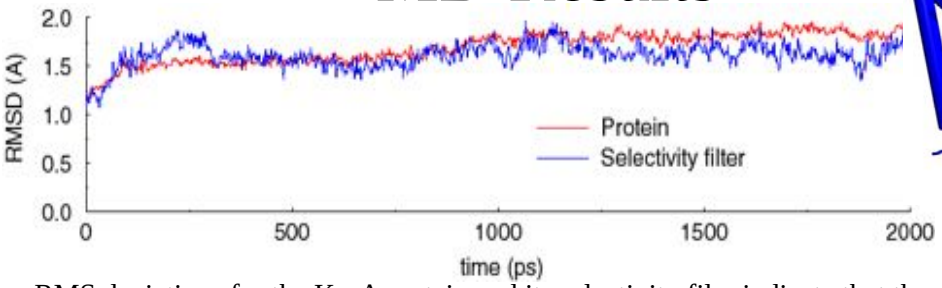
Simulating the system: Free MD

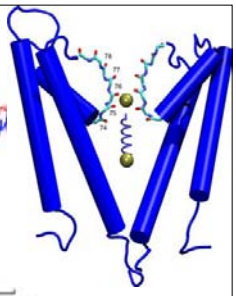
Summary of simulations:
• protein/membrane system contains 38,112 atoms, including 5117 water molecules, 100 POPE and 34 POPG lipids, plus K^+ counterions
• CHARMM26 forcefield
• periodic boundary conditions, PME electrostatics
• 1 ns equilibration at 310K, NpT
• 2 ns dynamics, NpT

Program: NAMD2

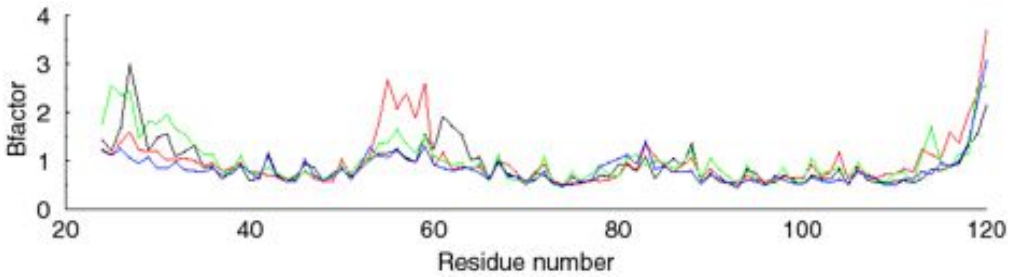
Platform: Cray T3E (Pittsburgh Supercomputer Center) or local computer cluster; choose ~1000 atoms per processor.

MD Results



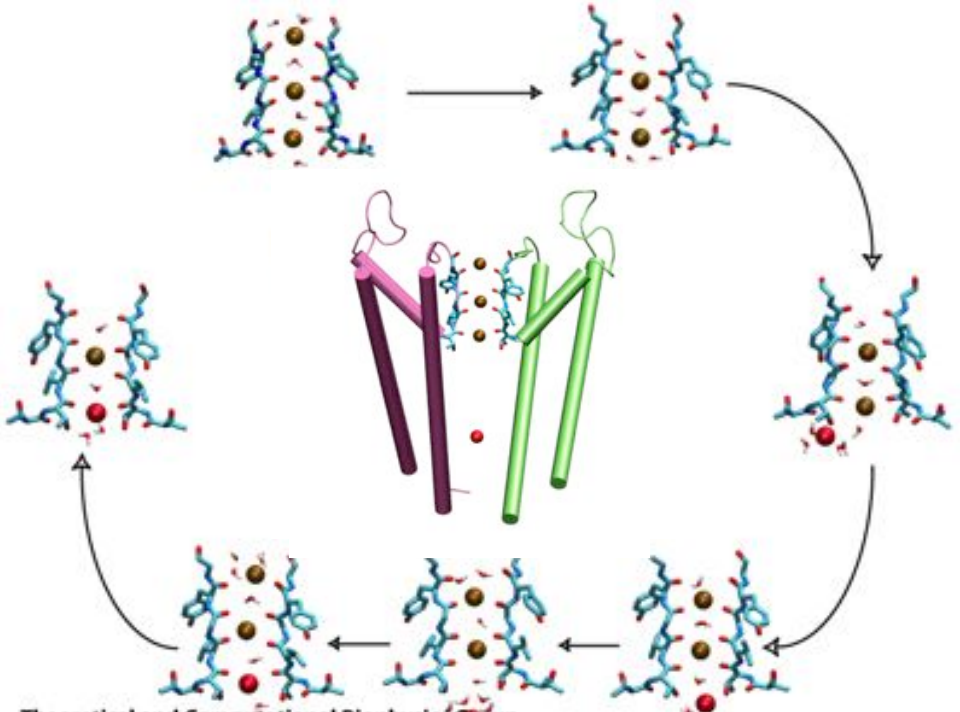


RMS deviations for the KcsA protein and its selectivity filer indicate that the protein is stable during the simulation with the selectivity filter the most stable part of the system.



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

NIH Center for Research Resources

Funding: NIH, NSF

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