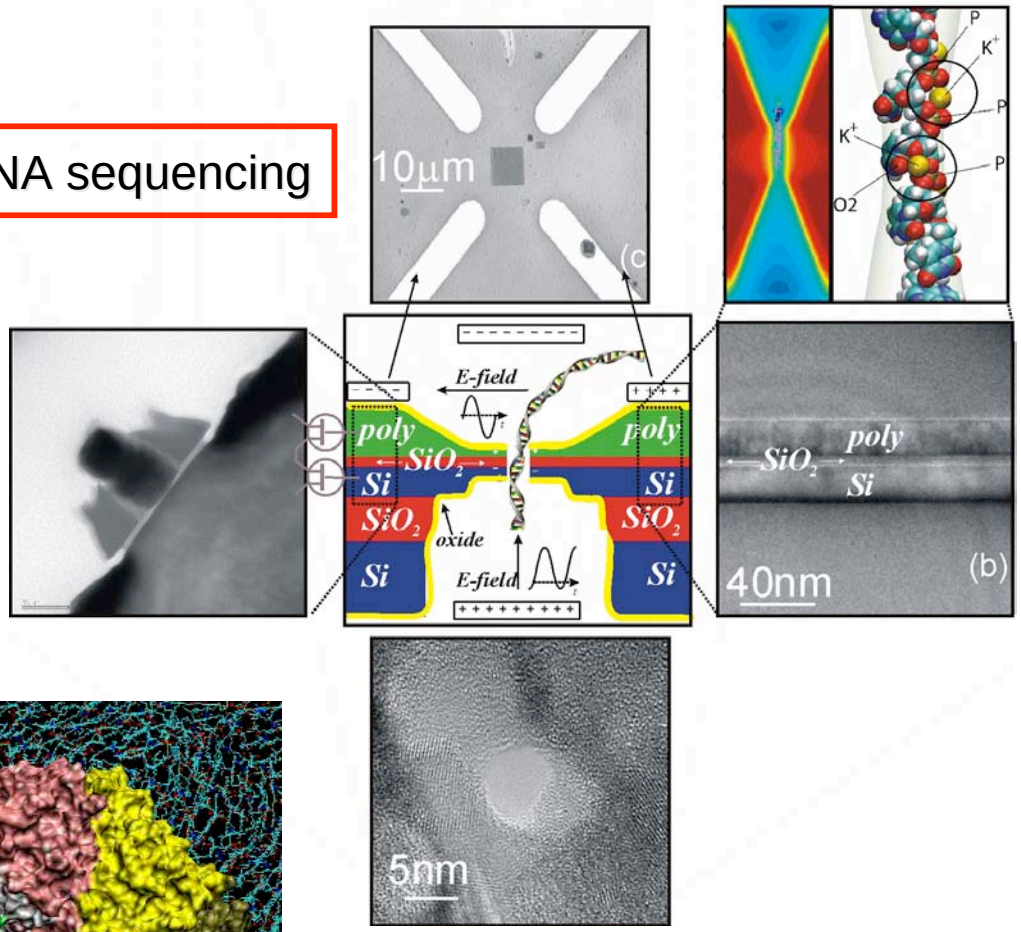
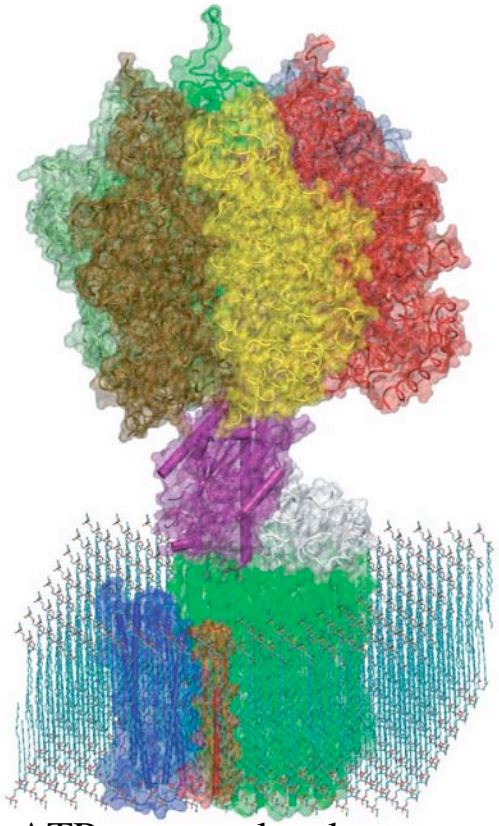
Introduction to:

NAnoscale Molecular Dynamics

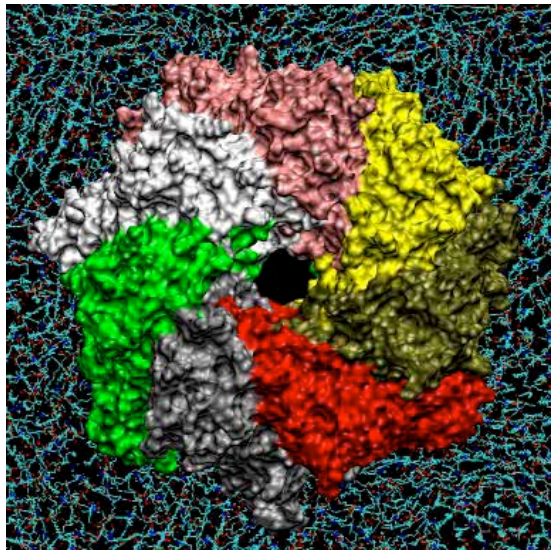
The Molecular Dynamics Method

Integrated chip for DNA sequencing



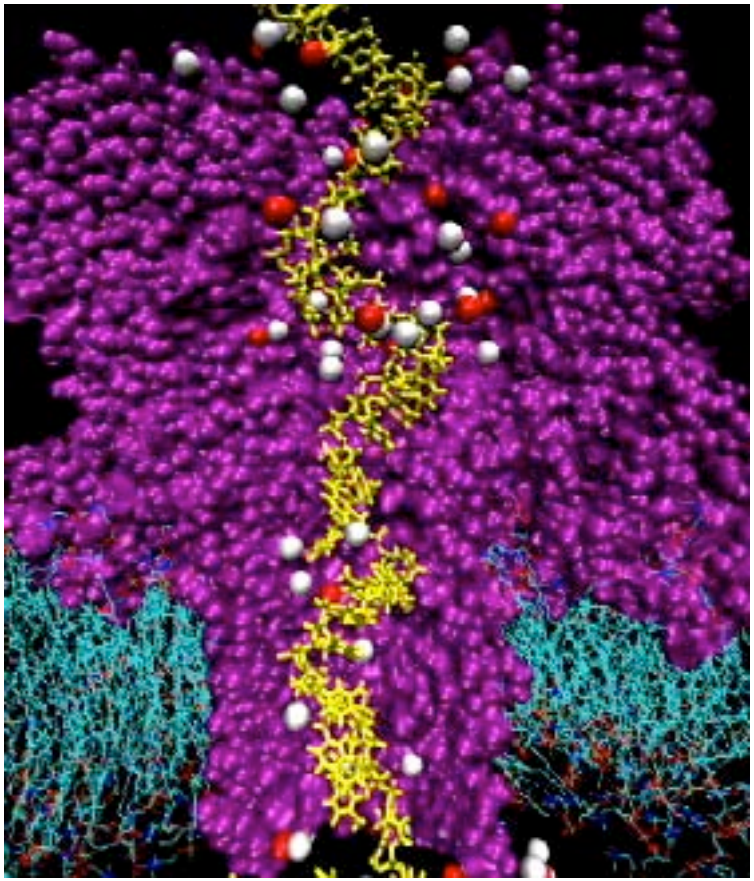


ATPase, a molecular motor that synthesizes the body's weight of ATP a day



Bacterial toxin alpha-hemolysin

Molecular Dynamics Example



5 ns, 360,000 atoms
(water not shown)

MD simulation of ionic current through membrane channel α-hemolysin blocked by DNA

What do you see?

Simulation conditions:

- 1.2 V electrostatic potential
- NVT ensemble
- 298 K
- PME 128x128x128

Classical Dynamics

F=ma 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.

Langevin Dynamics

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.

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

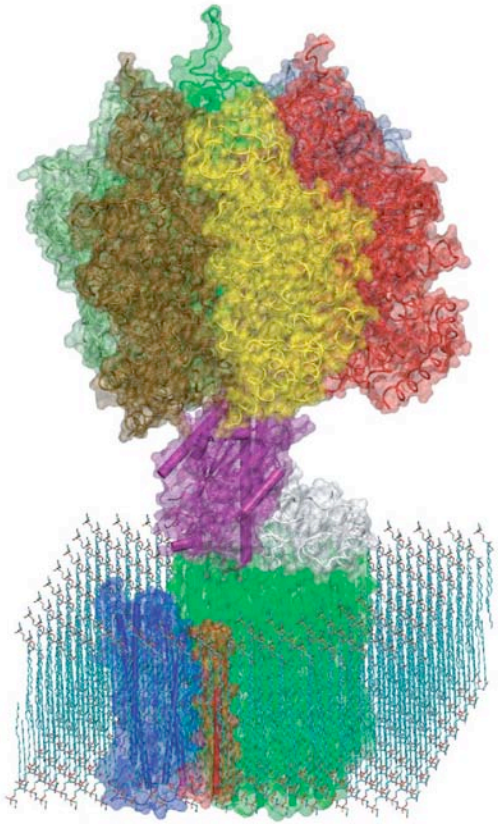
+

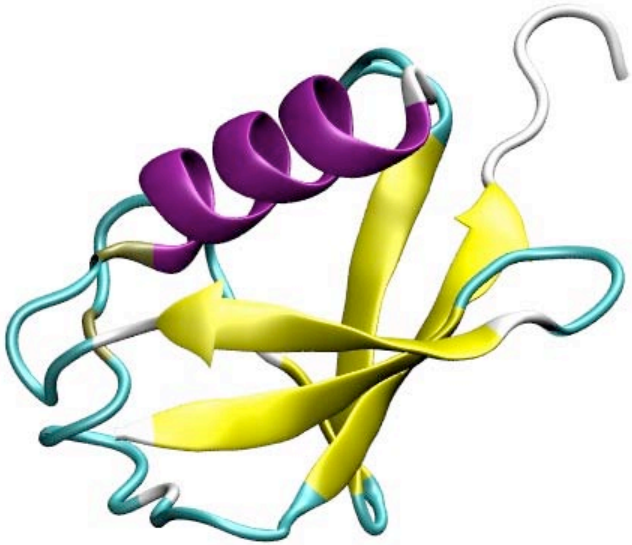
$$\mathbf{r}(t-\delta t) \approx \mathbf{r}(t) - \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2$$

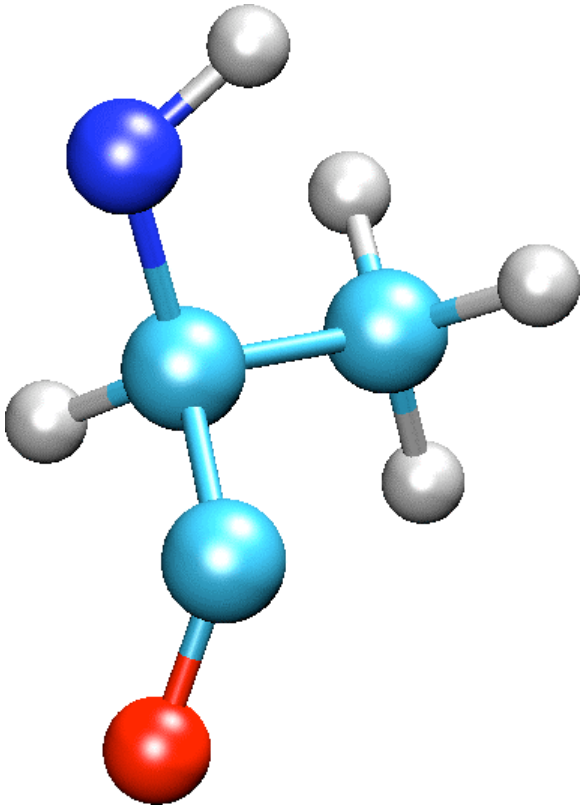
$-\vec{\nabla} U(\vec{R}) / m_i$

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

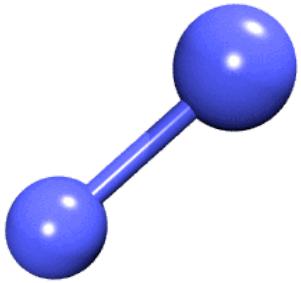
Protein Structure





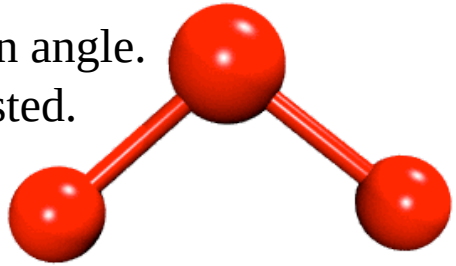


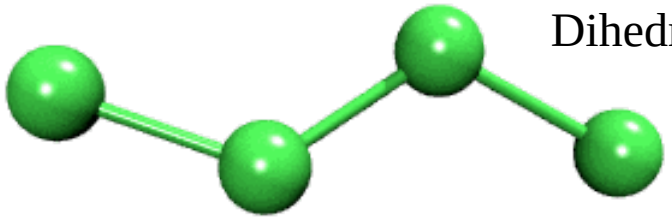
Molecular Structure (bonds, angles, etc.)



Bonds: Every pair of covalently bonded atoms is listed.

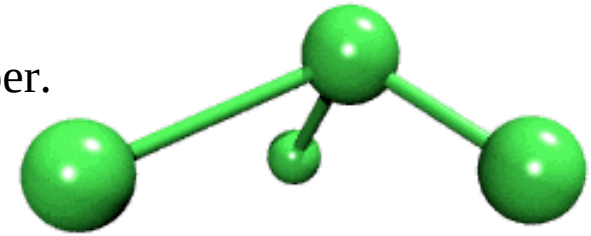
Angles: Two bonds that share a common atom form an angle.
Every such set of three atoms in the molecule is listed.





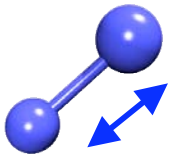
Dihedrals: Two angles that share a common bond form a dihedral.
Every such set of four atoms in the molecule is listed.

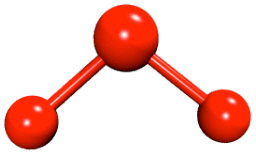
Impropers: Any *planar* group of four atoms forms an improper.
Every such set of four atoms in the molecule is listed.



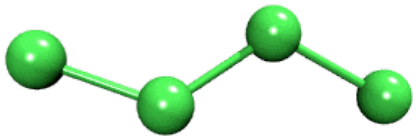
Potential Energy Function of Biopolymer

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





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

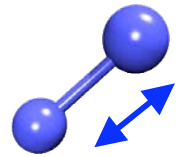


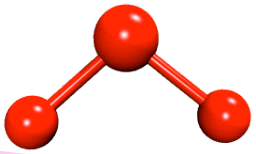
$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} +$$

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

Potential Energy Function of Biopolymer

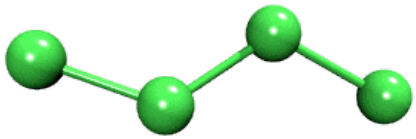
- Simple, fixed algebraic form for every type of interaction.
- Variable parameters depend on types of atoms involved.





heuristic

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

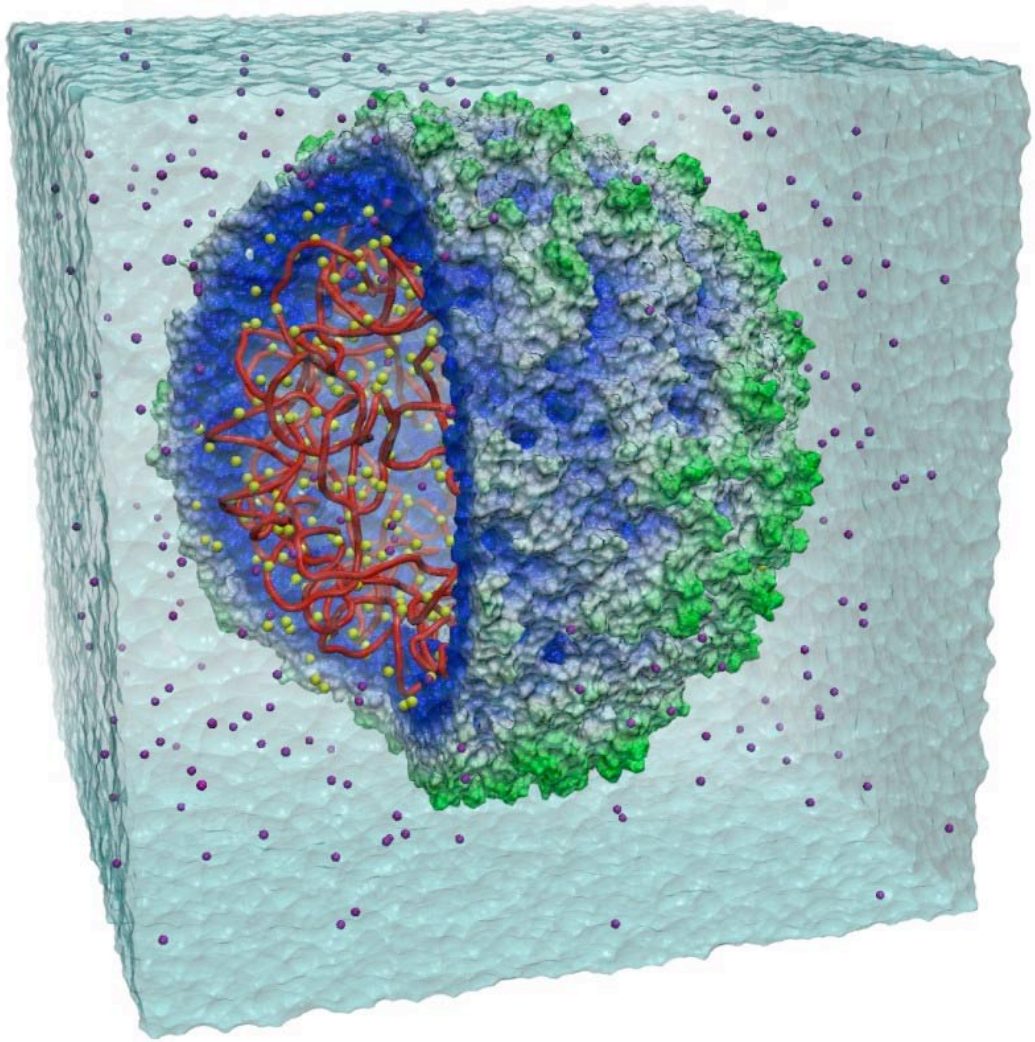


$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} +$$

from physics

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

Large is no problem. But …

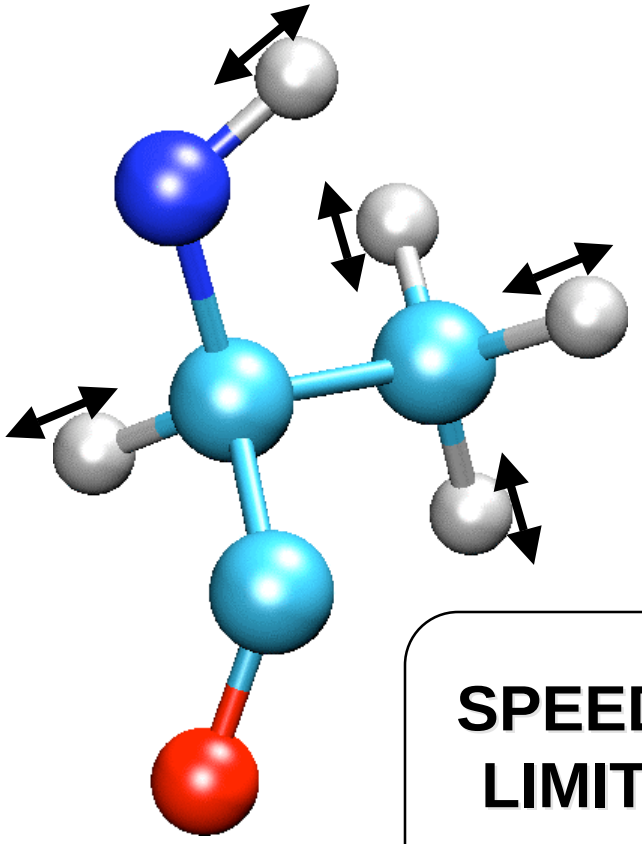


Molecular dynamics simulation of satellite tobacco mosaic virus with over 1,000,000 atoms

Massive parallel computer (PSC Lemieux)

But long is!

biomolecular timescale and timestep limits

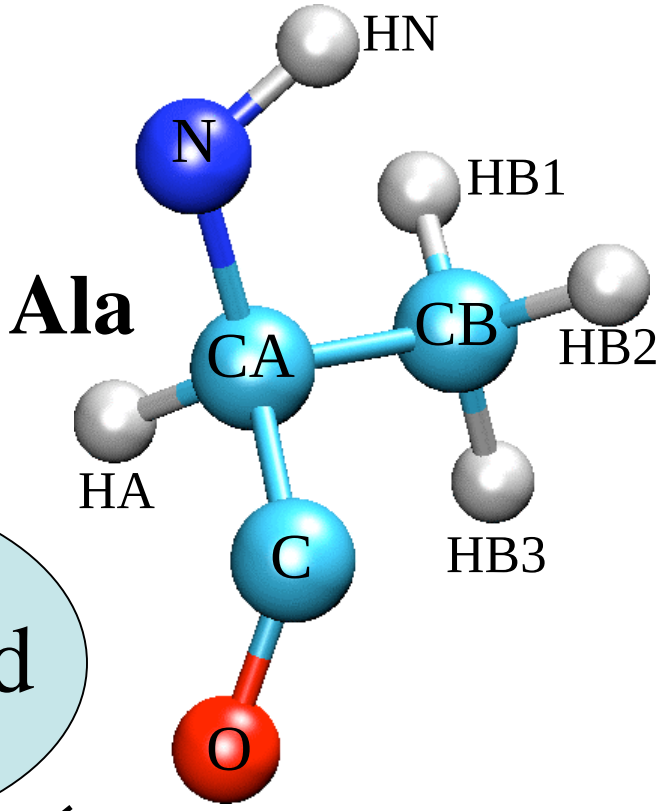


		steps
	s	10^{15}
Rotation of buried sidechains Local denaturations Allosteric transitions	ms	10^{12}
	μs	10^{9} (year)
Hinge bending Rotation of surface sidechains	ns	10^{6} (day)
Elastic vibrations	ps	10^{3}
Bond stretching		
Molecular dynamics timestep →	fs	10^{0}

SPEED LIMIT

δt = 1 fs

MD simulation basics



Atomic coordinates

Force Field

Connectivity

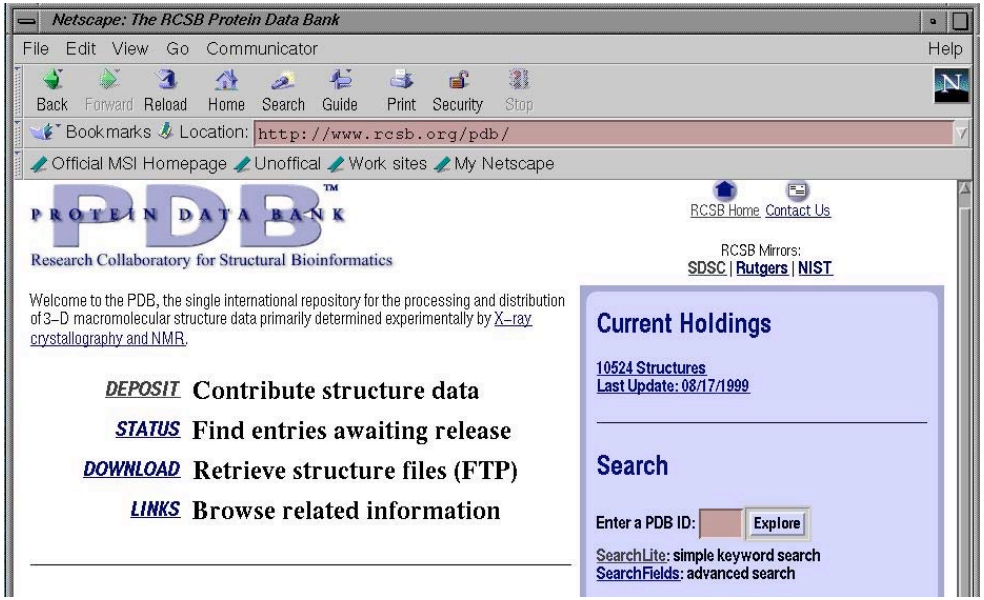
Simulation

PDB Files

a little information

- 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.)



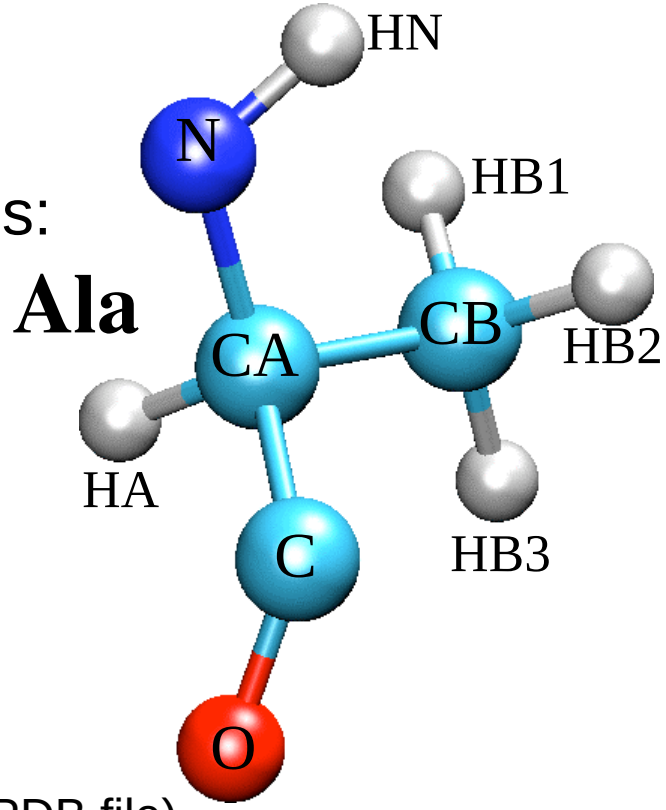
PDB File

(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
```

Topology Files

atomic properties (mass, charge, type)

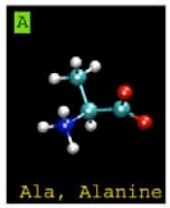


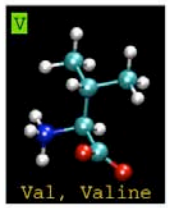
- Every possible residue is listed.
- Provides all static atom-specific values:
  - atom name (N, C, CA)
  - atom type (NH1, CT1)
  - residue name (ALA, HIS)
  - residue id (integer)
  - segment id (6PTI)
  - atomic mass (in atomic mass units)
  - partial charge (in electronic charge units)
- What is not in the topology 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)

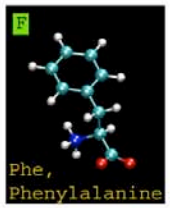
Protein Primary Structure

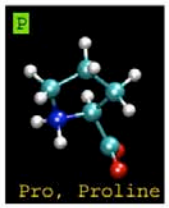
The twenty amino acids

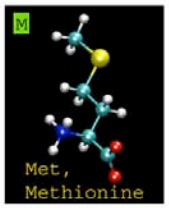
non-polar

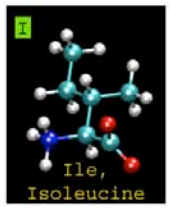


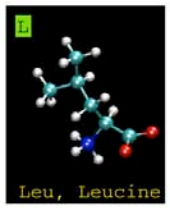




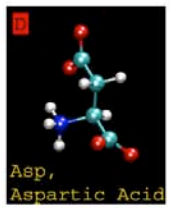


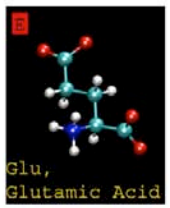


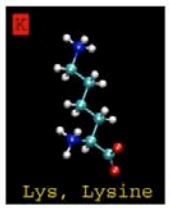


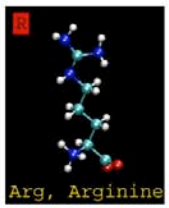


charged

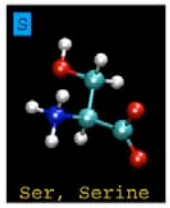


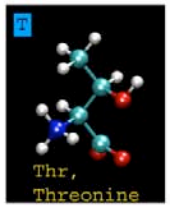


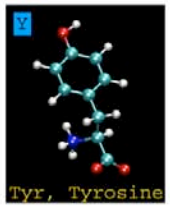


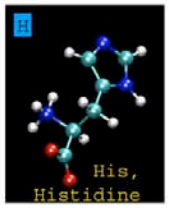


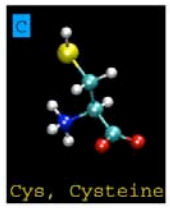
polar

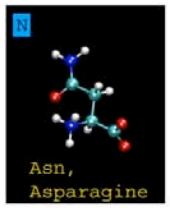


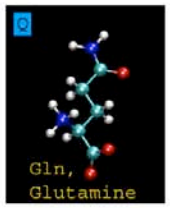


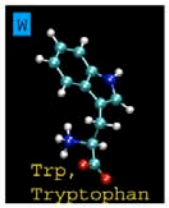


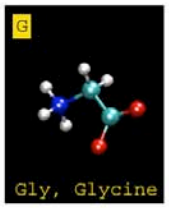












Introduction to Protein Structure, *2nd ed.*
Carl Branden & John Tooze, 1999

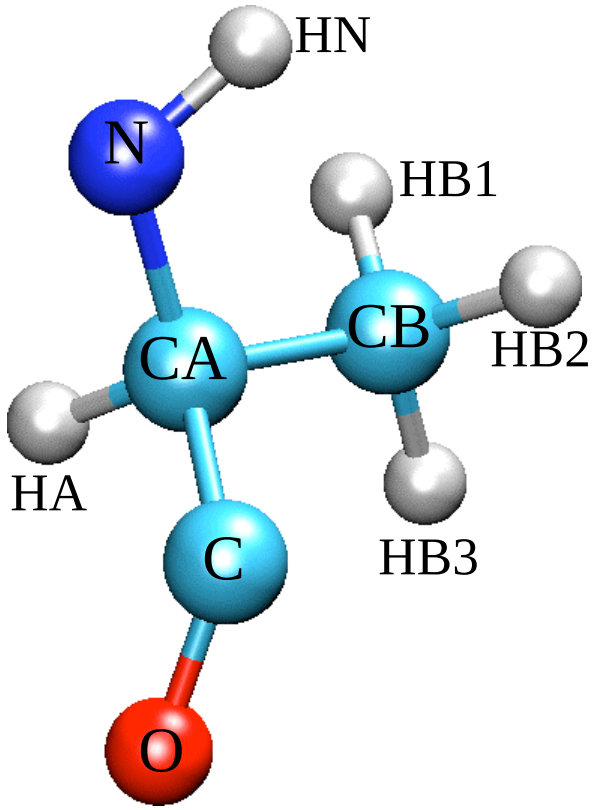
Protein Structure and Function,
Greg Petsko & Dagmar Ringe, 2003

Molecular Biology of The Cell
Alberts, Johnson, Lewis, Raff, Roberts, Walter, 2002

URL: http://lectures.molgen.mpg.de/ProteinStructure

Example of Topology File

```
...
MASS 121   CTL1 12.011000 C ! sp3 carbon with 1 H  (-CH1-)
MASS 122   CTL2 12.011000 C ! carbon of methylene group (-CH2-)
MASS 123   CTL3 12.011000 C ! carbon of methyl group (-CH3)
MASS 124   CTL5 12.011000 C ! carbon of methyl group (-CH3) for tetramethylammonium
MASS 125   CEL1 12.011000 C ! for alkene; RHC=CR
MASS 126   CEL2 12.011000 C ! for alkene; H2C=CR
MASS 140   NTL  14.007000 N ! ammonium nitrogen

...
RESI ALA         0.00
GROUP
ATOM N    NH1    -0.47         !    |
ATOM HN   H       0.31         !  HN-N
ATOM CA   CT1     0.07         !    |    HB1
ATOM HA   HB      0.09         !    |   /
GROUP                          !  HA-CA--CB-HB2
ATOM CB   CT3    -0.27         !    |   \
ATOM HB1  HA      0.09         !    |    HB3
ATOM HB2  HA      0.09         !   O=C
ATOM HB3  HA      0.09         !    |
GROUP                          !
ATOM C    C       0.51
ATOM O    O      -0.51
BOND CB CA  N  HN  N  CA
BOND C  CA  C  +N  CA HA  CB HB1  CB HB2  CB HB3
DOUBLE O  C
IMPR N -C CA HN  C CA +N O
```



PDB FILE (sequence of residues)

+

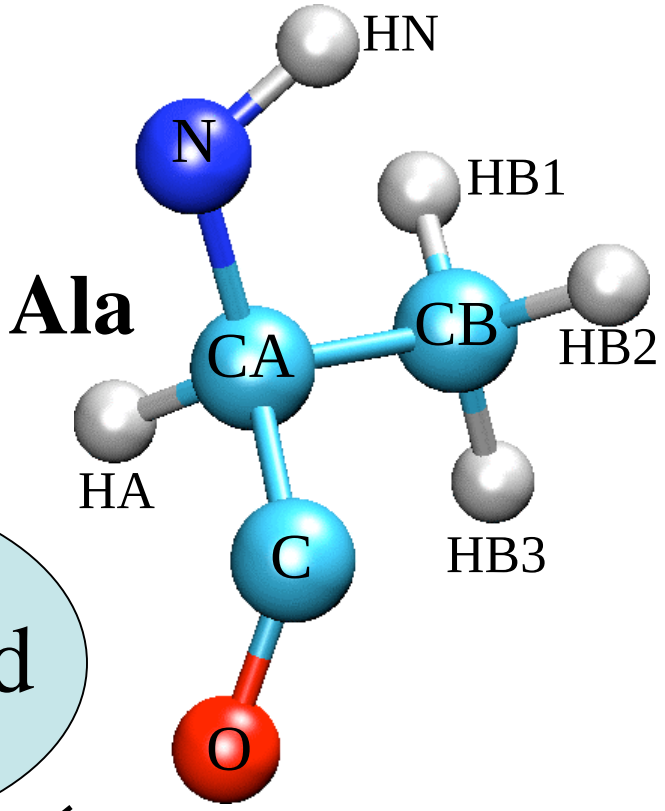
TOPOLOGY (of each residue)

=

CONNECTIVITY

(of entire protein)

MD simulation basics



Atomic coordinates

Force Field

Connectivity

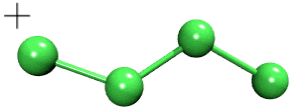
Simulation

Parameter Files

force constants for all types of interactions

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

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



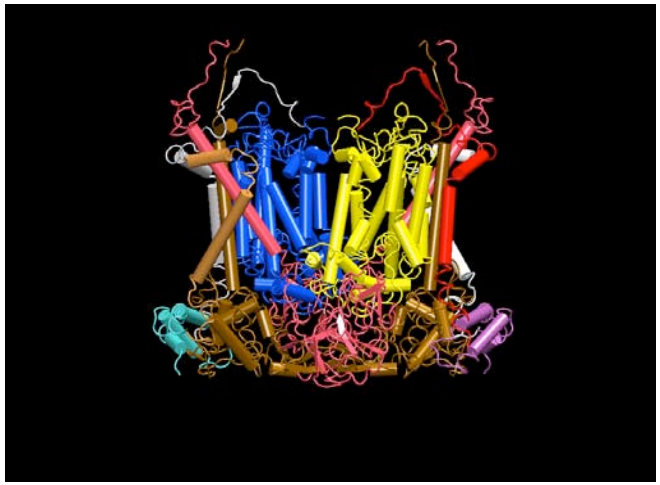
$$\underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$

```
BONDS
!V(bond) = Kb(b - b0)**2

!atom type Kb         b0
!
CE1  CE1   440.000     1.3400   !
                ! for butene; from propene, yin/adm jr., 12/95
CE1  CE2   500.000     1.3420   !
                ! for propene, yin/adm jr., 12/95

ANGLES
!
!V(angle) = Ktheta(Theta - Theta0)**2
!
!V(Urey-Bradley) = Kub(S - S0)**2
!
!
!atom types     Ktheta    Theta0   Kub     S0
!
CA   CA   CA    40.000    120.00   35.00   2.41620 ! ALLOW   ARO
            ! JES 8/25/89
CE1  CE1  CT3    48.00    123.50   !
                          ! for 2-butene, yin/adm jr., 12/95
CE1  CT2  CT3    32.00    112.20   !
                          ! for 1-butene; from propene, yin/adm jr., 12/95
CE2  CE1  CT2    48.00    126.00   !
```

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



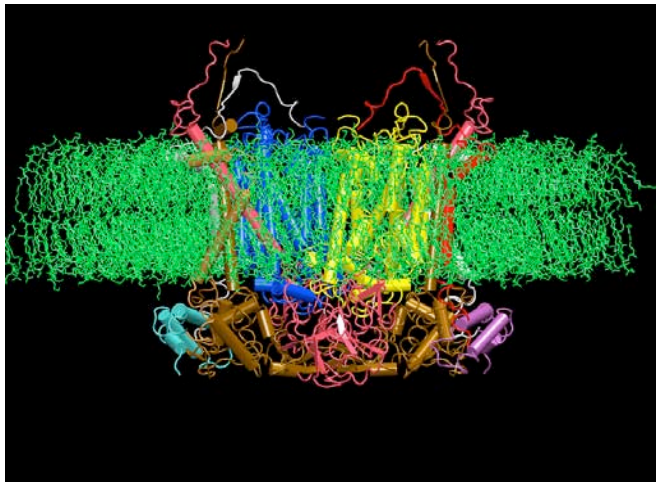
Why model solvation?
- many biological processes occur in aqueous solution
- solvation effects play a crucial role in determining molecular conformation, electronic properties, binding energies, etc

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



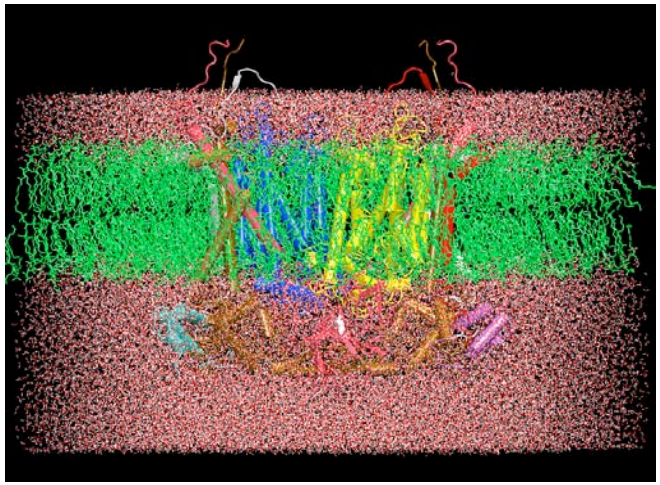
Why model solvation?
- many biological processes occur in aqueous solution
- solvation effects play a crucial role in determining molecular conformation, electronic properties, binding energies, etc

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

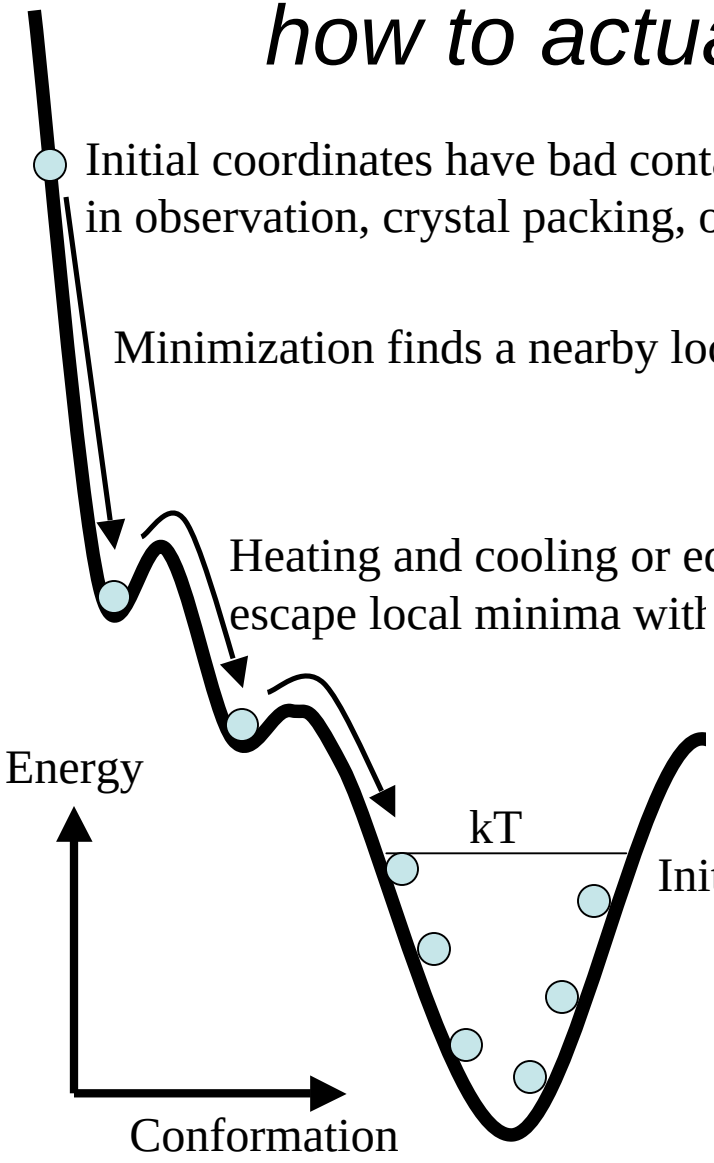


Why model solvation?
- many biological processes occur in aqueous solution
- solvation effects play a crucial role in determining molecular conformation, electronic properties, binding energies, etc

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

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)

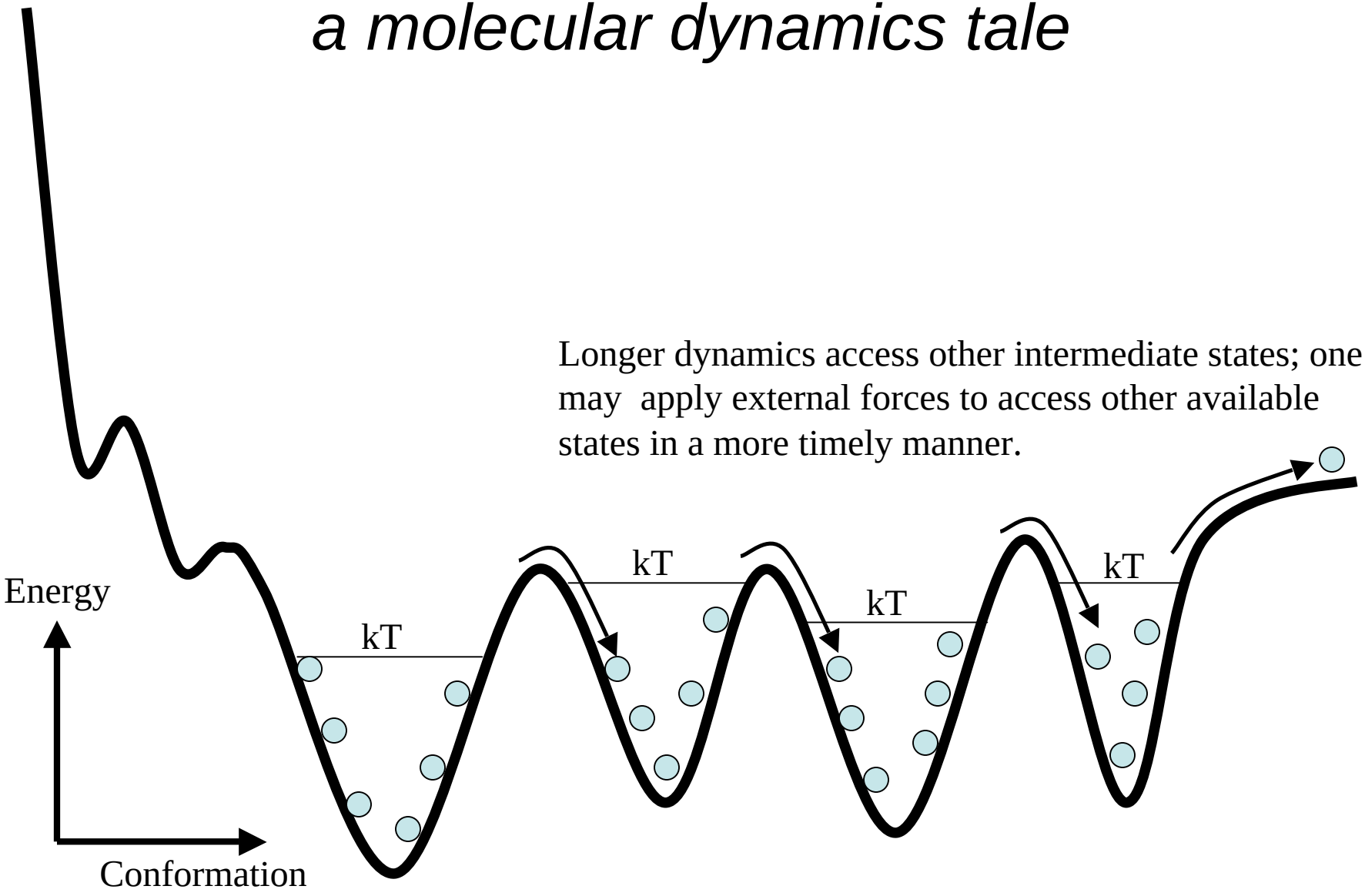
Minimization finds a nearby local minimum.

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

Initial dynamics samples thermally accessible states.

From the Mountains to the Valleys

a molecular dynamics tale



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

Molecular Dynamics Ensembles

Constant energy, constant number of particles (NE)

Constant energy, constant volume (NVE)

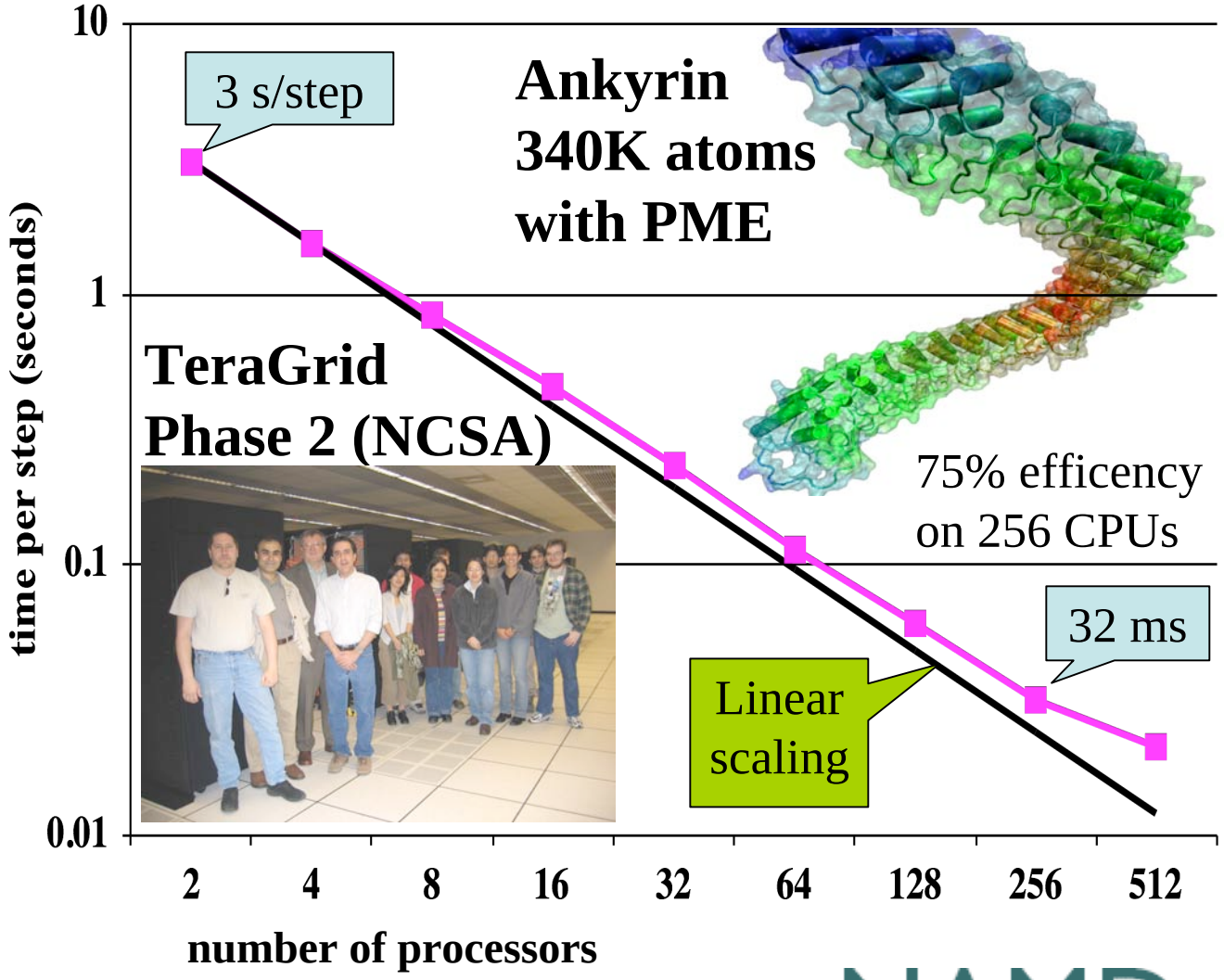
Constant temperature, constant volume (NVT)

Constant temperature, constant pressure (NPT)

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

NAMD: The Program we will Use

NAMD programmer J. Phillips Ph.D. UIUC Physics



NAMD
Scalable Molecular Dynamics

Simulation of large biomolecular systems

2002 Gordon Bell Award for parallel scalability.

Runs at NSF centers, on clusters, and on desktop.

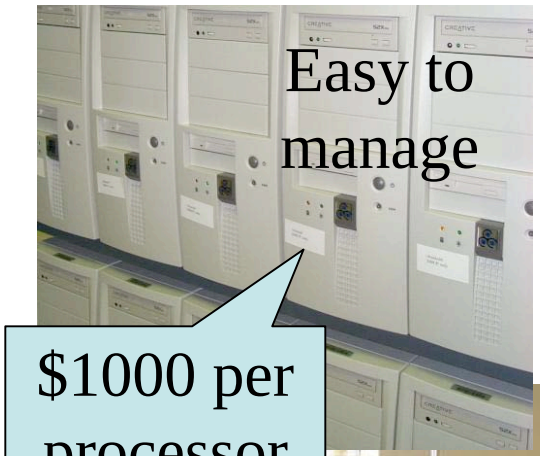
Available for **FREE** as precompiled binaries; includes source code.

10,000 registered users.

Linux Clusters 101

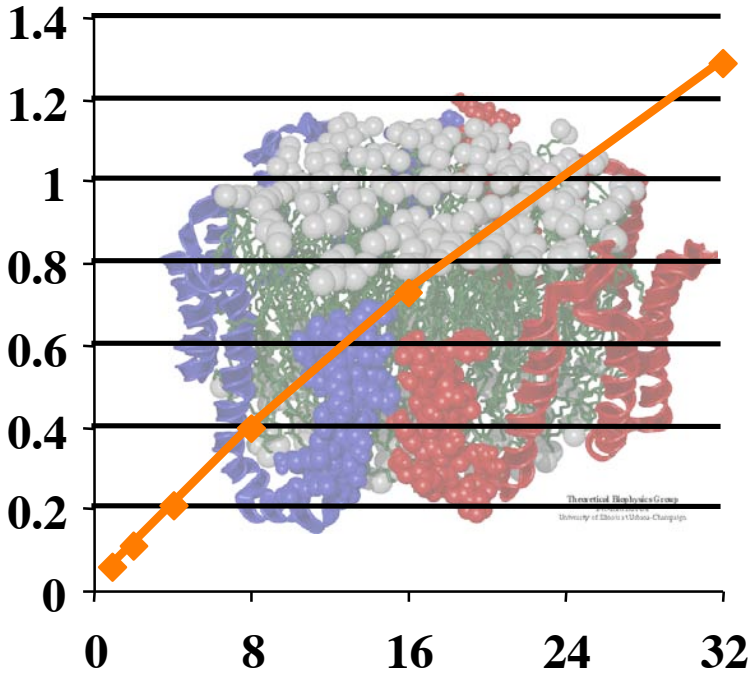
parallel computing on a professor's salary

Learn to build your own Linux cluster!



$1000 per processor

92K atoms with PME
(ns simulated per week)



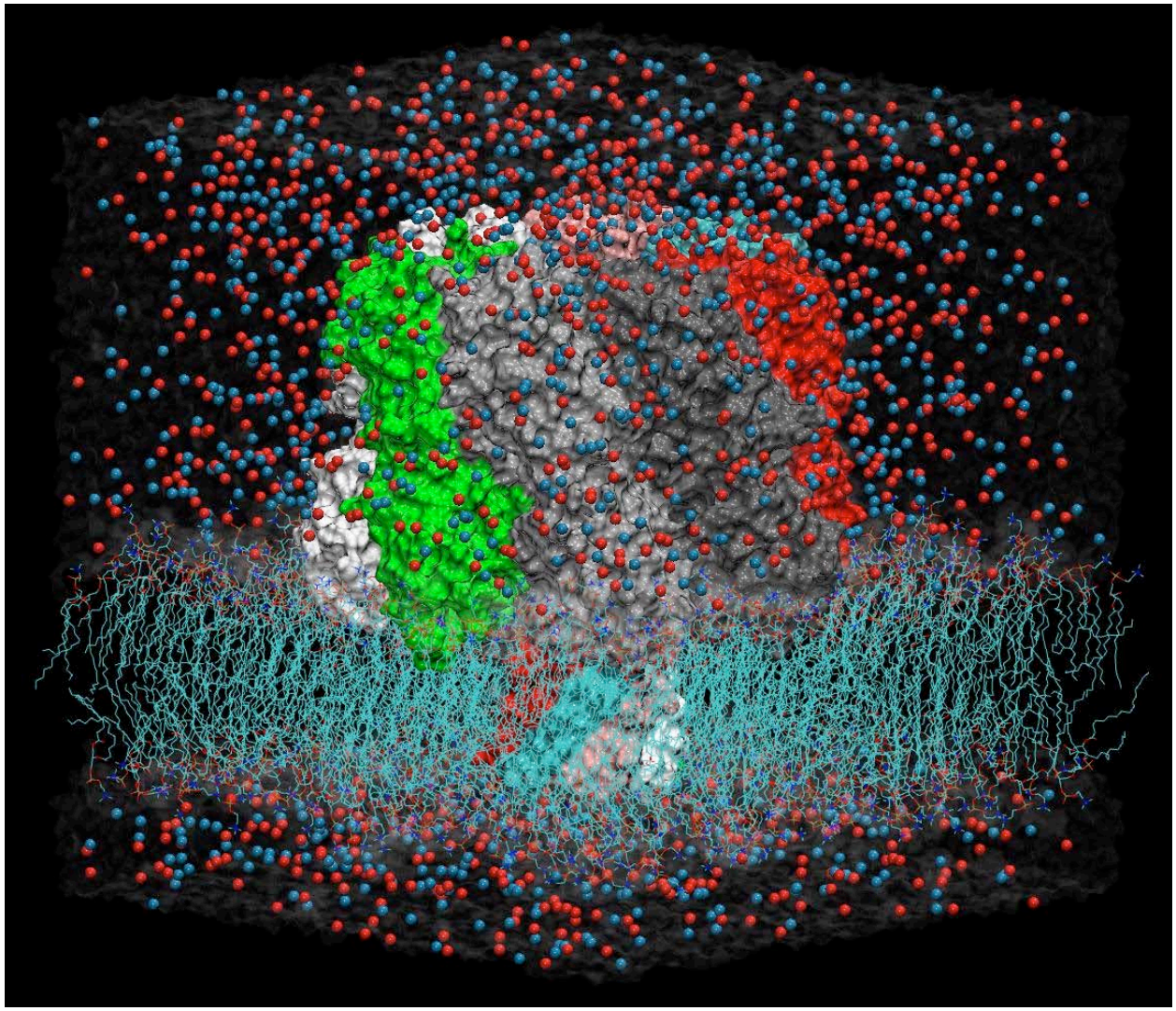
Cutting Corners

cutoffs, PME, rigid bonds, and multiple timesteps

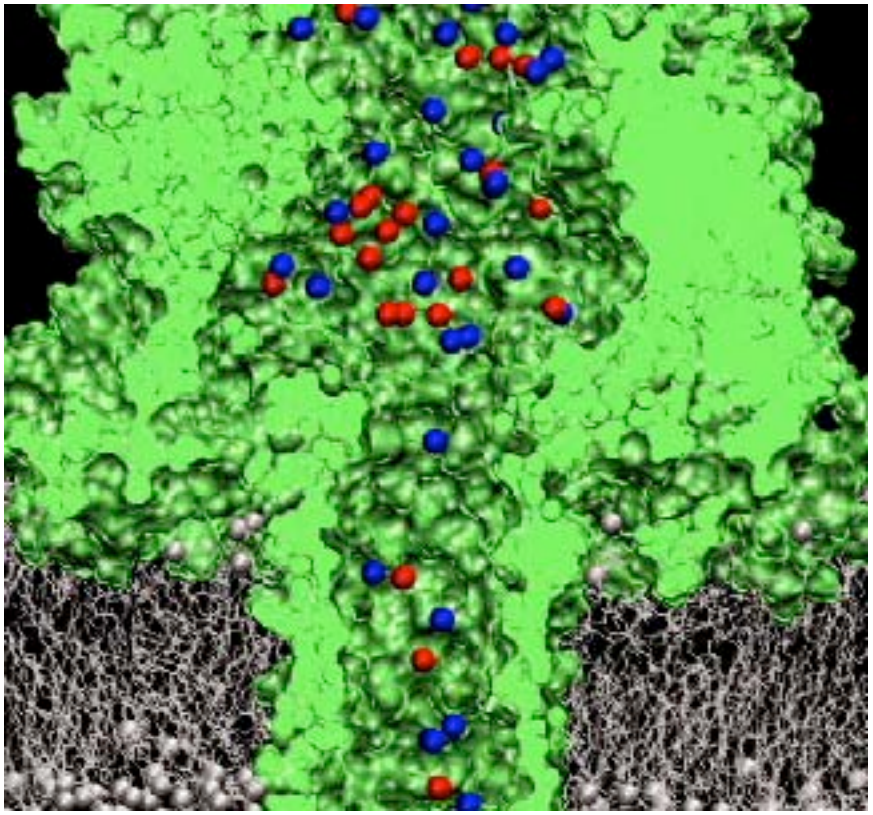
- Nonbonded interactions require order N^2 computer time!
  - Truncating at R_{cutoff} reduces this to order $N\ R_{cutoff}^3$
  - Particle mesh Ewald (PME) method adds long range electrostatics at order N log N, only minor cost compared to cutoff calculation.

- 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.

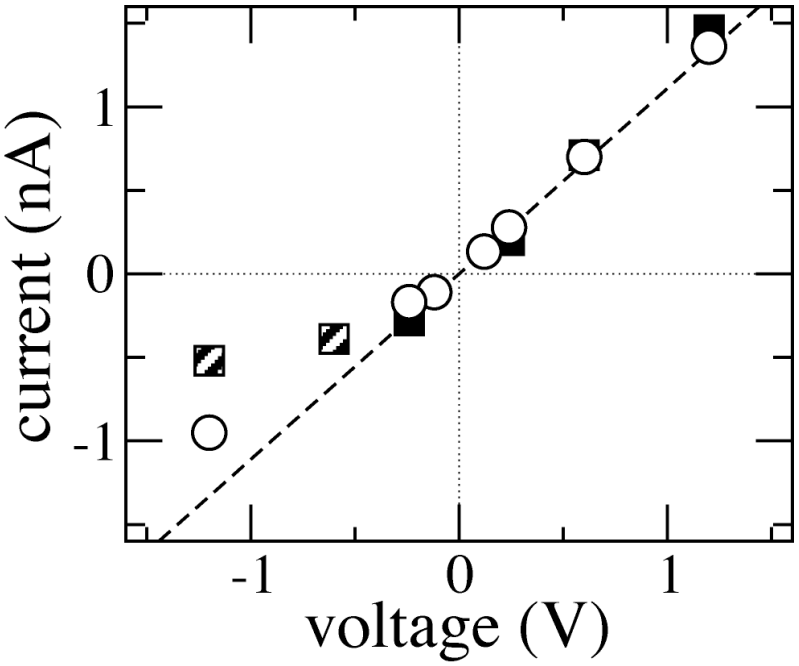
Setting up an α-hemolysin simulation



Computing current-voltage curve from MD



Ionic current at 120mV:
Experiment: 120 pA
Simulations: 130 pA

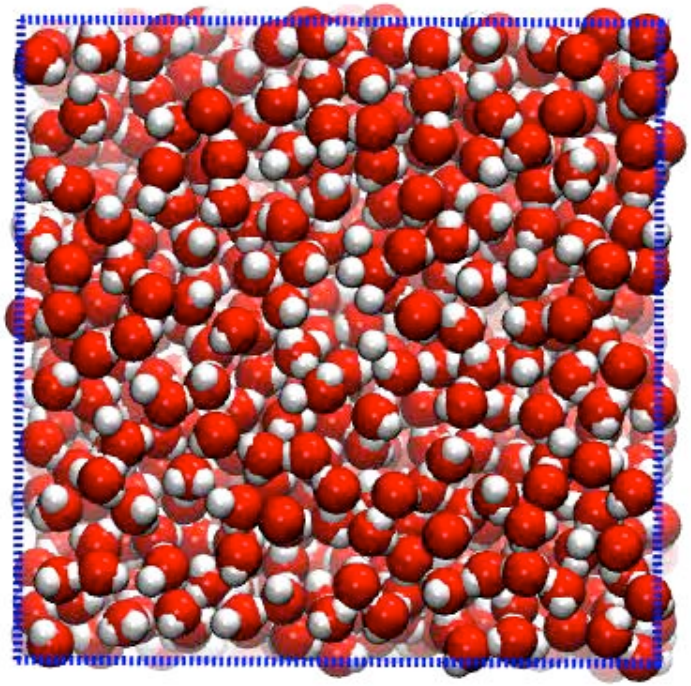


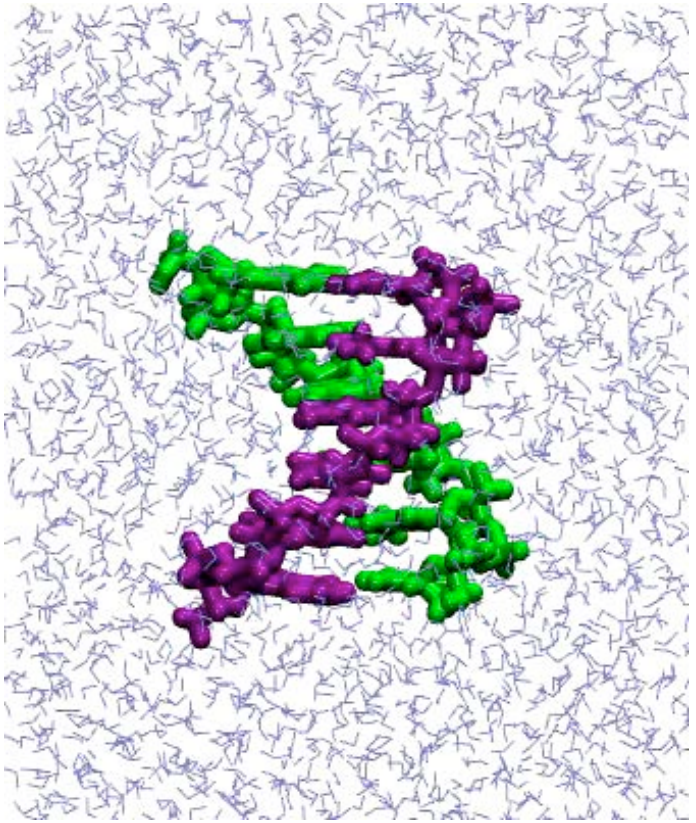
$$I(t) = \frac{1}{\Delta t L_z} \sum_{i=1}^{N} q_i (z_i(t + \Delta t) - z_i(t))$$

Instantaneous current

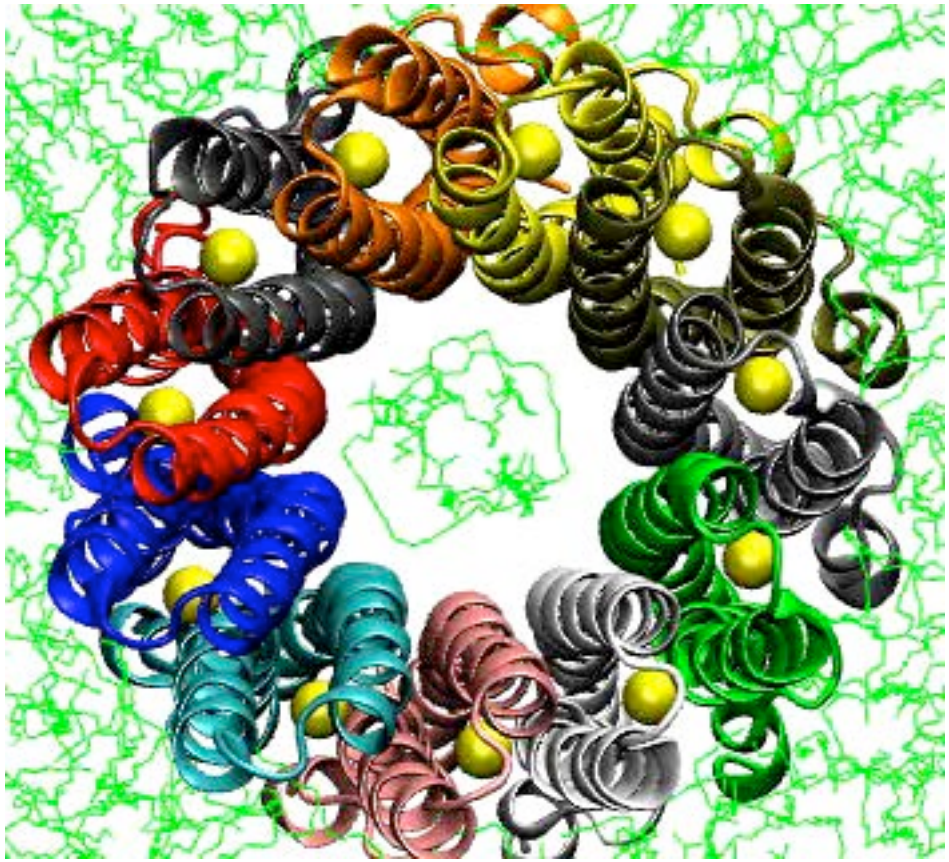
Current-voltage dependence

Apply external forces

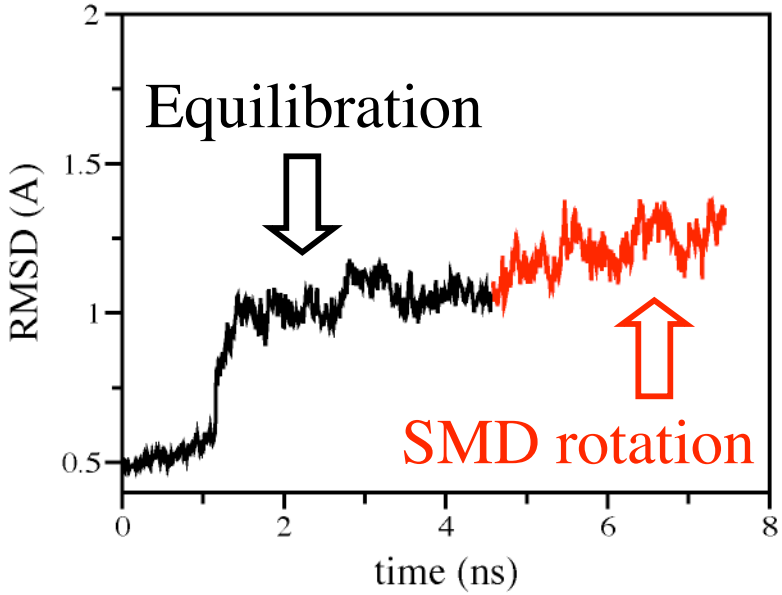




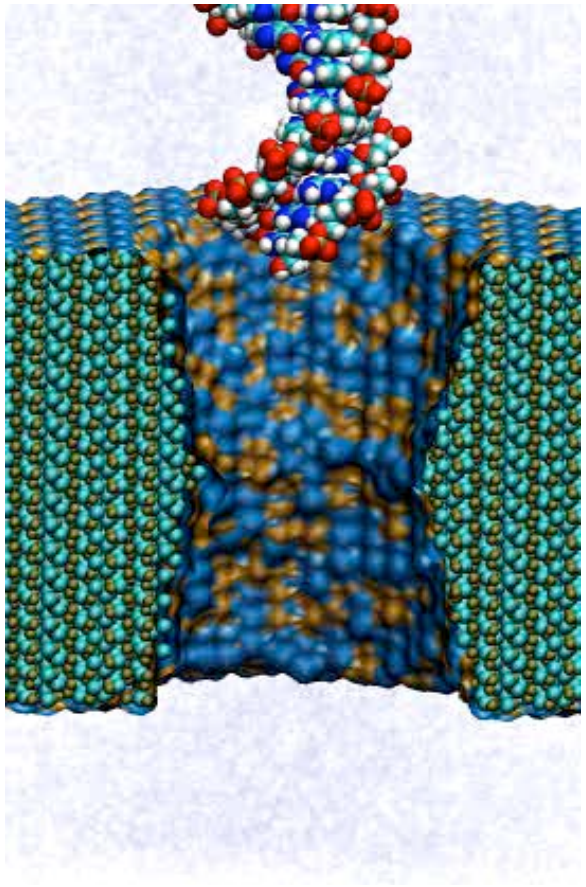
Forced Rotation of c_{11} in Na^+-ATP Synthase

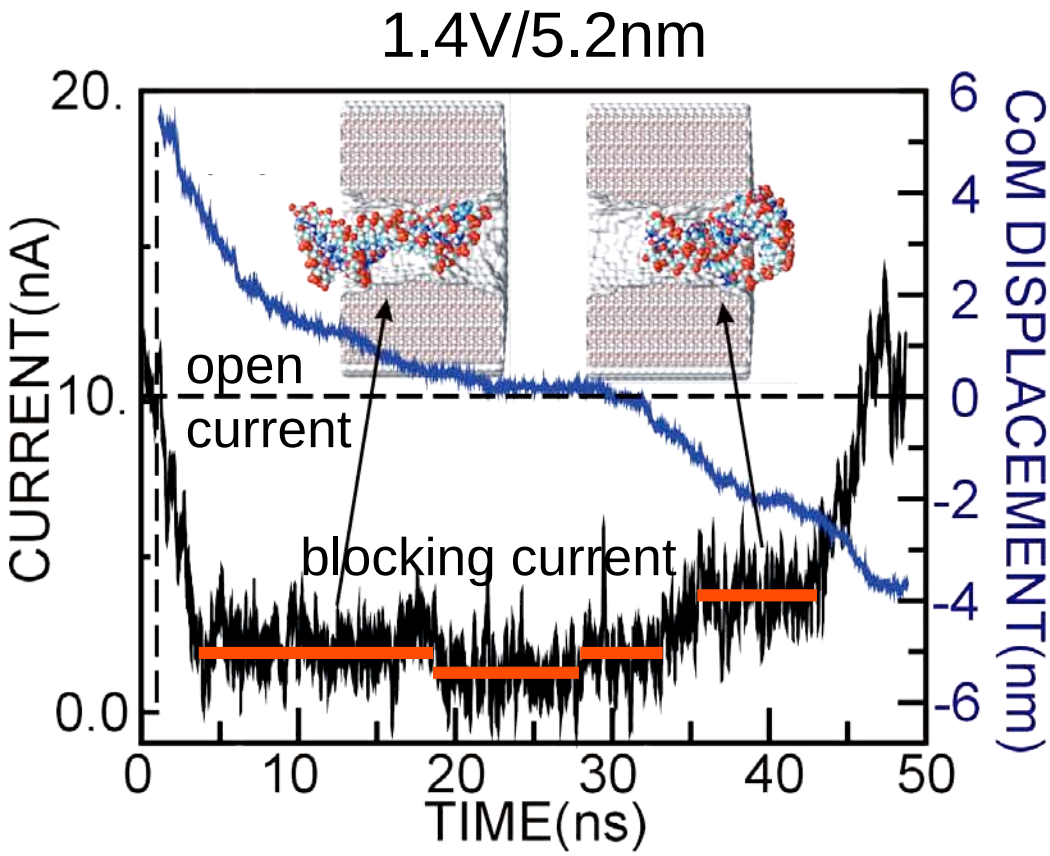


3-ns SMD simulation
torque: 1000 kcal/mol



DNA translocation through Si_3N_4 nanopore



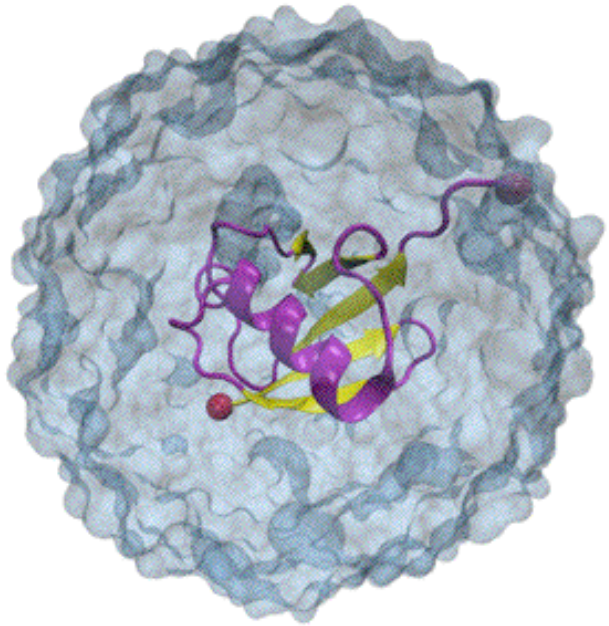


- at the end of the translocation *DNA* partially denatures
- blocking current correlates with molecular velocity
- translocation time: 10 ns - 3μs depending on the field

•Simulations: 1.4V/5.2nm ® *F*~400pN pore diameter ® d=2.5nm
•DNA sequence is CCCCCCCCCCCCCCCCCCCC

NAMD TUTORIAL

Unix/MacOSX Version



NAMD Developer: James Phillips

Timothy Isgro

James Phillips

Marcos Sotomayor

Elizabeth Villa

February 2006

The NAMD Configuration File / 1

Files needed:

```
structure          mypsf.psf
coordinates        mypdb.pdb
```

Define temperature

```
set temperature    310
     ;# target temperature used several times below
```

Starting simulation with random velocities

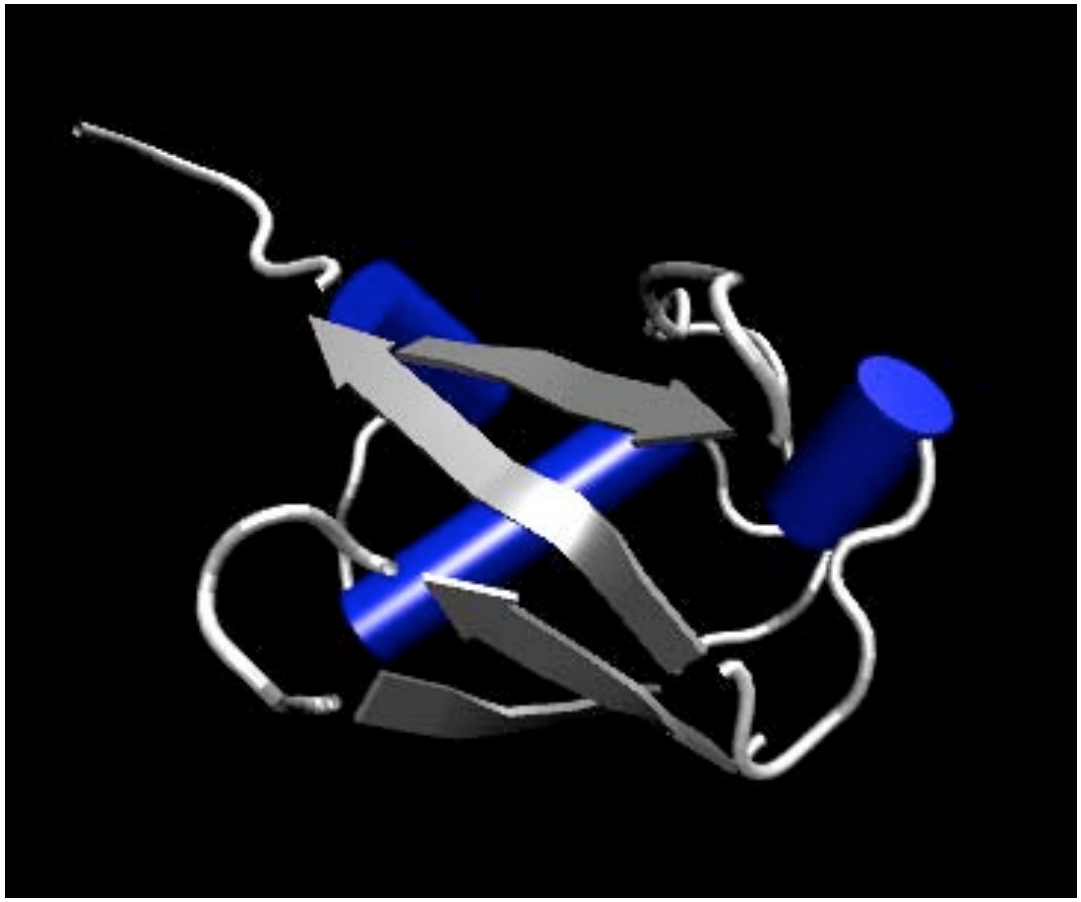
```
# starting from scratch
temperature         $temperature
       ;# initialize velocities randomly
```

The NAMD Output File / 1

Preamble

```
Info: NAMD 2.5b2ss03 for Linux-i686-Clustermatic
Info:
Info: Please visit http://www.ks.uiuc.edu/Research/namd/
Info: and send feedback or bug reports to namd@ks.uiuc.edu
Info:
Info: Please cite Phillips et al., J. Comp. Chem. 26: 1781-1802 (2005)
Info: in all publications reporting results obtained with NAMD.
Info:
Info: Built Fri May 30 13:09:06 CDT 2003 by jim on umbriel
Info: Sending usage information to NAMD developers via UDP.
Info: Sent data is: 1 NAMD  2.5b2ss03  Linux-i686-Clustermatic 47 umbriel jim
Info: Running on 47 processors.
```

Ubiquitin

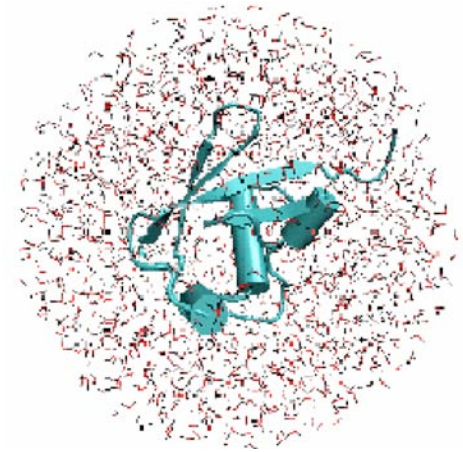


Fatemeh Araghi, Timothy Isgro, Marcos Sotomayor

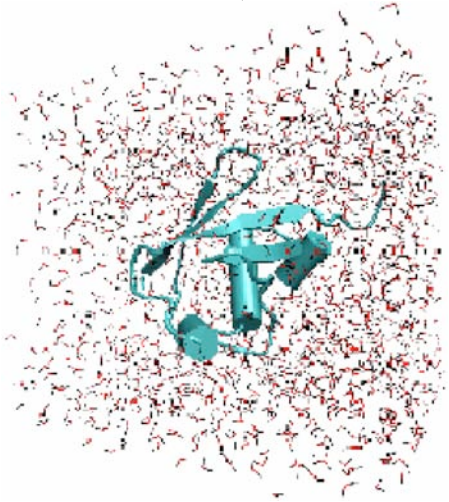
The NAMD Experience

You will first simulate ubiquitin in a water sphere and water box:

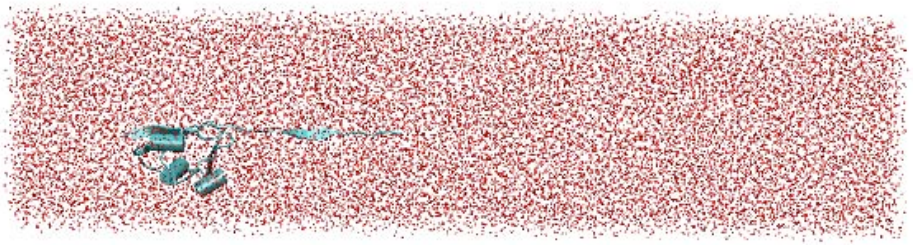
Solvate the protein in a water sphere (from VMD)

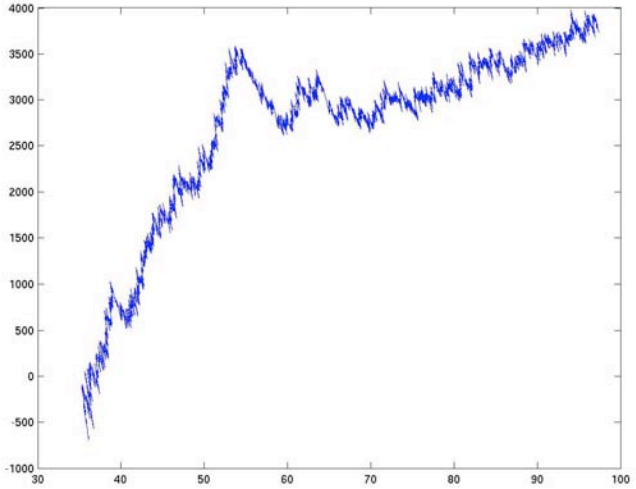


Solvate the protein in a water box (from VMD)



First SMD Simulation

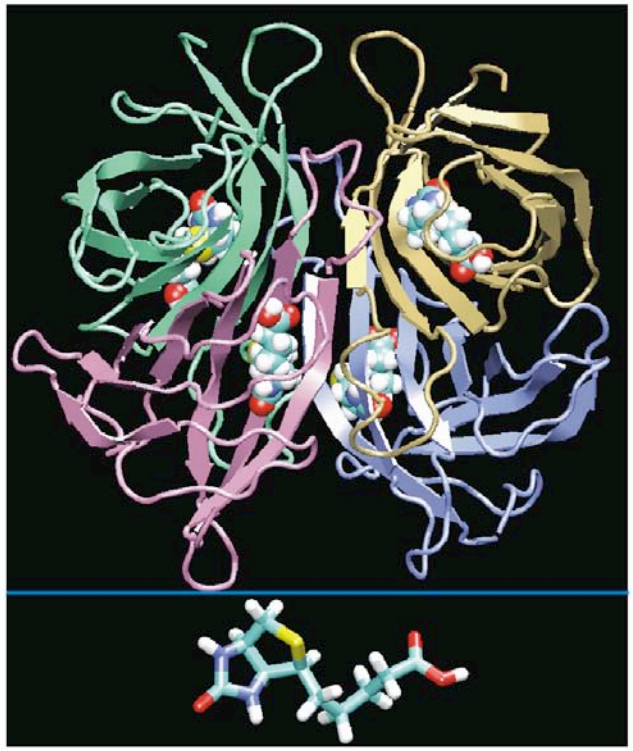




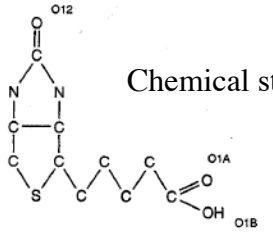
First peak when the first beta strand is stretched out

- SMD simulation, with constant velocity
- Box of water 70x240x70 A ~81K atoms
- smd velocity 0.4 A/ps
- smd spring constant 7 kcal/mol A^2

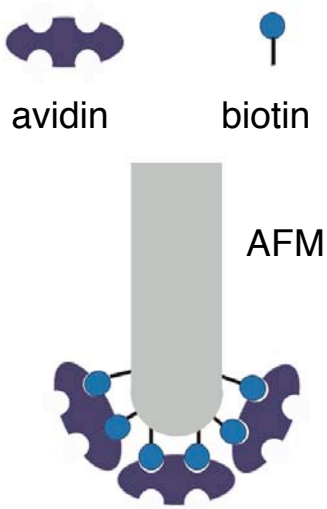
Atomic Force Microscopy Experiments of Ligand Unbinding

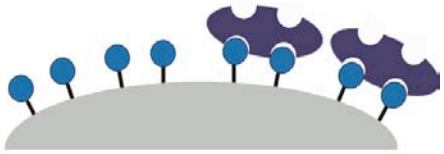


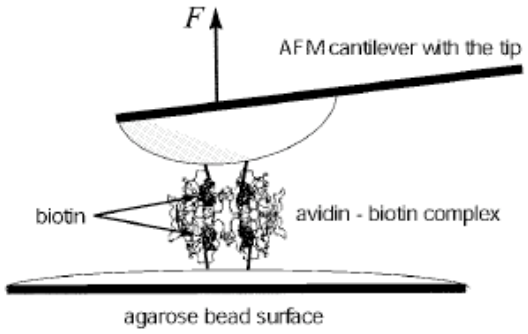
Biotin



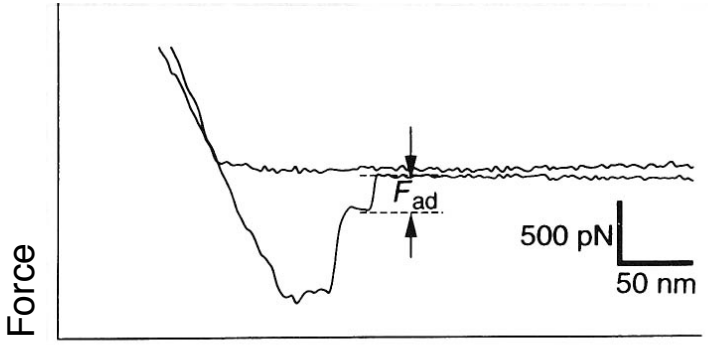
Chemical structure of biotin



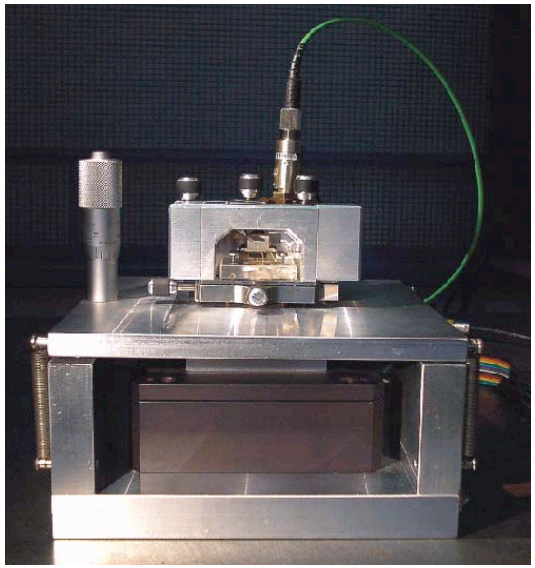




Florin et al., Science 264:415 (1994)

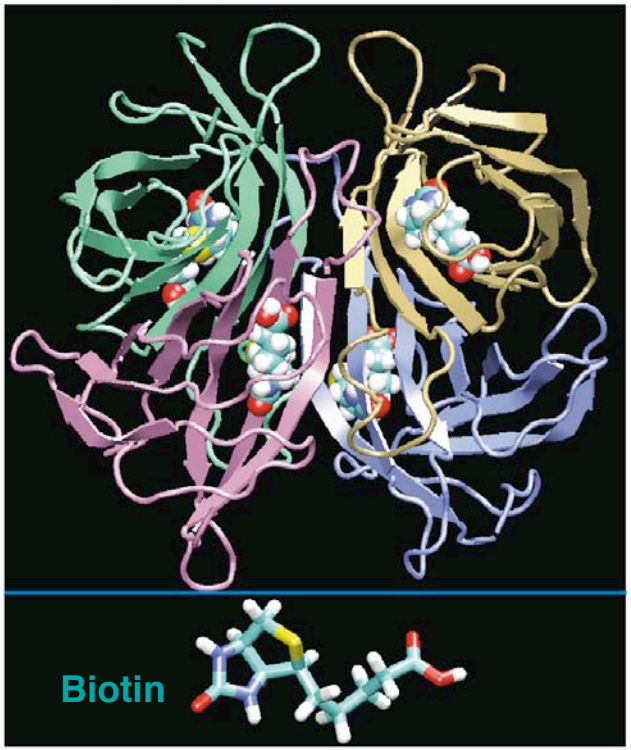


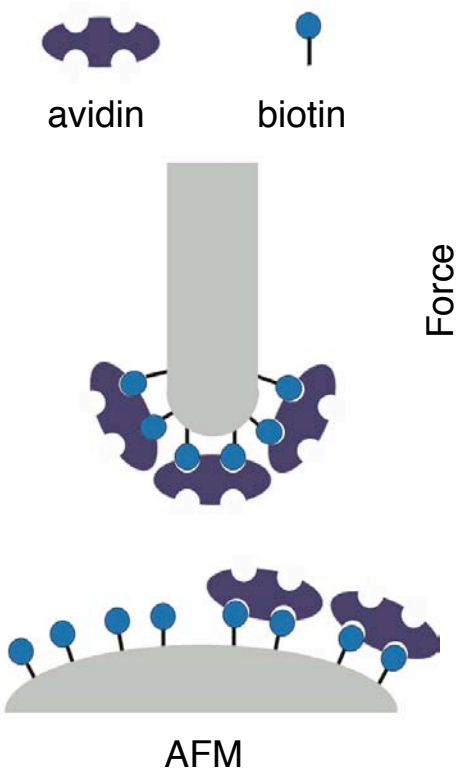
Displacement of AFM tip

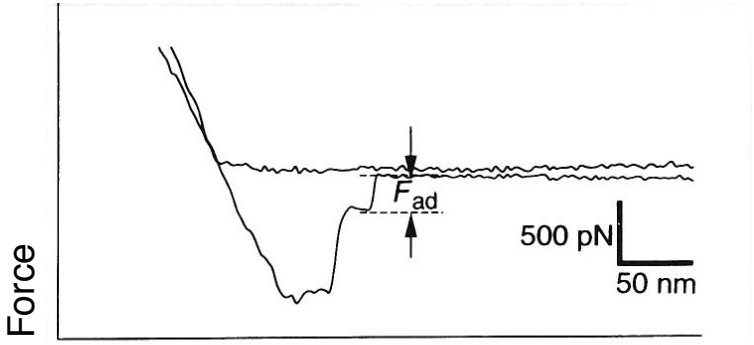


Atomic Force Microscopy Experiments of Ligand Unbinding

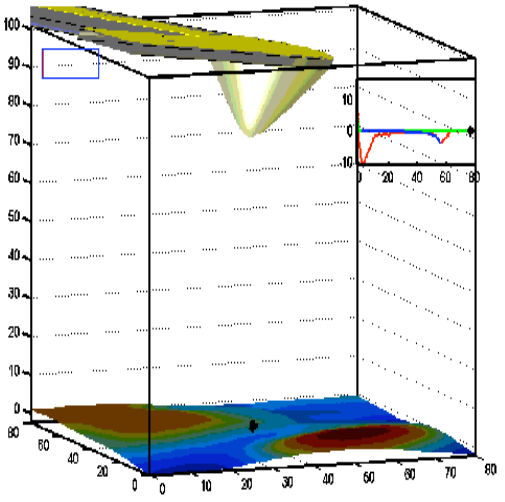
Florin et al., Science 264:415 (1994)







Displacement of AFM tip

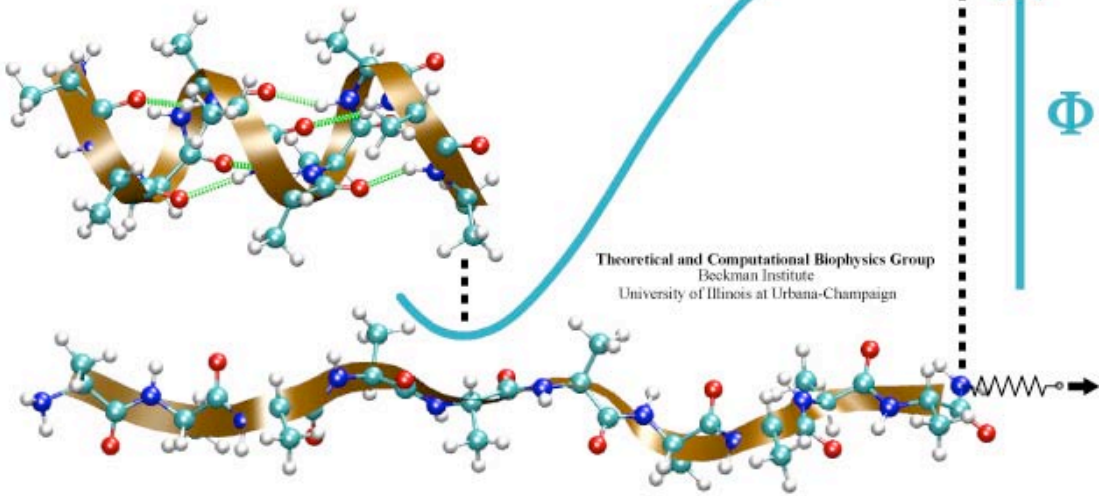


Free Energy of Stretched Alpha-Helix (Deca-alanin)

Thermodynamics: $\Delta G \le \langle W \rangle$

Jarzynski (1997): $e^{-\Delta G / k_B T} = \left\langle e^{-W / k_B T} \right\rangle$

$$\Phi = \langle W \rangle - \frac{\beta}{2}\left(\langle W^2 \rangle - \langle W \rangle^2\right)$$



Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten. *Journal of Chemical Physics* , 119:3559-3566, 2003

Calculating potentials of mean force from steered molcular dynamics simulations. S. Park and K. Schulten. *Journal of Chemical Physics* , 120: 5946-5961, 2004