The Molecular Dynamics Method

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

AQB filtering a bath tub of the body’s water a day

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

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

come on, feel the noise

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

\[
\begin{align*}
    m \ddot{\vec{r}} &= \vec{F}(\vec{r}) - \gamma m \dot{\vec{r}} + \vec{R}(t) \\
    \langle \vec{R}(t) \cdot \vec{R}(t') \rangle &= 6k_B T \gamma \delta(t - t')
\end{align*}
\]
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 \( t-\delta t \) to calculate new positions at time \( t+\delta t \).

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

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

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

Potential Energy Function of Biopolymer

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

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

From physics
Large is no problem. But ...

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

NCSA machine room
But long is!

*biomolecular timescale and timestep limits*

<table>
<thead>
<tr>
<th>Steps</th>
<th>Time (scale)</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>$10^{15}$</td>
</tr>
<tr>
<td>ms</td>
<td>$10^{12}$</td>
</tr>
<tr>
<td>μs</td>
<td>$10^{9}$</td>
</tr>
<tr>
<td>ns</td>
<td>$10^{6}$</td>
</tr>
<tr>
<td>ps</td>
<td>$10^{3}$</td>
</tr>
<tr>
<td>fs</td>
<td>$10^{0}$</td>
</tr>
</tbody>
</table>

- Rotation of buried sidechains
- Local denaturations
- Allosteric transitions

- Hinge bending
- Rotation of surface sidechains
- Elastic vibrations

Molecular dynamics timestep $\delta t = 1$ fs
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
PSF Files

*atomic properties (mass, charge, type)*

- Every atom in the simulation 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 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)
Example of 
PSF File

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}
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
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
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
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 with

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

TeraGrid Phase 2 (NCSA)

- 75% efficiency on 256 CPUs
- 32 ms
- Linear scaling
- 3 s/step
- 340K atoms with PME

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.

NAMD programmer
J. Phillips
Ph.D. UIUC
Physics

Ankyrin

NAMD

Scalable Molecular Dynamics
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.
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
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 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 $\Rightarrow$ long equilibration time required to fill the gaps.
Solution: manually adjust the position of lipids around the protein
The system

Kcsa channel protein (in blue) embedded in a (3:1) POPE/POPG lipid bilayer. Water molecules inside the channel are shown in vdw representation.
RMS deviations for the KcsA protein and its selectivity filter 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)
Simulation of Ion Conduction (here for Kv1.2)