Molecular Dynamics Method 1

James Phillips
Theoretical and Computational Biophysics Group
Beckman Institute, UIUC

The Road Ahead

- PDB, PSF, topology, and parameter files
- Molecular dynamics
  - …in an ideal world
    - …and in our world
      - …with computers
        - …using NAMD
- Justin prepares a protein using VMD
- You prepare a protein using VMD
  - …and simulate it using NAMD
    - …in the hands-on tomorrow afternoon

Don't worry, the written tutorial is very complete.
You will learn by doing. This talk is an overview.

The Road Ahead

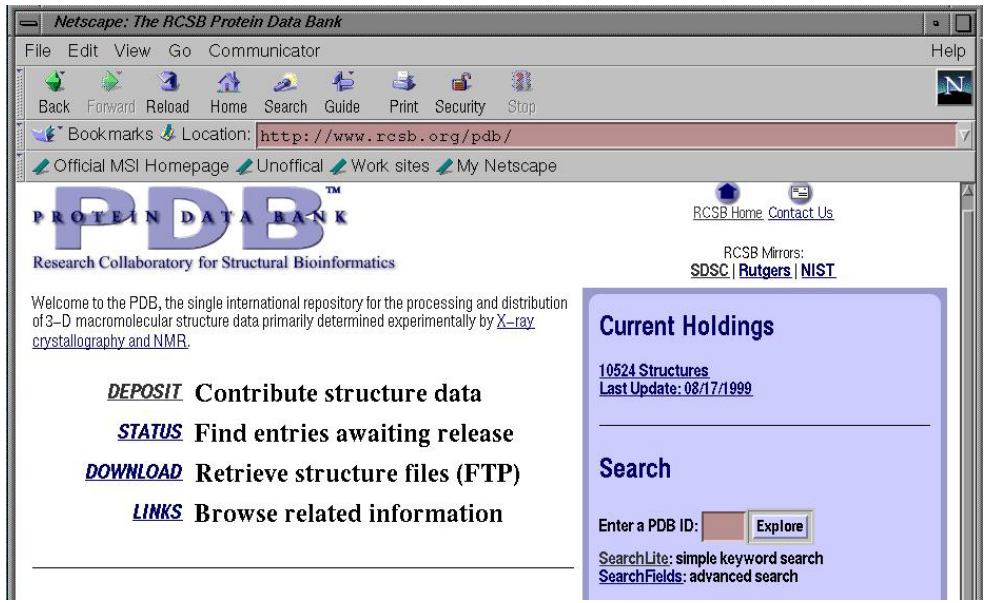
- PDB, PSF, topology, and parameter files
- Molecular dynamics
  - …in an ideal world
    - …and in our world
      - …with computers
        - …using NAMD
- Justin prepares a protein using VMD
- You prepare a protein using VMD
  - …and simulate it using NAMD
    - …in the hands-on tomorrow afternoon

Don't worry, the written tutorial is very complete.
You will learn by doing. This talk is an overview.

PDB Files

a little information (and a dangerous thing?)

- 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.)



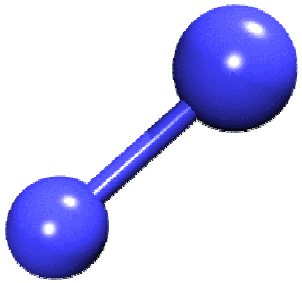
PSF Files

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, C, CT1)
  - residue name (ALA, HIS)
  - residue id (integer)
  - segment id (6PTI)
  - atomic mass (in atomic mass units)
  - partial charge (in electronic charge units)
- What is not in the PSF file?
  - coordinates (dynamic data, initially read from PDB file)
  - velocities (dynamic data, initially from Boltzmann distribution)
  - force field parameters (non-specific, used for many molecules)

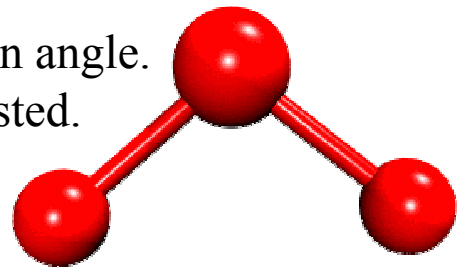
PSF Files

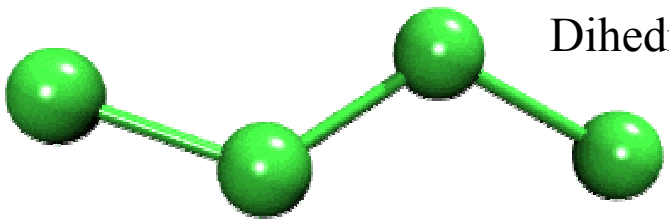
molecular structure (bonds, angles, etc.)



Bonds: Every pair of covalently bonded atoms is listed.

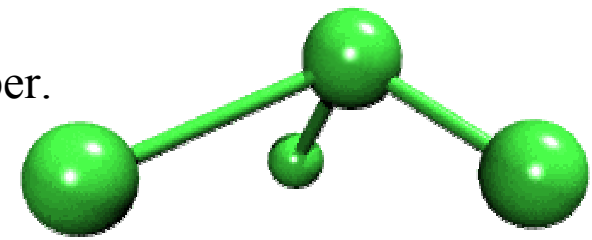
Angles: Two bonds that share a common atom form an angle.
Every such set of three atoms in the molecule is listed.





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

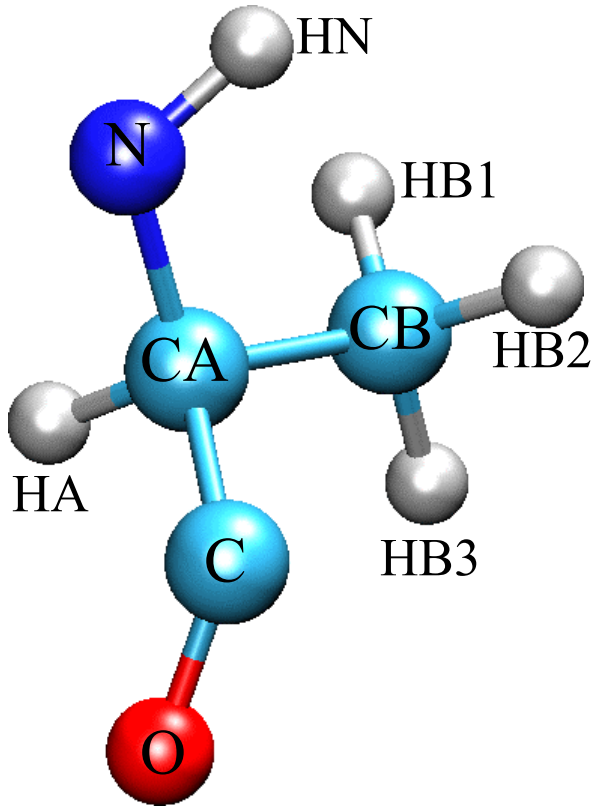
Impropers: Any *planar* group of four atoms forms an improper.
Every such set of four atoms in the molecule is listed.



Topology Files

blueprints for building a PSF file

- For every type of residue known:
  - atom name, type, mass, and charge
  - bonds within the residue
  - bonds to other residues
  - any planar impropers (rare)
- Additional "patches" for:
  - terminating protein segments
  - joining protein segments
  - modifying protonation states
  - adding disulphide bonds
  - deoxygenating nucleic acids



CHARMM Potential Function

form without substance

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

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

Parameter Files

biomolecular paint by numbers

- Equilibrium value and spring constant for
  - every pair of atom types that can form and bond
  - every triple of atom types that can form an angle
  - every quad of atom types that can form a dihedral or improper (many wildcard cases)
- vdW radius and well depth for every atom type
  - actually need these for every pair of atoms types!
  - pair radius calculated from arithmetic mean
  - pair well depth calculated from geometric mean
- Closely tied to matching topology file!

The Road Ahead

- PDB, PSF, topology, and parameter files
- Molecular dynamics
  - …in an ideal world
    - …and in our world
      - …with computers
        - …using NAMD
- Justin prepares a protein using VMD
- You prepare a protein using VMD
  - …and simulate it using NAMD
    - …in the hands-on tomorrow afternoon

Don't worry, the written tutorial is very complete.
You will learn by doing. This talk is an overview.

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.

The Flying Ice Cube

a molecular dynamics horror story

- Velocity rescaling controls temperature by periodically rescaling all velocities to match a target temperature:

$$T = \left\langle m\left|\vec{v}\right|^2 / 3k_B \right\rangle \qquad \vec{v}_{\text{new}} = \sqrt{T_{\text{target}} / T} \times \vec{v}_{\text{old}}$$

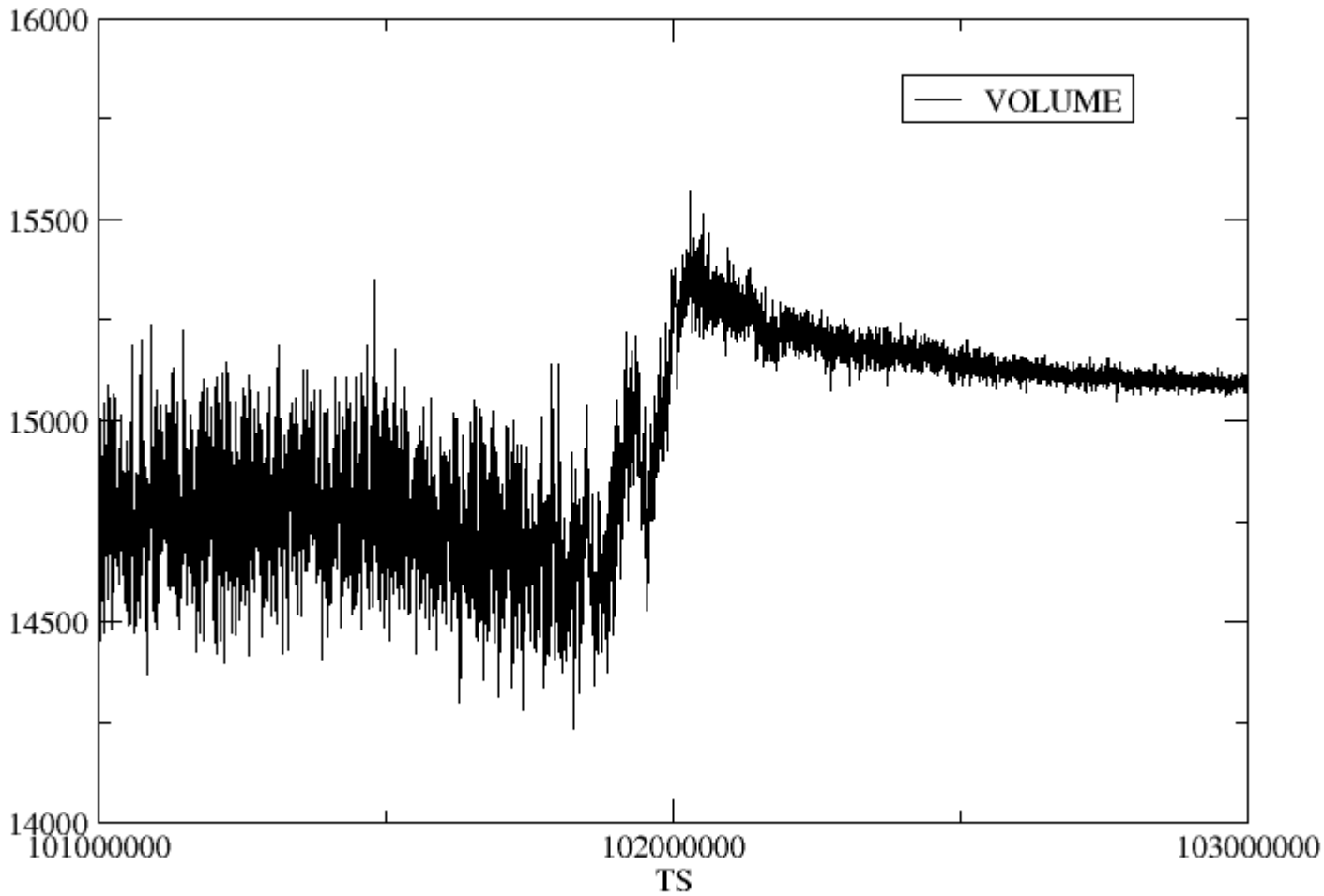
- Temperature coupling makes this a continuous process by adding a temperature-dependent friction term:

$$\dot{\vec{v}} = \vec{F}(\vec{r}) / m - (T / T_{\text{target}} - 1)\gamma\vec{v}$$

- Unfortunately, these methods drive energy into the lowest-frequency normal modes in the system. So what, you ask?

The Flying Ice Cube

a molecular dynamics horror story



Langevin Dynamics

come on, feel the noise

If the protein is at 200K, and the water is at 350K, do you:

a) heat the water

b) cool the protein

c) none of the above

Unfortunately, with temperature control based on measuring the temperature of the entire system, those are your only choices!

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

$$\dot{\vec{v}} = \vec{F}(\vec{r})/m - \gamma\vec{v} + \vec{F}_{\text{random}}(t)$$

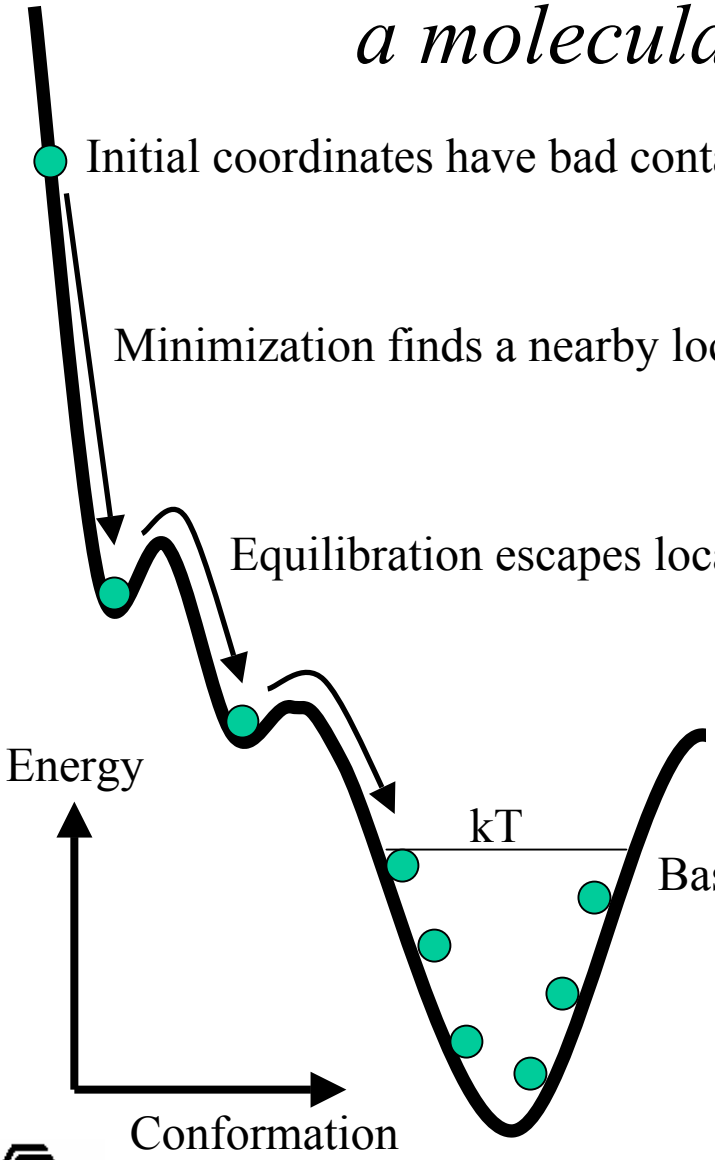
From the Mountains to the Valleys

a molecular dynamics fairy tale

Initial coordinates have bad contacts, causing high energies and forces.

Minimization finds a nearby local minimum.

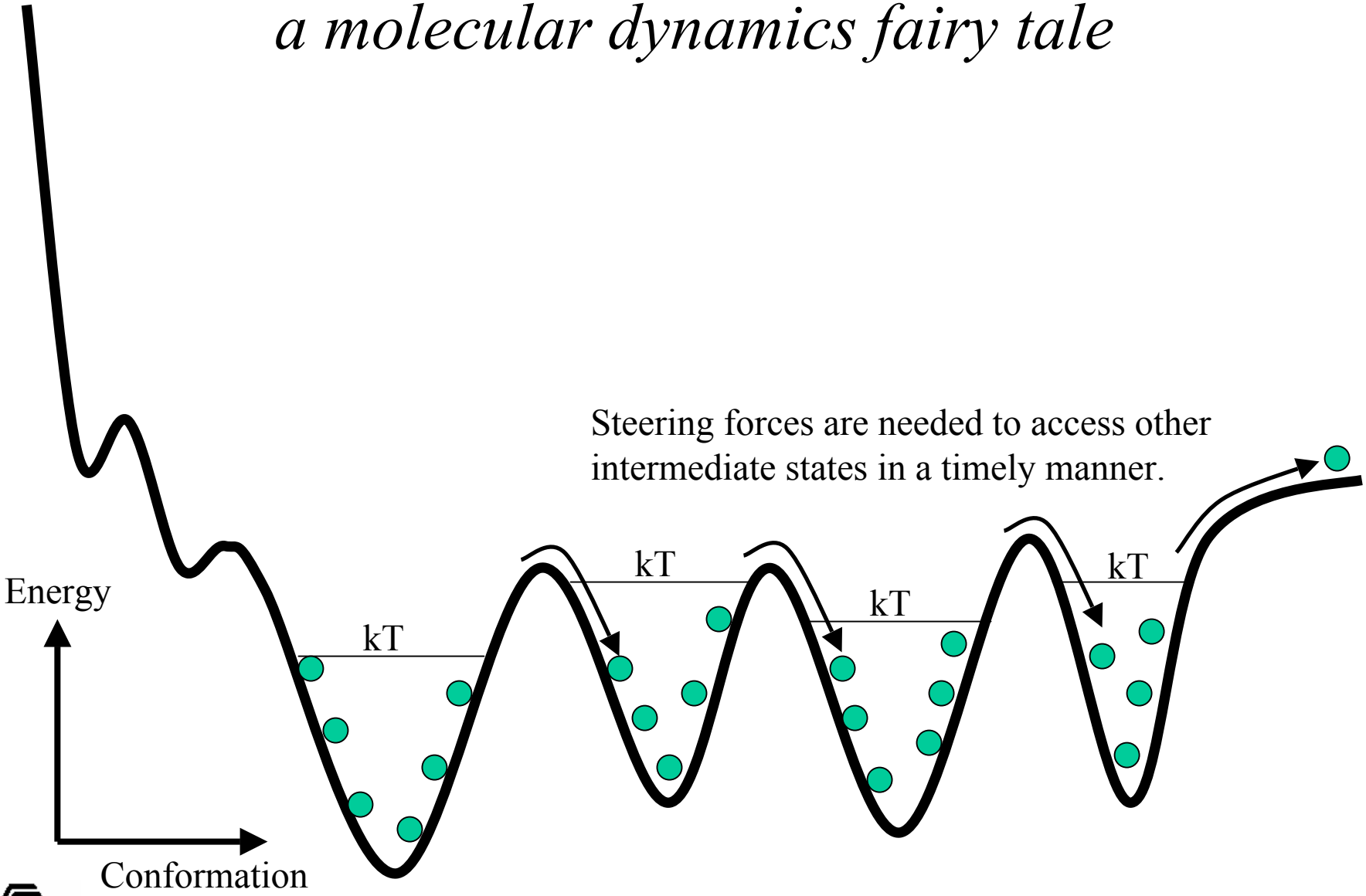
Equilibration escapes local minima with low energy barriers.



Basic simulation samples thermally accessible states.

From the Mountains to the Valleys

a molecular dynamics fairy tale



The Road Ahead

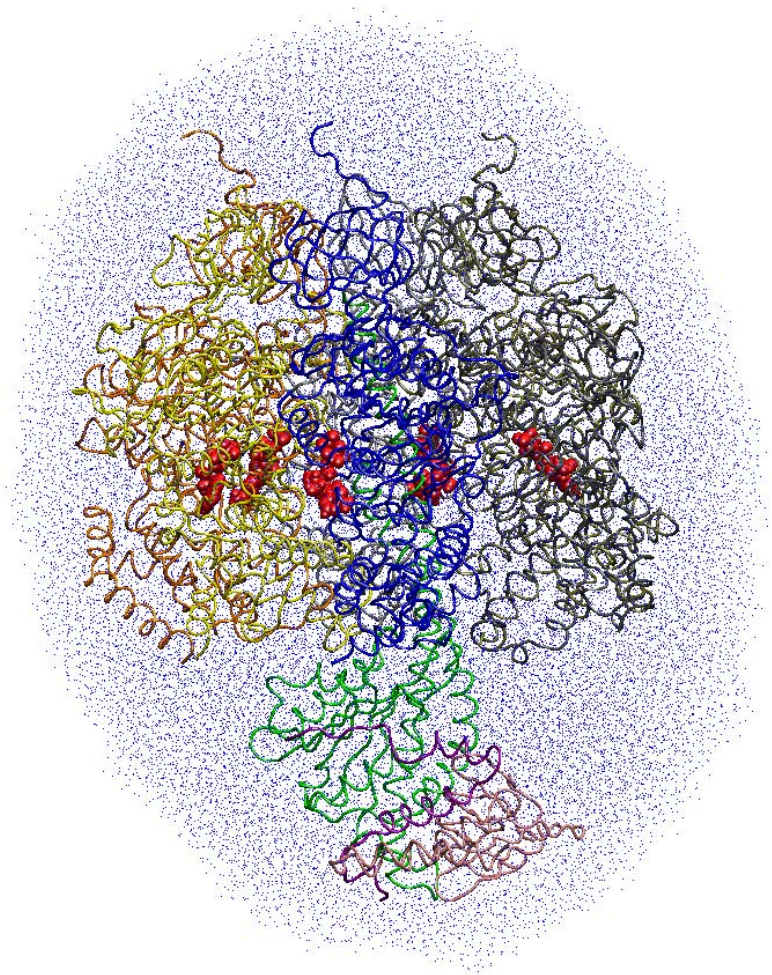
- PDB, PSF, topology, and parameter files
- Molecular dynamics
  - …in an ideal world
    - …and in our world
      - …with computers
        - …using NAMD
- Justin prepares a protein using VMD
- You prepare a protein using VMD
  - …and simulate it using NAMD
    - …in the hands-on tomorrow afternoon

Don't worry, the written tutorial is very complete.
You will learn by doing. This talk is an overview.

Life in a Drop of Water

a study in surface tension

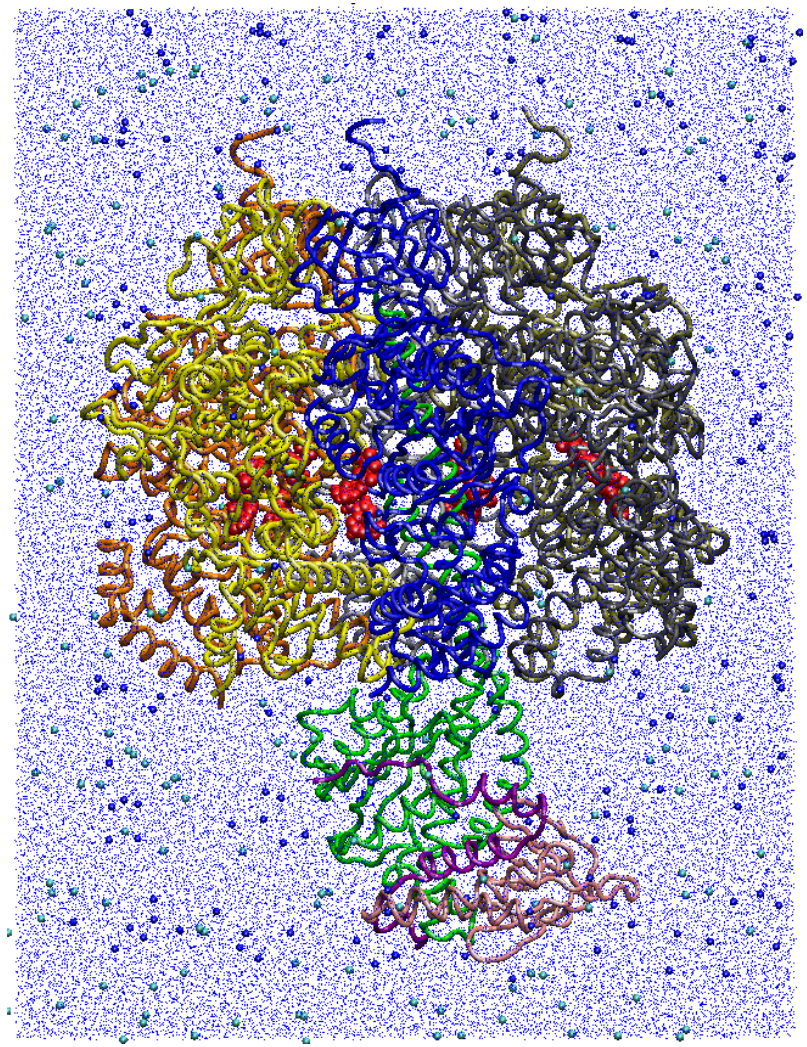
- The structure of water optimizes the network of hydrogen bonds between individual molecules.
- At a liquid-gas interface these bonds orient parallel to the interface, generating surface tension.
- This causes any blob of water to form a sphere with internal pressure inversely proportional to its radius.
- But should this matter to us?



Delusions of Grandeur

periodic boundary conditions

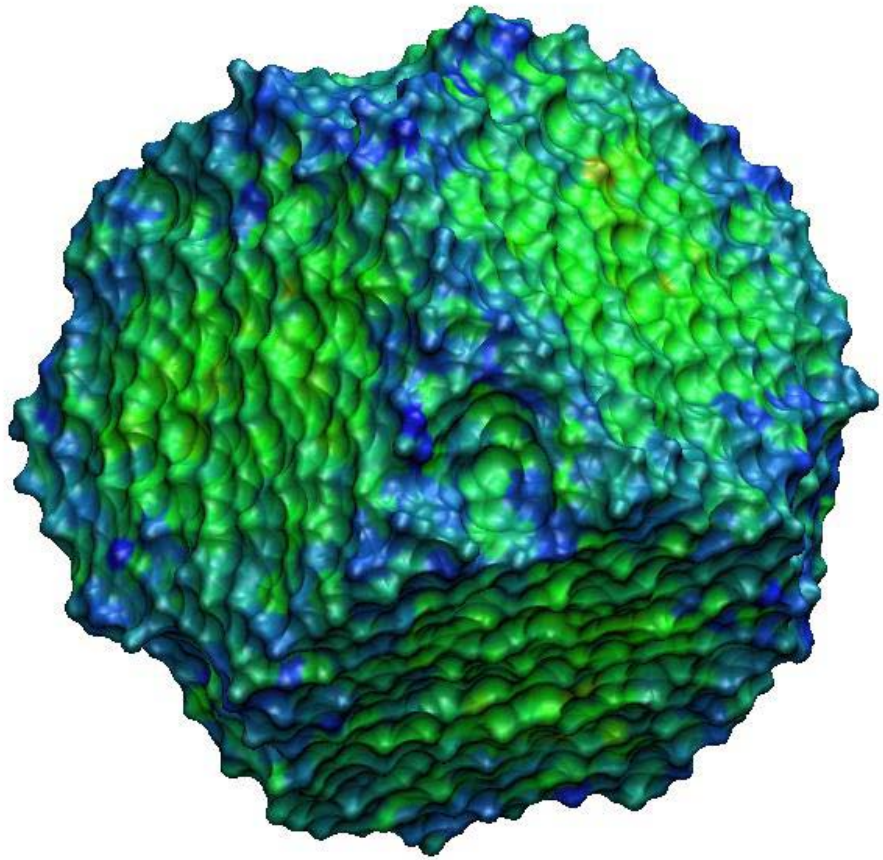
- Problem: How to simulate an infinite amount of solvent with a minimal number of atoms.
- Solution: Define a space-filling “cell” surrounded on all sides by identical images of itself.
- As atoms leave one side of the cell, they re-enter from the opposite side.



Tiny Bubbles

another molecular dynamics horror story

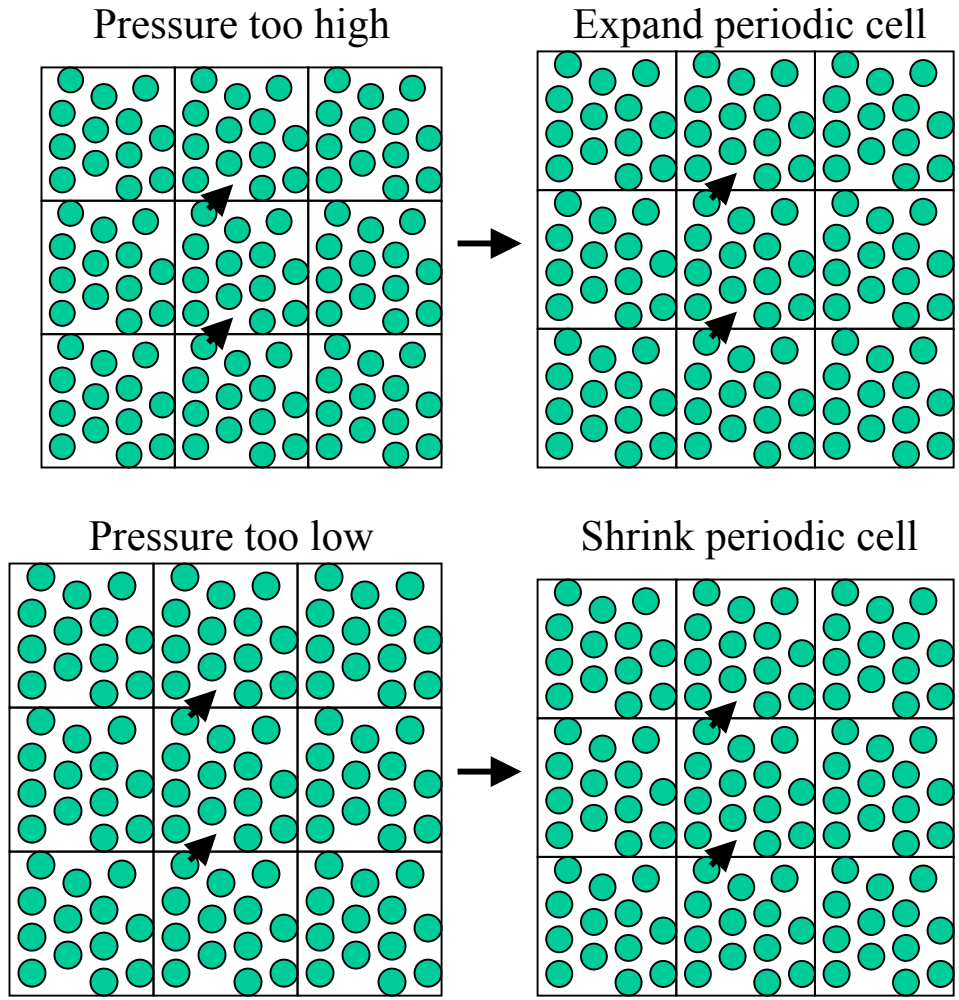
- Periodic boundaries are used to eliminate surface effects.
- This assumes that the simulation completely fills the periodic cell.
- A gas can expand to fill any container, but water has a narrow range of densities.
- What happens if the volume we choose for the periodic cell is too large?



Elastic in the Waist

basics of constant pressure simulation

- The pressure of a molecular system depends on its volume and temperature.
- Non-periodic systems can adjust themselves in infinite volume and are at zero pressure.
- Periodic systems must use a barostat to vary cell volume and maintain constant pressure.
- Atomic coordinates are rescaled along with cell.



The Road Ahead

- PDB, PSF, topology, and parameter files
- Molecular dynamics
  - …in an ideal world
    - …and in our world
      - …with computers
        - …using NAMD
- Justin prepares a protein using VMD
- You prepare a protein using VMD
  - …and simulate it using NAMD
    - …in the hands-on tomorrow afternoon

Don't worry, the written tutorial is very complete.
You will learn by doing. This talk is an overview.

Step by Step

discretization in time

Use positions and accelerations at time t and the positions from time t-δt to calculate new positions at time t+δt.

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

+

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

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

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

Hurry Up and Wait

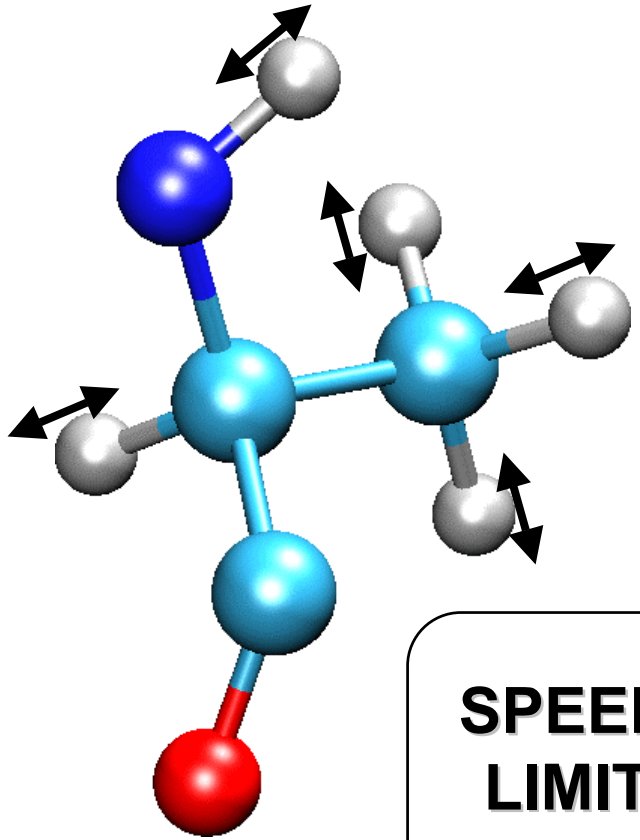
biomolecular timescales and timestep limits

Rotation of buried sidechains
Local denaturations
Allosteric transitions

Hinge bending
Rotation of surface sidechains
Elastic vibrations

Bond stretching
Molecular dynamics timestep →

s
ms
μs
ns
ps
fs



SPEED LIMIT

δt = 1 fs

Cutting Corners

cutoffs, PME, rigid bonds, and multiple timesteps

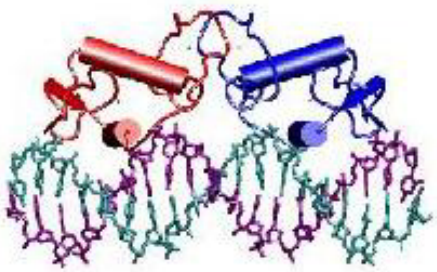
- Nonbonded interactions require order N^2 computer time!
  - Truncating at R_{cutoff} reduces this to order $N\ R_{cutoff}^3$
  - Particle mesh Ewald (PME) method adds long range electrostatics at order N log N, only minor cost compared to cutoff calculation.

- 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.

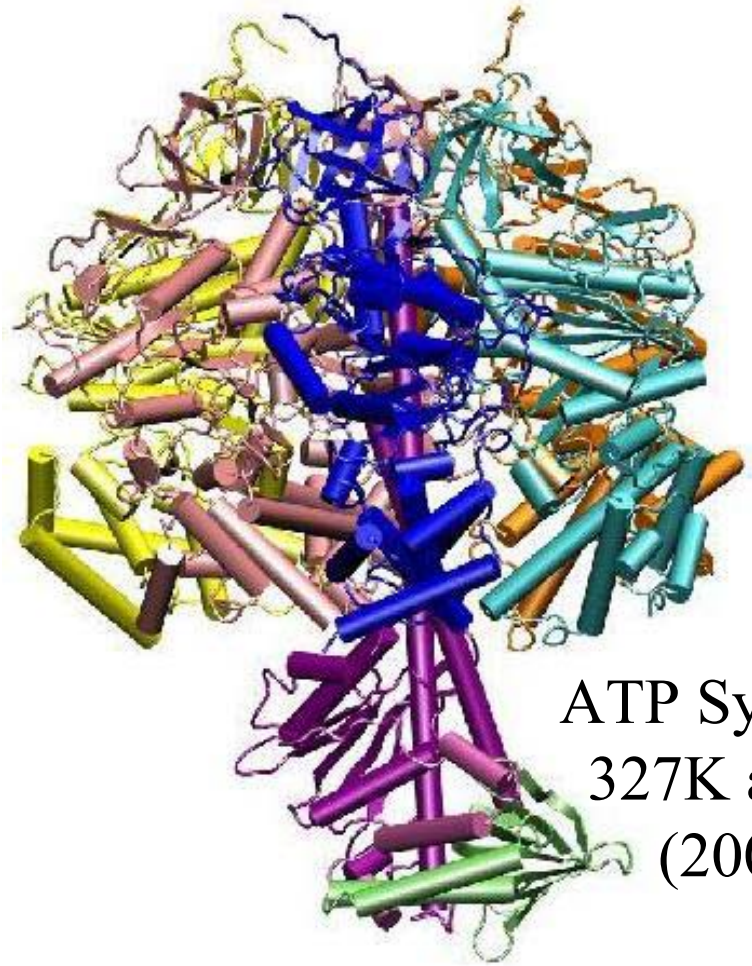
Give a Mouse a Protein

exponential growth in simulation size

BPTI
3K atoms



Estrogen Receptor
36K atoms (1996)



ATP Synthase
327K atoms
(2001)

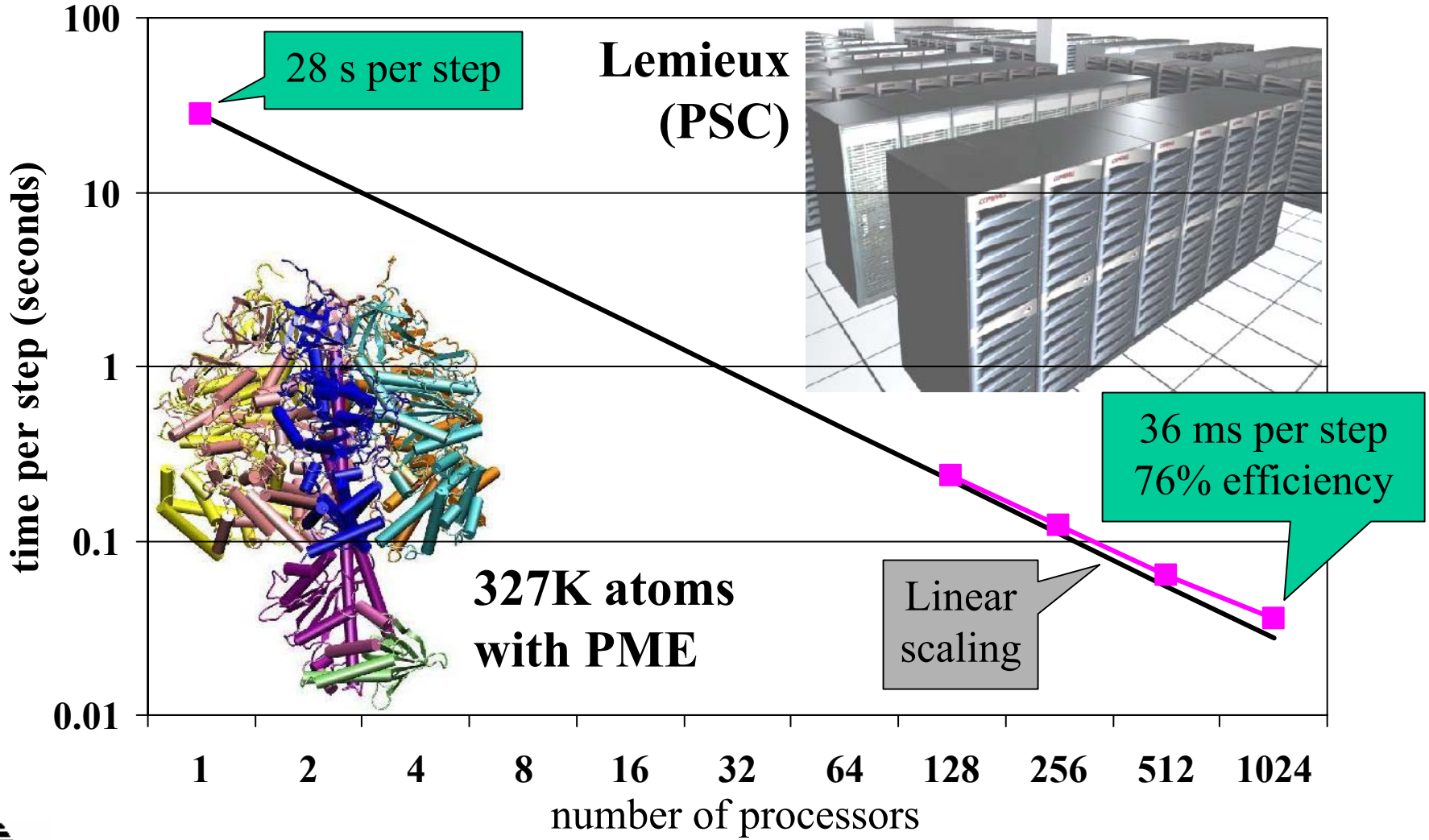
The Road Ahead

- PDB, PSF, topology, and parameter files
- Molecular dynamics
  - …in an ideal world
    - …and in our world
      - …with computers
        - …using NAMD
- Justin prepares a protein using VMD
- You prepare a protein using VMD
  - …and simulate it using NAMD
    - …in the hands-on tomorrow afternoon

Don't worry, the written tutorial is very complete.
You will learn by doing. This talk is an overview.

A Cast of Thousands

NAMD and parallel computing



How to Waste Computer Time

tips for postponing your Ph.D.

- Use large cutoffs instead of PME full electrostatics.

- Run NAMD on more than one processor per 1000 atoms.
- Don't bother measuring parallel efficiency and speedup.

- Compute for several weeks before checking your results.
- Ignore NAMD warnings that you don't understand.

- Try to use NAMD for things it was never meant to do.

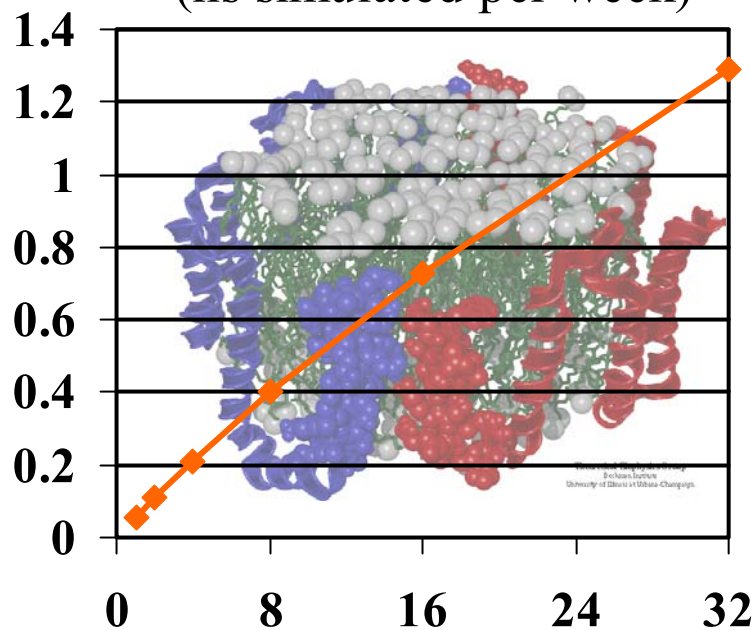
- Build and maintain a Linux cluster for your group. 

Linux Clusters 101

parallel computing on a professor's salary

Learn to build your own Linux cluster in a special hands-on session next week!

92K atoms with PME
(ns simulated per week)



The Road Ahead

- PDB, PSF, topology, and parameter files
- Molecular dynamics
  - …in an ideal world
    - …and in our world
      - …with computers
        - …using NAMD
- Justin prepares a protein using VMD
- You prepare a protein using VMD
  - …and simulate it using NAMD
    - …in the hands-on tomorrow afternoon

Don't worry, the written tutorial is very complete.
You will learn by doing. This talk is an overview.