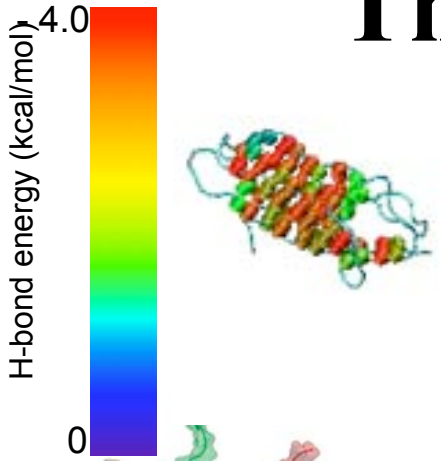
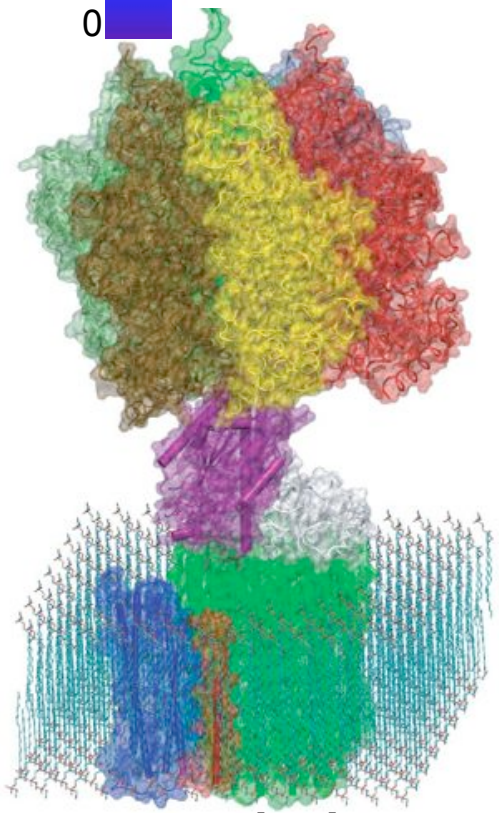
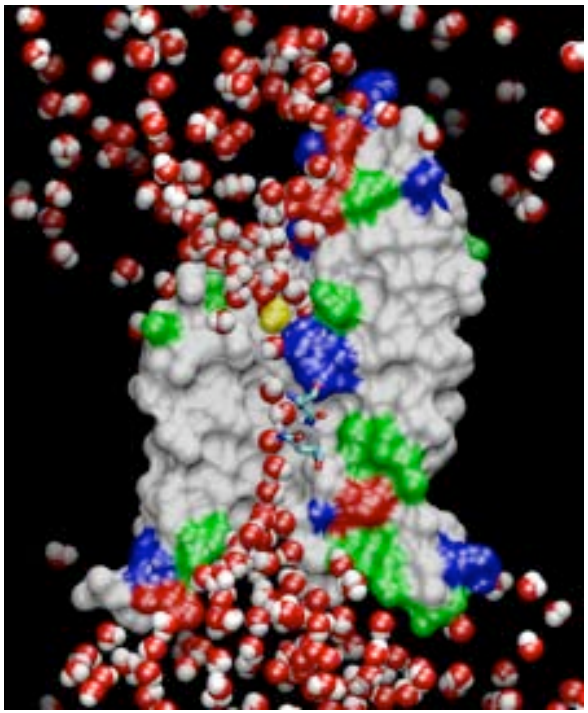
The Molecular Dynamics Method



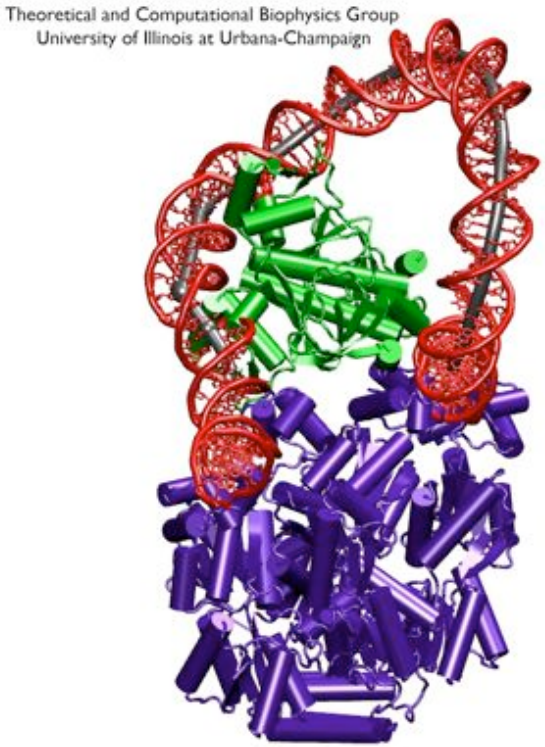
Fibronectin III_1, a mechanical protein that glues cells together in wound healing and in preventing tumor metastasis



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



AQP filtering a bath tub of the body's water a day



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

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

come on, feel the noise

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

$$m\ddot{\vec{r}} = \vec{F}(\vec{r}) - \gamma m\dot{\vec{r}} + \vec{R}(t)$$

$$\langle \vec{R}(t) \cdot \vec{R}(t') \rangle = 6k_B T\gamma\, \delta(t - t')$$

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.

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

+

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

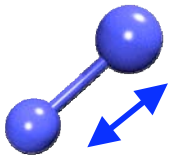
$\cup$

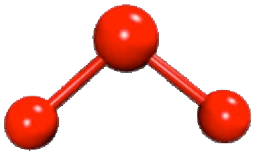
$-\vec{\nabla} U(\vec{R}) / m_i$

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

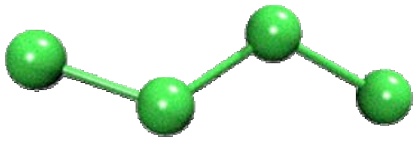
Potential Energy Function of Biopolymer

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





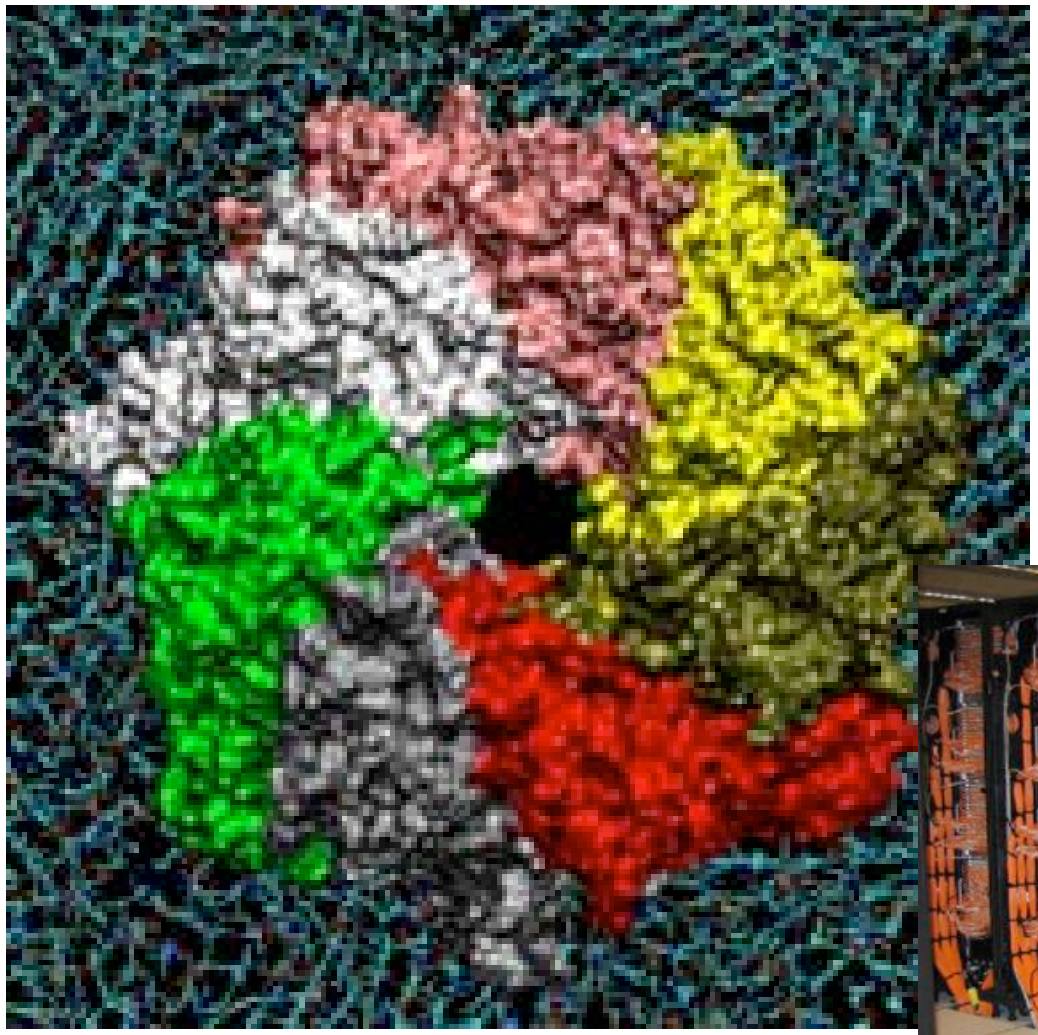
$$U(\vec{R}) = \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} +$$



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

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

Large is no problem. But …

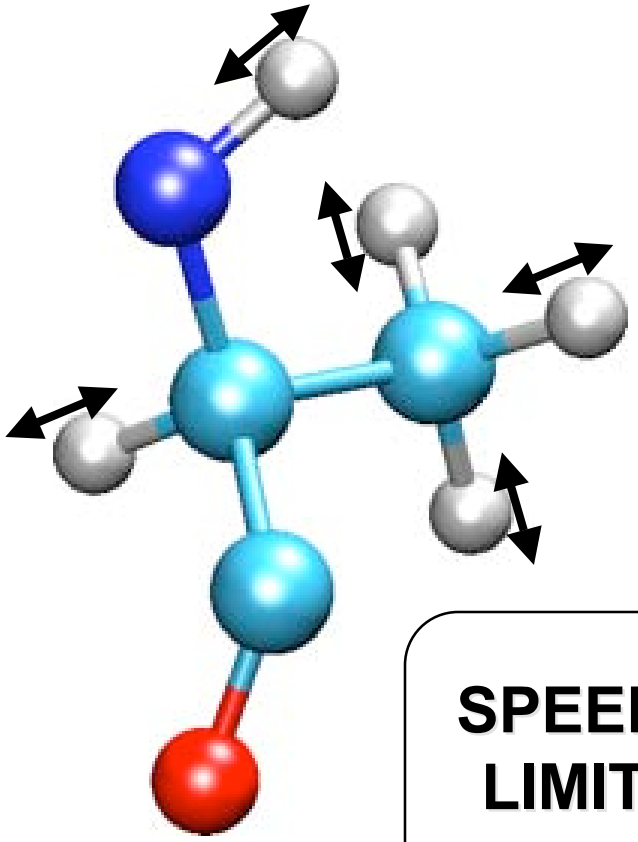


Molecular dynamics simulation of alpha-hemolysin with about 300,000 atoms

NCSA machine room

But long is!

biomolecular timescale and timestep limits



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

SPEED LIMIT

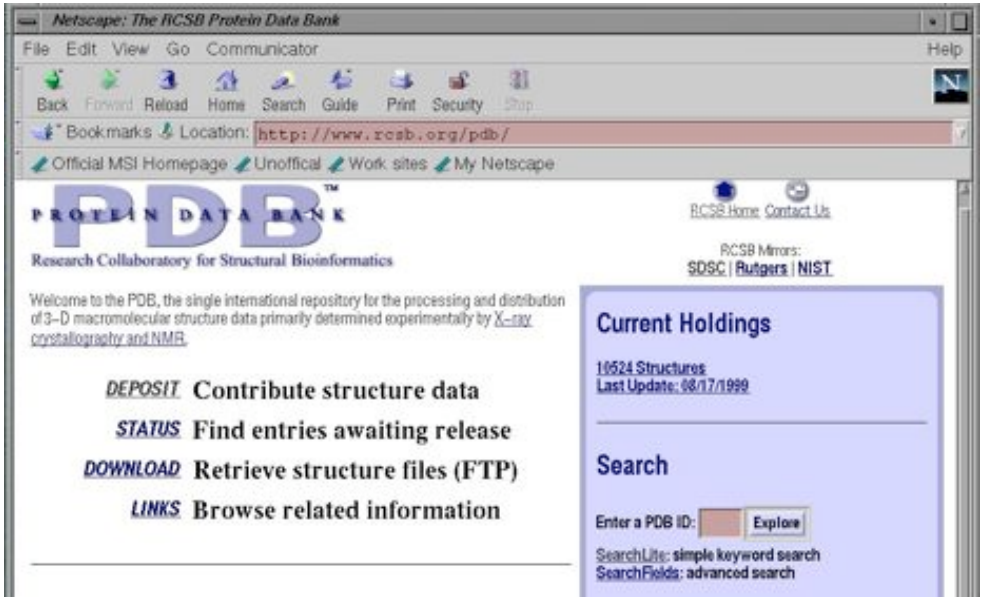
δt = 1 fs

PDB Files

a little information

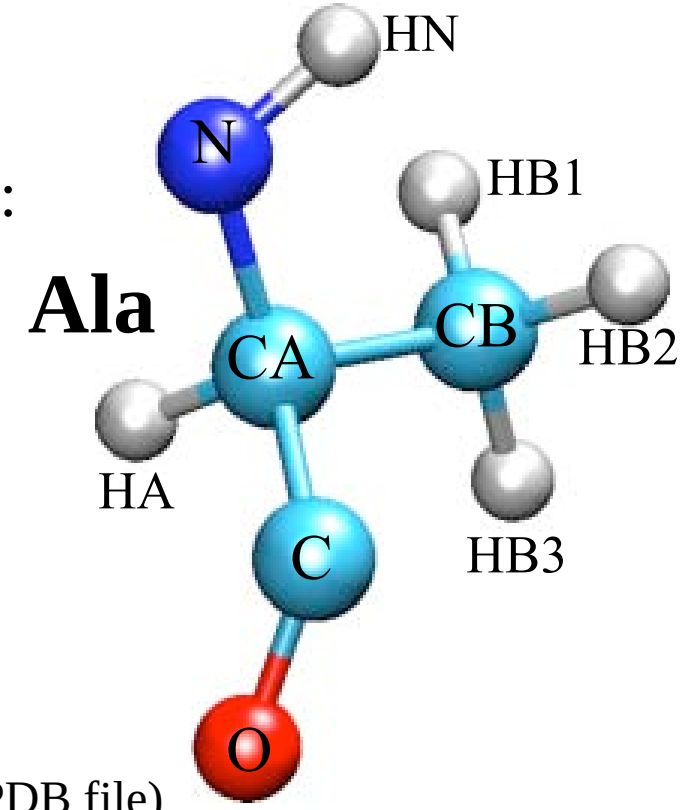
- Simulations start with a crystal structure from the Protein Data Bank, in the standard PDB file format.
- PDB files contain standard records for species, tissue, authorship, citations, sequence, secondary structure, etc.
- We only care about the atom records…
  - atom name (N, C, CA)
  - residue name (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.)



PSF Files

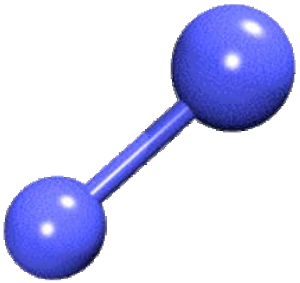
atomic properties (mass, charge, type)



- 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)

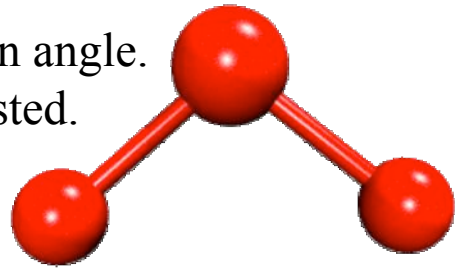
PSF Files

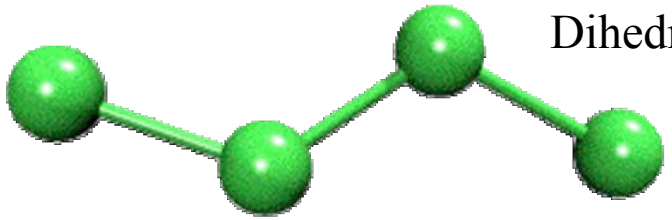
molecular structure (bonds, angles, etc.)



Bonds: Every pair of covalently bonded atoms is listed.

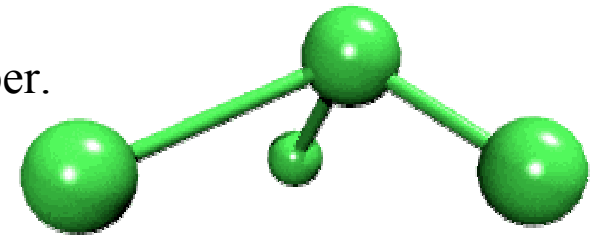
Angles: Two bonds that share a common atom form an angle.
Every such set of three atoms in the molecule is listed.





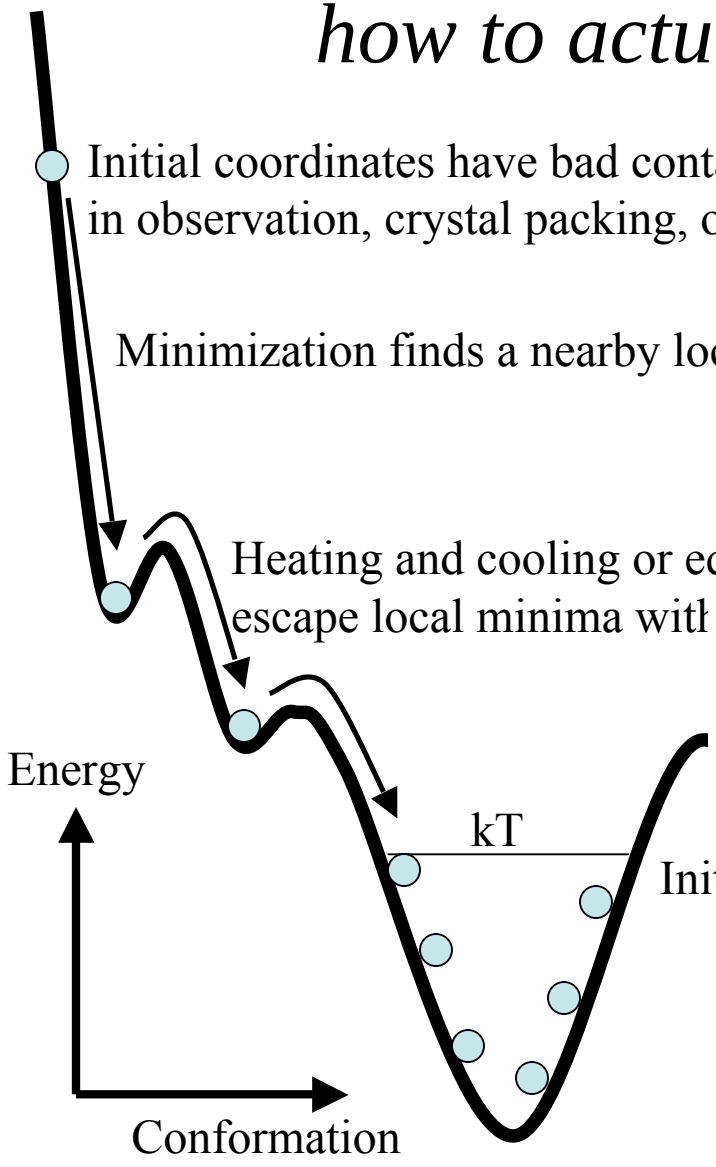
Dihedrals: Two angles that share a common bond form a dihedral.
Every such set of four atoms in the molecule is listed.

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



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)

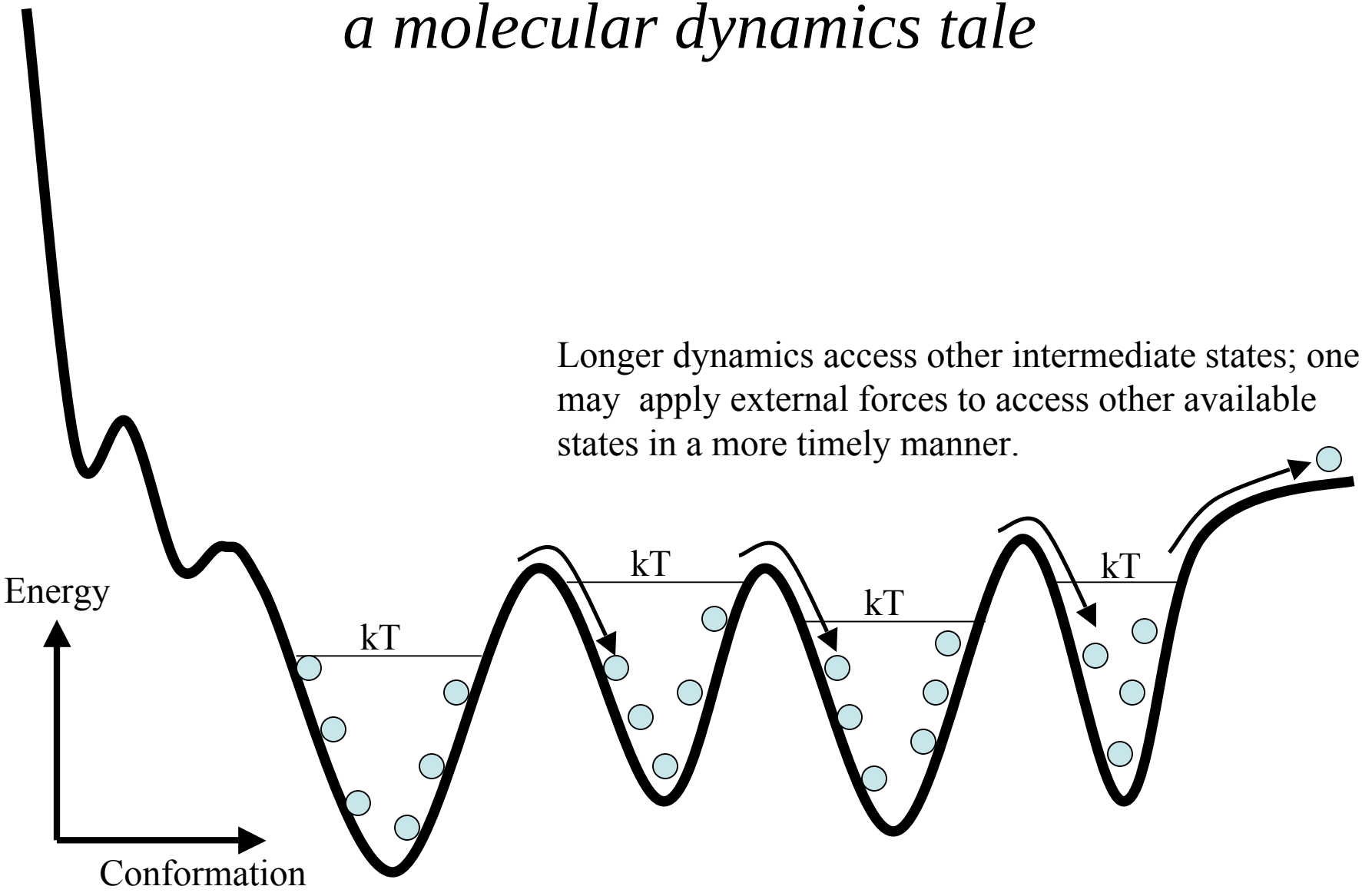
Minimization finds a nearby local minimum.

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

Initial dynamics samples thermally accessible states.

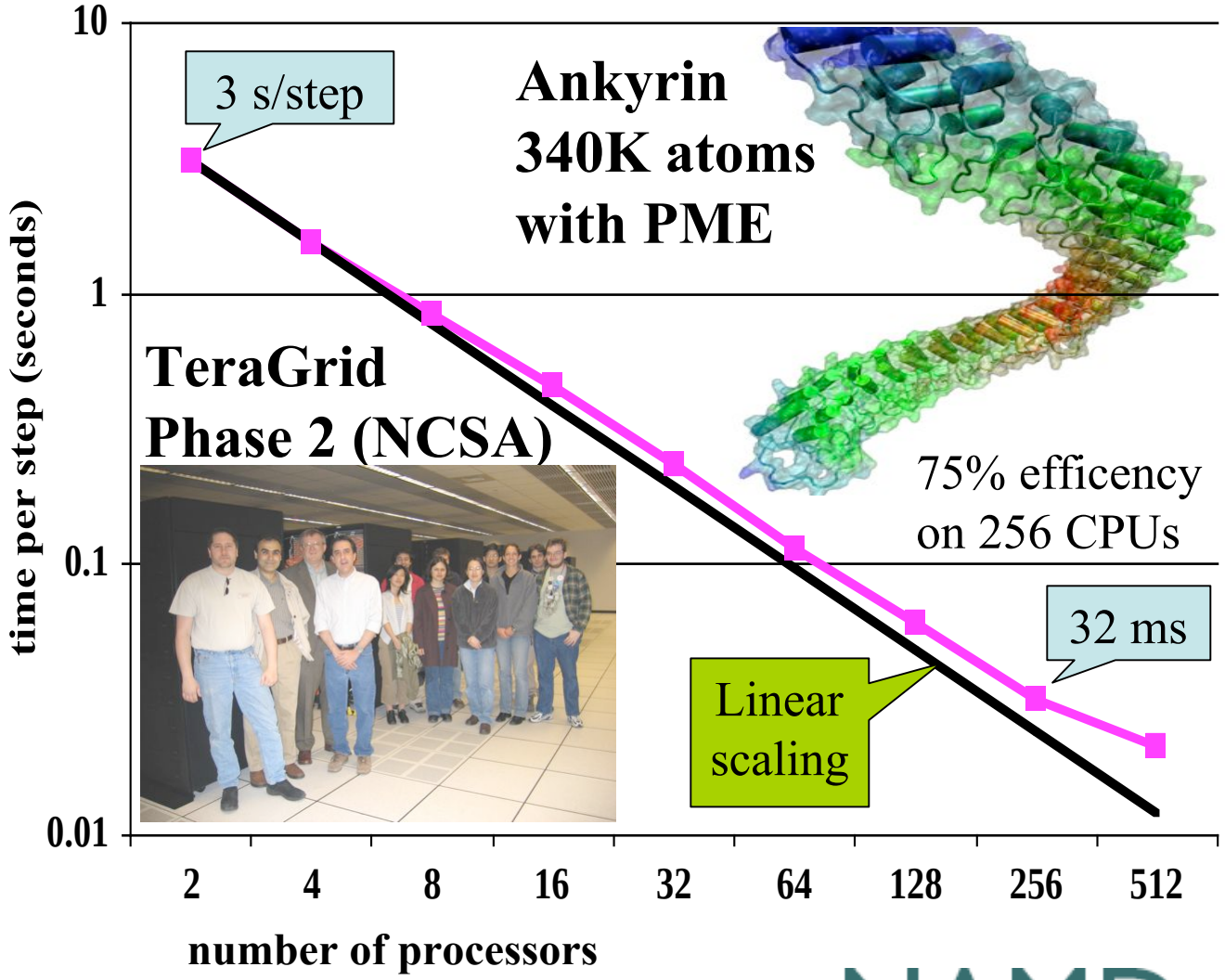
From the Mountains to the Valleys

a molecular dynamics tale



NAMD: The Program we will Use

NAMD programmer J. Phillips Ph.D. UIUC Physics



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.

NAMD

Scalable Molecular Dynamics

CHARMM Potential Function

- Simple, fixed algebraic form for every type of interaction.
- Form stems from compromise between accuracy and speed.
- 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!

Molecular Dynamics Method

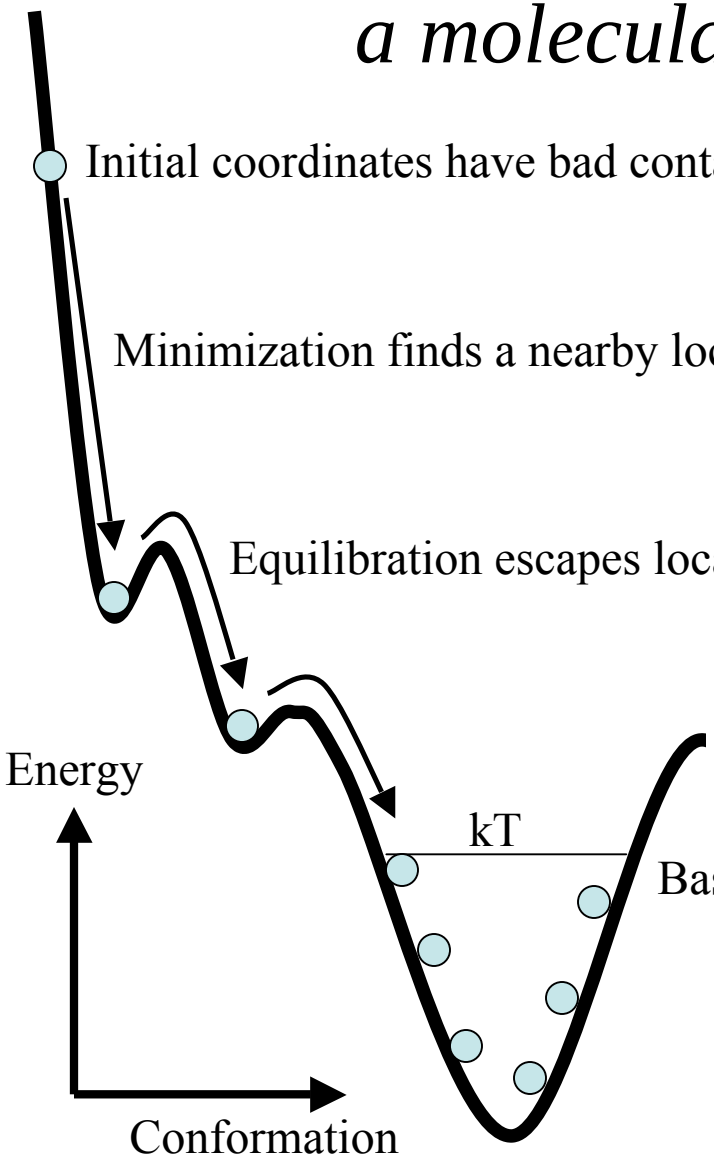
- PDB, PSF, topology, and parameter files
- Molecular dynamics
  - …in an ideal world
    - …and in our world
      - …with computers
        - …using NAMD
- Preparing a protein using VMD
- You prepare a protein using VMD
  - …and simulate it using NAMD
    - …in the hands-on Tuesday afternoon

Don't worry, the written tutorial is very complete.

You will learn by doing. This talk is an overview.

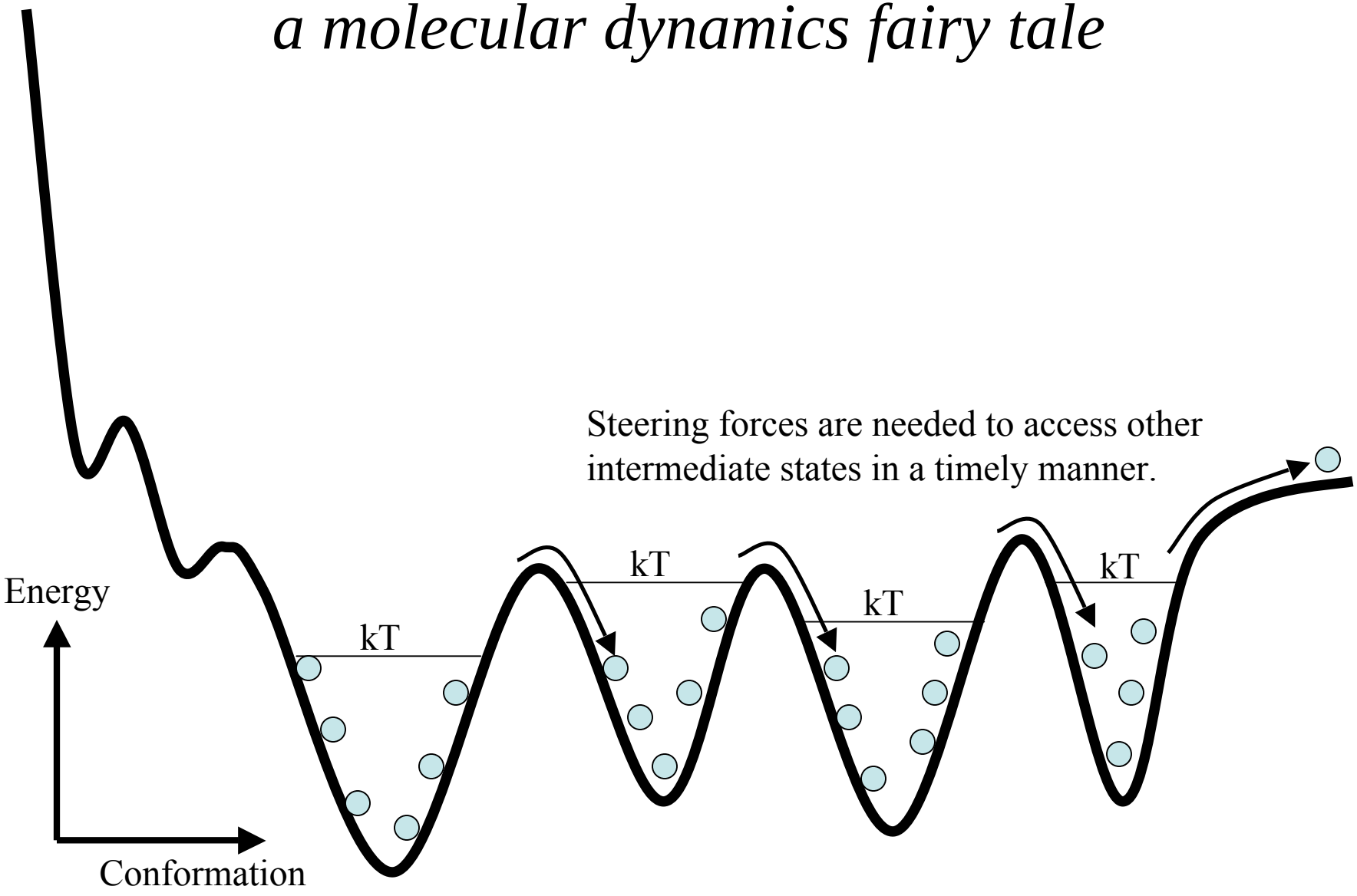
From the Mountains to the Valleys

a molecular dynamics fairy tale



From the Mountains to the Valleys

a molecular dynamics fairy tale



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.

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

+

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

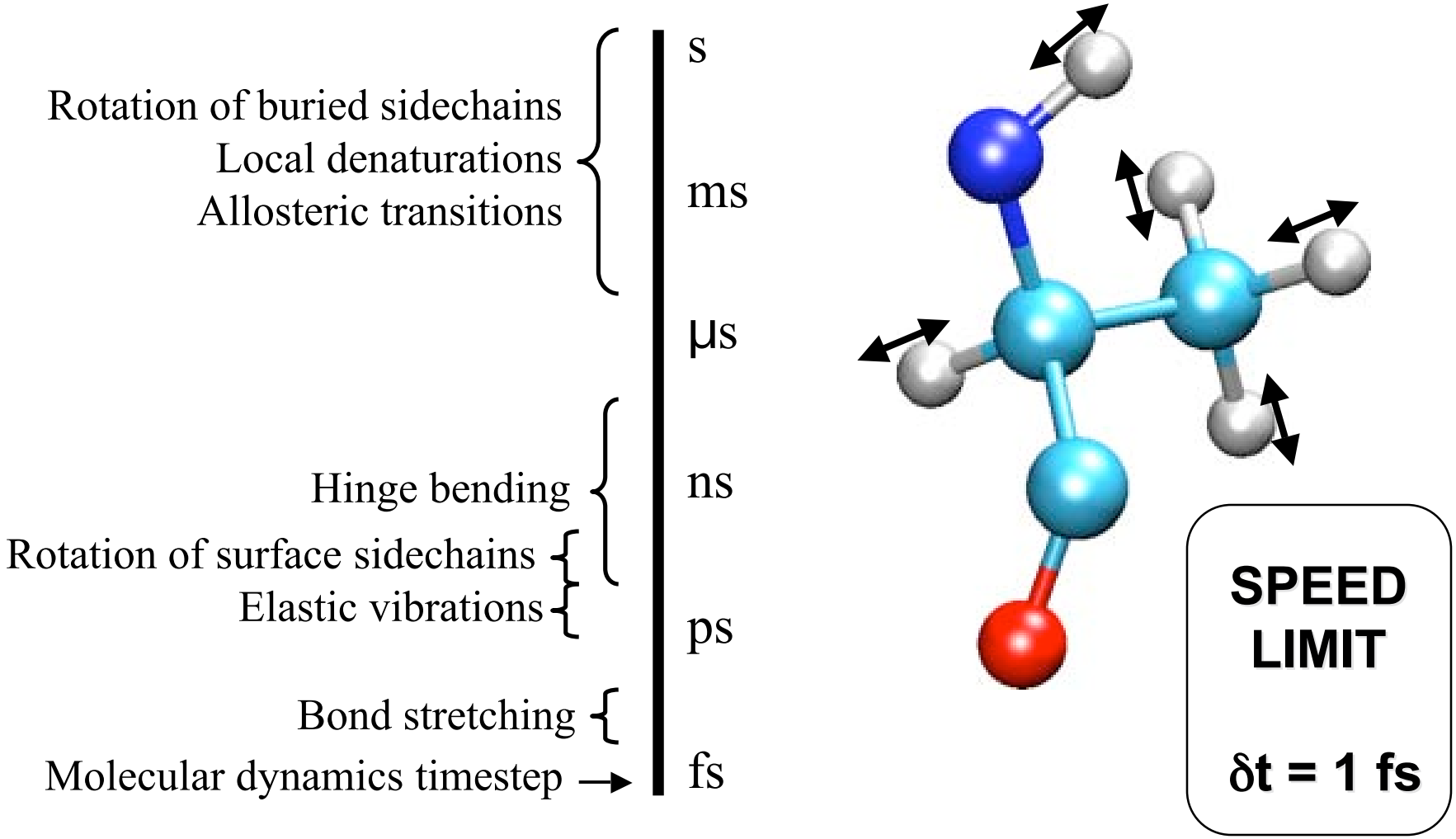
$\cup$

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

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

Hurry Up and Wait

biomolecular timescale and timestep limits



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.

Linux Clusters 101

parallel computing on a professor's salary

Learn to build your own Linux cluster!