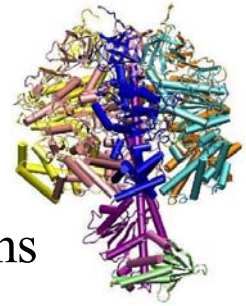
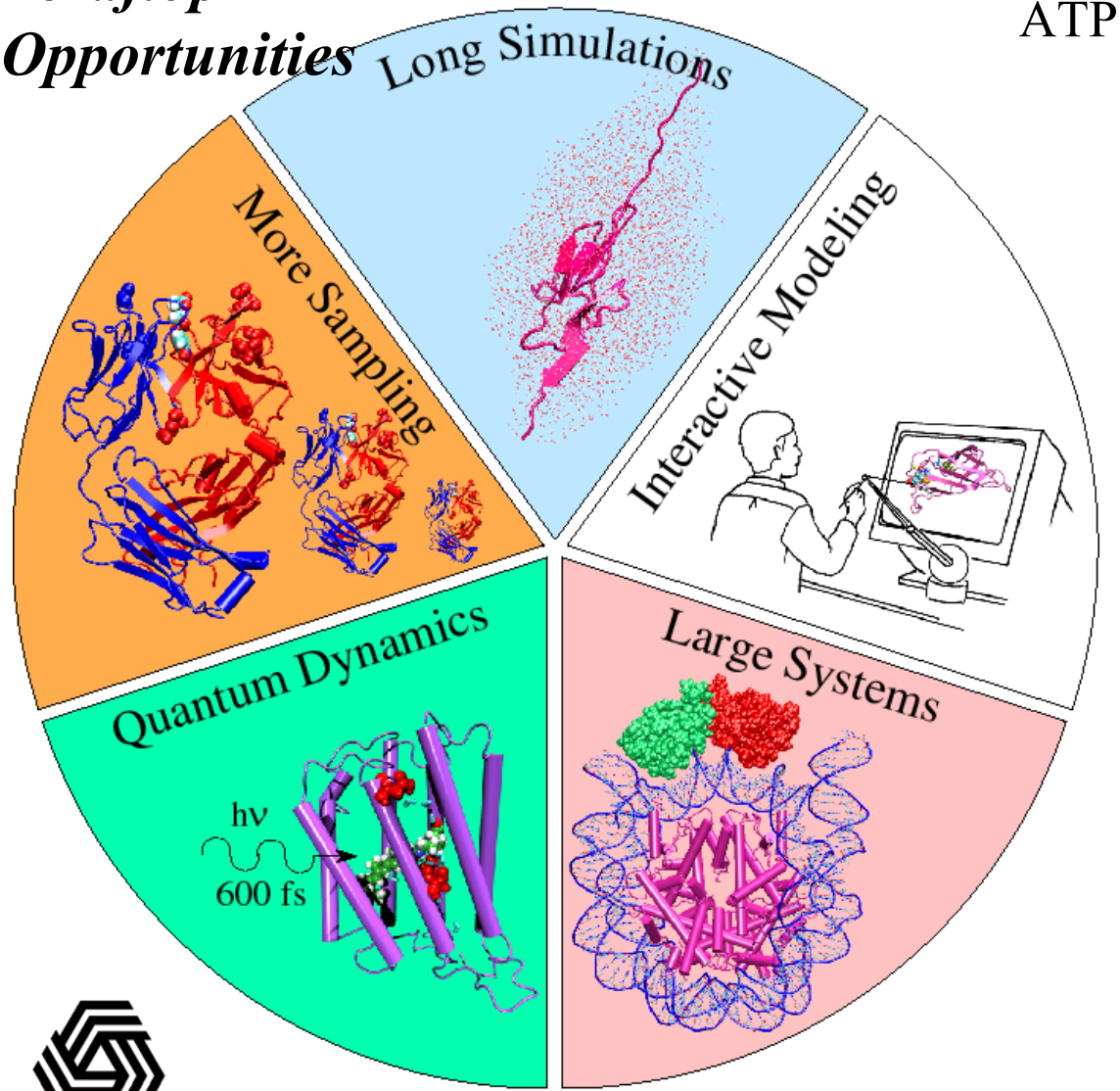
Introduction to Molecular Dynamics

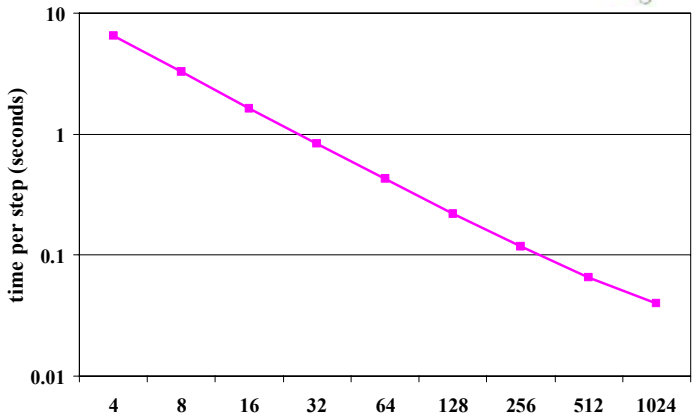
Theor. Biophys. Group, Beckman Inst.., U. Illinois



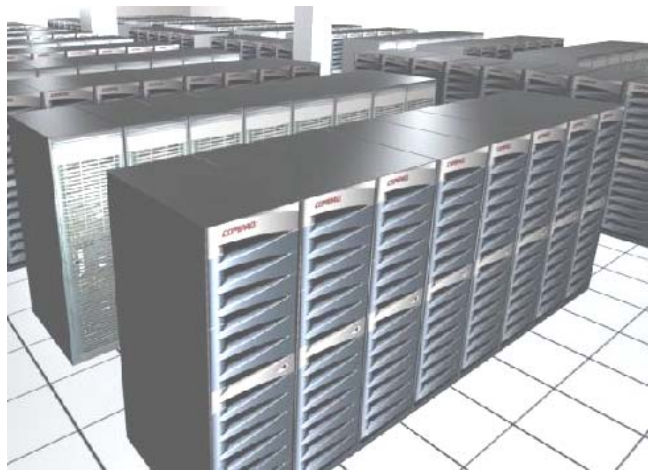
Teraflop Opportunities



ATP synthase 327,000 atoms



PSC LeMieux

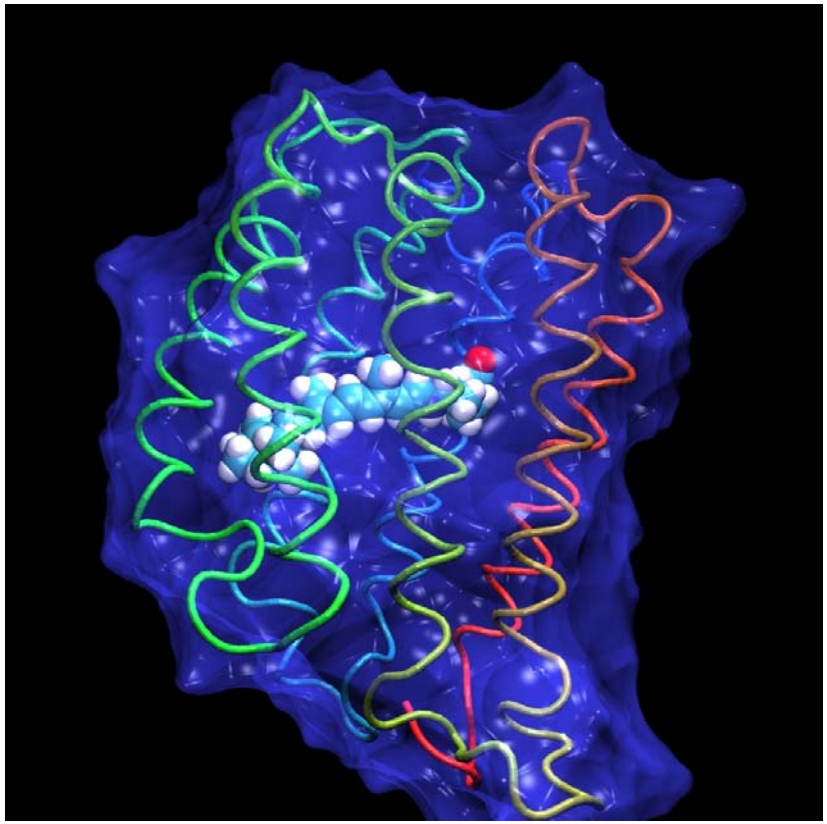


National Center for Research Resources

NAMD

Scalable Molecular Dynamics

VMD: Visual Molecular Dynamics



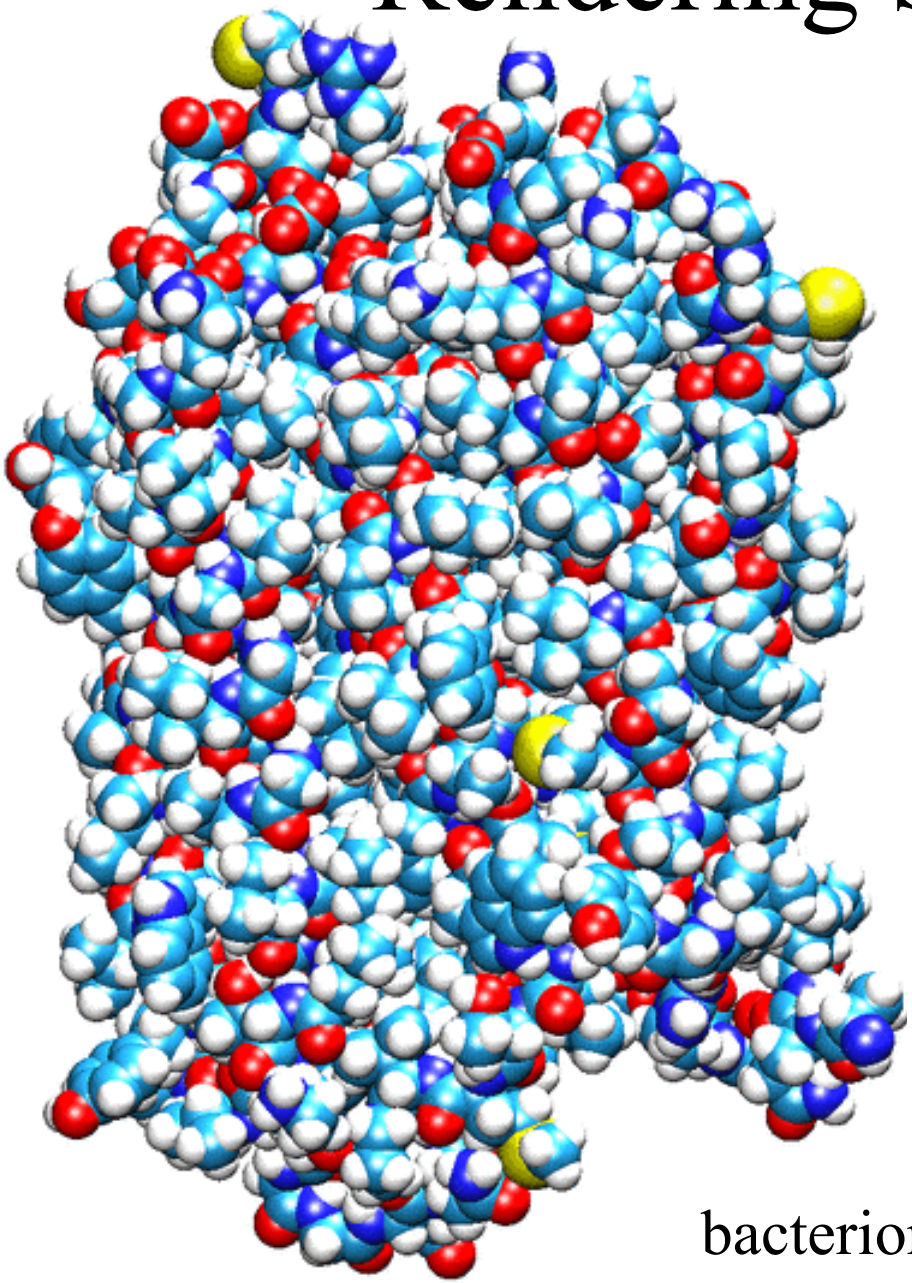
- Platforms:
  - Unix (16 builds)
  - Windows
  - MacOS X
- Display of large biomolecules and simulation trajectories
- Sequence browsing and structure highlighting
- User-extensible scripting interfaces for analysis and customization
- Interactive MD

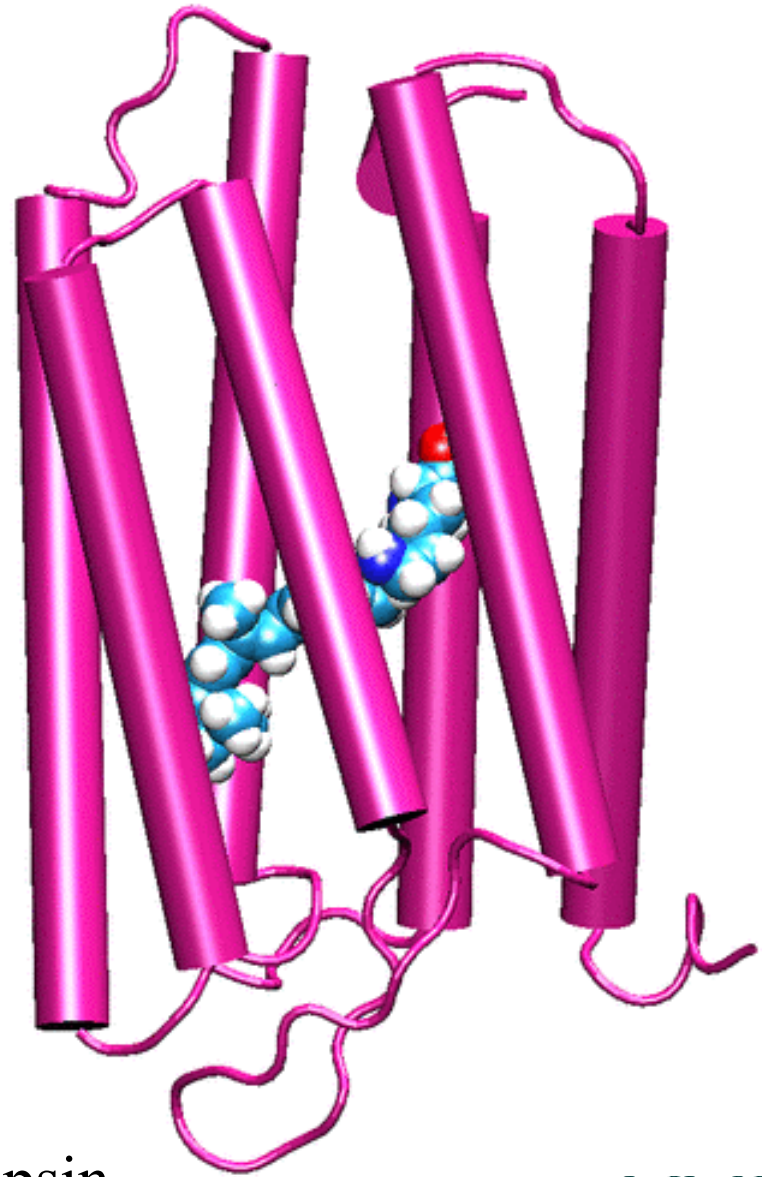
http://www.ks.uiuc.edu/Research/vmd/

Humphrey *et al.*, J. Molec. Graphics, 14.1, 33 (1996)

VMD
Visual Molecular Dynamics

Rendering Styles in VMD

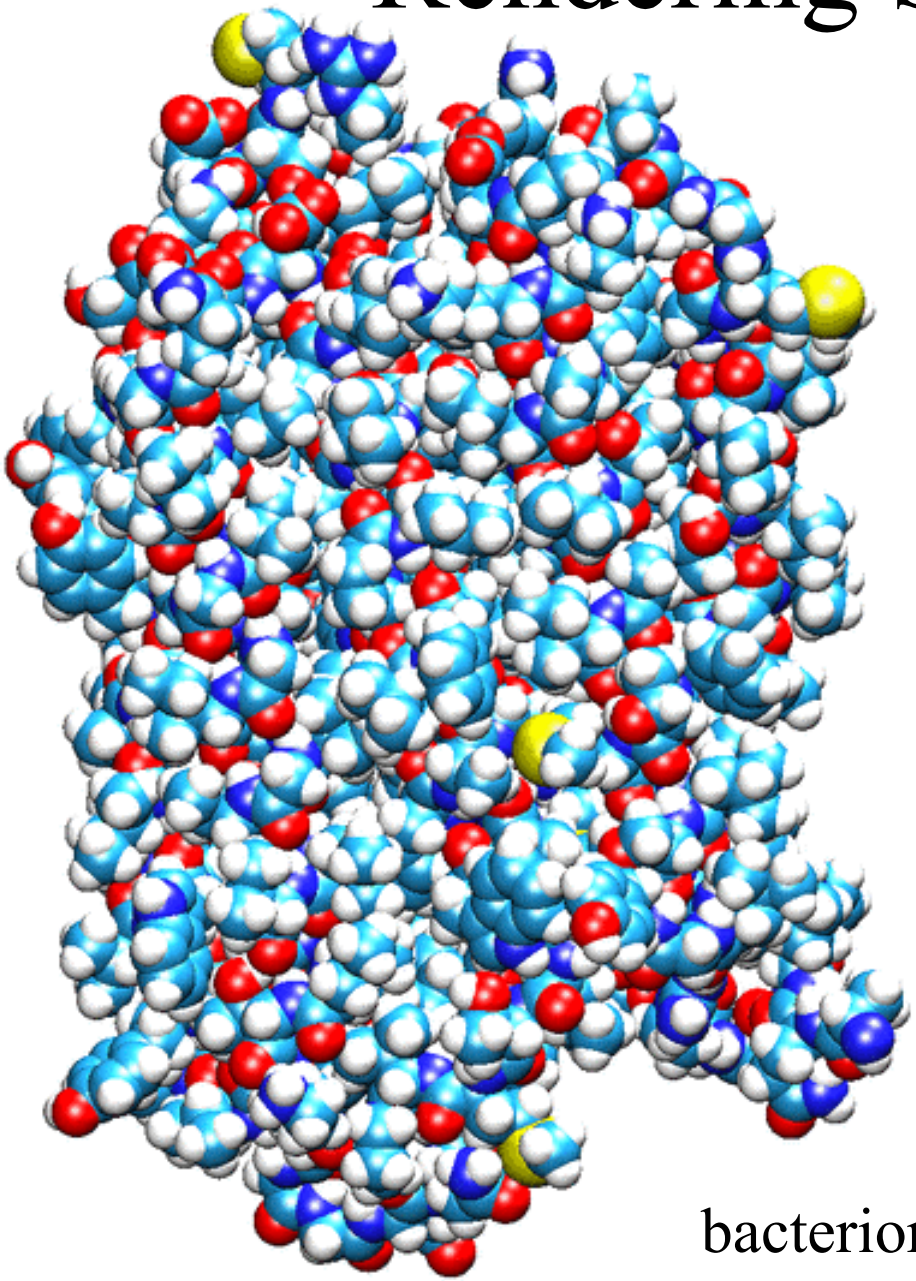


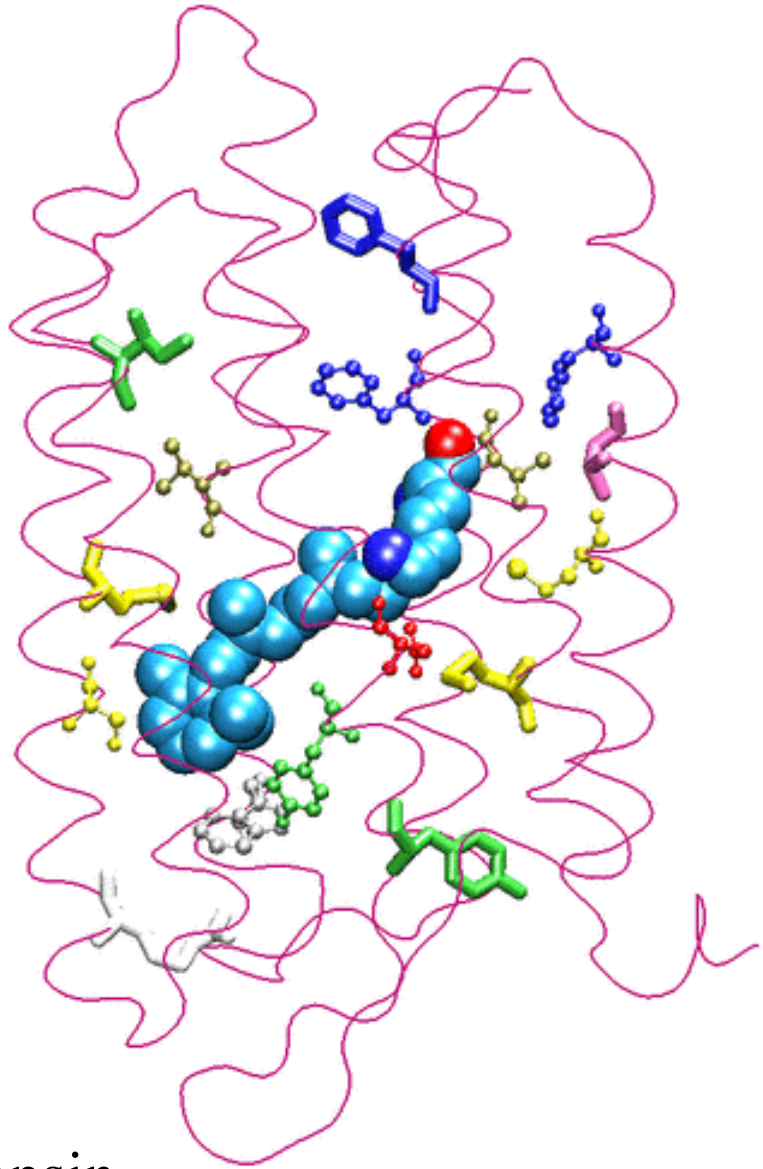


bacteriorhodopsin

VMD
Visual Molecular Dynamics

Rendering Styles in VMD

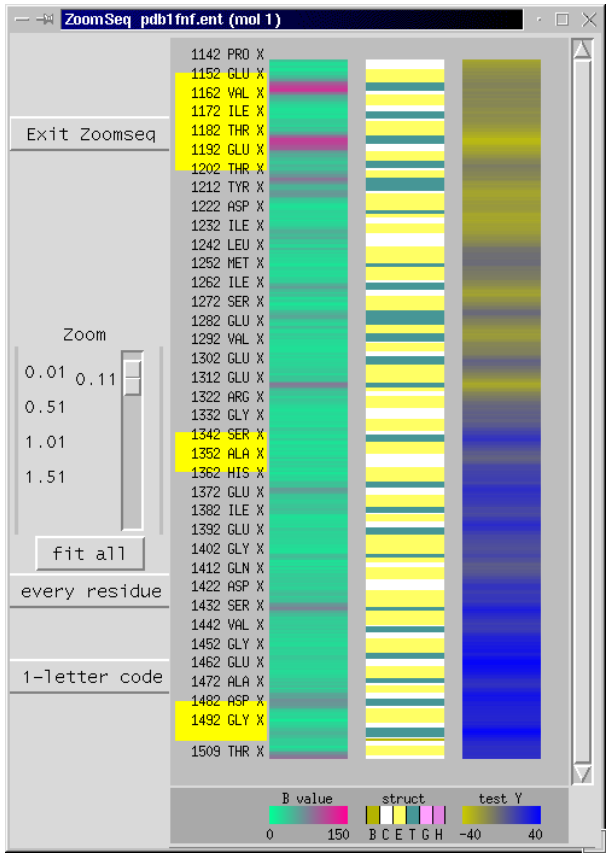


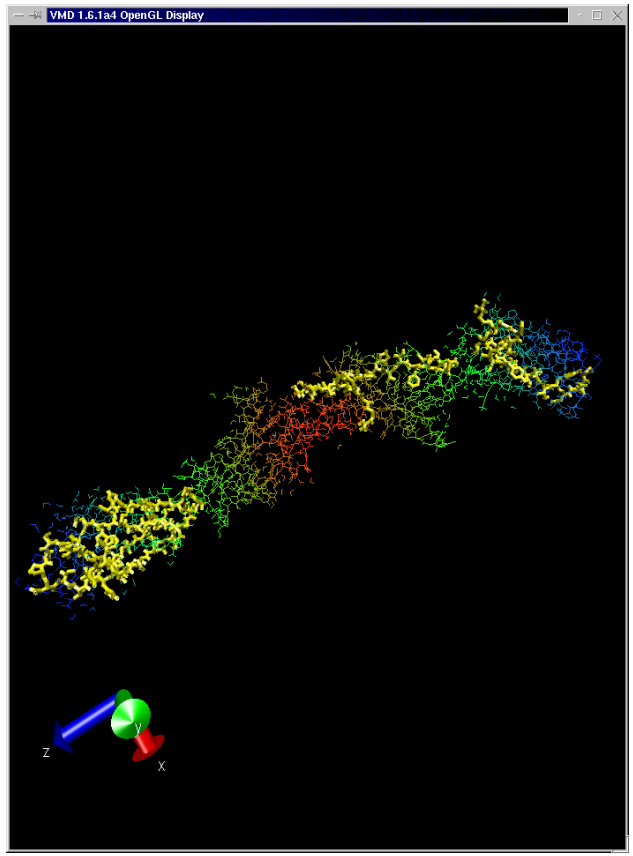


bacteriorhodopsin

VMD
Visual Molecular Dynamics

VMD Sequence Window





Map multiple columns of sequence data onto structure.

An Introduction to Molecular Dynamics Simulations

Macroscopic properties are often determined by molecule-level behavior.

Quantitative and/or qualitative information about macroscopic behavior of macromolecules can be obtained from simulation of a system at atomistic level.

Molecular dynamics simulations calculate the motion of the atoms in a molecular assembly using Newtonian dynamics to determine the net force and acceleration experienced by each atom. Each atom $\boldsymbol{i}$ at position $\boldsymbol{r_i}$, is treated as a point with a mass $\boldsymbol{m_i}$ and a fixed charge $\boldsymbol{q_i}$.

MD: Verlet Method

Energy function: $U(\vec{r}_1, \vec{r}_2, \cdots \vec{r}_N) = U(\vec{R})$

used to determine the force on each atom:

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

Newton's equation represents a set of N second order differential equations which are solved numerically at discrete time steps to determine the trajectory of each atom.

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

Deriving the Verlet algorithm

- Uses positions and accelerations at time t and the positions from time t-δt to calculate new positions at time t+ δ t.
- Uses no explicit velocities.

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

+

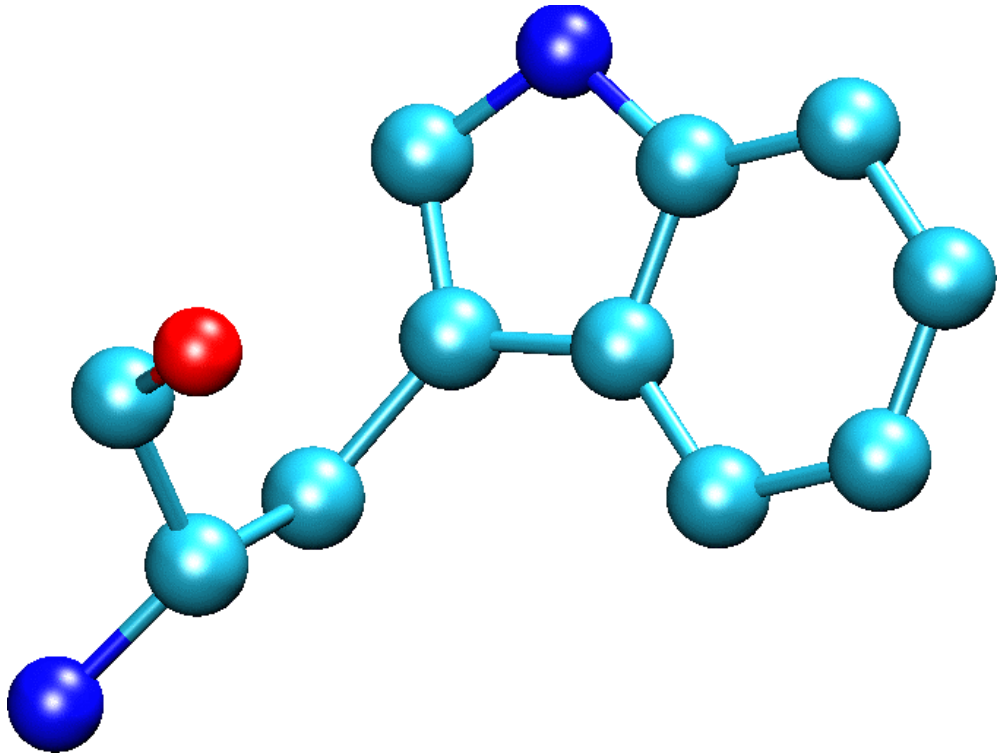
$$\mathbf{r}(t-\delta t) \approx \mathbf{r}(t) - \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2$$

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

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

What is the Force Field $-\vec{\nabla}U(\vec{R})$?

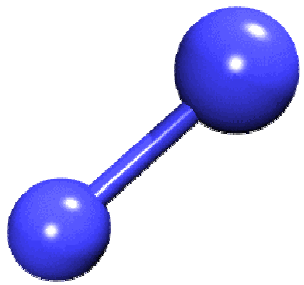
In molecular dynamics a molecule is described as a series of charged points (atoms) linked by springs (bonds).



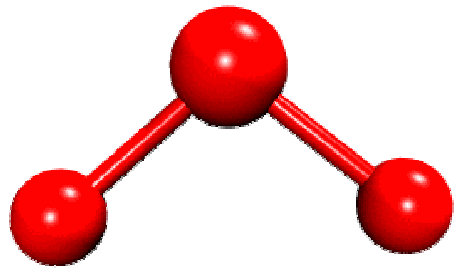
To describe the time evolution of bond lengths, bond angles and torsions, also the non-bonding van der Waals and elecrostatic interactions between atoms, one uses a forcefield.
The forcefield is a collection of equations and associated constants designed to reproduce molecular geometry and selected properties of tested structures.

Energy Terms Described in the CHARMm Force Field

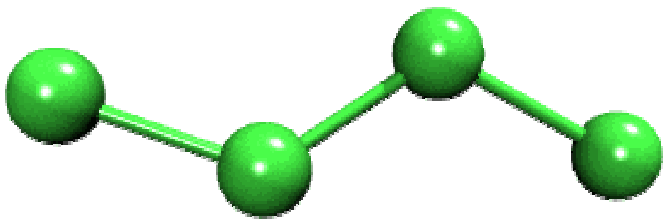
Bond



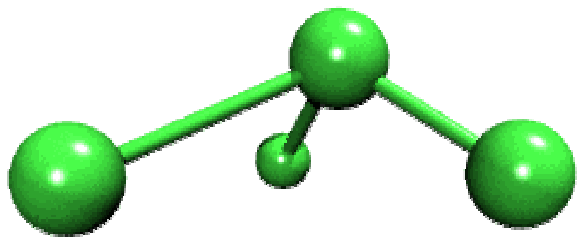
Angle



Dihedral



Improper



Energy Functions

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

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

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

U_{bond} = oscillations about the equilibrium bond length
U_{angle} = oscillations of 3 atoms about an equilibrium angle
$U_{dihedral}$ = torsional rotation of 4 atoms about a central bond
$U_{nonbond}$ = non-bonded energy terms (electrostatics and Lenard-Jones)

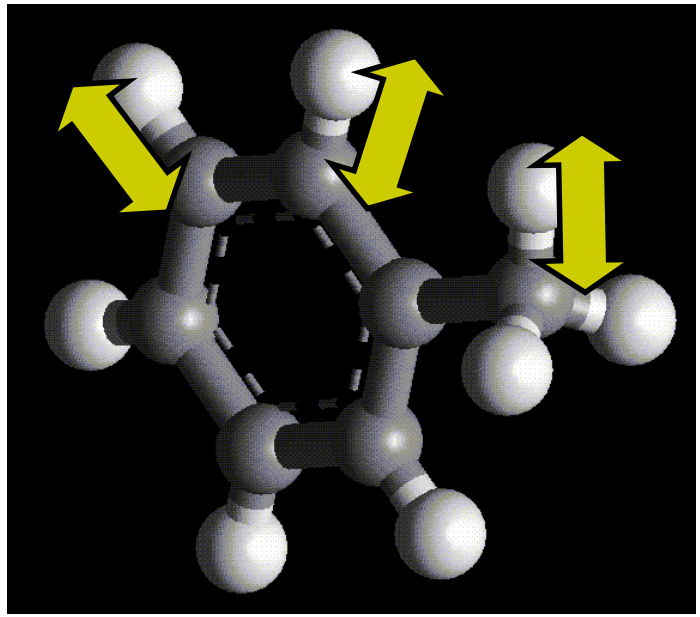
Time scales

- Time scale of biological events

Motion	Time Scale (sec)
Bond stretching	**10^{-14} to 10^{-13}**
Elastic vibrations	**10^{-12} to 10^{-11}**
Rotations of surface sidechains	**10^{-11} to 10^{-10}**
Hinge bending	**10^{-11} to 10^{-7}**
Rotation of buried side chains	**10^{-4} to 1 sec**
Allosteric transistions	**10^{-5} to 1 sec**
Local denaturations	**10^{-5} to 10 sec**

The 1 fs Time Barrier

- Dynamics simulations are limited by the highest frequency vibration
- Ideally the timestep should be 1/10 highest frequency
- In most cases C-H bond stretching (10^{-14} s) is the fastest mode



SPEED LIMIT

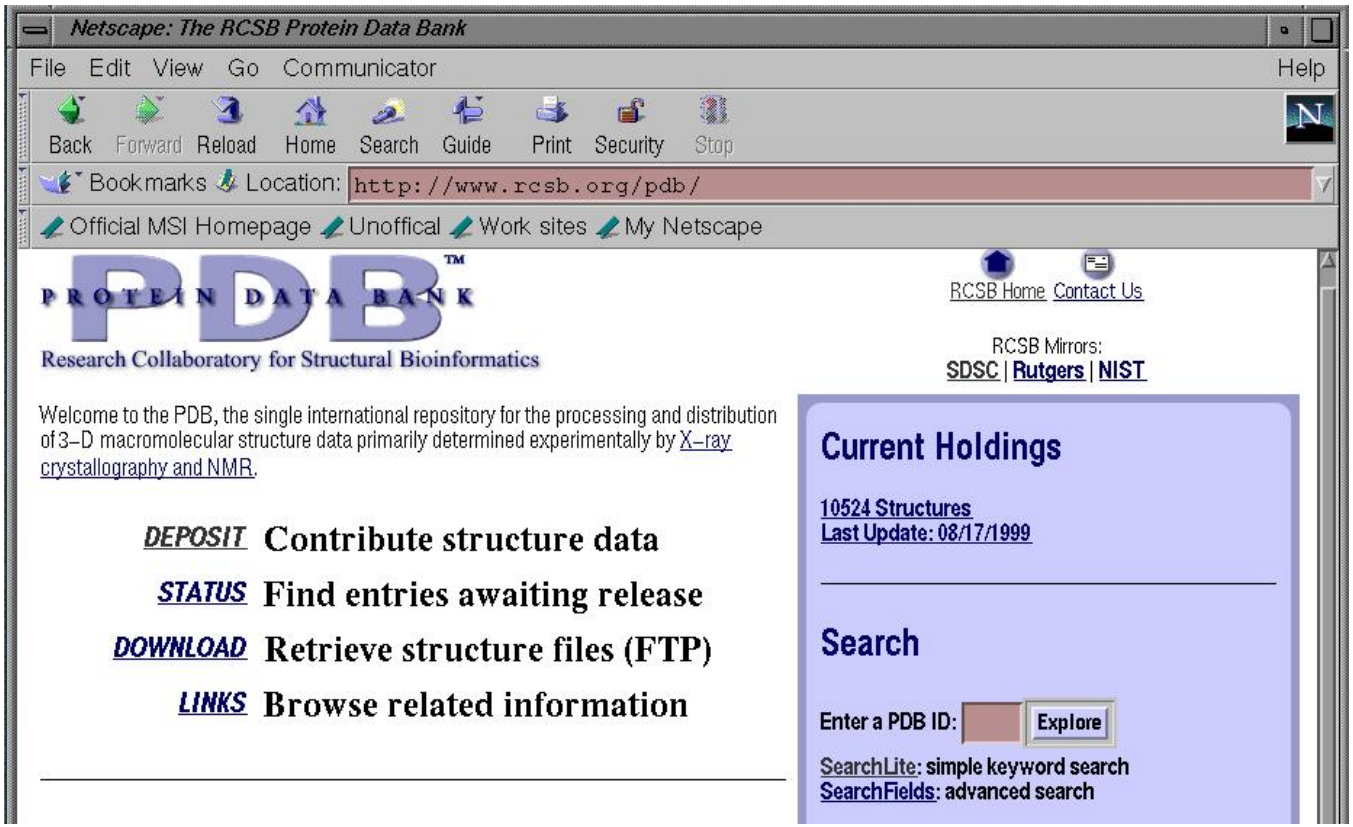
1 fs

Steps in a Typical MD Simulation

- 1. Prepare molecule
  - Read in pdb and psf file
- 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
  - Collect your data
  - Evaluate observables (macroscopic level properties)
  - Or relate to single molecule experiments

Obtaining files

- Files can be downloaded through the Web



First, You Need a PDB File

(available from **www.rcsb.org** if structure of biopolymer solved)

```
REMARK FILENAME="bpti19.pdb“
REMARK    PROTEINASE INHIBITOR (TRYPSIN)        13-MAY-87   6PTI
REMARK    BOVINE PANCREATIC TRYPSIN INHIBITOR
REMARK    BOVINE (BOS TAURUS) PANCREAS
REMARK    A.WLODAWER
REMARK DATE:26-Jun-00  21:34:42       created by user:
ATOM      1  HT1 ARG     1      13.150  -7.331  10.849  1.00  0.00      BPTI
ATOM      2  HT2 ARG     1      11.747  -7.115  11.780  1.00  0.00      BPTI

etc etc etc

ATOM    554  CA   GLY    56      15.319   0.828  11.790  1.00 17.33      BPTI
ATOM    555  C    GLY    56      16.029  -0.385  12.375  1.00 18.91      BPTI
ATOM    556  OT1 GLY    56      15.443  -1.332  12.929  1.00 21.00      BPTI
ATOM    557  OT2 GLY    56      17.308  -0.138  12.617  1.00 21.95      BPTI
END
```

What you need to know to build a realistic atomistic model of your system

- What is a force field?
- How to prepare your system for MD?
- What specific conditions (temperature, pressure, volume, etc) will be used in MD?

Energy Functions

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

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

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

U_{bond} = oscillations about the equilibrium bond length
U_{angle} = oscillations of 3 atoms about an equilibrium angle
$U_{dihedral}$ = torsional rotation of 4 atoms about a central bond
$U_{nonbond}$ = non-bonded energy terms (electrostatics and Lenard-Jones)

Topology and Parameter Files

Topology files contain:

- atom types are assigned to identify different elements and different molecular orbital environments
- charges are assigned to each atom
- connectivities between atoms are established

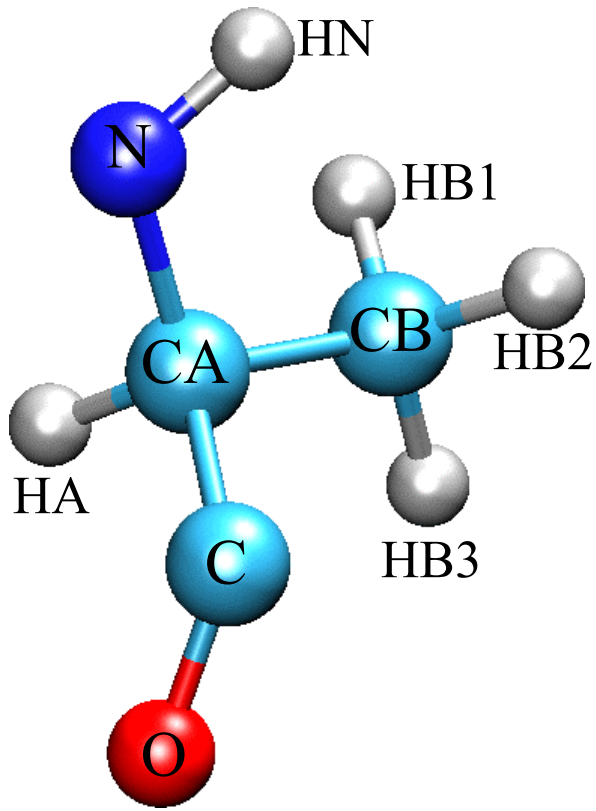
Parameter files contain:

- force constants necessary to describe the bond energy, angle energy, torsion energy, nonbonded interactions (van der Waals and electrostatics)
- suggested parameters for setting up the energy calculations

Example of Topology File

```
MASS HS     1.0080 ! thiol hydrogen
MASS C     12.0110 ! carbonyl C, peptide backbone
MASS CA    12.0110 ! aromatic C
..... (missing data here)
!-----------------------------------------------------------
AUTOGENERATE ANGLES=TRUE  DIHEDRALS=TRUE  END
!-----------------------------------------------------------
RESIDUE ALA

GROUP
    ATOM  N    TYPE=NH1  CHARGE=   -.4700  END !
    ATOM  HN   TYPE=H    CHARGE=    .3100  END  !
    ATOM  CA   TYPE=CT1  CHARGE=    .0700  END !
    ATOM  HA   TYPE=HB   CHARGE=    .0900  END !
GROUP                                          !
    ATOM  CB   TYPE=CT3  CHARGE=   -.2700  END !
    ATOM  HB1  TYPE=HA   CHARGE=    .0900  END !
    ATOM  HB2  TYPE=HA   CHARGE=    .0900  END !
    ATOM  HB3  TYPE=HA   CHARGE=    .0900  END !
GROUP                  !
    ATOM  C    TYPE=C    CHARGE=    .5100  END
    ATOM  O    TYPE=O    CHARGE=   -.5100  END
!END GROUP
   BOND  CB   CA
   BOND  N    HN
   BOND  N    CA
   BOND  O    C
   BOND  C    CA
   BOND  CA   HA
   BOND  CB   HB1
   BOND  CB   HB2
   BOND  CB   HB3
   DONOR  HN   N
   ACCEPTOR  O    C
 END {ALA }
```

```
          |
        N--HN
          |   HB1
          |  /
    HA-CA--CB-HB2
          |  \
          |   HB3
        O=C
          |
```

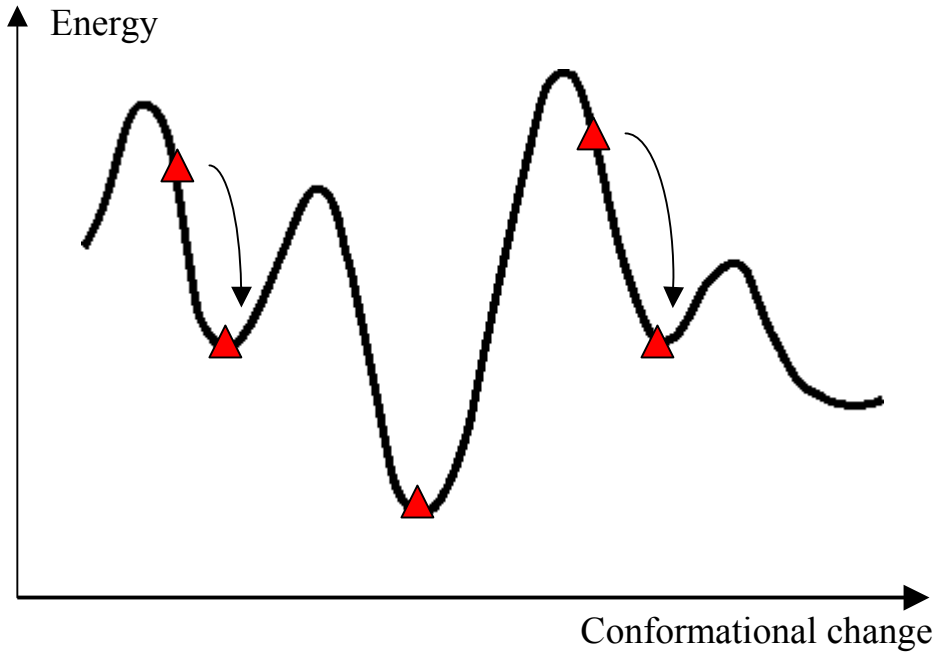


Example of Parameter File

```
!BOND PARAMETERS:  Force Constant, Equilibrium Radius
 BOND  C    C       600.000 {SD=.022}      1.335 ! ALLOW ARO HEM
 BOND  CA   CA    305.000 {SD=.031}      1.375 ! ALLOW   ARO

!ANGLE PARAMETERS:  Force Constant, Equilibrium Angle,
Urie-Bradley Force Const., U.-B. equilibrium  (if any)
 ANGLE  CA   CA   CA       40.00 {SD=.086}    120.0000    UB  35.000  2.416
 ANGLE  CP1  N    C        60.00 {SD=.070}    117.0000 ! ALLOW PRO

!DIHEDRAL PARAMETERS:  Energy Constant, Periodicity, Phase Shift, Multiplicity
 DIHEDRAL  C    CT2  NH1  C          1.60 {SD=.430}    1    180.0000 ! ALLOW PEP
 DIHEDRAL  C    N     CP1   C          .80 {SD=.608}    3       .0000 ! ALLOW PRO PEP

!IMPROPER PARAMETERS:  Energy Constant, Periodicity(0), Phase Shift(0)
!   Improper angles are introduced for PLANARITY maintaining
 IMPROPER  HA   C    C    HA        20.00 {SD=.122}    0       .0000 ! ALLOW   PEP POL ARO
 IMPROPER  HA   HA   C    C         20.00 {SD=.122}    0    180.0000 ! ALLOW   PEP POL ARO

! -----NONBONDED-LIST-OPTIONS-----------------------------
  CUTNB=  13.000 TOLERANCE=    .500 WMIN=   1.500 ATOM
  INHIBIT=    .250
! -----ELECTROSTATIC OPTIONS------------------------------
  EPS=   1.000 E14FAC=   1.000 CDIELECTRIC SHIFT
! -----VAN DER WAALS OPTIONS------------------------------
  VSWITCH
! -----SWITCHING /SHIFTING PARAMETERS----------------------
  CTONNB=  10.000 CTOFNB=  12.000
! -----EXCLUSION LIST OPTIONS-----------------------------
  NBXMOD=   5
! ------------
!               EPS     SIGMA    EPS(1:4)  SIGMA(1:4)

 NONBONDED  C        .1100   4.0090    .1100   4.0090 ! ALLOW   PEP POL ARO
 NONBONDED  CA       .0700   3.5501    .0700   3.5501 ! ALLOW   ARO
```

Preparing Your System for MD

Minimization



The energy of the system can be calculated using the forcefield. The conformation of the system can be altered to find lower energy conformations through a process called minimization.

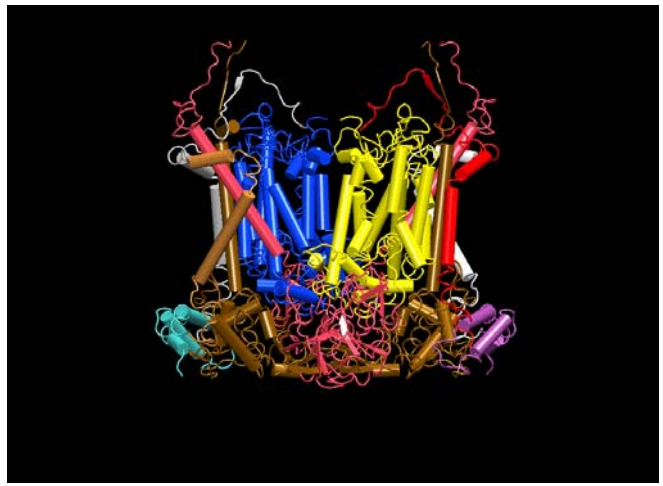
Minimization algorithms:

- steepest descent (slowly converging – use for highly restrained systems
- conjugate gradient (efficient, uses intelligent choices of search direction – use for large systems)
- BFGS (quasi-newton variable metric method)
- Newton-Raphson (calculates both slope of energy and rate of change)

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



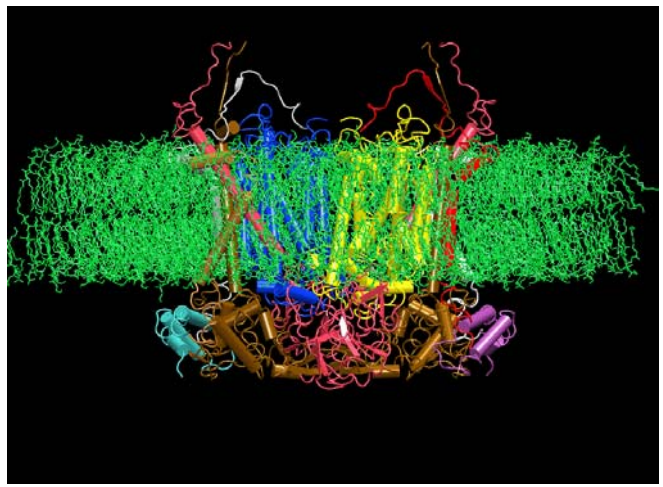
Why model solvation?
- many biological processes occur in aqueous solution
- solvation effects play a crucial role in determining molecular conformation, electronic properties, binding energies, etc

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



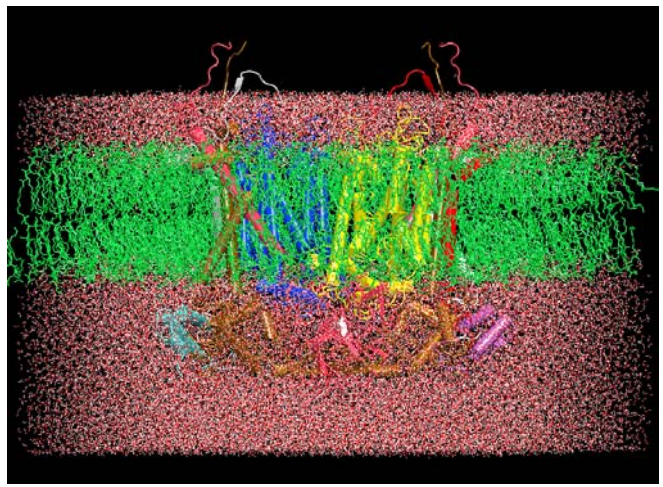
Why model solvation?
- many biological processes occur in aqueous solution
- solvation effects play a crucial role in determining molecular conformation, electronic properties, binding energies, etc

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

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

What you need to know to build a realistic atomistic model of your system

- What is a force field?
- How to prepare your system for MD?
- What specific conditions (temperature, pressure, volume, etc) will be used in MD?

Molecular Dynamics Ensembles

Constant energy, constant number of particles (NE)

Constant energy, constant volume (NVE)

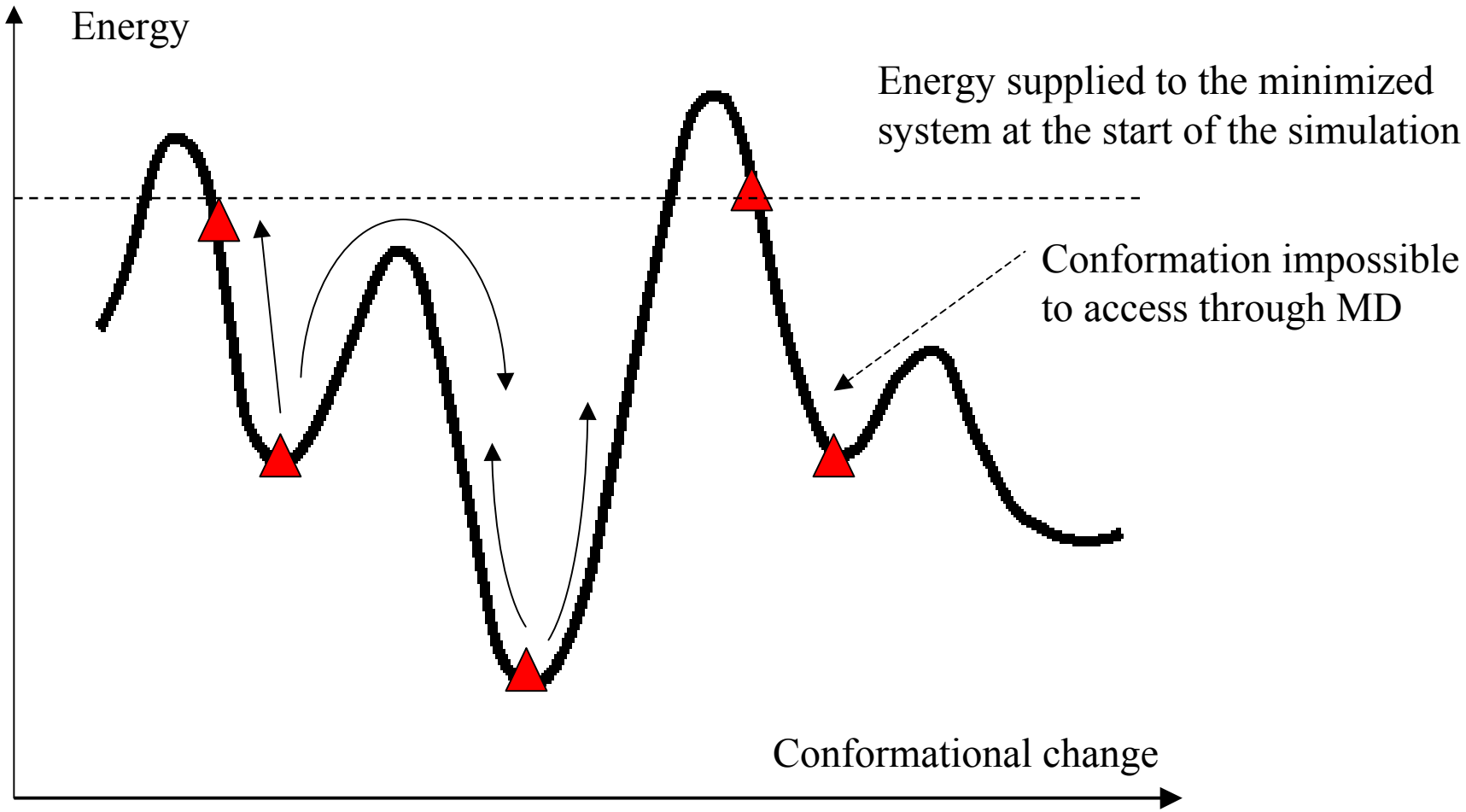
Constant temperature, constant volume (NVT)

Constant temperature, constant pressure (NPT)

Choose the ensemble that best fits your system and start the simulations

Molecular Dynamics

MD = change in conformation over time using a forcefield



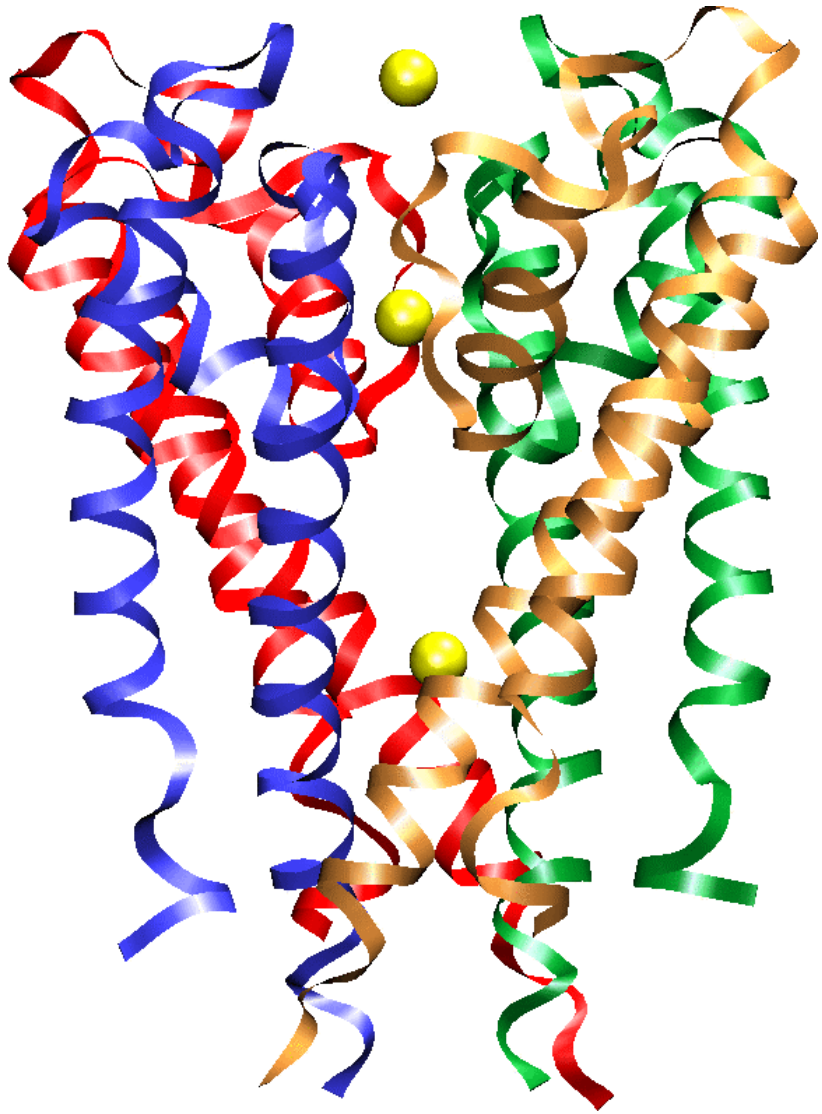
Steps in a Typical MD Simulation

- 1. Prepare molecule
  - Read in pdb and psf file
- 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

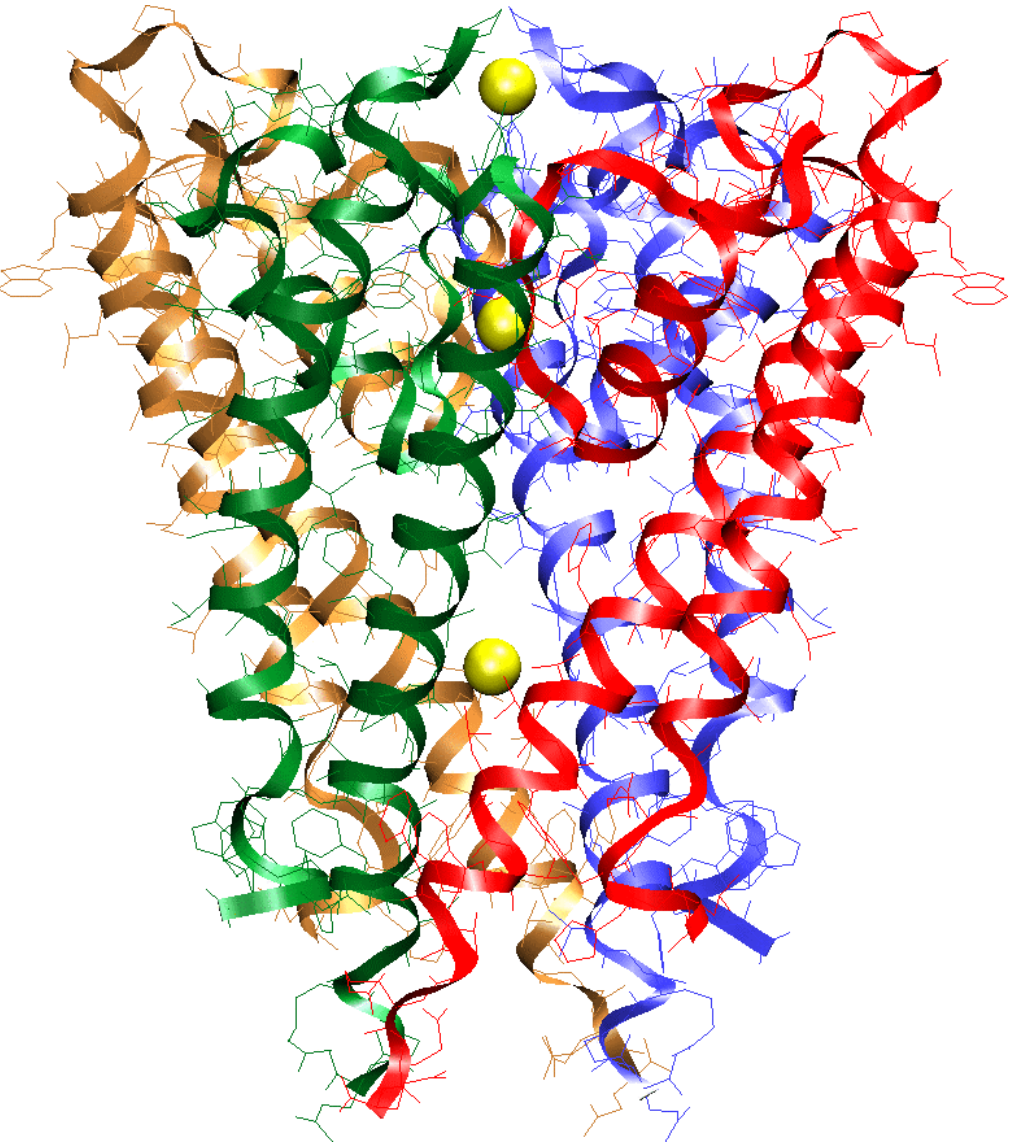
1st Example: MD Simulations of the K^+ Channel Protein

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.

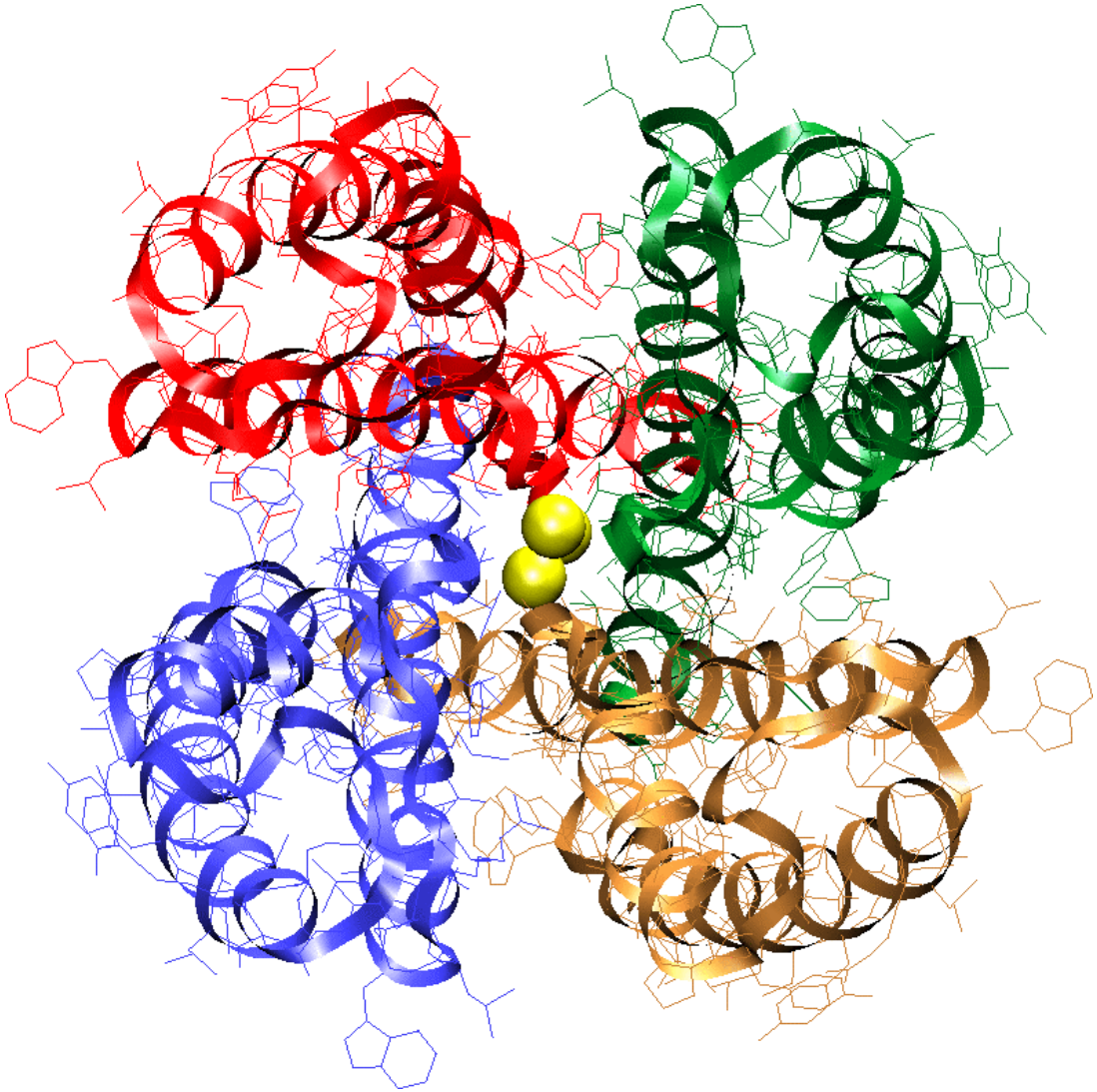


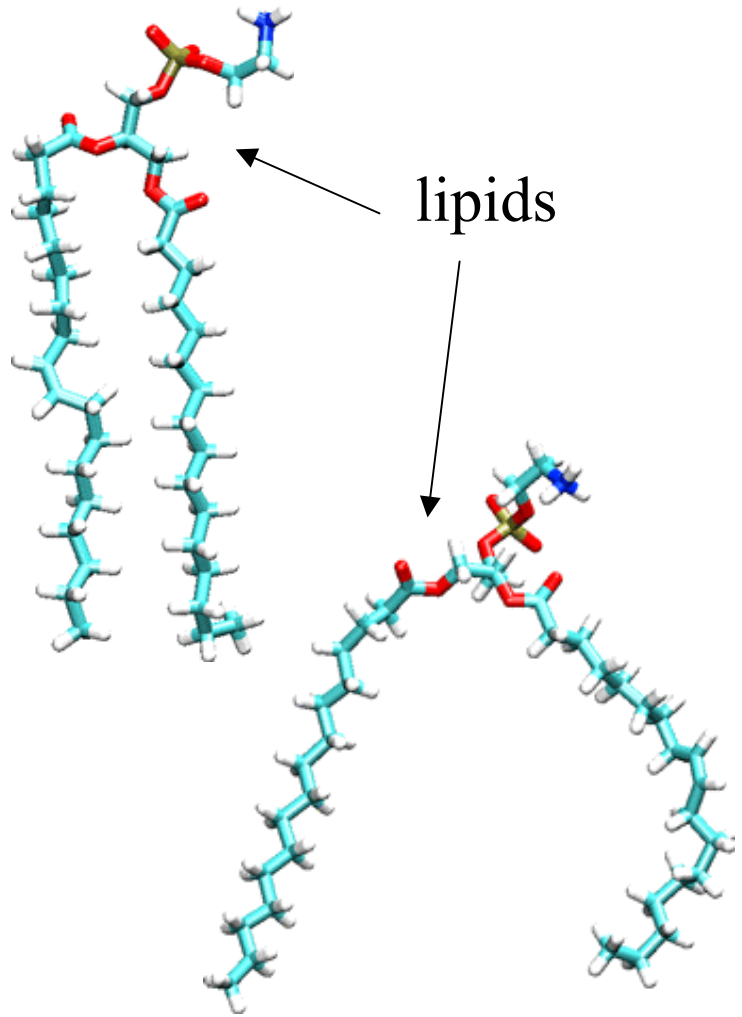
Setting up the system (1)



- retrieve the PDB (coordinates) file from the Protein Data Bank
- add hydrogen atoms using X-PLOR
- use topology and parameter files to set up the structure
- minimize the protein structure using NAMD2

Setting up the system (2)

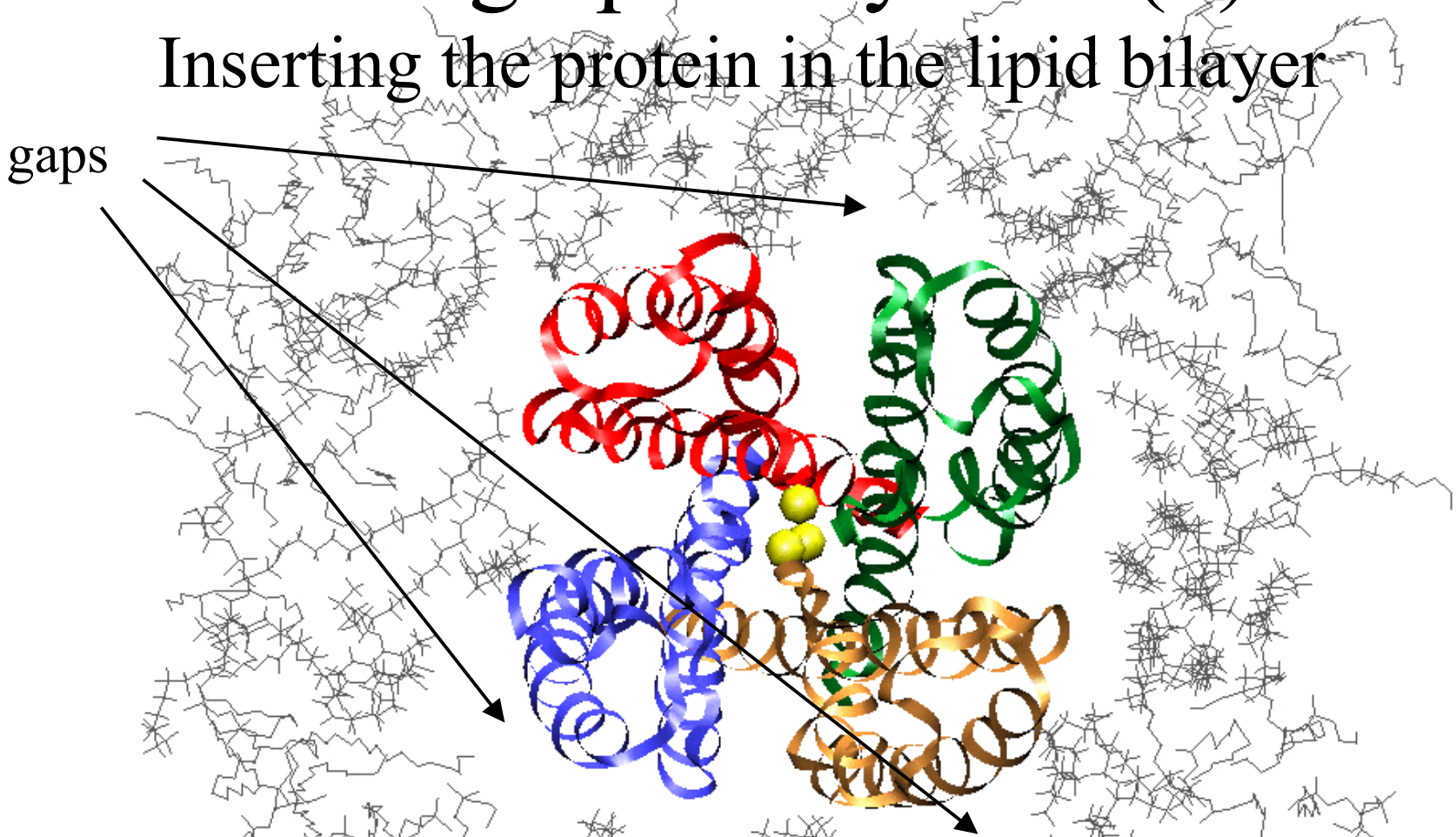




Simulate the protein in its natural environment: solvated lipid bilayer

Setting up the system (3)

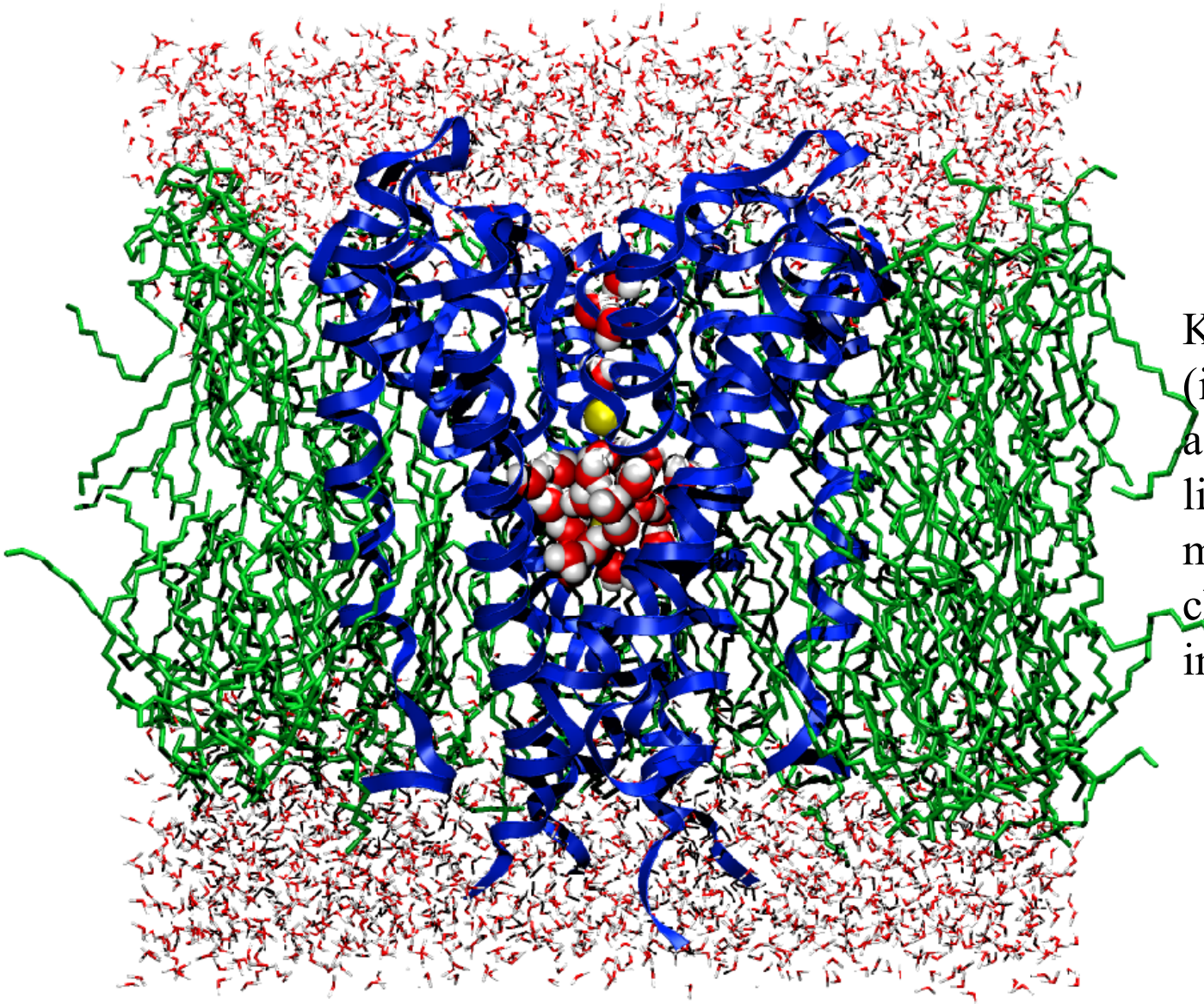
Inserting the protein in the lipid bilayer



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

The system

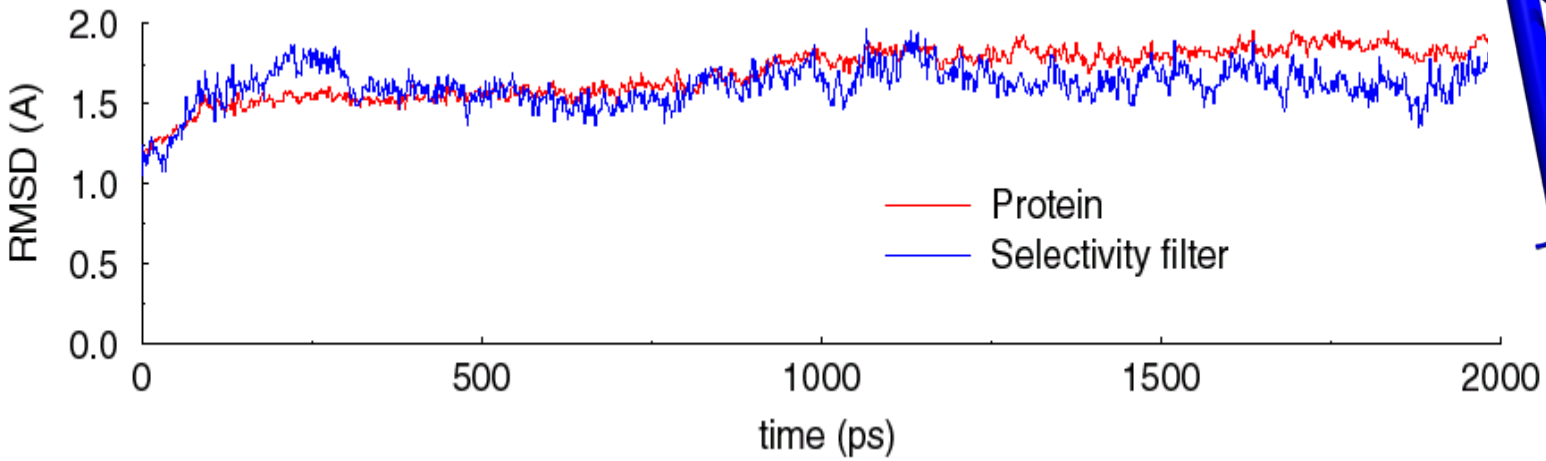


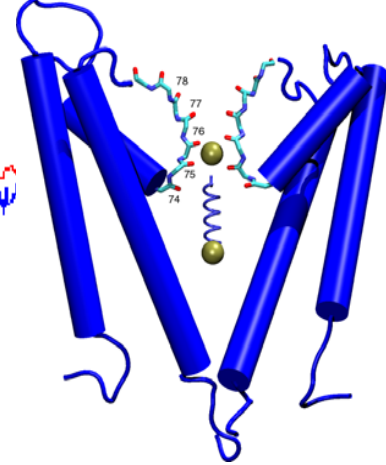
solvent

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

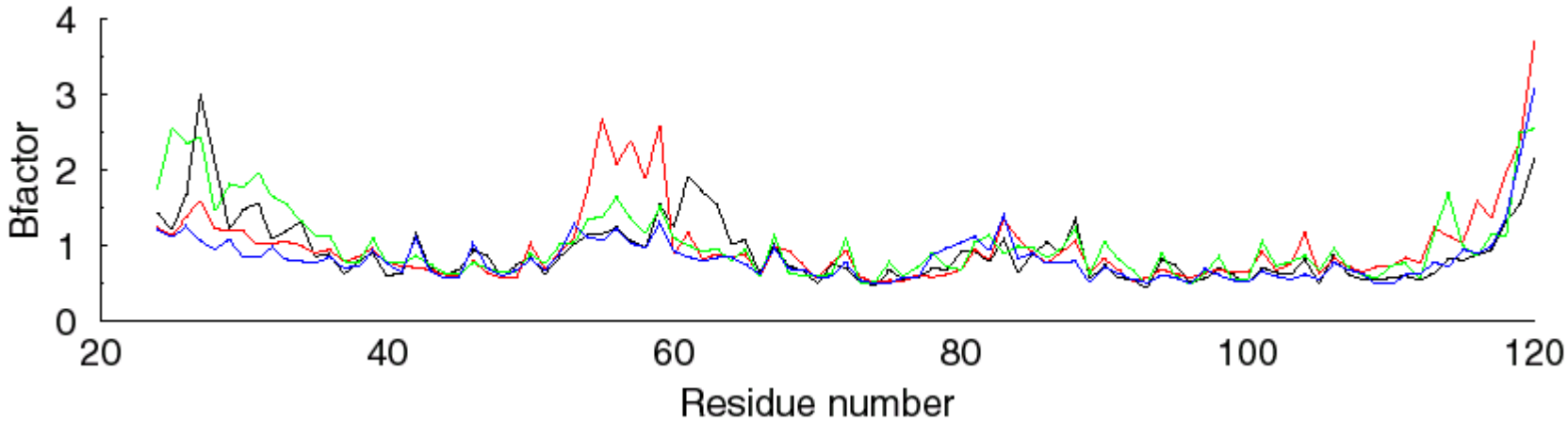
solvent

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.

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

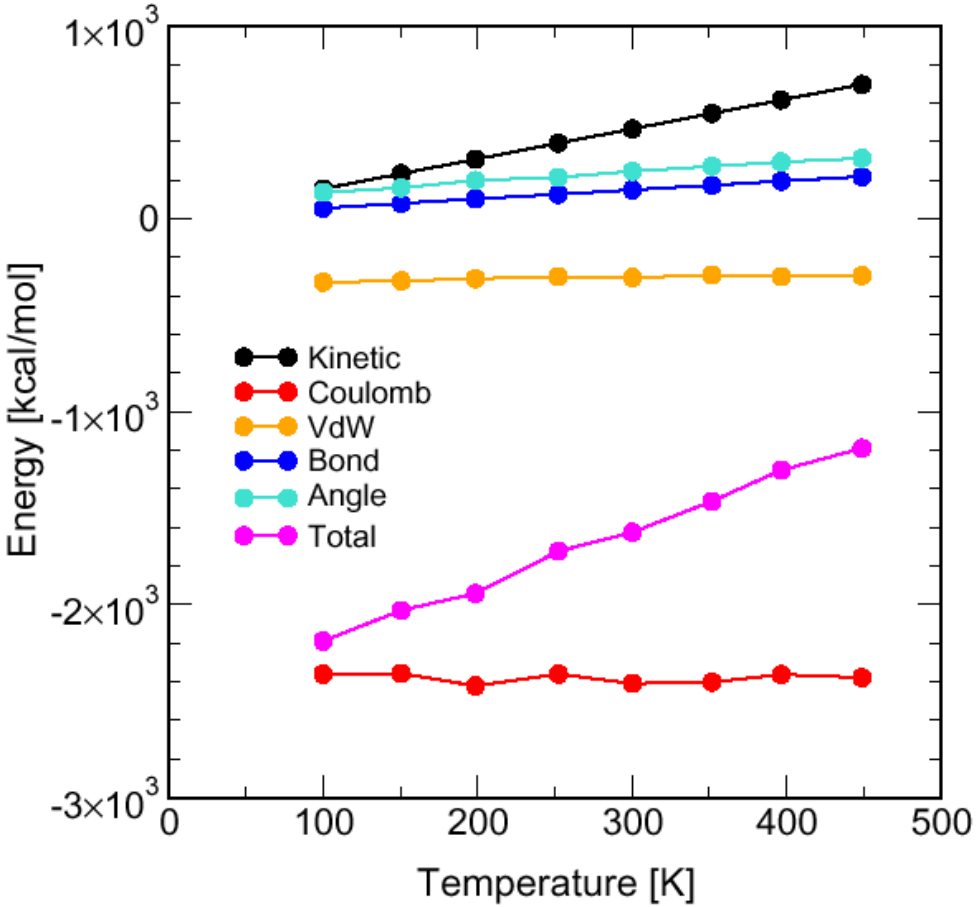
Program: NAMD2

Platform: Cray T3E (Pittsburgh Supercomputer Center)

Analysis of Trajectories

- Energetic analysis
  - Kinetic energy
  - Potential energy
  - Total energy
  - Generate ensemble averages
    - Assumption that trajectory properly sampled distribution
  - Intermolecular interactions

Energies: kinetic and potential

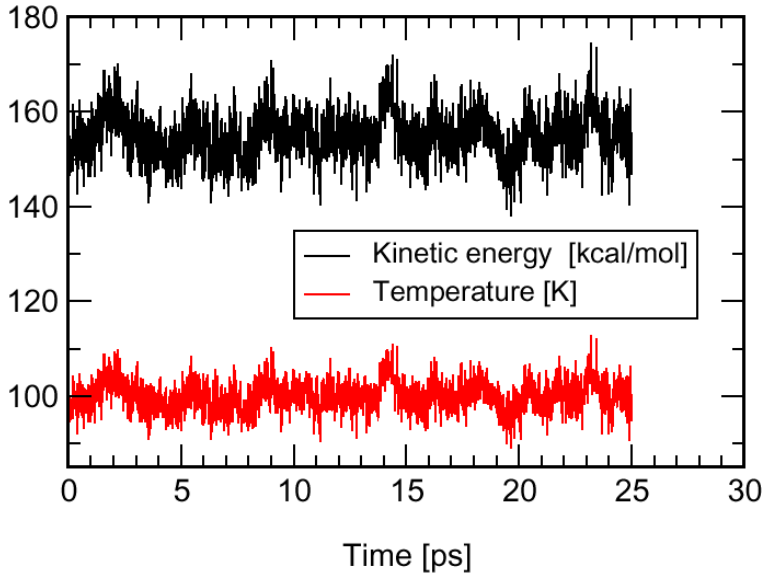


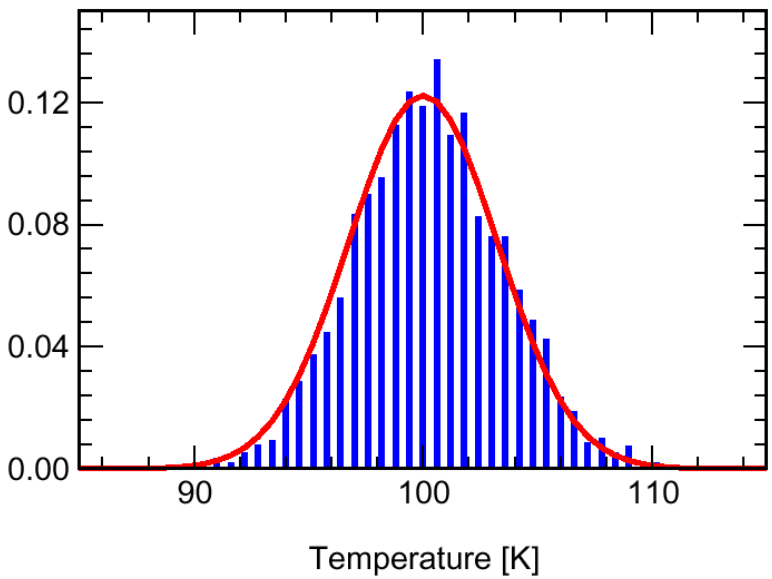
Langevin dynamics

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

$$\begin{aligned} \langle f_n(t) \rangle &= 0 \\ \langle f_n(t)\, f_n(0) \rangle &= 2 k_B\, T_0\, b_n\, m_n\, \delta(t)\,. \end{aligned}$$

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





Definition of Temperature

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

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

Temperatur Fluctuations

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 $\sigma \;=\; 3.6$.

A Brief Tutorial on NAMD

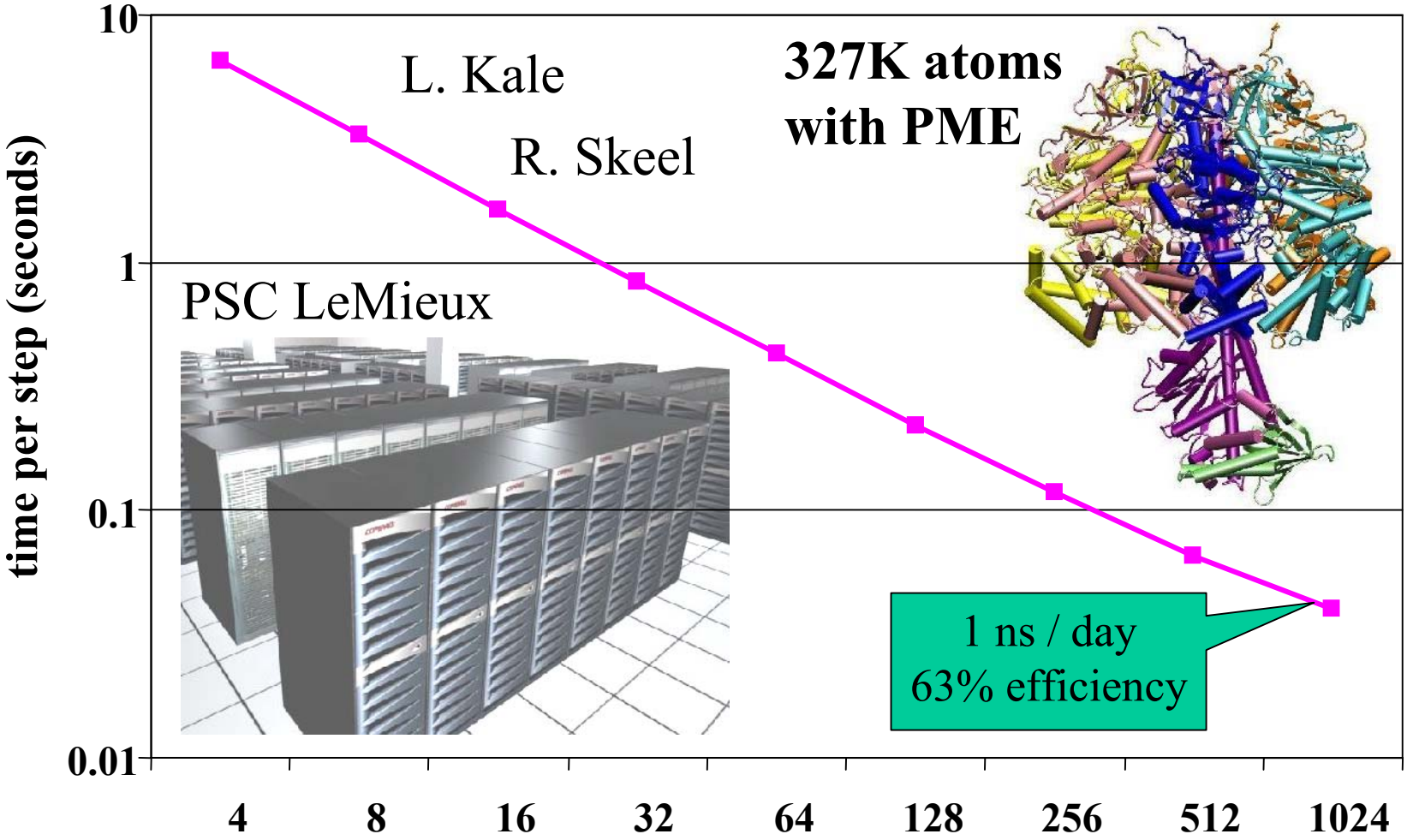
Not (Just) Another MD Program

Theoretical Biophysics Group
University of Illinois

What Is NAMD Designed For?

- Classical molecular dynamics simulations
- CHARMM energy function
- Large systems (10,000 to 1,000,000 atoms)
- Parallel supercomputers (T3E, TCS)
- Clusters of Unix workstations (Beowulf)
- Full electrostatics (PME recommended)
- Multiple timestep integration

NAMD Parallel Scalability

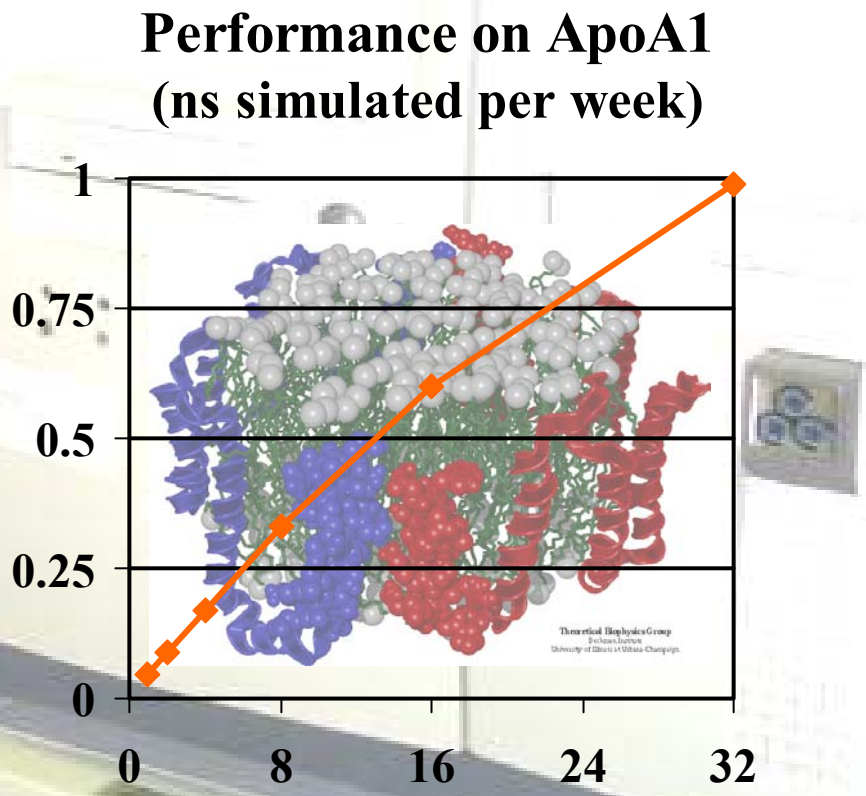


What Is NAMD Designed For?

- Classical molecular dynamics simulations
- CHARMM energy function
- Large systems (10,000 to 1,000,000 atoms)
- Parallel supercomputers (T3E, TCS)
- Clusters of Unix workstations (Beowulf)
- Full electrostatics (PME recommended)
- Multiple timestep integration

NAMD 2.3 on Athlon Cluster

- 67% efficiency on 32, *commodity hardware*.
- Equivalent to owning a 100 CPU Cray T3E for only $30K.
- Available in the lab



What Is NAMD Designed For?

- Classical molecular dynamics simulations
- CHARMM energy function
- Large systems (10,000 to 1,000,000 atoms)
- Parallel supercomputers (T3E, TCS)
- Clusters of Unix workstations (Beowulf)
- Full electrostatics (PME recommended)
- Multiple timestep integration