Introduction to:

Nanoscale Molecular Dynamics
ATPase, a molecular motor that synthesizes the body’s weight of ATP a day

The Molecular Dynamics Method

Integrated chip for DNA sequencing

Bacterial toxin alpha-hemolysin
Molecular Dynamics Example

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

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

\[ F=ma \text{ at } 300K \]

Energy function:

\[ U(\vec{r}_1, \vec{r}_2, \cdots \vec{r}_N) = U(\vec{R}) \]

used to determine the force on each atom:

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

yields a set of 3N coupled 2\textsuperscript{nd}-order differential equations that can be propagated forward (or backward) in time.

Initial coordinates obtained from crystal structure, velocities taken at random from Boltzmann distribution.

Maintain appropriate temperature by adjusting velocities.
Langevin Dynamics

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

\[ m \ddot{\mathbf{r}} = \mathbf{F}(\mathbf{r}) - \gamma m \dot{\mathbf{r}} + \dot{\mathbf{R}}(t) \]

\[ \langle \dot{\mathbf{R}}(t) \cdot \dot{\mathbf{R}}(t') \rangle = 6k_B T \gamma \delta(t - t') \]
Classical Dynamics

discretization in time for computing

\[ m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i = -\mathbf{\nabla}U(\mathbf{R}) \]

Use positions and accelerations at time \( t \) and the positions from time \( t-\delta t \) to calculate new positions at time \( t+\delta t \).

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

\[
\mathbf{r}(t + \delta t) \approx 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + \mathbf{a}(t)\delta t^2
\]
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.

Improper: 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.

\[
U(\vec{R}) = \sum_{\text{bonds}} k_{i}^{\text{bond}}(r_{i} - r_{0})^2 + \sum_{\text{angles}} k_{i}^{\text{angle}}(\theta_{i} - \theta_{0})^2 + \sum_{\text{dihedrals}} k_{i}^{\text{dihedral}}[1 + \cos(n_{i}\phi_{i} + \delta_{i})] + \sum_{i} \sum_{j \neq i} 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \sum_{i} \sum_{j \neq i} \frac{q_{i}q_{j}}{\epsilon r_{ij}}
\]
Potential Energy Function of Biopolymer

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

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

<table>
<thead>
<tr>
<th>Time Scale</th>
<th>Steps</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>fs (femtosecond)</td>
<td>$10^0$</td>
<td>Bond stretching, Molecular dynamics timestep</td>
</tr>
<tr>
<td>ps (picosecond)</td>
<td>$10^3$</td>
<td>Bond stretching, Elastic vibrations</td>
</tr>
<tr>
<td>ns (nanosecond)</td>
<td>$10^6$</td>
<td>Rotation of surface sidechains, Hinge bending</td>
</tr>
<tr>
<td>μs (microsecond)</td>
<td>$10^9$</td>
<td>Local denaturations, Allosteric transitions</td>
</tr>
<tr>
<td>ms (millisecond)</td>
<td>$10^{12}$</td>
<td>Rotation of buried sidechains</td>
</tr>
<tr>
<td>s (second)</td>
<td>$10^{15}$</td>
<td></td>
</tr>
</tbody>
</table>

$\delta t = 1$ fs

SPEED LIMIT
MD simulation basics

Atomic coordinates

Connectivity

Force Field

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)
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)
The twenty amino acids

- **non-polar**
  - Ala, Alanine
  - Val, Valine
  - Phe, Phenylalanine
  - Pro, Proline
  - Met, Methionine
  - Ile, Isoleucine
  - Leu, Leucine

- **charged**
  - Asp, Aspartic Acid
  - Glu, Glutamic Acid
  - Lys, Lysine
  - Arg, Arginine
  - Ser, Serine
  - Thr, Threonine
  - Tyr, Tyrosine
  - His, Histidine

- **polar**
  - Cys, Cysteine
  - Asn, Asparagine
  - Gln, Glutamine
  - Trp, Tryptophan
  - Gly, Glycine

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

*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
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

Ala

CA

CB

N

HN

HA

C

O

HB1

HB2

HB3
Parameter Files

force constants for all types of interactions

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

<table>
<thead>
<tr>
<th>atom type</th>
<th>Kb</th>
<th>b0</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE1 CE1</td>
<td>440.000</td>
<td>1.3400</td>
</tr>
<tr>
<td></td>
<td></td>
<td>! for butene; from propene, yin/adm jr., 12/95</td>
</tr>
<tr>
<td>CE1 CE2</td>
<td>500.000</td>
<td>1.3420</td>
</tr>
<tr>
<td></td>
<td></td>
<td>! for propene, yin/adm jr., 12/95</td>
</tr>
</tbody>
</table>

ANGLES

!V(angle) = Ktheta(Theta - Theta0)**2

<table>
<thead>
<tr>
<th>atom types</th>
<th>Ktheta</th>
<th>Theta0</th>
<th>Kub</th>
<th>S0</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA CA CA</td>
<td>40.000</td>
<td>120.00</td>
<td>35.00</td>
<td>2.41620</td>
</tr>
<tr>
<td></td>
<td>! ALLOW ARO</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>JES 8/25/89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE1 CE1 CT3</td>
<td>48.00</td>
<td>123.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>! for 2-butene, yin/adm jr., 12/95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE1 CT2 CT3</td>
<td>32.00</td>
<td>112.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>! for 1-butene; from propene, yin/adm jr., 12/95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE2 CE1 CT2</td>
<td>48.00</td>
<td>126.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
U(\vec{R}) = \sum_{\text{bonds}} k_{i}^{\text{bond}} (r_{i} - r_{0})^{2} + \sum_{\text{angles}} k_{i}^{\text{angle}} \left(\theta_{i} - \theta_{0}\right)^{2} + \sum_{\text{dihedrals}} k_{i}^{\text{dihedral}} \left[1 + \cos \left(n_{i} \phi_{i} + \delta_{i}\right)\right] + \sum_{i \neq j} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{i \neq j} q_{i} q_{j} \epsilon r_{ij} \]

<table>
<thead>
<tr>
<th>atom type</th>
<th>Ktheta</th>
<th>Theta0</th>
<th>Kub</th>
<th>S0</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA CA CA</td>
<td>40.000</td>
<td>120.00</td>
<td>35.00</td>
<td>2.41620</td>
</tr>
<tr>
<td></td>
<td>! ALLOW ARO</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>JES 8/25/89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE1 CE1 CT3</td>
<td>48.00</td>
<td>123.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>! for 2-butene, yin/adm jr., 12/95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE1 CT2 CT3</td>
<td>32.00</td>
<td>112.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>! for 1-butene; from propene, yin/adm jr., 12/95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE2 CE1 CT2</td>
<td>48.00</td>
<td>126.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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
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
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 with

  \[ kT \]

- Initial dynamics samples thermally accessible states.
From the Mountains to the Valleys

*a molecular dynamics tale*

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

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.

Ankyrin
340K atoms
with PME

NAMD
programmer
J. Phillips
Ph.D. UIUC
Physics

TeraGrid
Phase 2 (NCSA)

3 s/step

75% efficiency on 256 CPUs

Linear scaling

32 ms

32
64
128
256
512

number of processors

time per step (seconds)

0.01
0.1
1
10
Linux Clusters 101

parallel computing on a professor’s salary

Learn to build your own Linux cluster!

Easy to manage

$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 $R_{\text{cutoff}}$ reduces this to order $N R_{\text{cutoff}}^3$
  – Particle mesh Ewald (PME) method adds long range electrostatics at order $N \log N$, only minor cost compared to cutoff calculation.

• Can we extend the timestep, and do this work fewer times?
  – Bonds to hydrogen atoms, which require a 1fs timestep, can be held at their equilibrium lengths, allowing 2fs steps.
  – Long range electrostatics forces vary slowly, and may be evaluated less often, such as on every second or third step.
Setting up an $\alpha$-hemolysin simulation
Computing current-voltage curve from MD

Instantaneous current

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

Current-voltage dependence

\[ I(t) = \frac{1}{\Delta t L_z} \sum_{i=1}^{N} q_i(z_i(t + \Delta t) - z_i(t)) \]
Apply external forces
Forced Rotation of $c_{11}$ in Na$^+$-ATP Synthase

3-ns SMD simulation
torque: 1000 kcal/mol
DNA translocation through Si$_3$N$_4$ nanopore

- at the end of the translocation DNA partially denatures
- blocking current correlates with molecular velocity
- translocation time: 10 ns - 3$\mu$s depending on the field

Simulations: 1.4V/5.2nm $\sim$ 400pN  pore diameter $\sim$ d=2.5nm
DNA sequence is CCCCCCCCCCCCCCCCCCCCCC
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: 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

- SMD simulation, with constant velocity

- Box of water 70x240x70 Å  ~81K atoms

- smd velocity 0.4 Å/ps
- smd spring constant 7 kcal/mol Å^2

First peak when the first beta strand is stretched out
Atomic Force Microscopy Experiments of Ligand Unbinding


Chemical structure of biotin

AFM cantilever with the tip

Displacement of AFM tip

Biotin
Atomic Force Microscopy Experiments of Ligand Unbinding

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

**Thermodynamics:** \[ \Delta G \leq \langle W \rangle \]

**Jarzynski (1997):** \[ e^{-\Delta G/k_B T} = \langle e^{-W/k_B T} \rangle \]

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