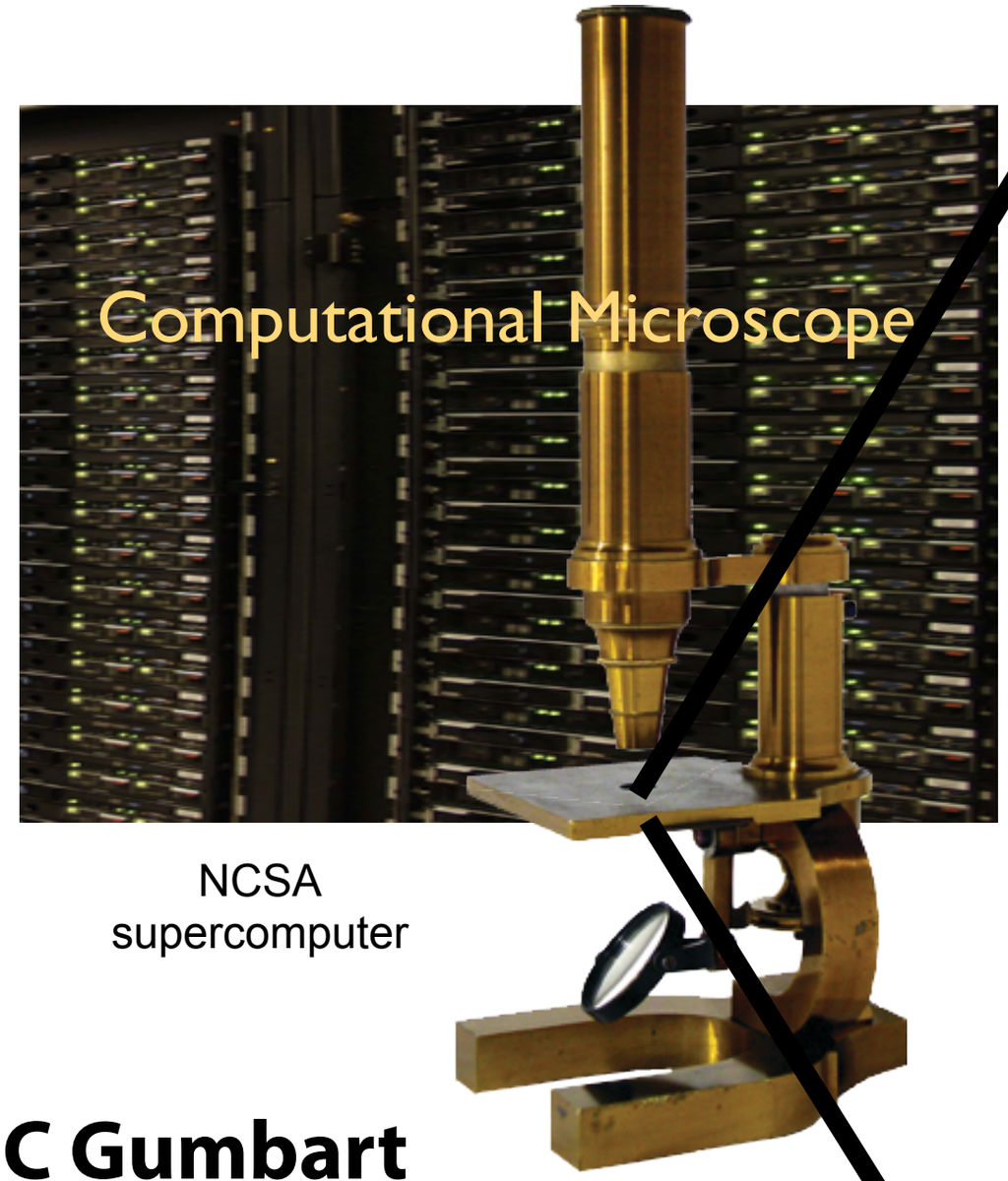
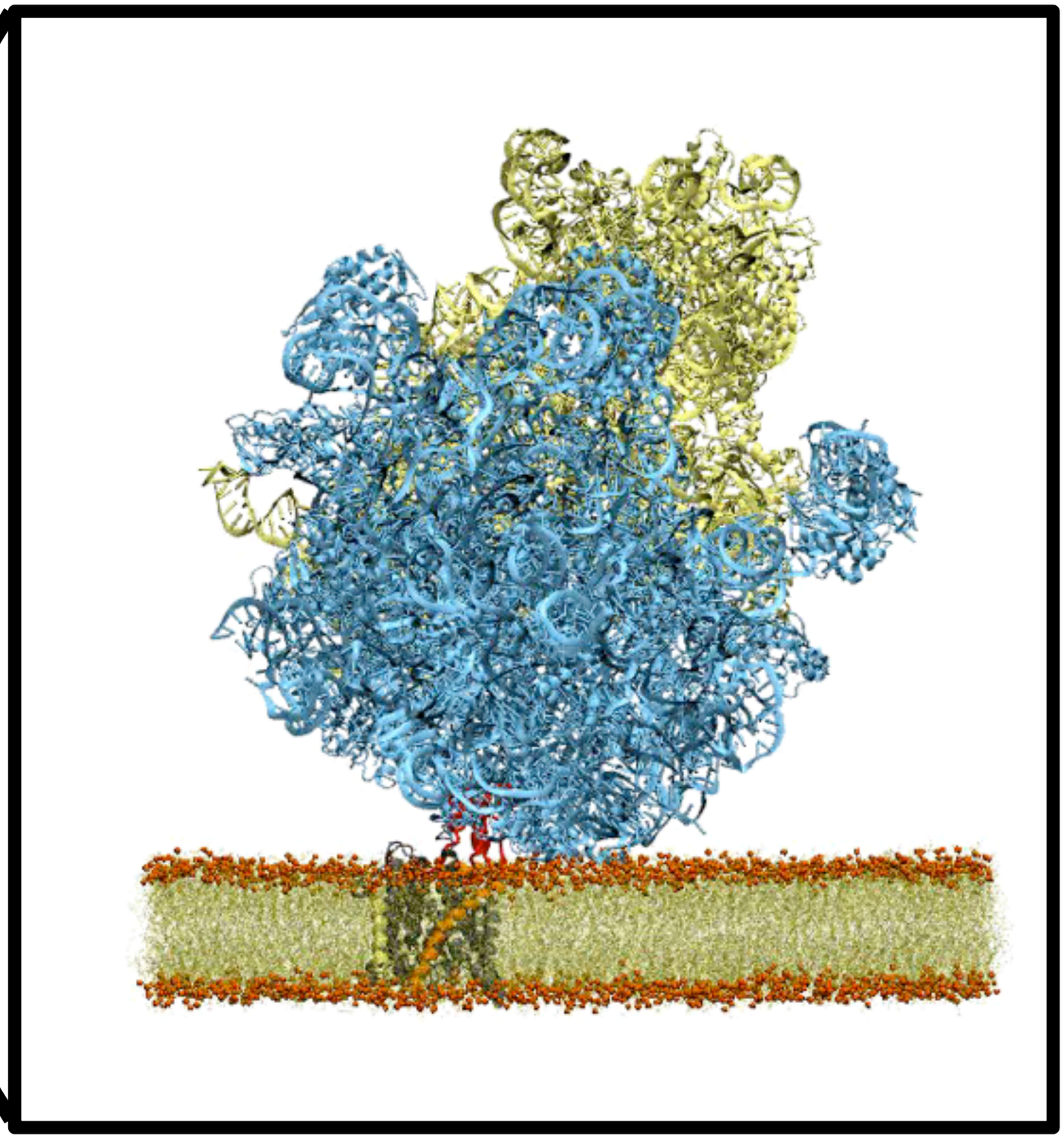
Principles and Applications of Molecular Dynamics Simulations with NAMD

Nov. 14, 2016



NCSA
supercomputer



JC Gumbart
Assistant Professor of Physics
Georgia Institute of Technology
gumbart@physics.gatech.edu
simbac.gatech.edu

These workshops started as one person's dream

Klaus Schulten **1947-2016**

"We really feel that you cannot teach just by lecturing; you have to teach through hands-on examples."

This is workshop **#45**

I've taught at workshops **4, 5, 8, 9, 10, 12, 14, 15, 20, 21, 34, 35**

Biophysics Research at Georgia Tech

Physics of Living Systems @ GT

km

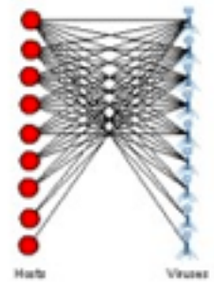
m

mm

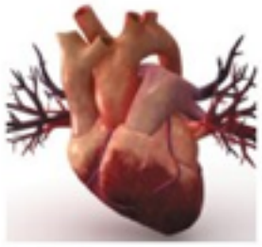
µm

nm

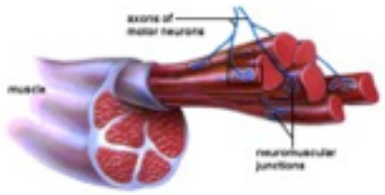
Weitz*
(population dynamics)



Goldman/Hu*
(biomechanics)



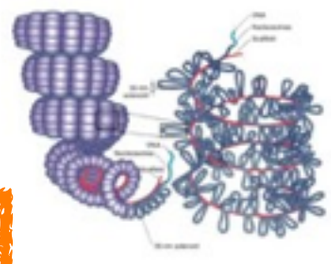
Fenton
(tissue mechanics)



Sponberg/Goldman
(neuromechanics)

Curtis
(cell physics)

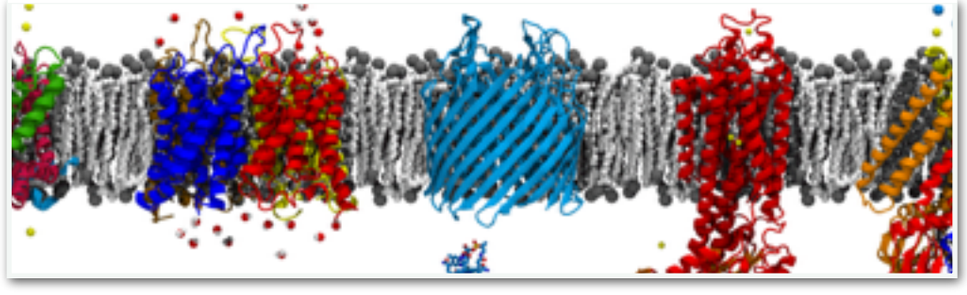
Kim
(genome biophysics)



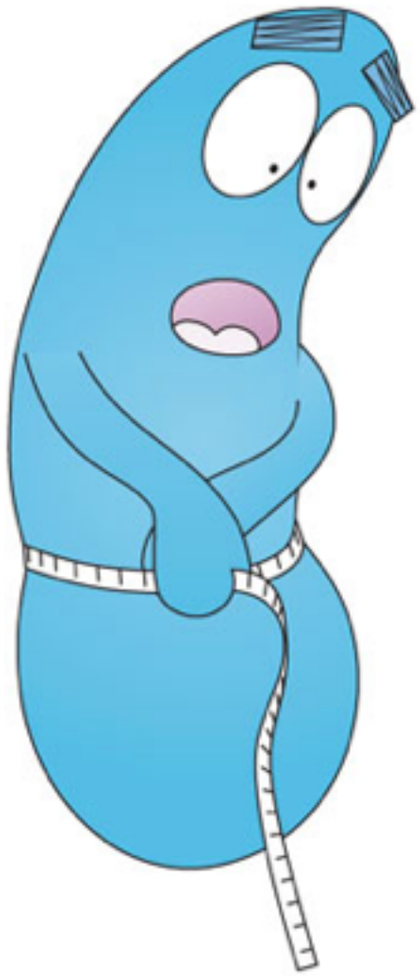
Gumbart
(molecular biophysics)

*Denotes adjunct in SoP, but core PoLS faculty

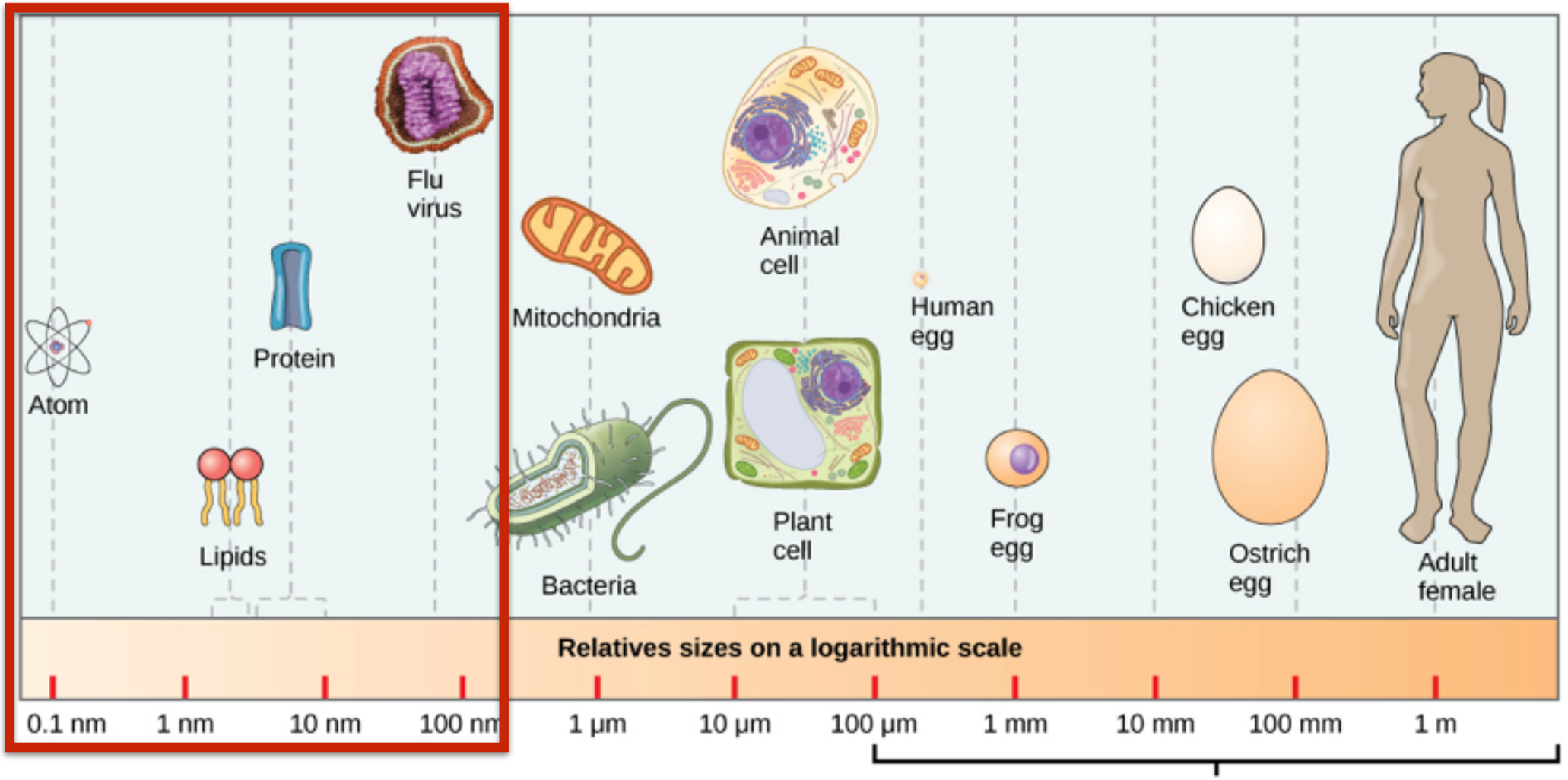
SIMBAC(**sim**ulations of **bac**terial systems)
Lab at Georgia Tech



How small are the things we simulate?



VERY, VERY SMALL!!!!

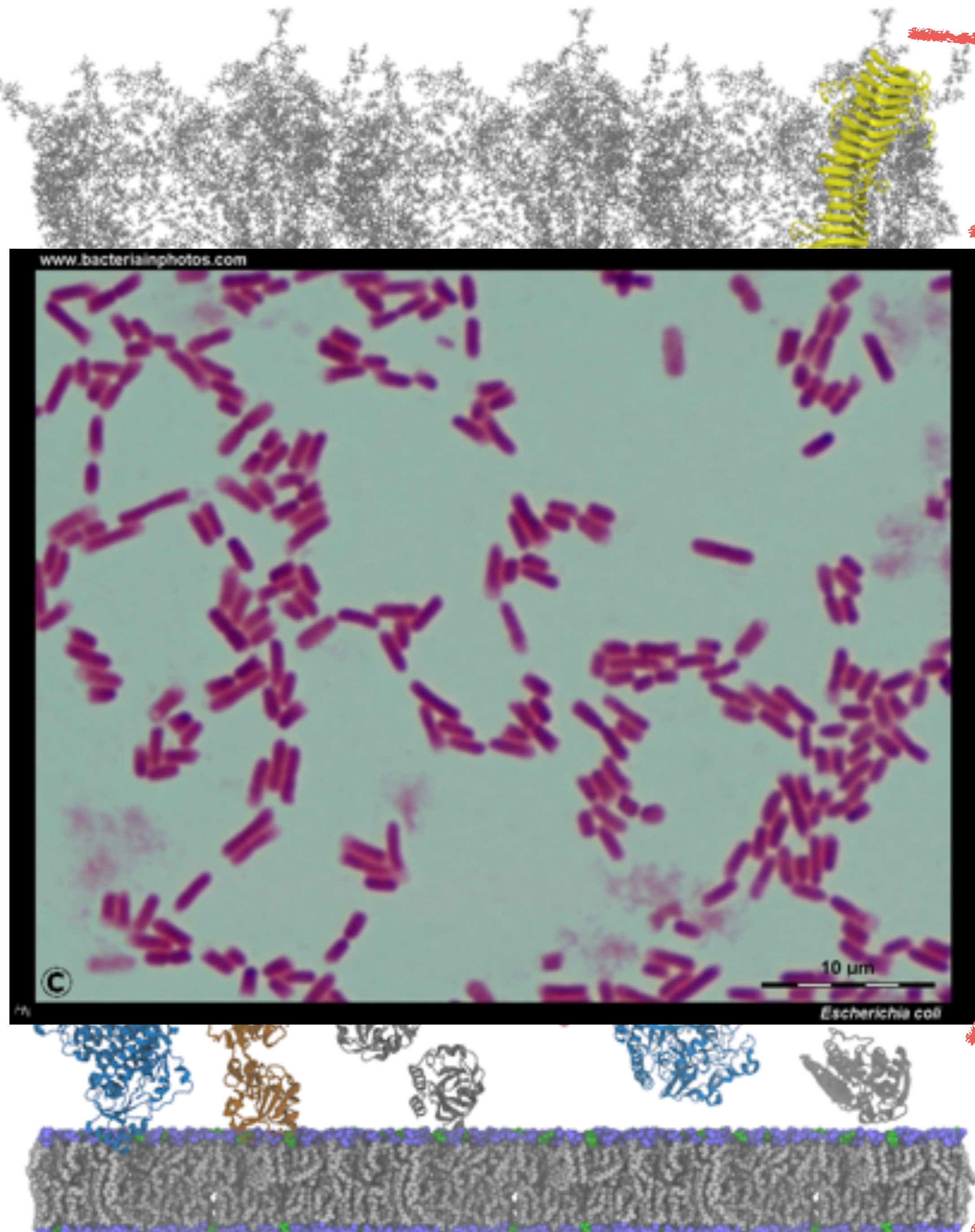


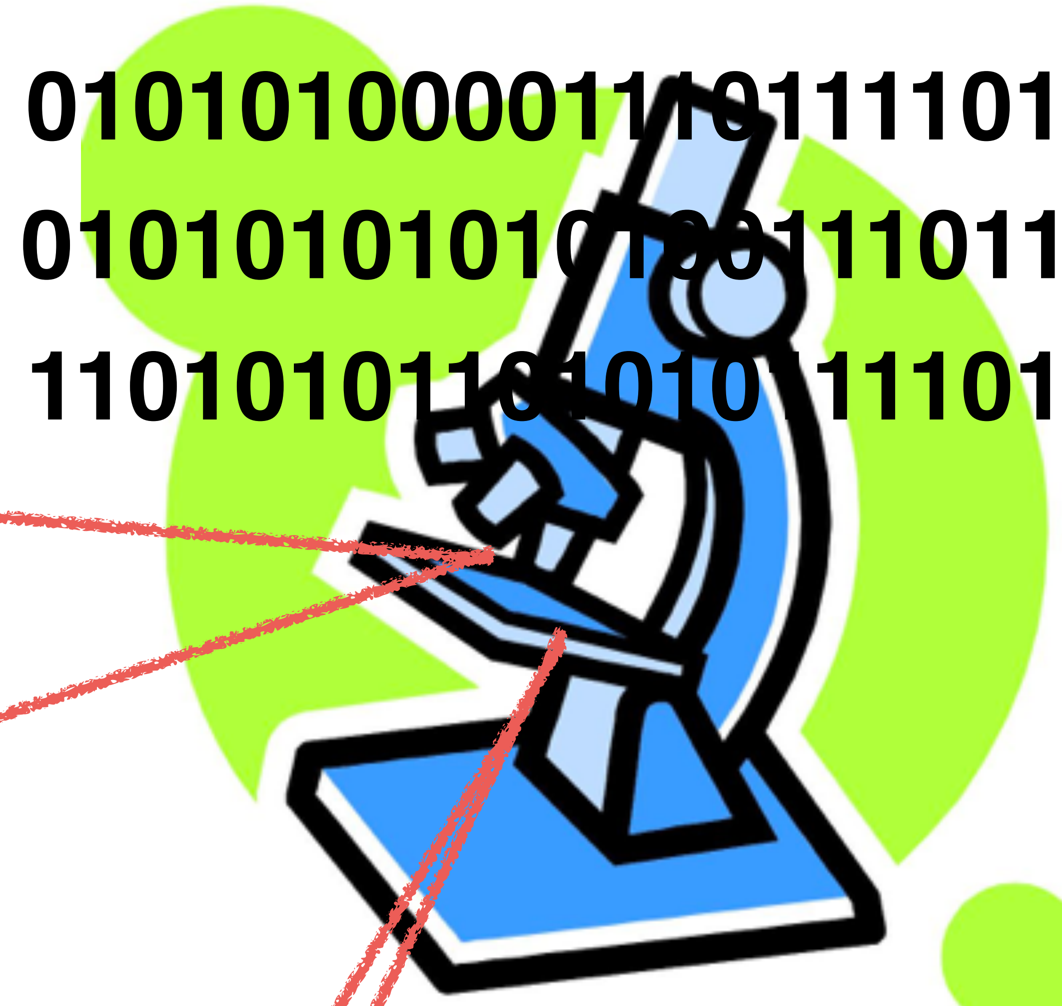
can be counted in atoms!

can not even be seen with visible light!

MD simulations as a computational microscope

the traditional light microscope can **barely** see bacteria

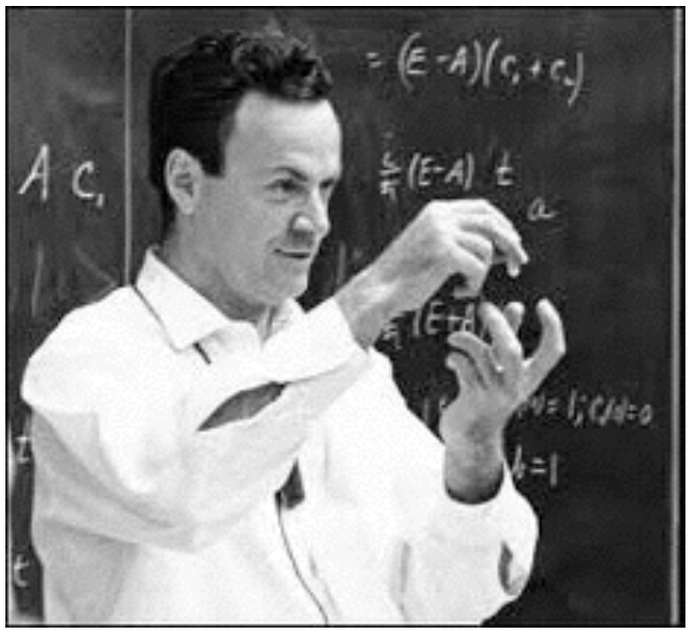




Scalable molecular dynamics with NAMD. J. C. Phillips, R. Braun, W. Wang, J. Gumbart et al. *J. Comp. Chem.*, 26:1781-1802, 2005.

the computational microscope can "see" **everything** with infinite resolution

"Everything that living things do can be reduced to wiggling and jiggling of atoms."



Richard Feynman (1963)

The Nobel Prize in Chemistry 2013

Photo: A. Mahmoud
Martin Karplus
Prize share: 1/3

Photo: A. Mahmoud
Michael Levitt
Prize share: 1/3

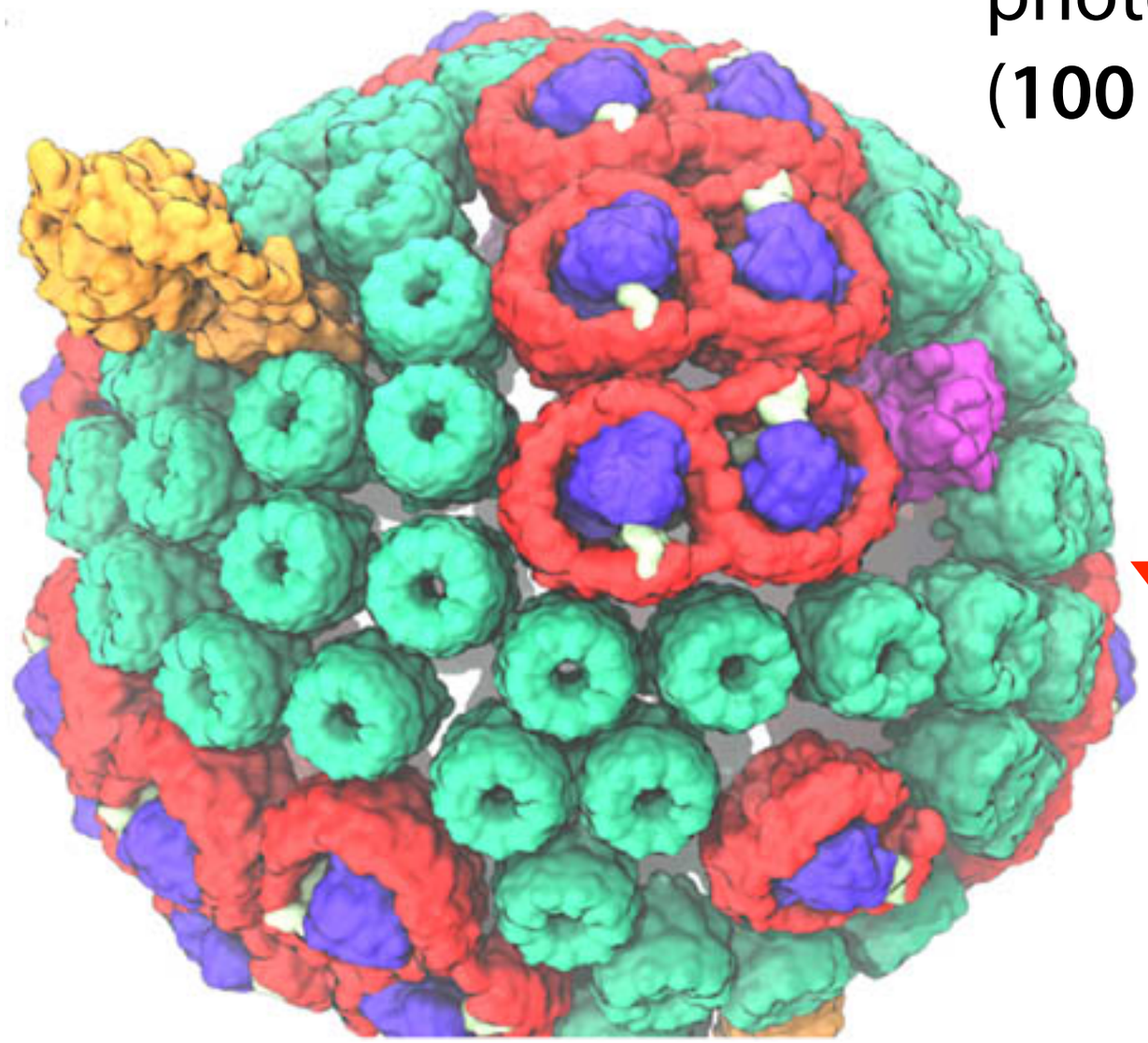
Photo: A. Mahmoud
Arieh Warshel
Prize share: 1/3

The Nobel Prize in Chemistry 2013 was awarded jointly to Martin Karplus, Michael Levitt and Arieh Warshel *"for the development of multiscale models for complex chemical systems"*.

50 years later…

But we are far from done!

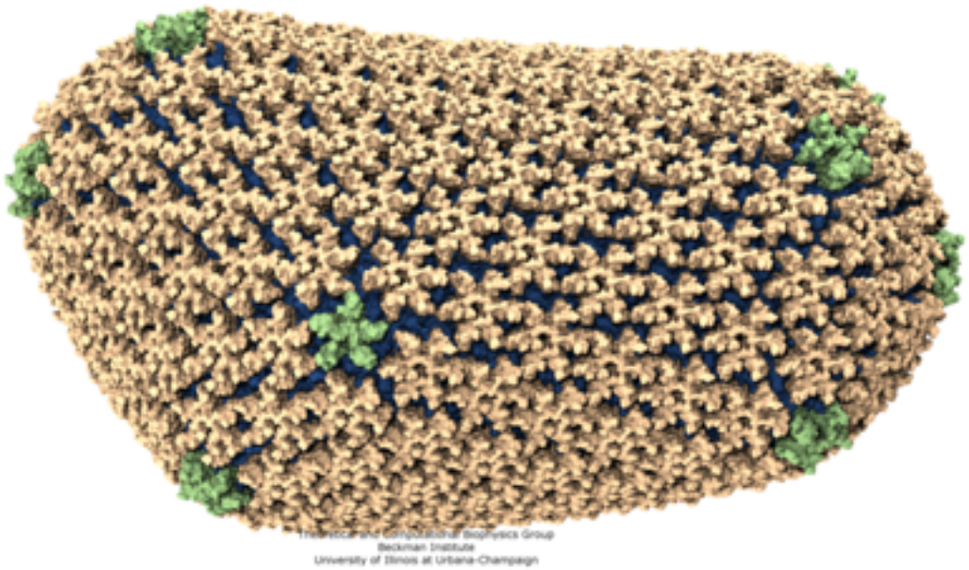
State of the art



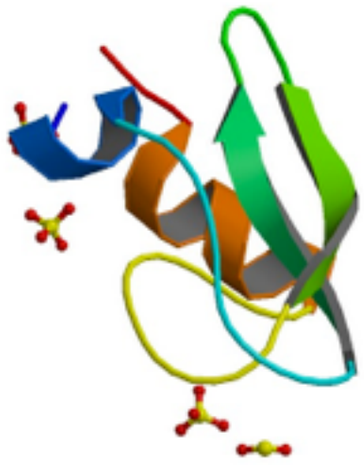
photosynthetic chromatophore (**100 million** atoms, 2016)

Sener, Strumpfer, Singharoy, Hunter, Schulten. Overall energy conversion efficiency of a photosynthetic vesicle. *eLife*, 5:e09541, 2016.

HIV virus capsid (**64 million** atoms, 2013)



Zhao, Perilla, Yufenyuy, Meng, Chen, Ning, Ahn, Gronenborn, Schulten, Aiken, Zhang. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*, 497:643-646, 2013.



individual protein (**500 atoms,** 1977)

McCammon, Gelin, Karplus. Dynamics of folded proteins. *Nature*, 267:585-590, 1977.

Why do we use MD?

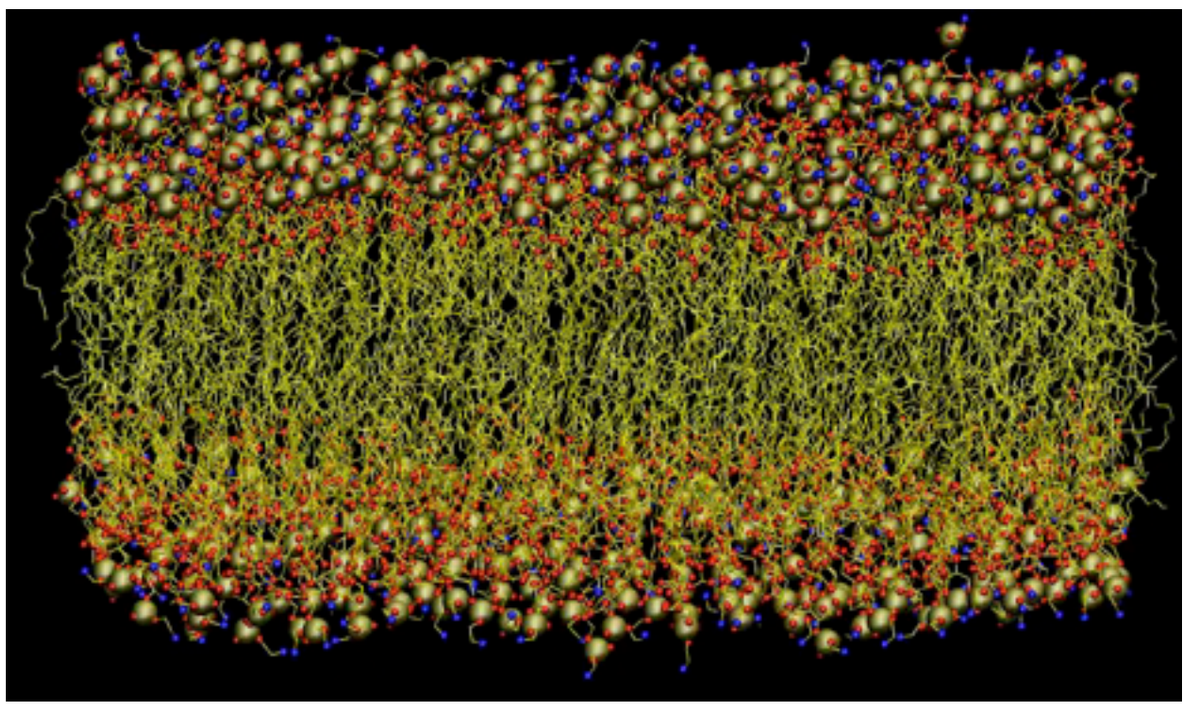
"It is nice to know that the computer understands the problem. But I would like to understand it, too."

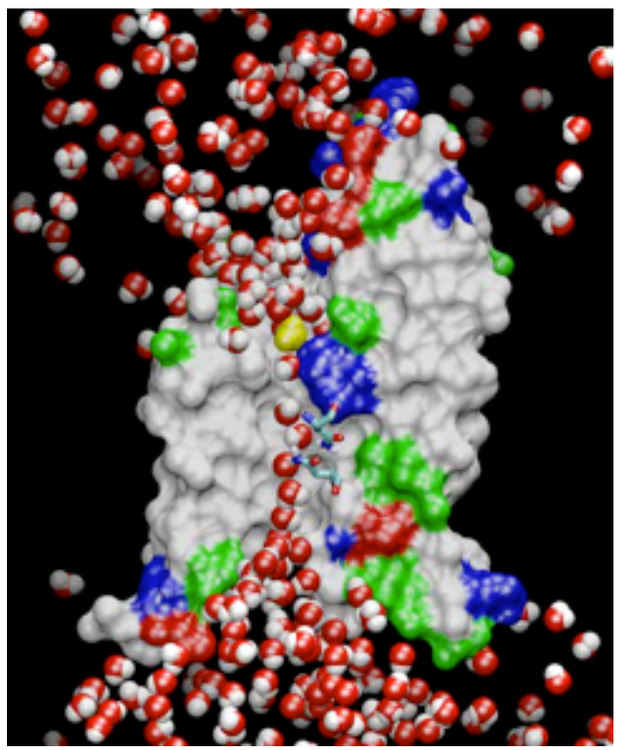
Eugene Wigner

We don't simulate just to watch! But also to measure, analyze, and understand

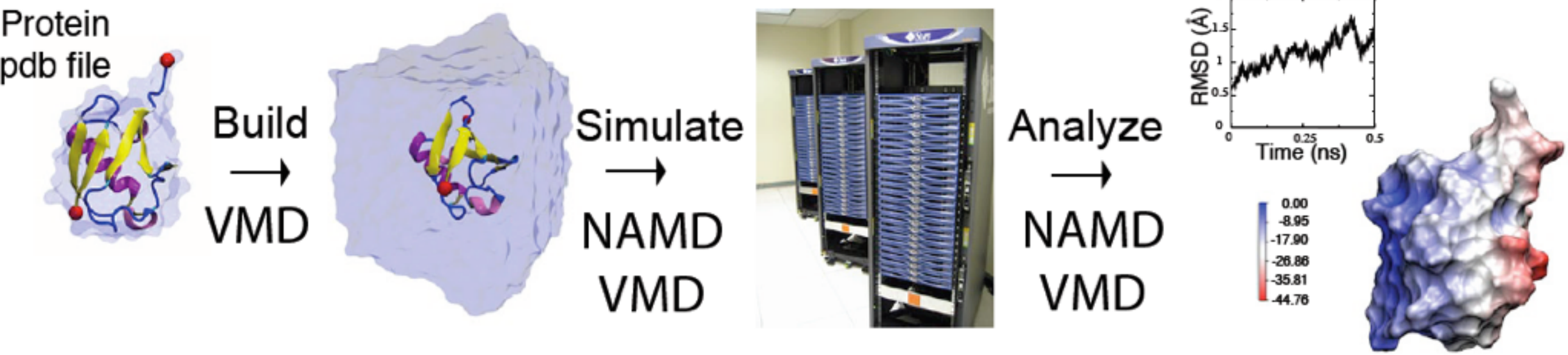
Why do we use MD?

- Generating a thermodynamic ensemble (sampling / statistics)
- Taking into account fluctuations/dynamics in interpretation of experimental observables
- Describing molecular processes + free energy
- Help with molecular modeling





The Molecular Dynamics Simulation Process



For textbooks see:

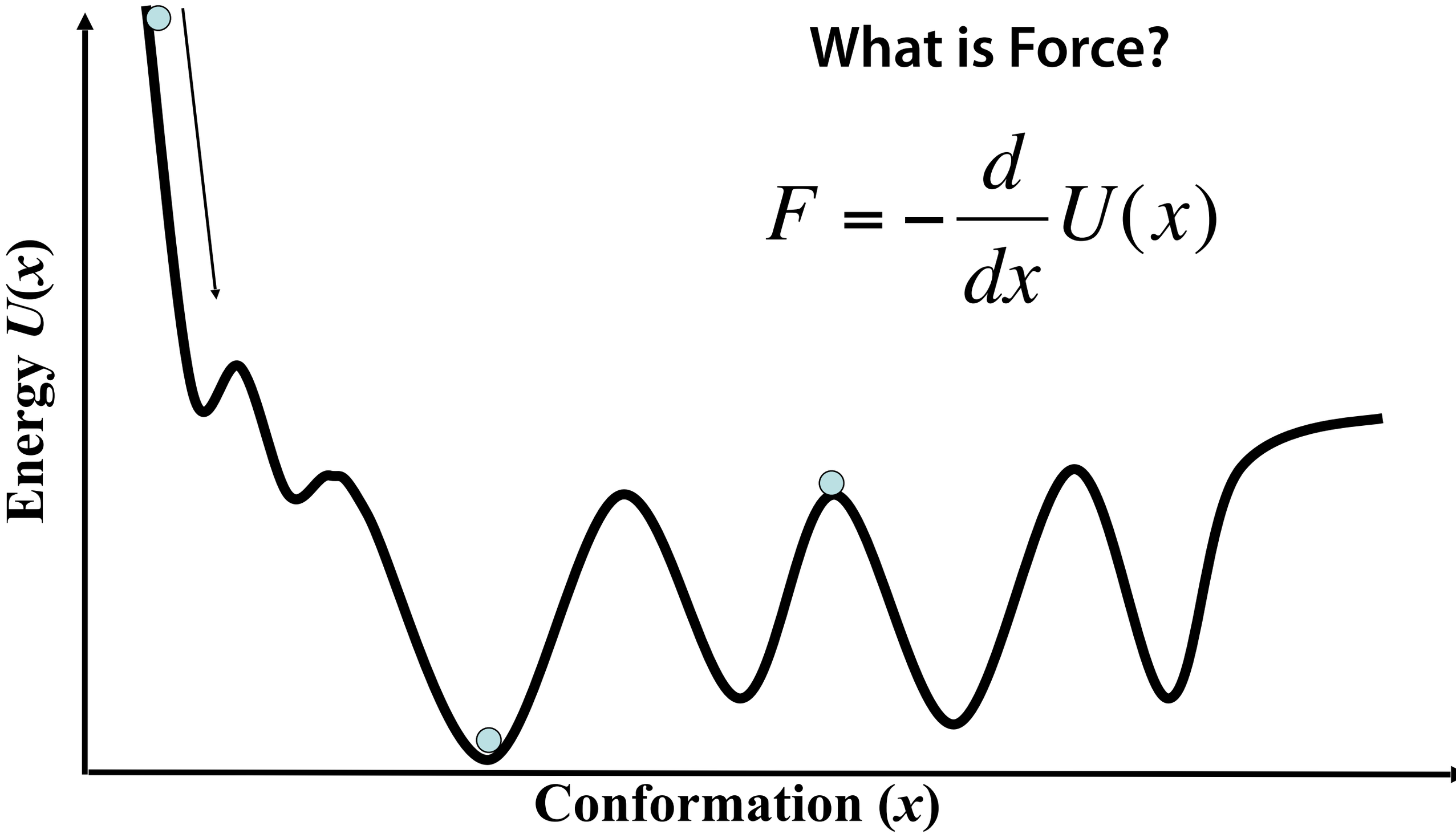
M.P. Allen and D.J. Tildesley. ***Computer Simulation of Liquids.***Oxford University Press, New York, 1987.

D. Frenkel and B. Smit. ***Understanding Molecular Simulations. From Algorithms to Applications.*** Academic Press, San Diego, California, 1996.

A. R. Leach. ***Molecular Modelling. Principles and Applications.***Addison Wesley Longman, Essex, England, 1996.

More at http://www.biomath.nyu.edu/index/course/99/textbooks.html

Potential Energy (hyper)Surface



What is Force?

$$F = -\frac{d}{dx}U(x)$$

Classical Molecular Dynamics *at 300 K*

Energy function: $U(\vec{r}_1, \vec{r}_2, \cdots \vec{r}_N) = U(\vec{R})$

used to determine the force on each atom:

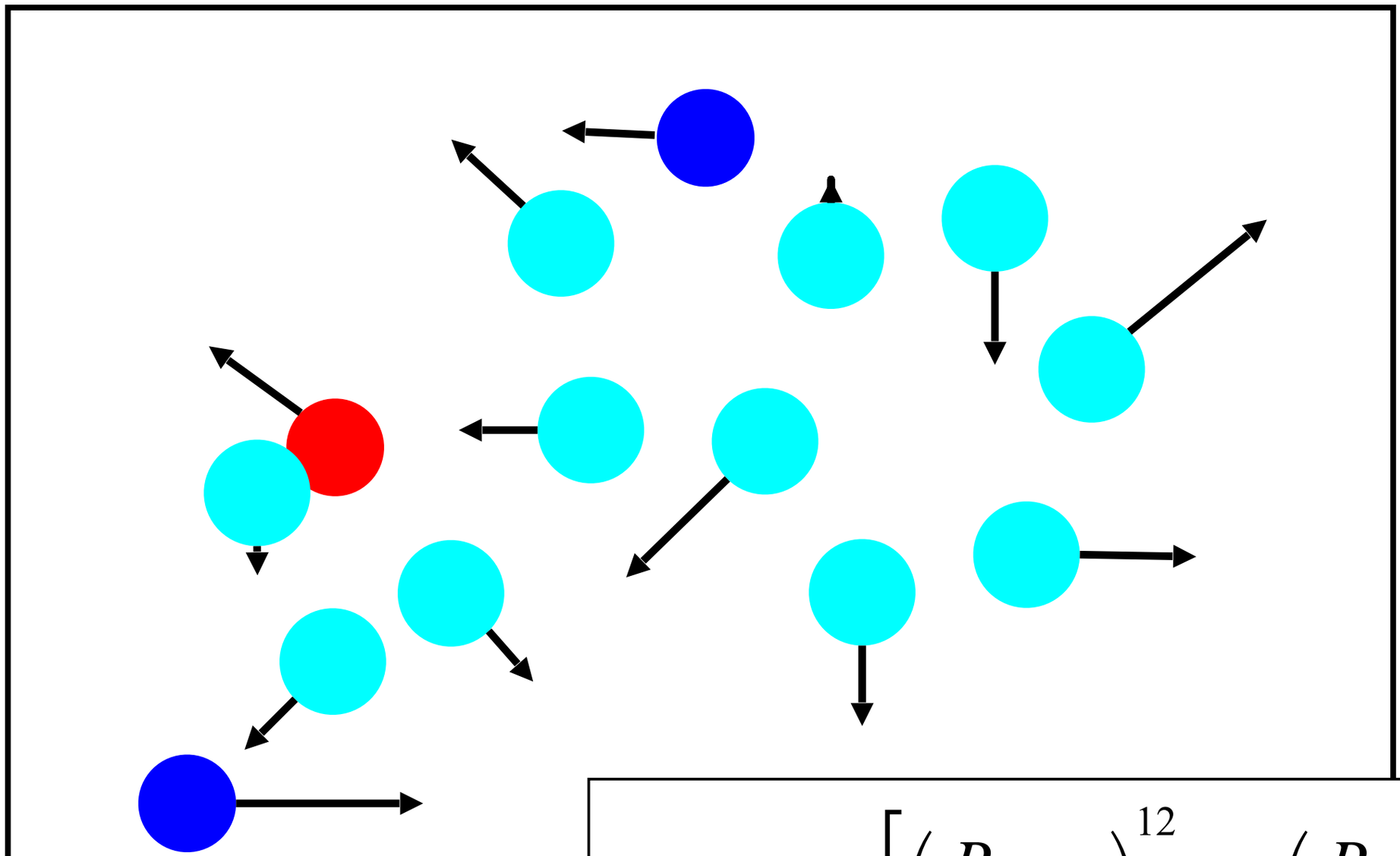
$$m_i \frac{d^2\vec{r}_i}{dt^2} = \vec{F}_i = -\vec{\nabla} U(\vec{R})$$

yields a set of 3*N* 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.

Classical Molecular Dynamics

tryptophan



$$U(r) = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}}$$

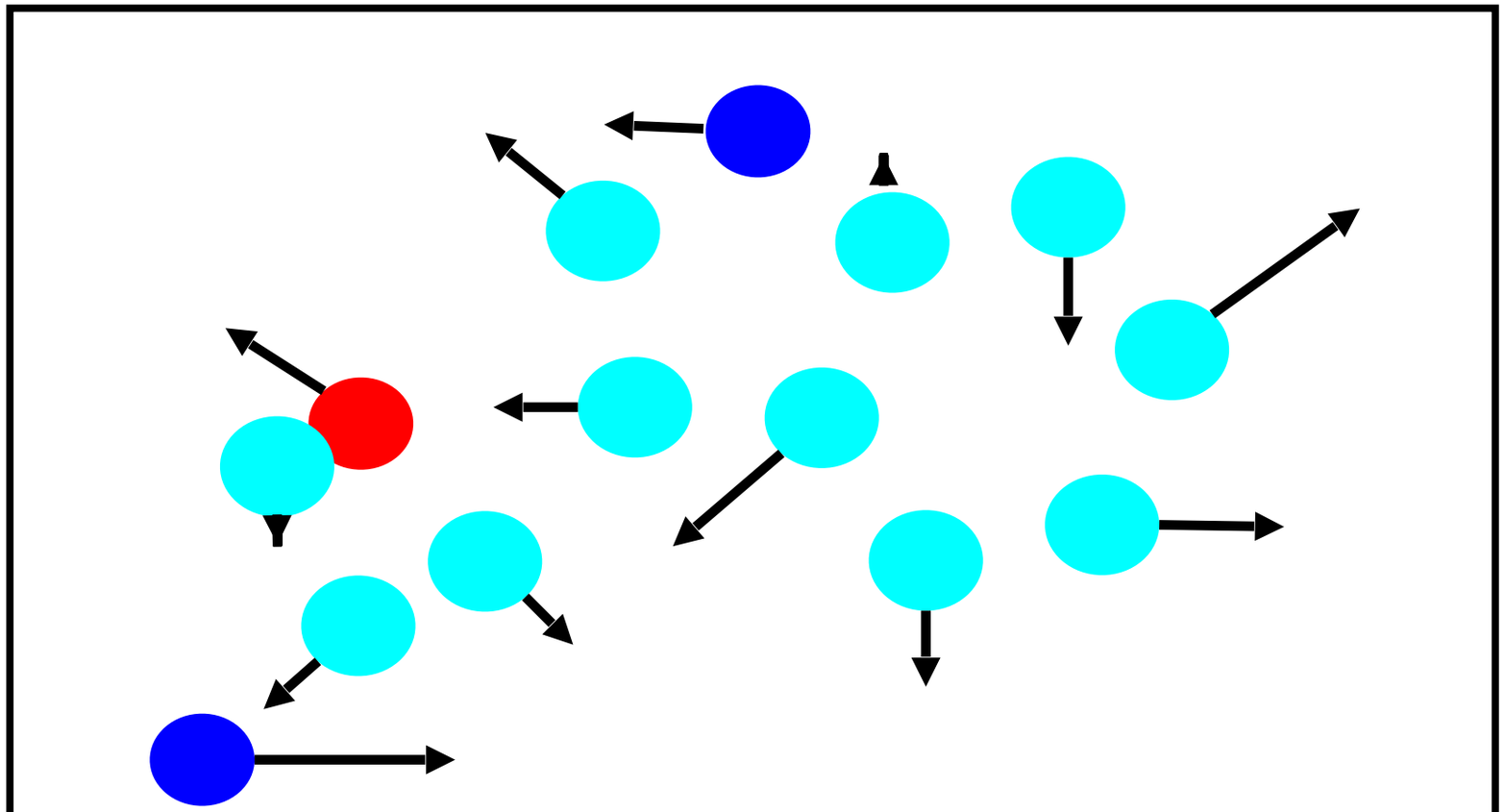
Coulomb interaction

$$U(r) = \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$

van der Waals interaction

Classical Molecular Dynamics

tryptophan

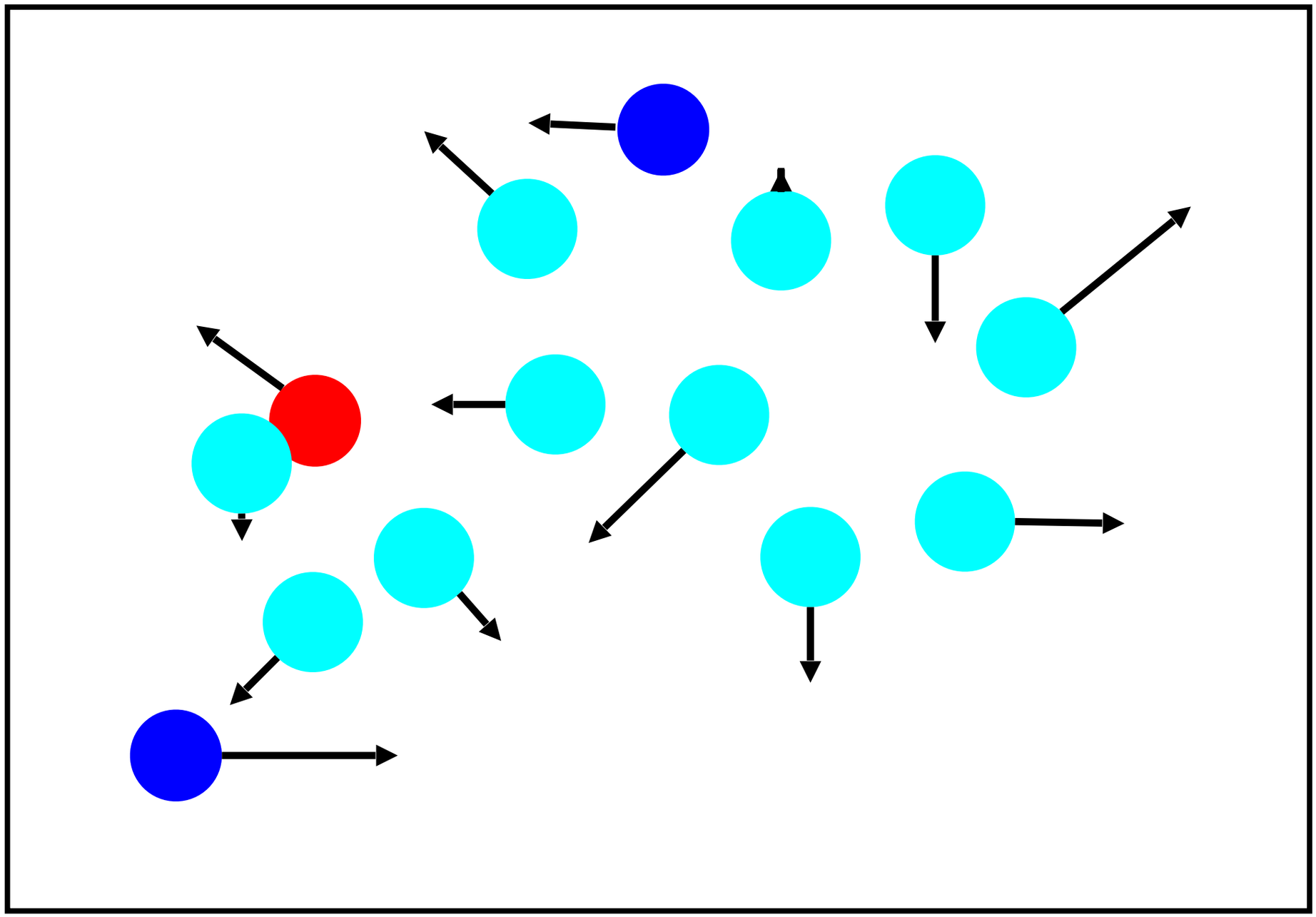


$$U(r) = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}} + \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$

$$\mathbf{F}(\mathbf{r}) = \left(-\frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}^{\,2}} - 12\frac{\varepsilon_{ij}}{|r_{ij}|}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - \left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]\right)\hat{\mathbf{r}}_{ij}$$

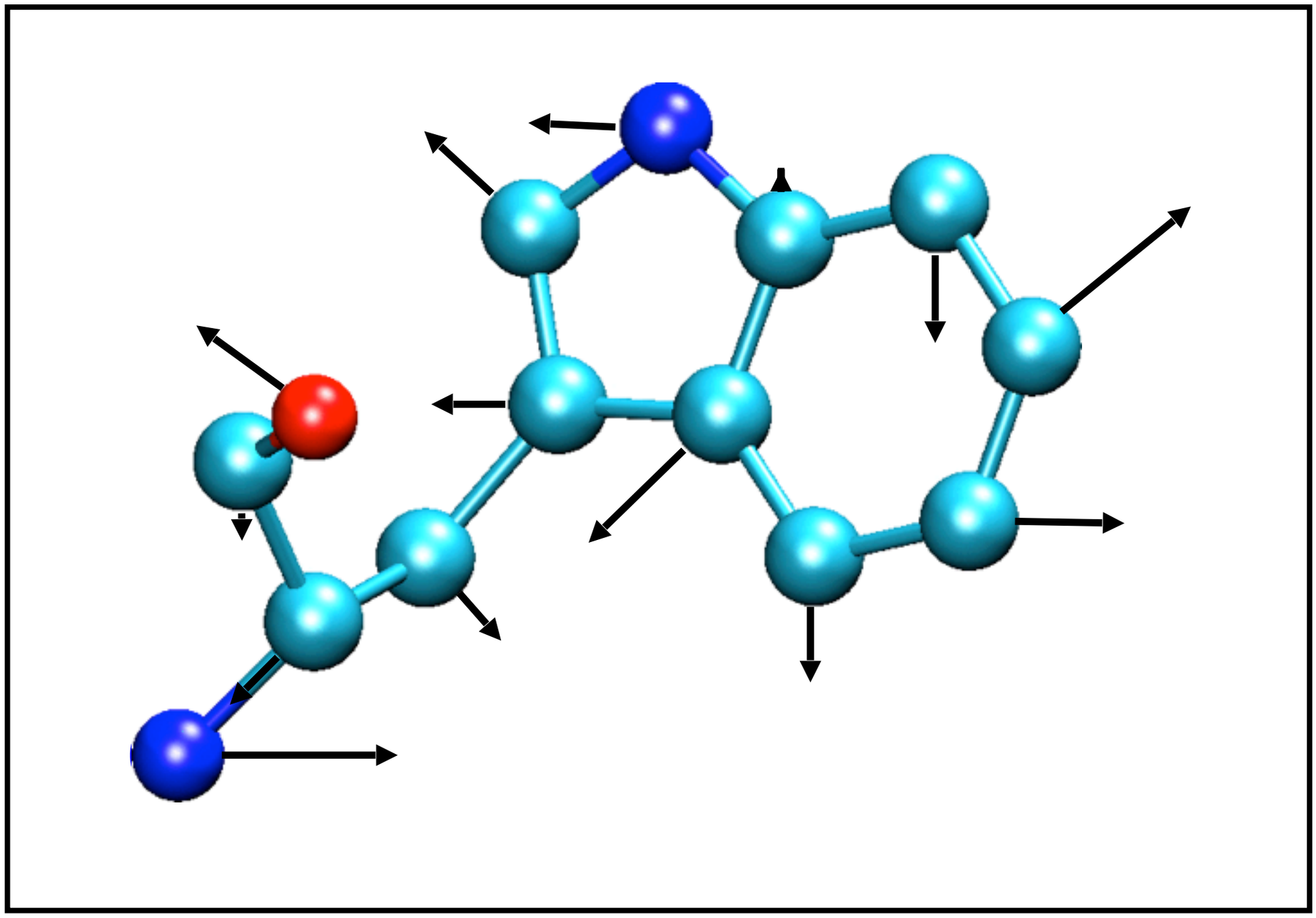
Classical Molecular Dynamics

tryptophan



Classical Molecular Dynamics

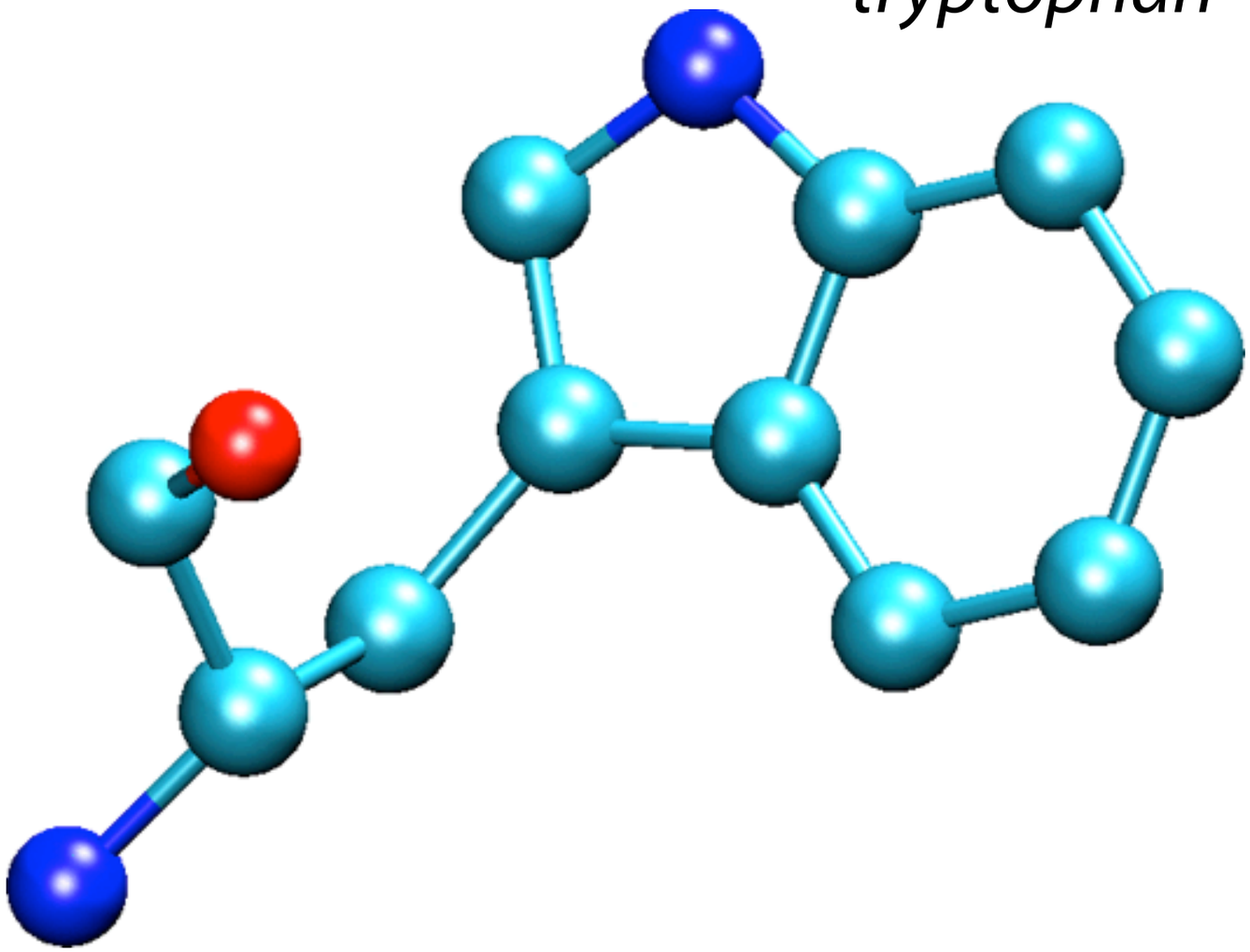
tryptophan



Bond definitions, atom types, atom names, parameters, ….

What is a Force Field?

tryptophan

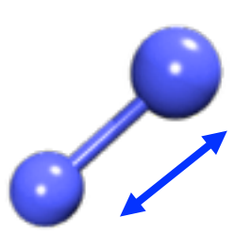


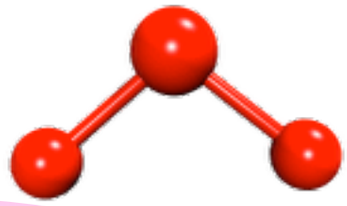
In molecular dynamics a molecule is described as a series of **charged points** (atoms) **linked by springs** (bonds).

To describe the time evolution of bond lengths, bond angles and torsions, also the non-bonding van der Waals and elecrostatic interactions between atoms, one uses a **force field**.
The **force field** is a collection of equations and associated constants designed to reproduce molecular geometry and selected properties of tested structures.

Potential Energy Function of Biopolymers

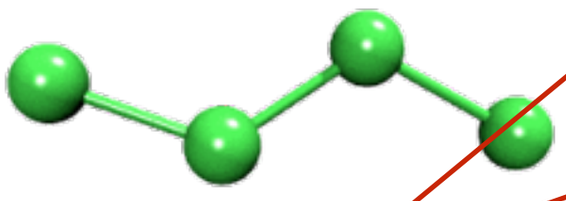
- Simple, fixed algebraic form for every type of interaction.
- Variable parameters depend on types of atoms involved.





heuristic

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

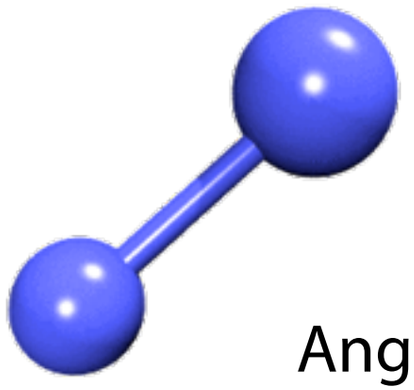


from physics

Parameters: "force fields" like Amber, Charmm (note version number!)

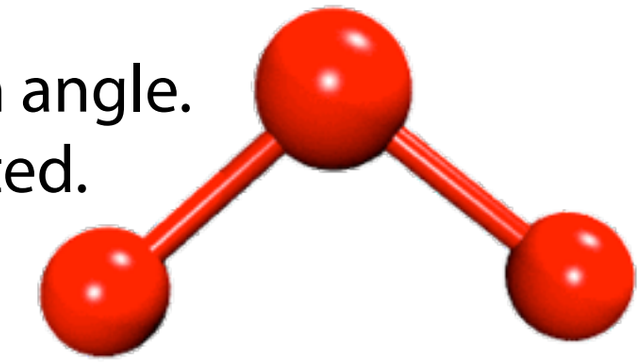
We currently use CHARMM36

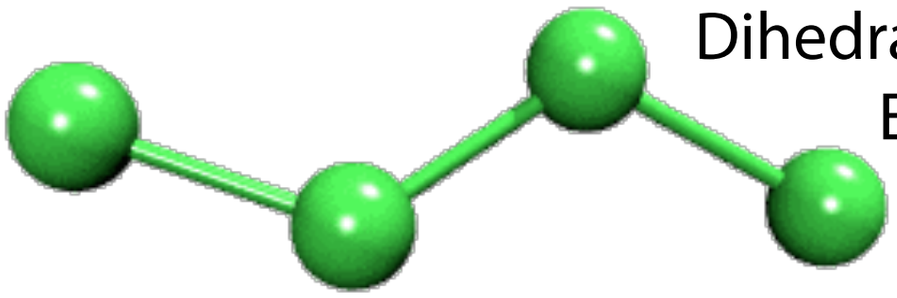
Potential Energy Function of Biopolymers



Bonds: Every pair of covalently bonded atoms is listed.

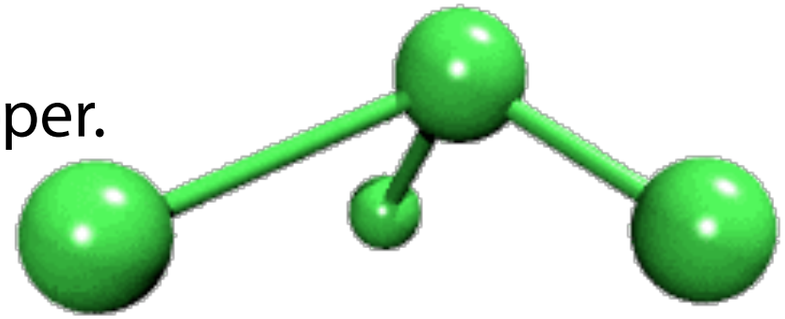
Angles: Two bonds that share a common atom form an angle.
Every such set of three atoms in the molecule is listed.





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

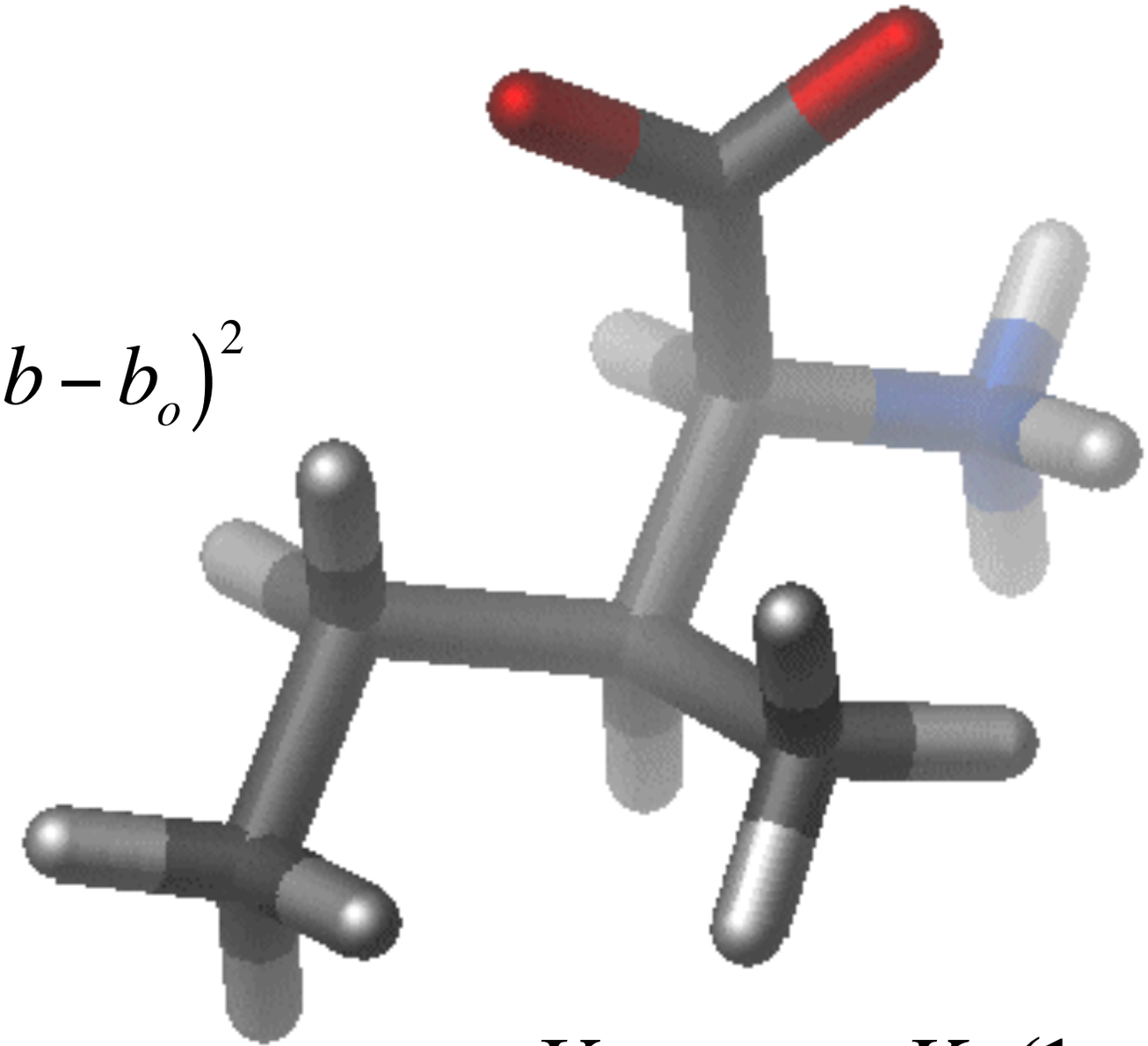
Impropers: Any *planar* group of four atoms forms an improper.
Specific sets of four atoms in the molecule are listed.



Interactions between bonded atoms

$$V_{angle} = K_\theta \left(\theta - \theta_o\right)^2$$

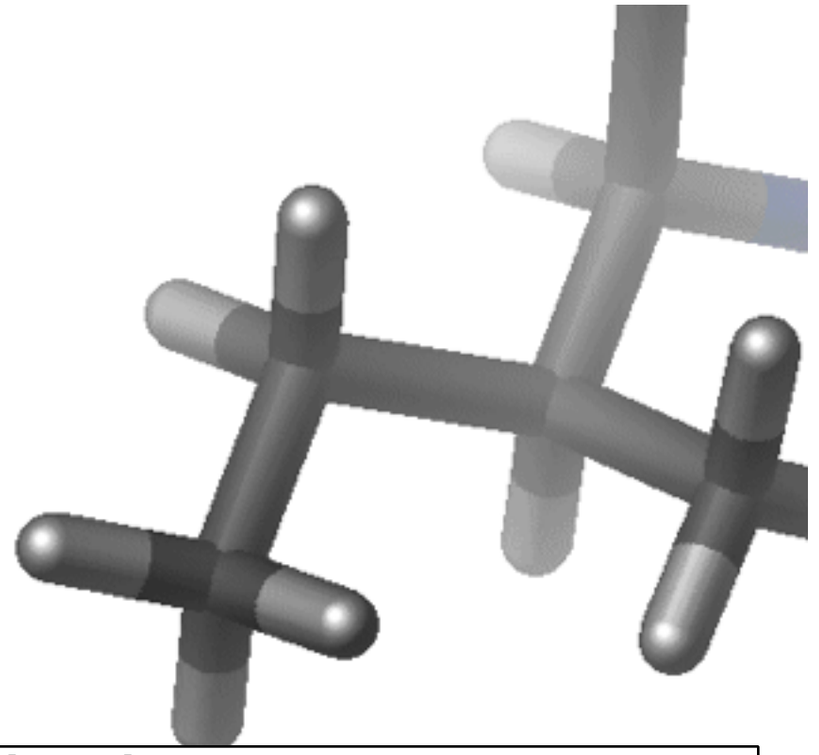
$$V_{bond} = K_b \left(b - b_o\right)^2$$



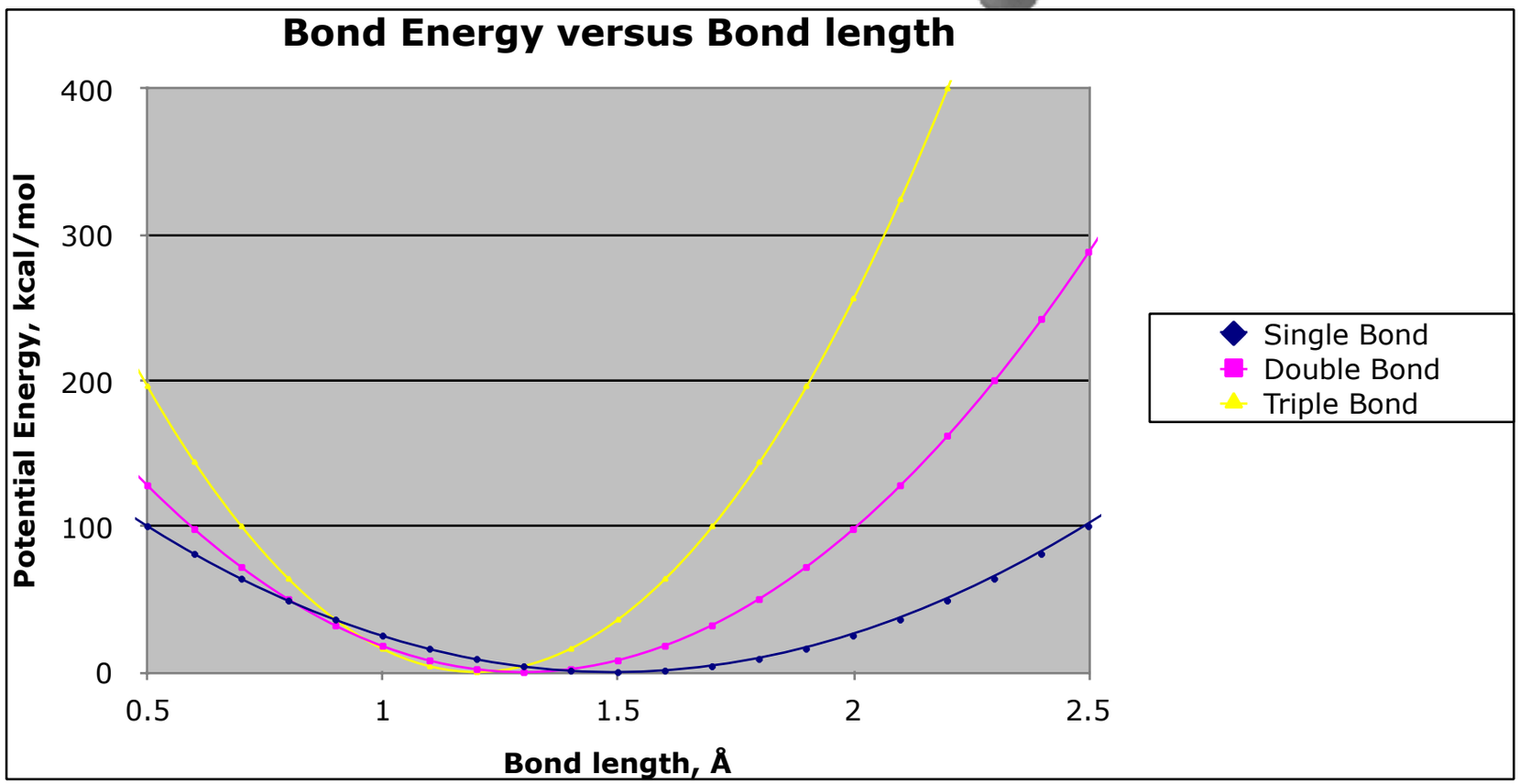
$$V_{dihedral} = K_\phi (1 + \cos(n\phi - \delta))$$

Bond potential

Chemical type	K_{bond}	b_o
C-C	100 kcal/mole/Å ^{2}	1.5 Å
C=C	200 kcal/mole/Å ^{2}	1.3 Å
C≡C	400 kcal/mole/Å ^{2}	1.2 Å



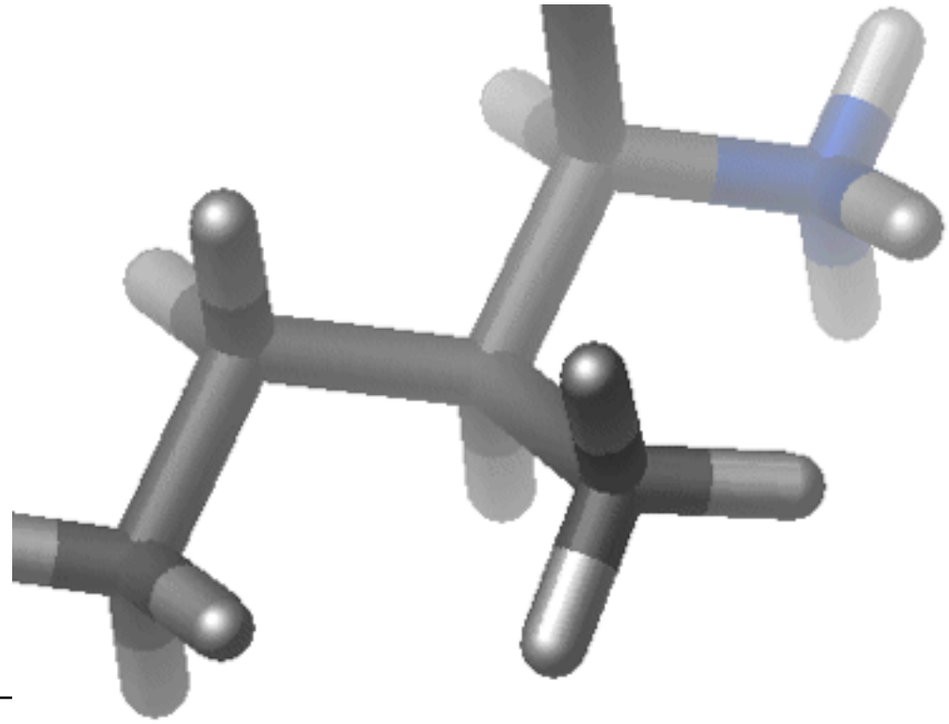
$$V_{bond} = K_b(b - b_0)^2$$

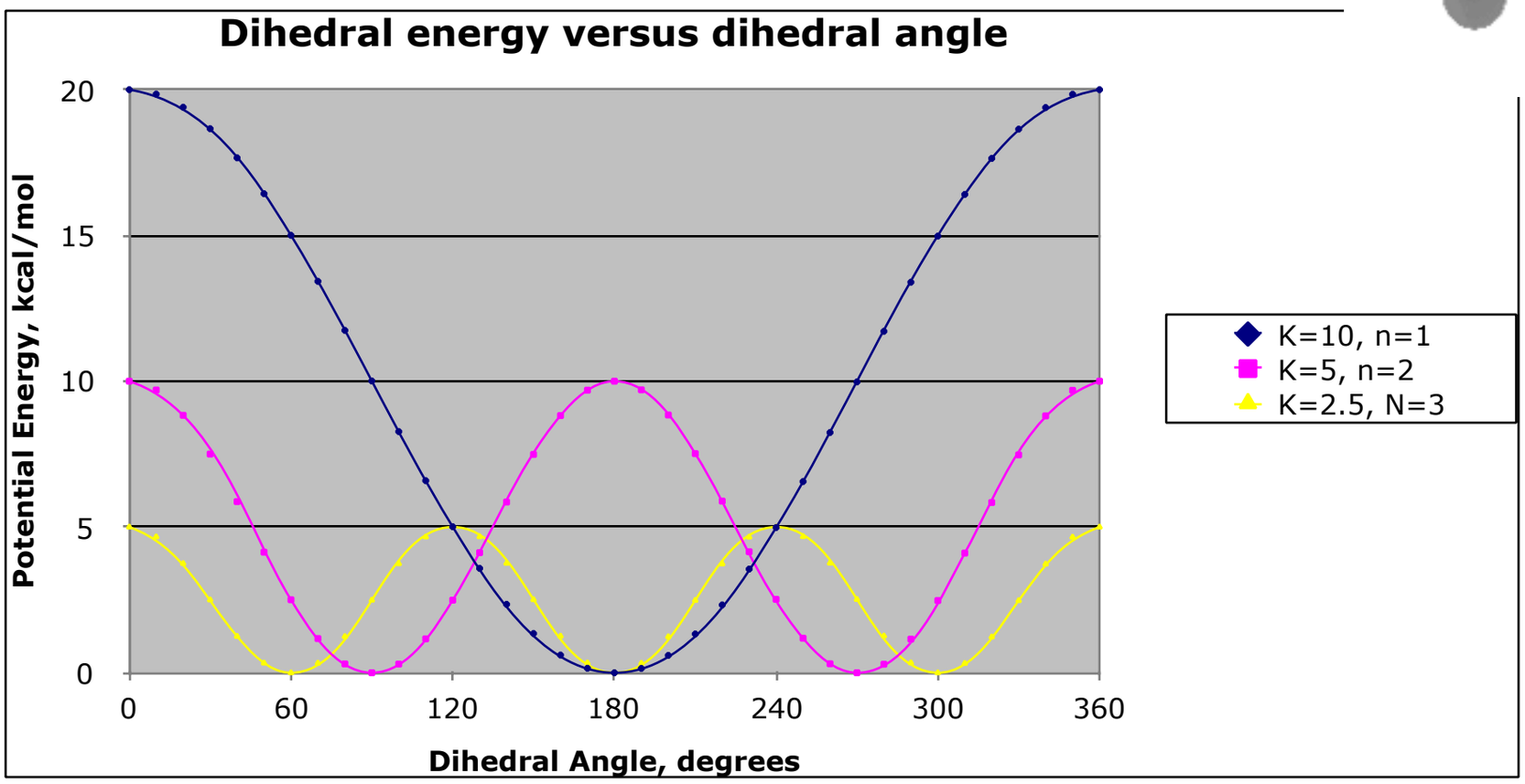


Bond-angle (3-body) and *improper (4-body about a center)* terms have similar quadratic forms, but with softer spring constants. The force constants can be obtained from vibrational analysis of the molecule (experimentally or theoretically).

Dihedral potential

$$V_{dihedral} = K_{\phi}(1 + \cos(n\phi - \delta))$$

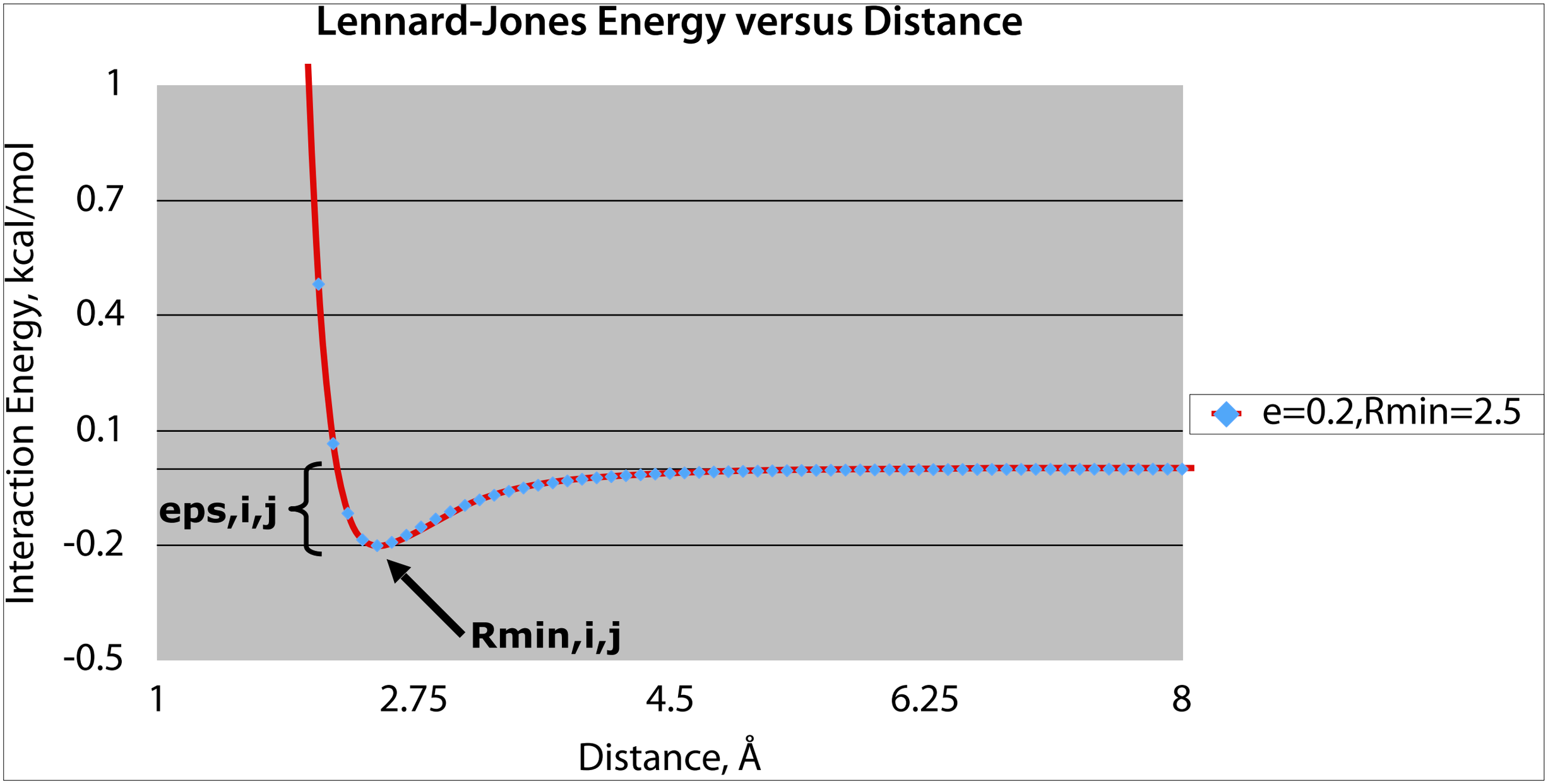




$\delta = 0$ for all three

dihedral-angle (4-body) terms come from symmetry in the electronic structure. Cross-term map (CMAP) terms in CHARMM force field are a refinement to this part of the potential

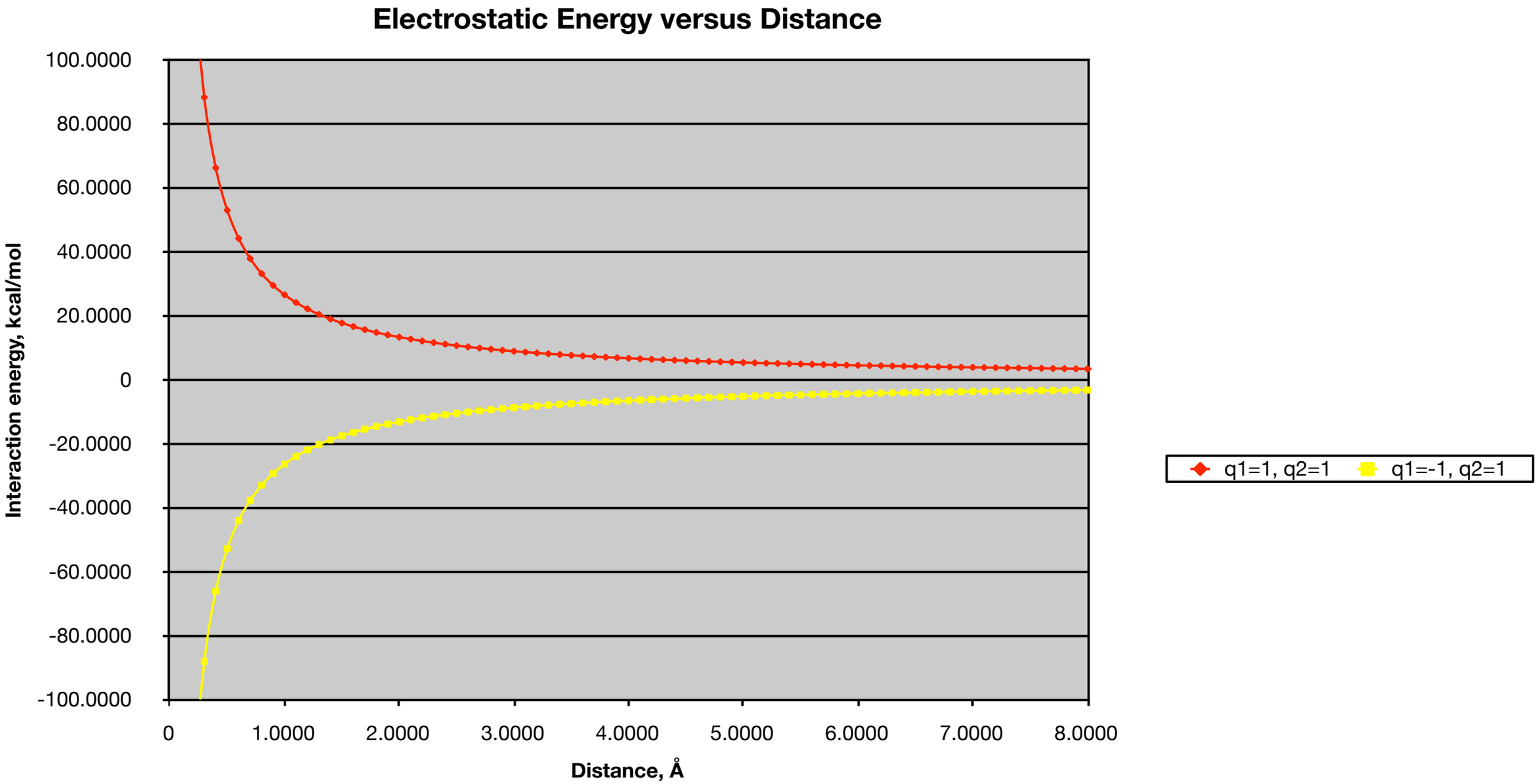
van der Waals potential



$$\varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$

Short range; in CHARMM, cutoff at 12 Å other FFs will cutoff at lower values (even 8 Å!) but beware the easy "**shortcut**"

Coulomb potential



Note that the effect is long range.

From MacKerell

Equilibrium Properties of Proteins

Energies: kinetic and potential

?

temperature dependence

$$\langle \sum_j \frac{1}{2} m_j v_j^2 \rangle = \frac{3}{2} N k_B T$$

Kinetic energy (quadratic)

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

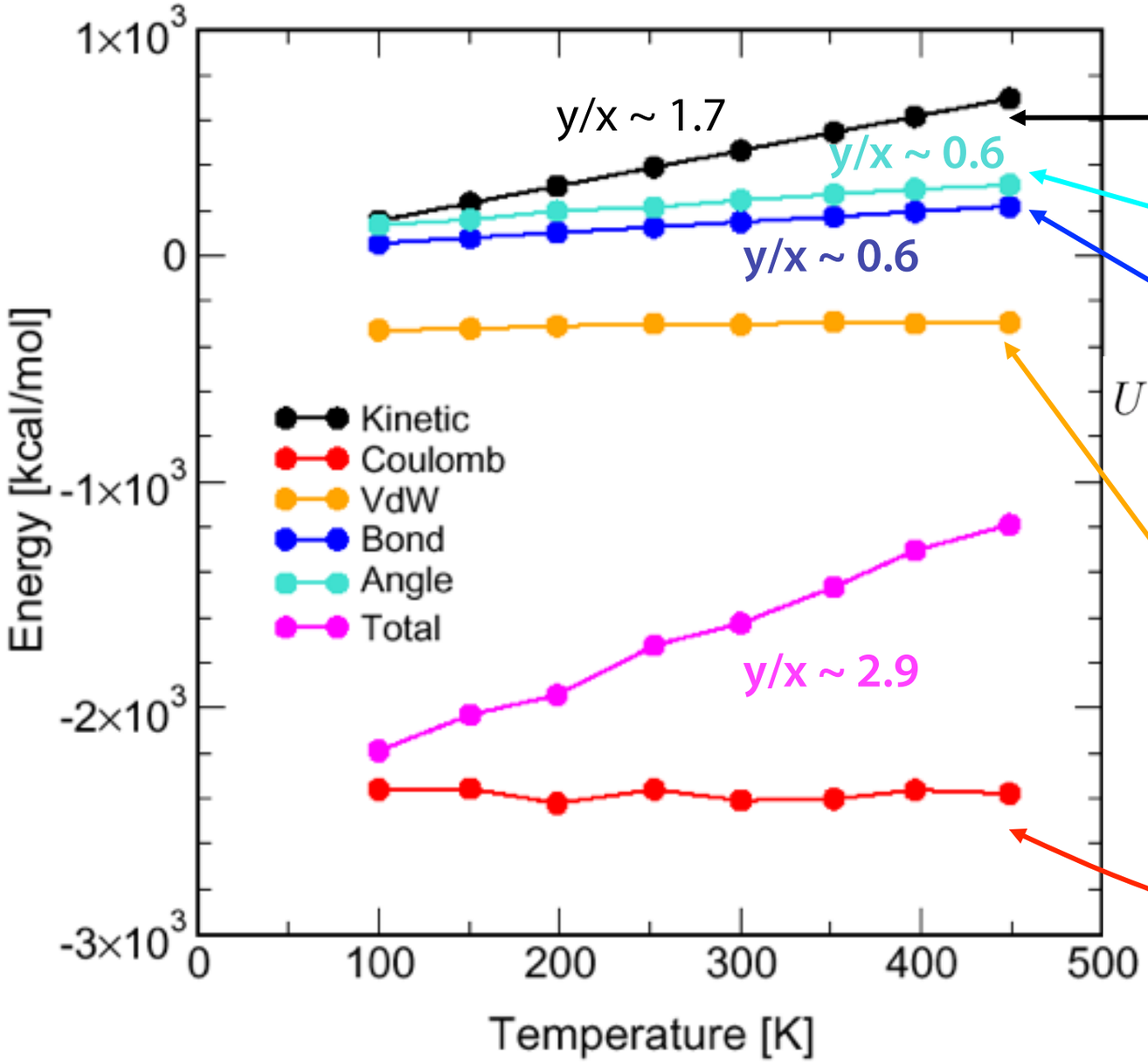
$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i \phi_i + \delta_i)]}_{U_{dihedral}} +$$

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

Potential energy (not all quadratic)

Equilibrium Properties of Proteins

Energies: kinetic and potential



$$\langle \sum_j \frac{1}{2} m_j v_j^2 \rangle = \frac{3}{2} N k_B T$$

Kinetic energy (quadratic)

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

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

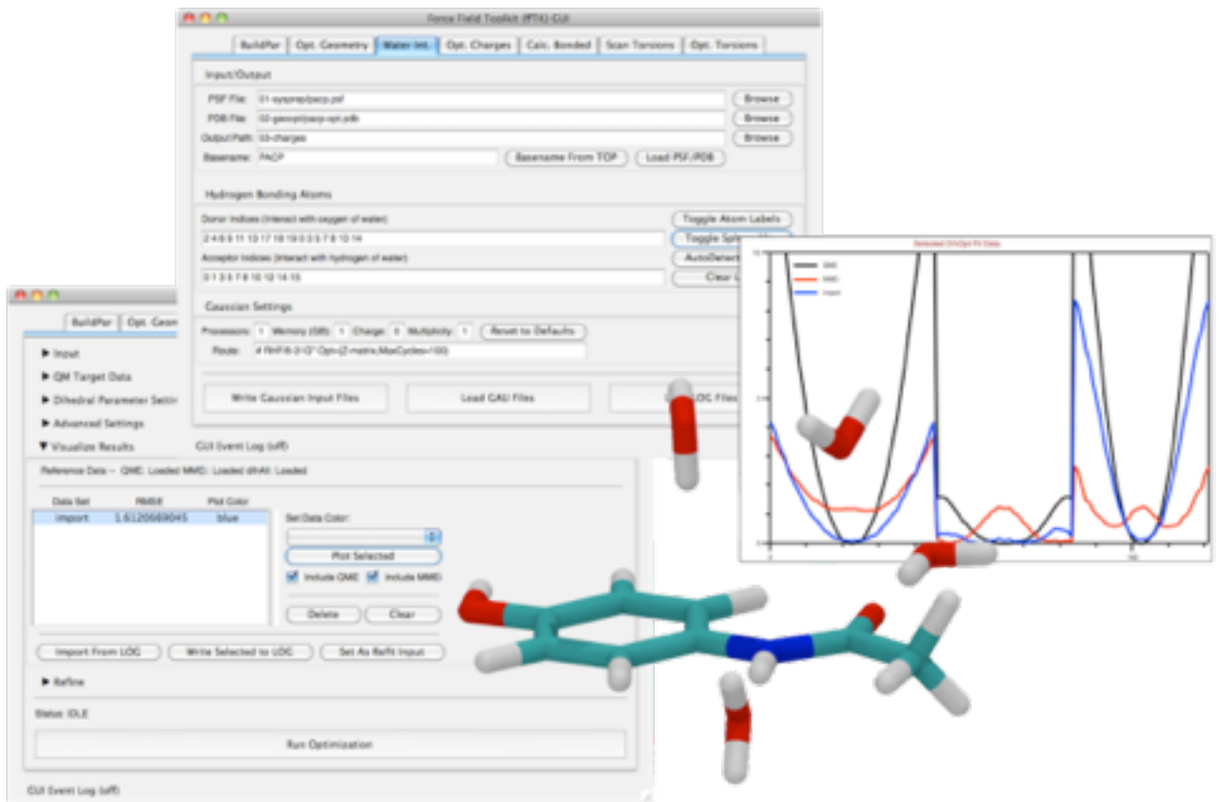
$$\underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$

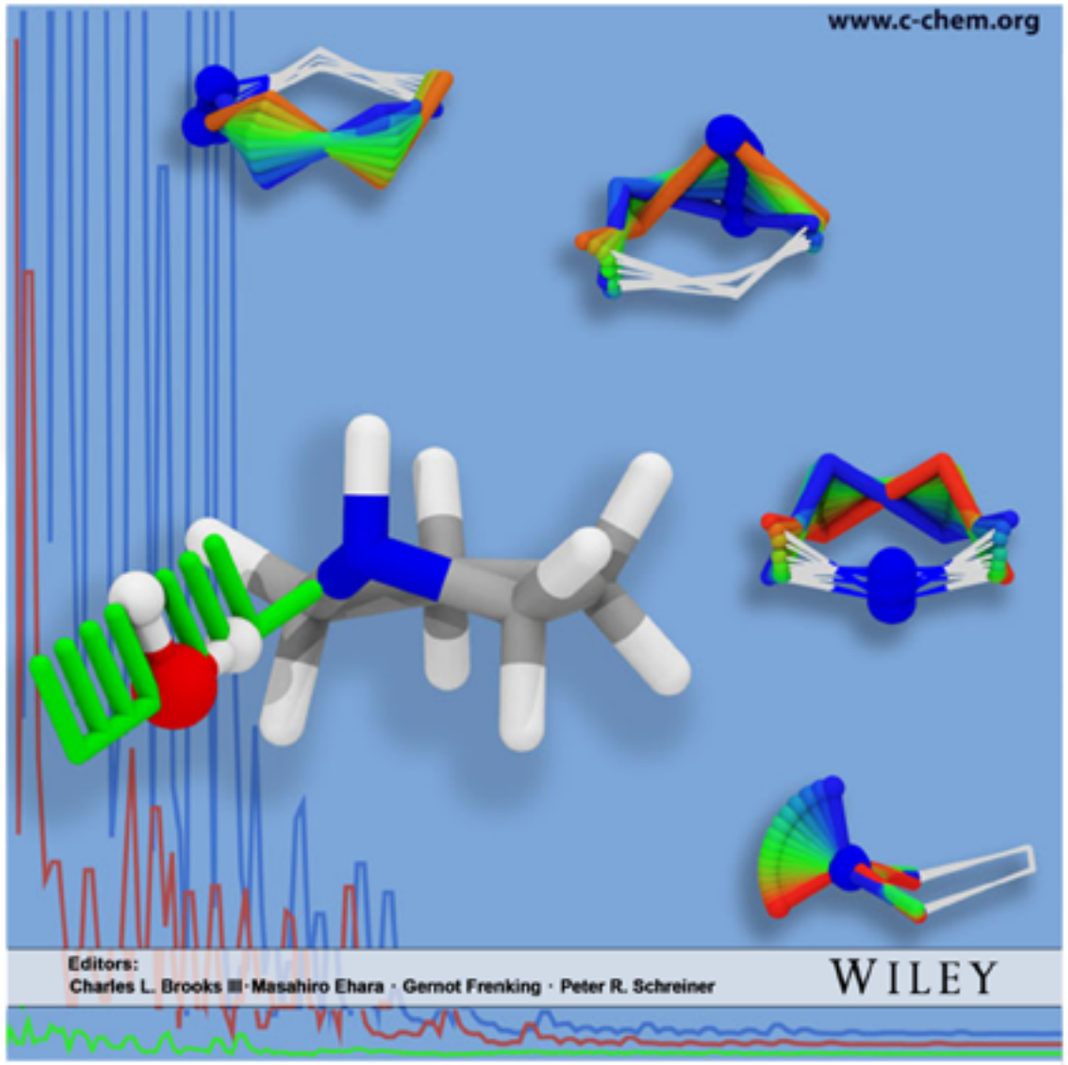
Potential energy (not all quadratic)

Force field toolkit (FFTK)

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

FFTK aids in the development of parameters in the **MD potential function** for novel molecules, ligands





Volume 34 | Issues 13–14 | 2013
Included in this print edition:
Issue 13 (May 15, 2013)
Issue 14 (May 30, 2013)
Journal of COMPUTATIONAL CHEMISTRY
Organic · Inorganic · Physical
Biological · Materials

Mayne, Saam, Schulten, Tajkhorshid, and Gumbart. **Rapid parameterization of small molecules using the force field toolkit.** (2013) *J. Comp. Chem.* 34:2757-2770.

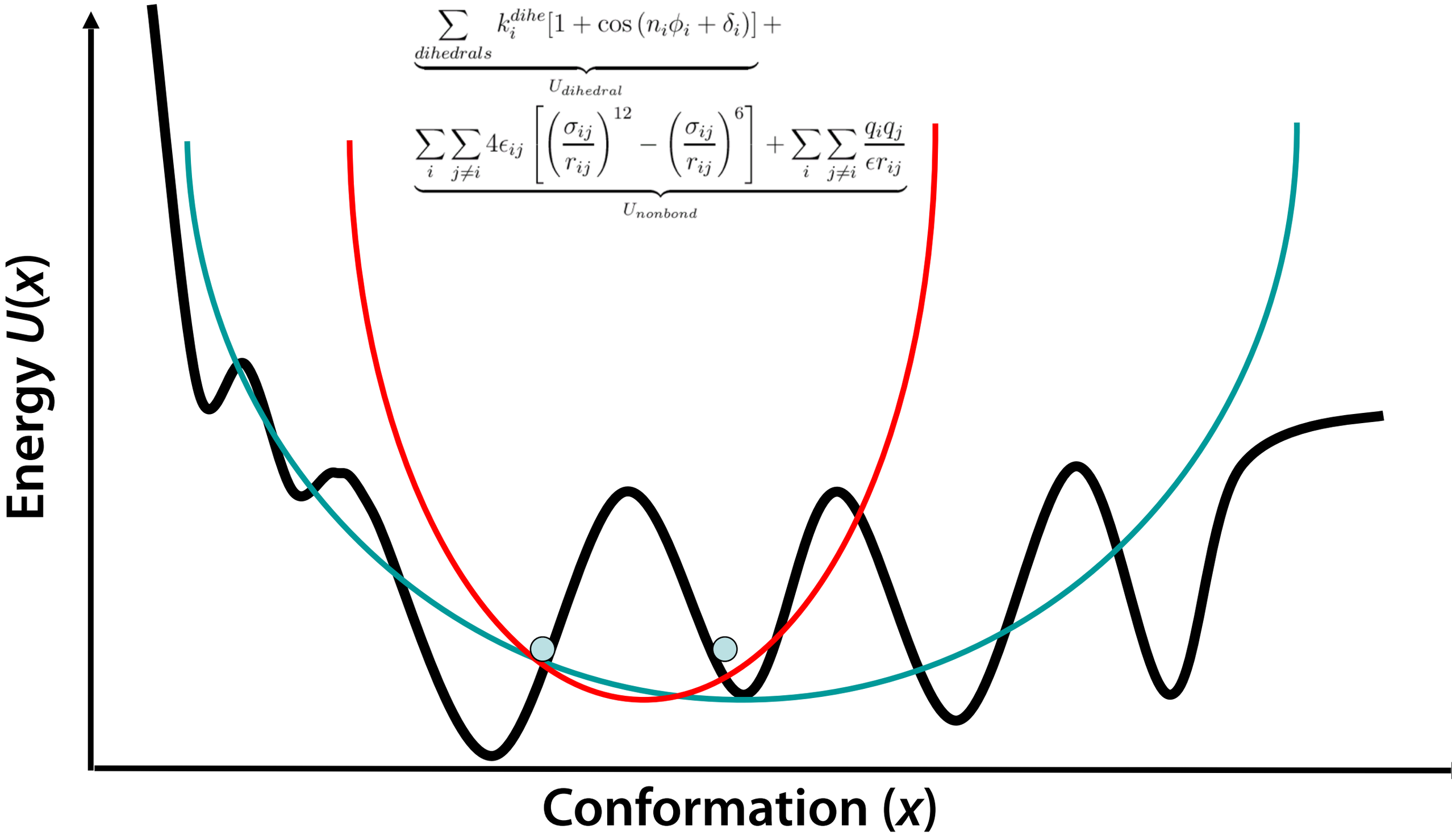
http://www.ks.uiuc.edu/Research/vmd/plugins/fftk/

Potential Energy (hyper)Surface

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

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

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



Classical Molecular 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.

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

+

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

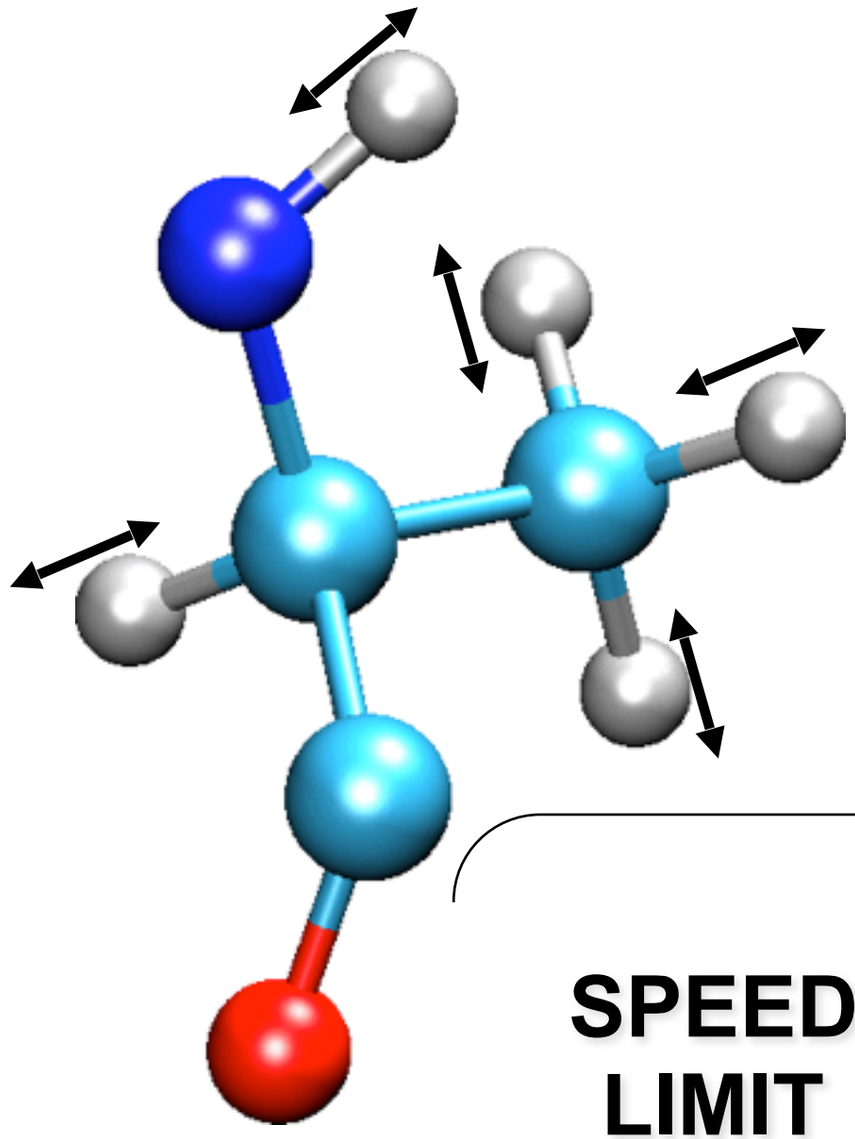
"Verlet algorithm"

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

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

The most serious bottleneck

	steps	
Protein folding (typical)	s	10^{15}
Rotation of buried sidechains	ms	10^{12}
Allosteric transitions		
Protein folding (fastest)	µs	10^{9}
Hinge bending	ns	10^{6}
Rotation of surface sidechains	ps	10^{3}
bond stretching		
Molecular dynamics timestep →	fs	10^{0}



SPEED LIMIT

δt = 1-2 fs*

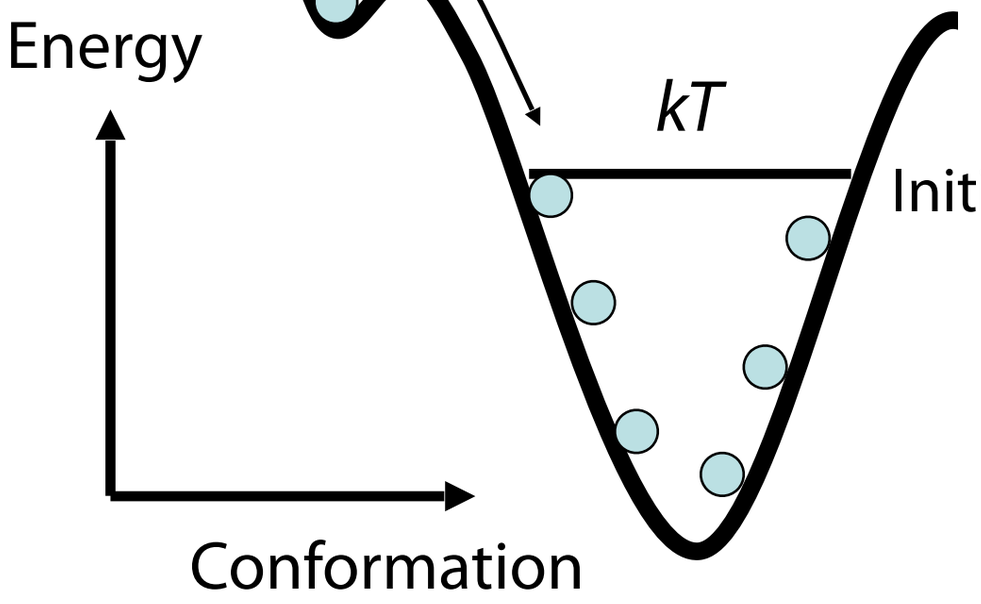
*experimental 4-fs time steps with "hydrogen mass repartitioning" exists, but not in NAMD yet

Molecular Dynamics to Sample Energy Landscape

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.

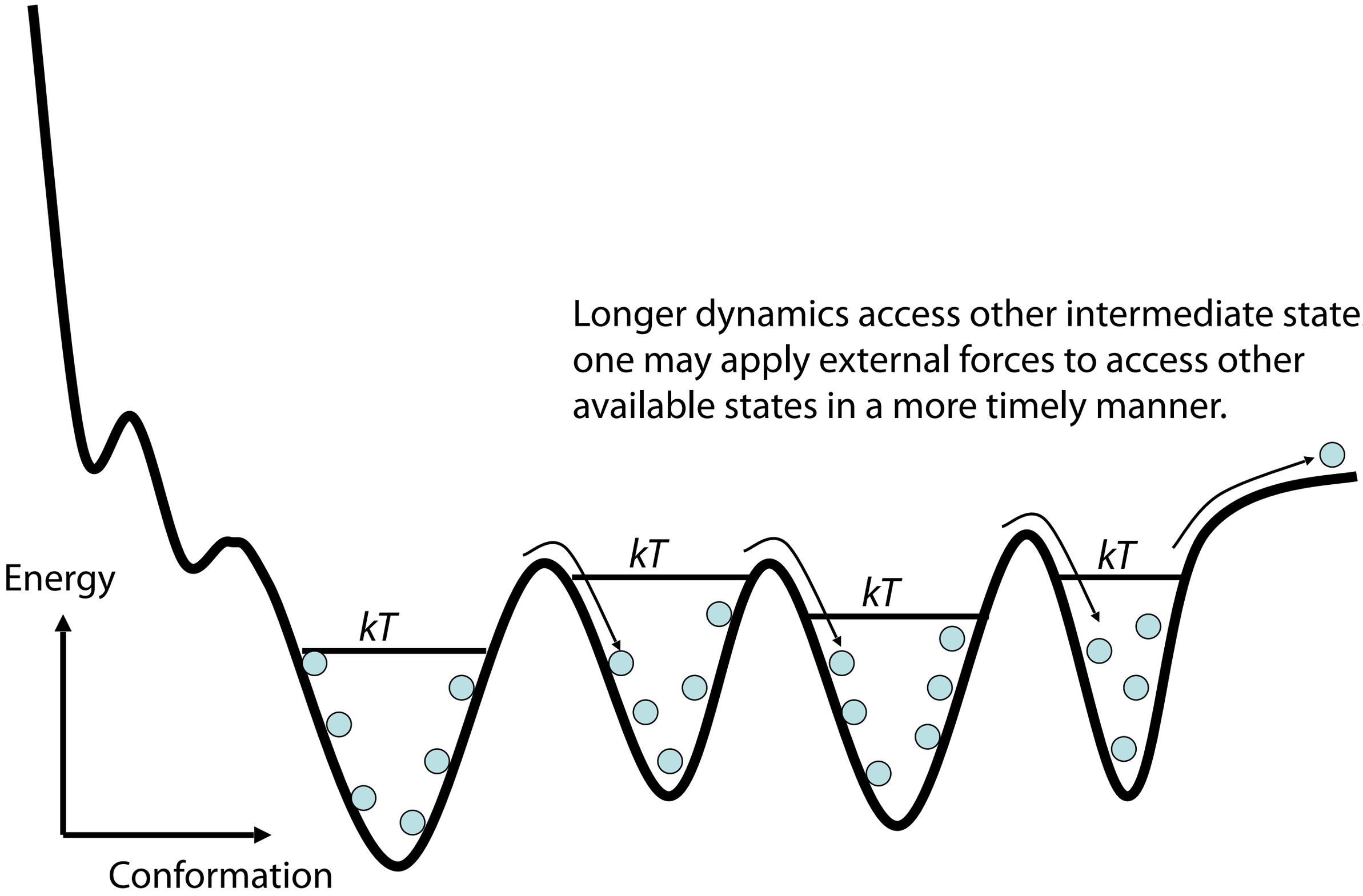
Heating and cooling or equilibration at fixed temperature permits biopolymer to escape local minima with low energy barriers.



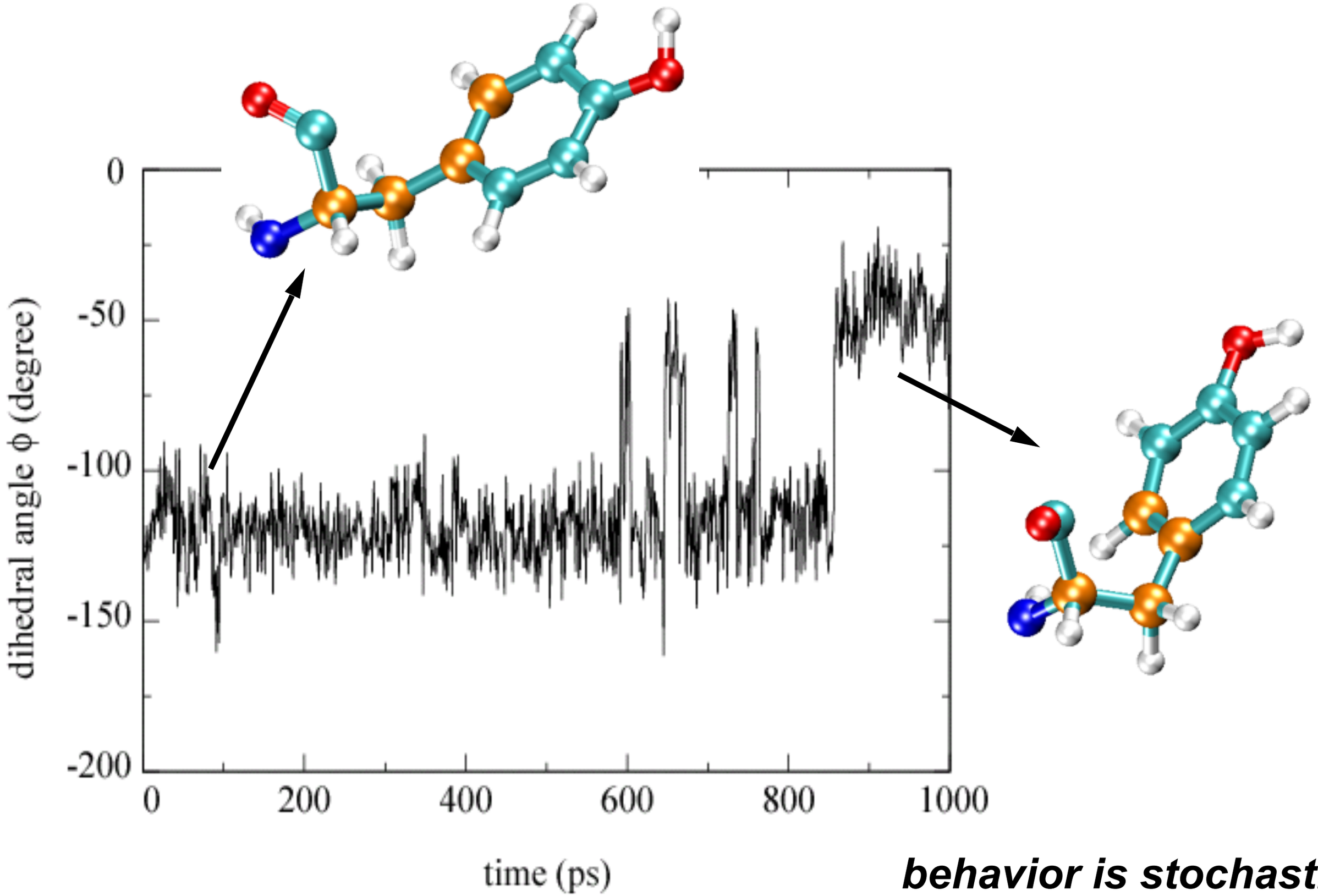
Initial dynamics samples thermally accessible states.

Molecular Dynamics to Sample Energy Landscape

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



Patience is required to observe molecular events



behavior is stochastic

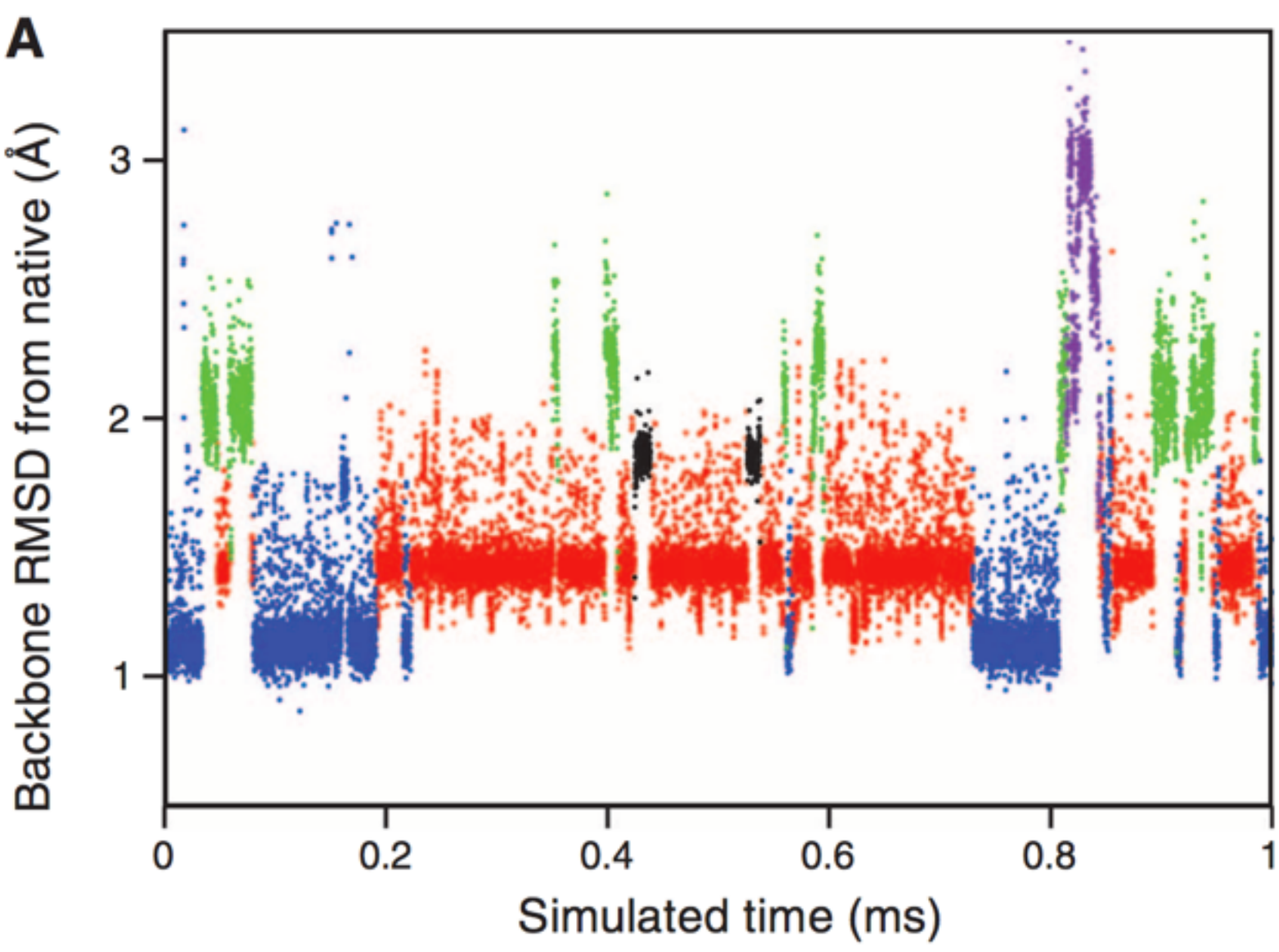
Patience is required to observe molecular events

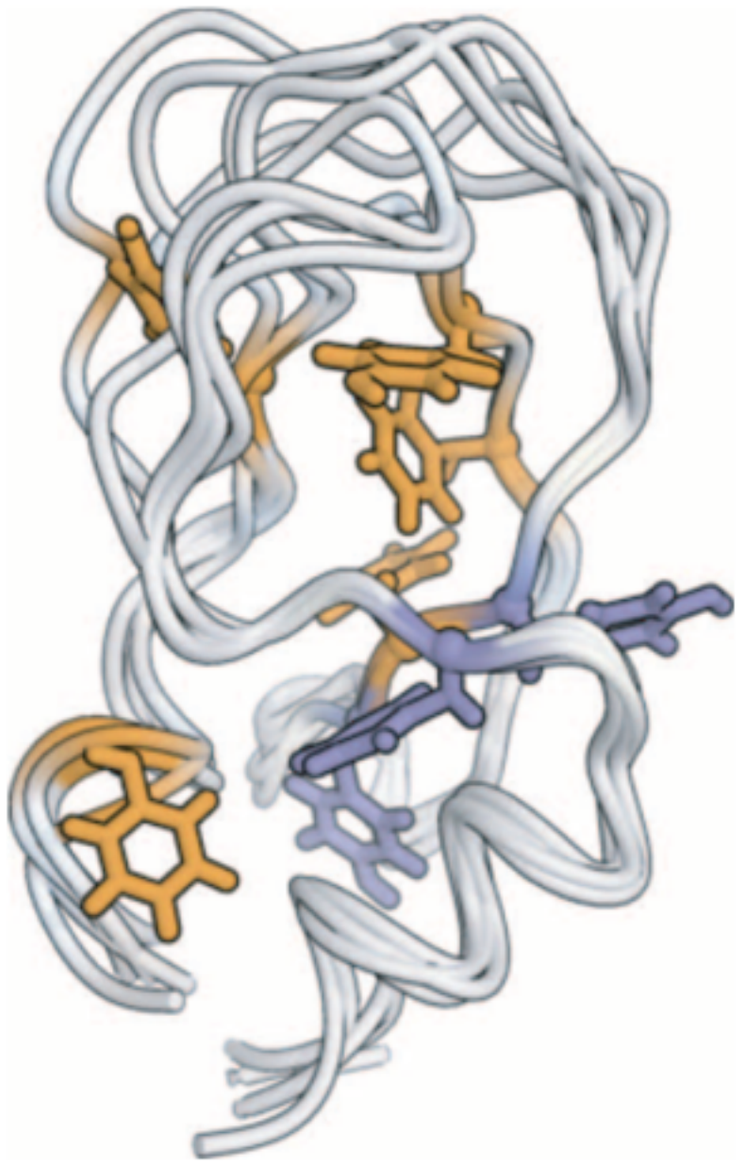
RESEARCH ARTICLE

Atomic-Level Characterization of the Structural Dynamics of Proteins

David E. Shaw[1,2,*], Paul Maragakis[1,†], Kresten Lindorff-Larsen[1,†], Stefano Piana[1,†], Ron O. Dror[1], Michael P. Eastwood[1], Joseph A. Bank[1], John M. Jumper[1], John K. Salmon[1], Yibing Shan[1], Willy Wriggers[1]

And there will always be a scale of dynamics you don't observe!





DE Shaw et al. (2010) *Science* 330:341-346.

Molecular Dynamics Ensembles

thinking in terms of statistical mechanics

Constant energy, constant number of particles (NE; *only if no periodic boundary conditions*)

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 - for most biomolecular systems, we choose **NPT**

Temperature and Pressure control methods

-to simulate **NVT** (canonical) ensemble, need to duplicate the effect of a large thermal bath around the system

Dynamics governed by the Langevin equation;
gives the **correct ensemble**

damping term

random term

$$M\dot{v} = F(x) - \gamma M v - \sqrt{2\gamma k_b T M}\,R(t)$$

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

-Damping (1/ps) should be enough to maintain temperature without significantly perturbing dynamics (**5 is too much, 1 is probably okay**)

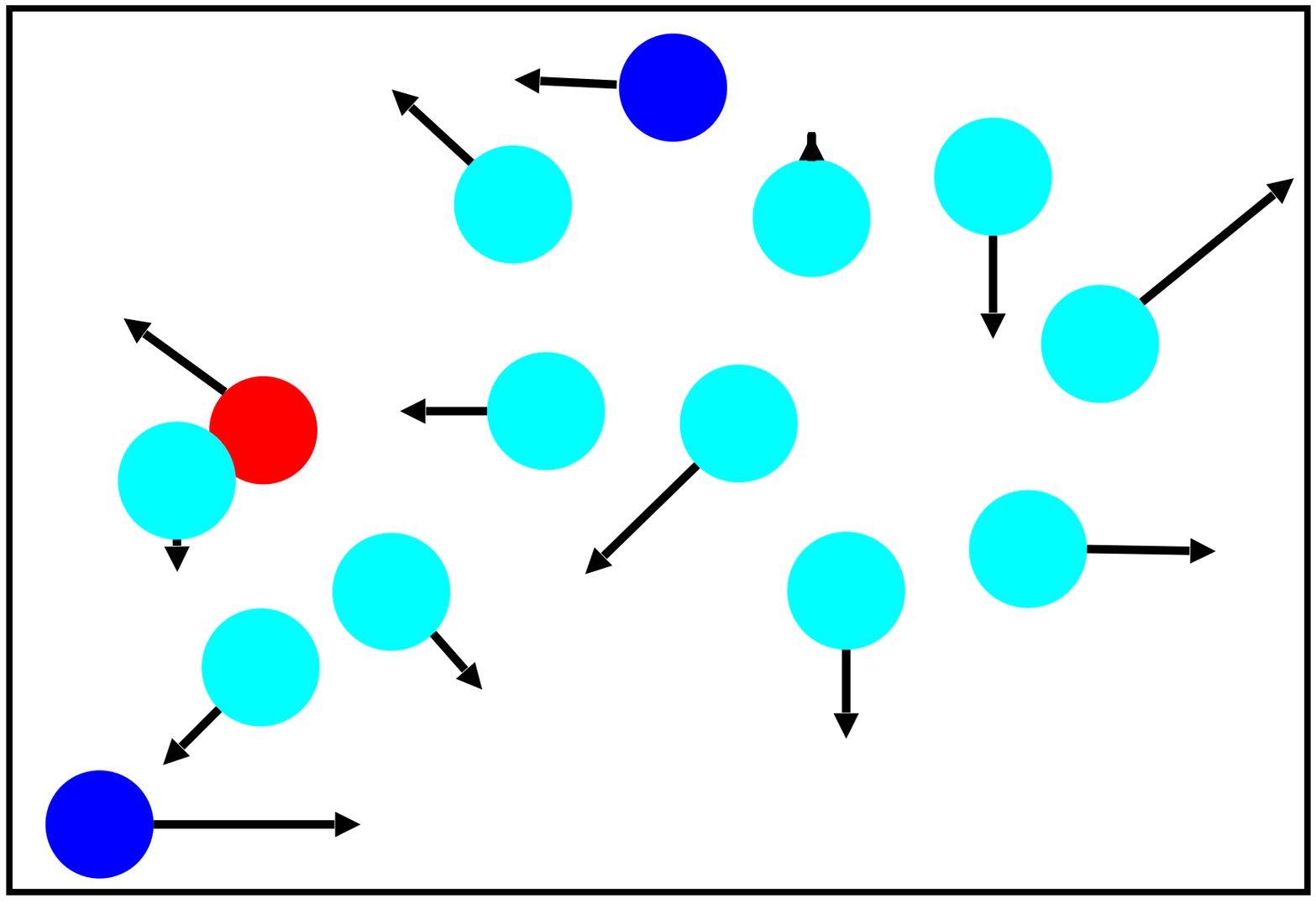
**For non-equilibrium simulations the Lowe-Andersen thermostat, which conserves momentum and does not suppress flows, may be preferred*

NPT (isobaric-isothermal) ensemble adds additional variables to control temperature, pressure which are ultimately integrated out to generate the correct distribution

$$P = \rho k_B T + \frac{1}{dV}\left\langle \sum_{i<j} \mathbf{f}(\mathbf{r}_{ij}) \cdot \mathbf{r}_{ij} \right\rangle$$

In the simulation, pressure is calculated from the virial expansion

Boundary Conditions?

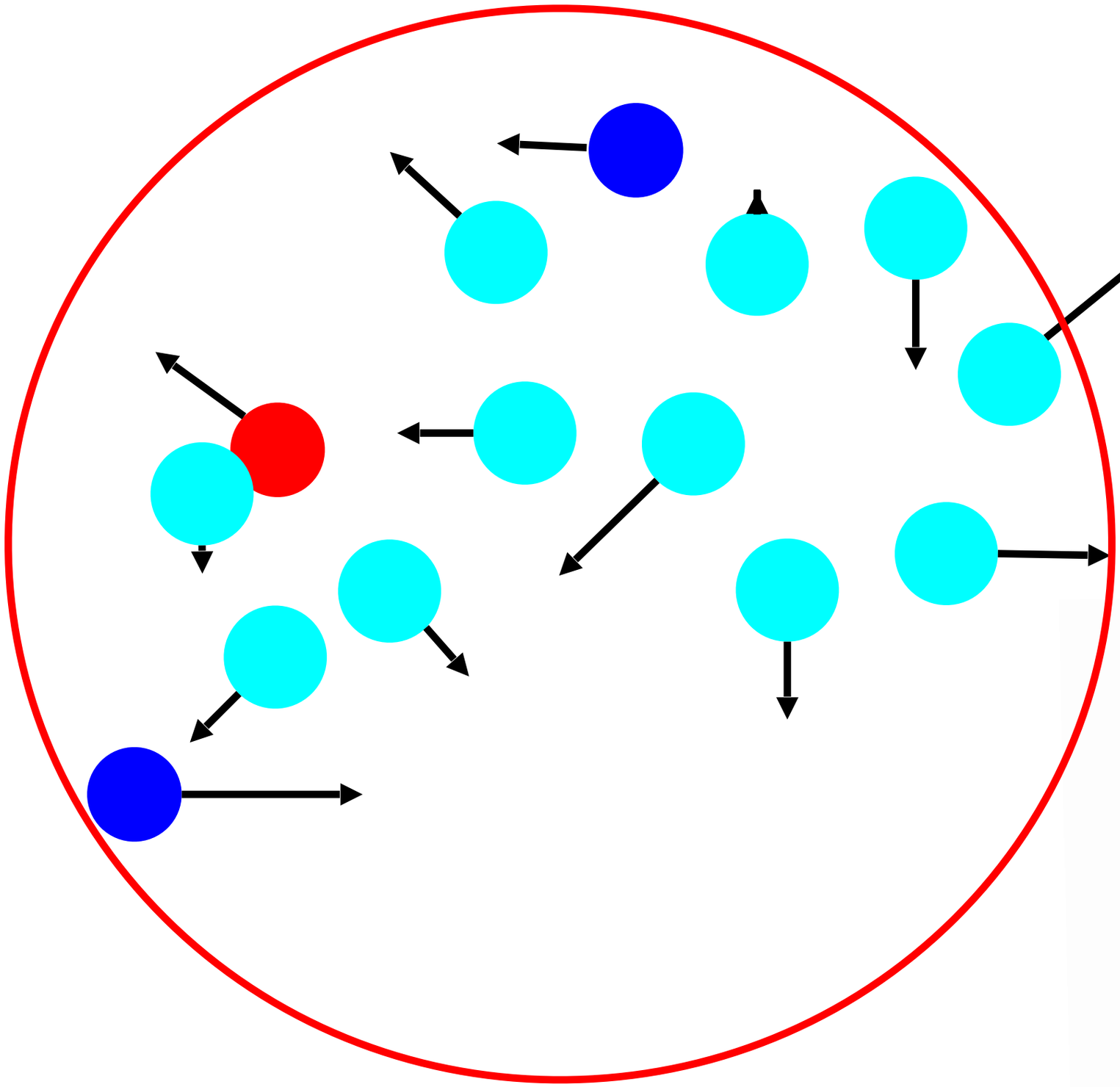


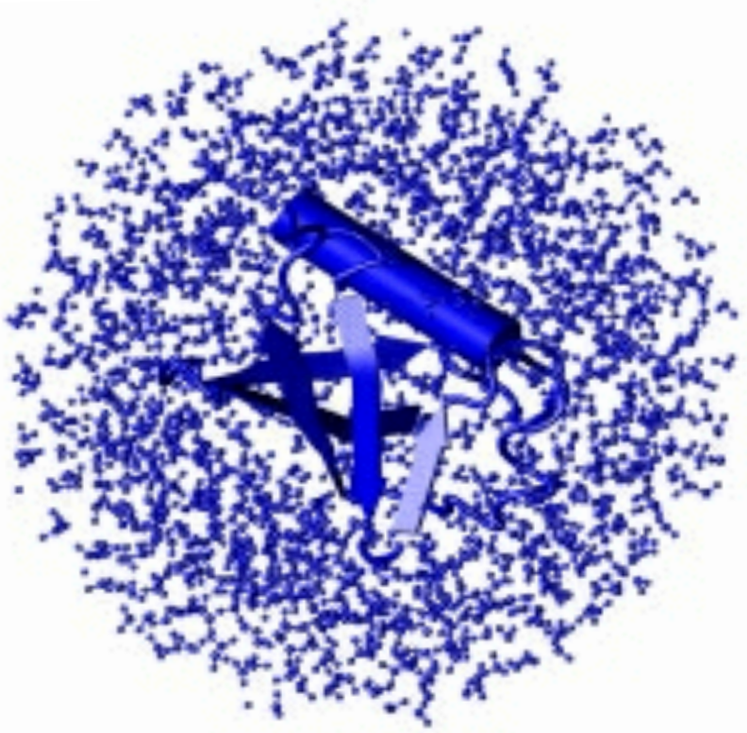
What happens if you put water under vacuum!?

Problems: Density, pressure, boundary effects, …

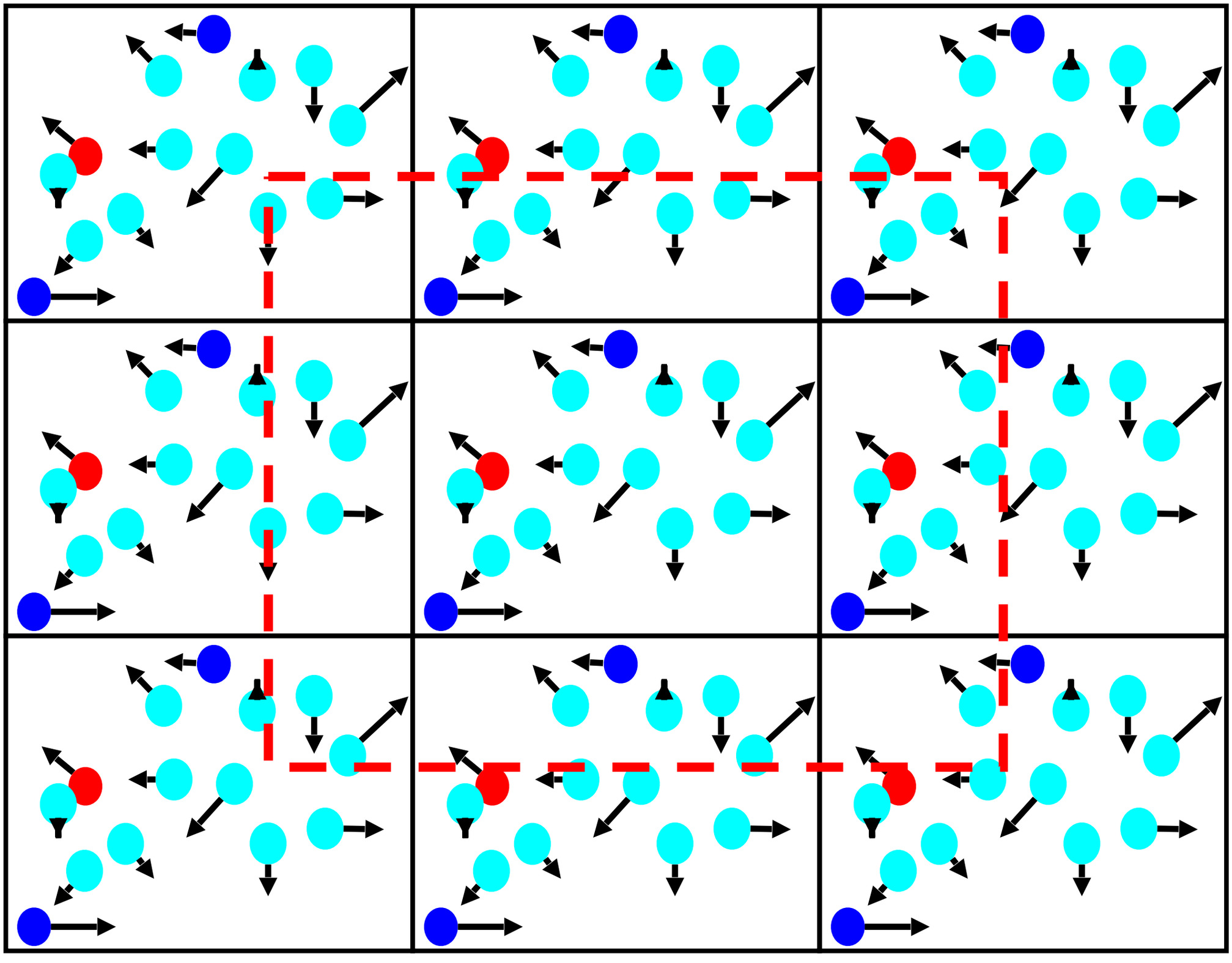
One solution: reflective boundaries, not quite good.

Spherical boundary conditions

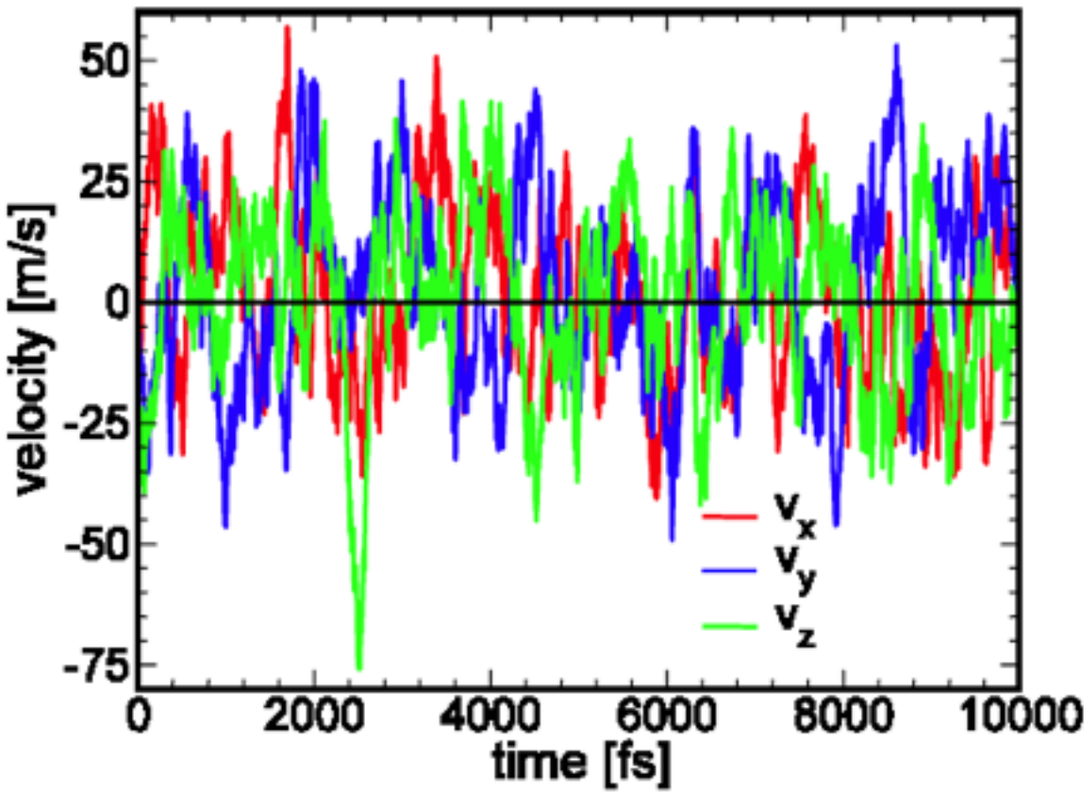


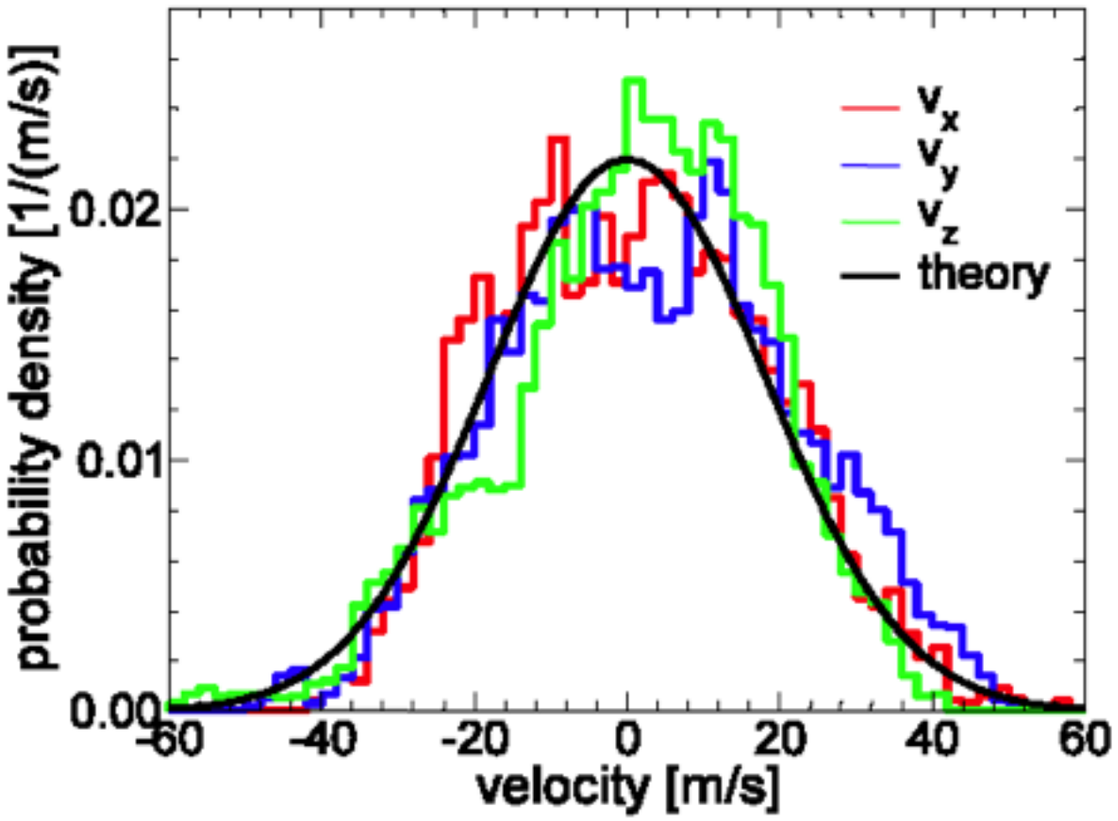


Periodic Boundary Conditions



Maxwell Distribution of Atomic Velocities

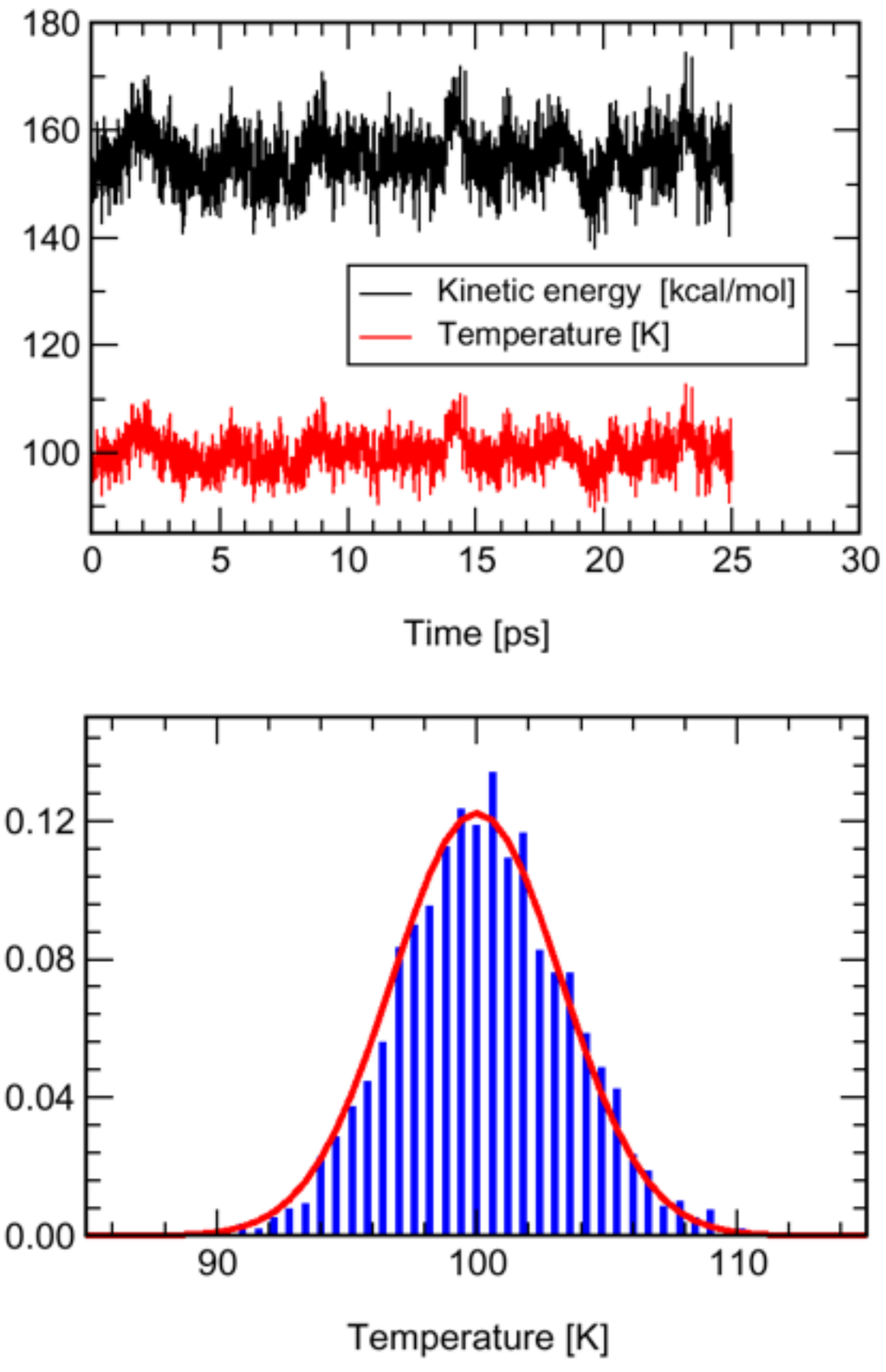




$$p(v_\sigma) = \sqrt{\frac{m}{2\pi k_B T}} \exp\left[-\frac{m v_\sigma^2}{2 k_B T}\right]$$

$$\sigma = x,\, y,\, z$$

Analysis of *K*, *T* (free dynamics)

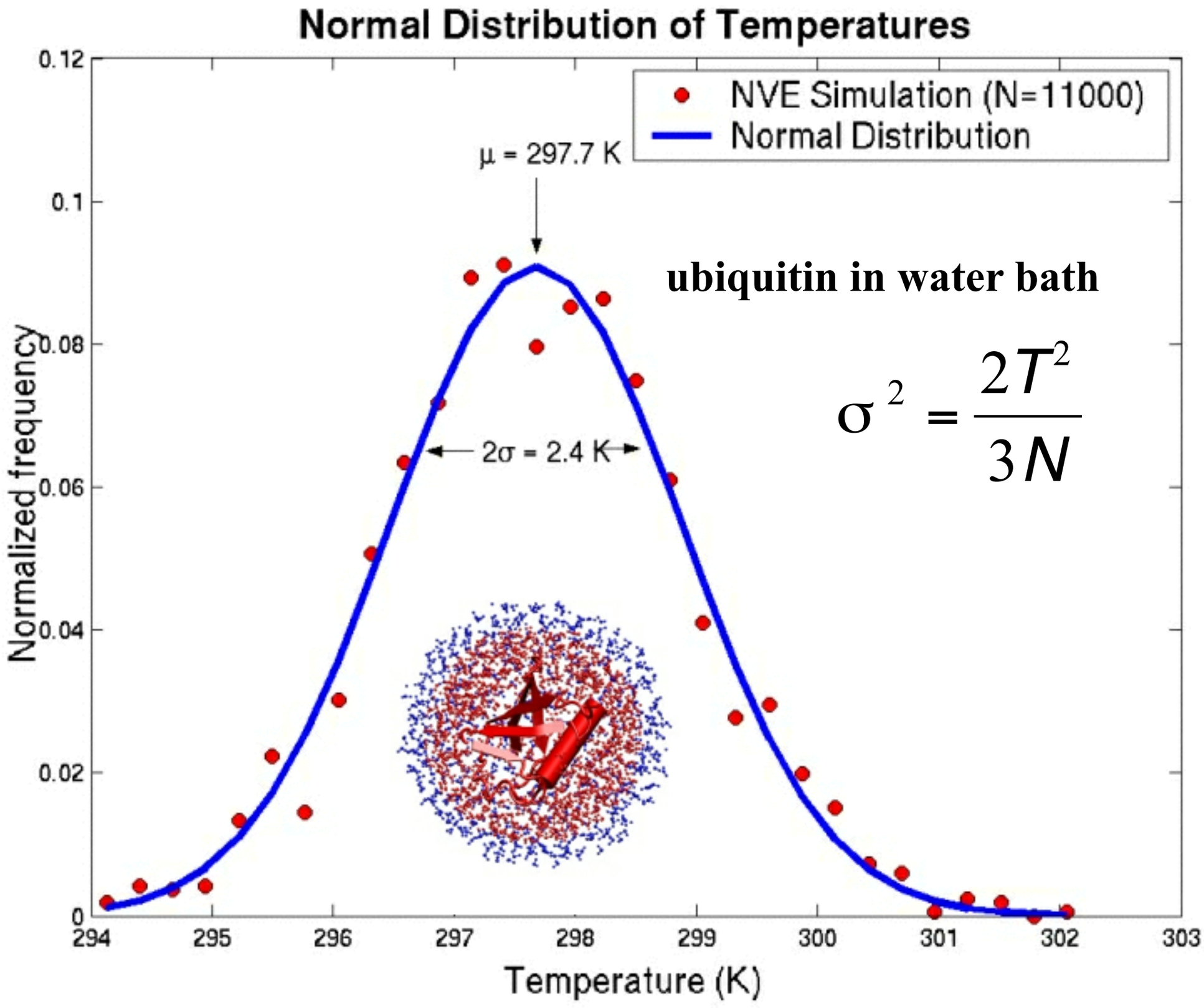


Definition of temperature

$$\langle \sum_j \frac{1}{2} m_j v_j^2 \rangle \;=\; \frac{3}{2} N\, k_B\, T$$

$$T \;=\; \frac{2}{3N\, k_B} \langle \sum_j \frac{1}{2} m_j v_j^2 \rangle$$

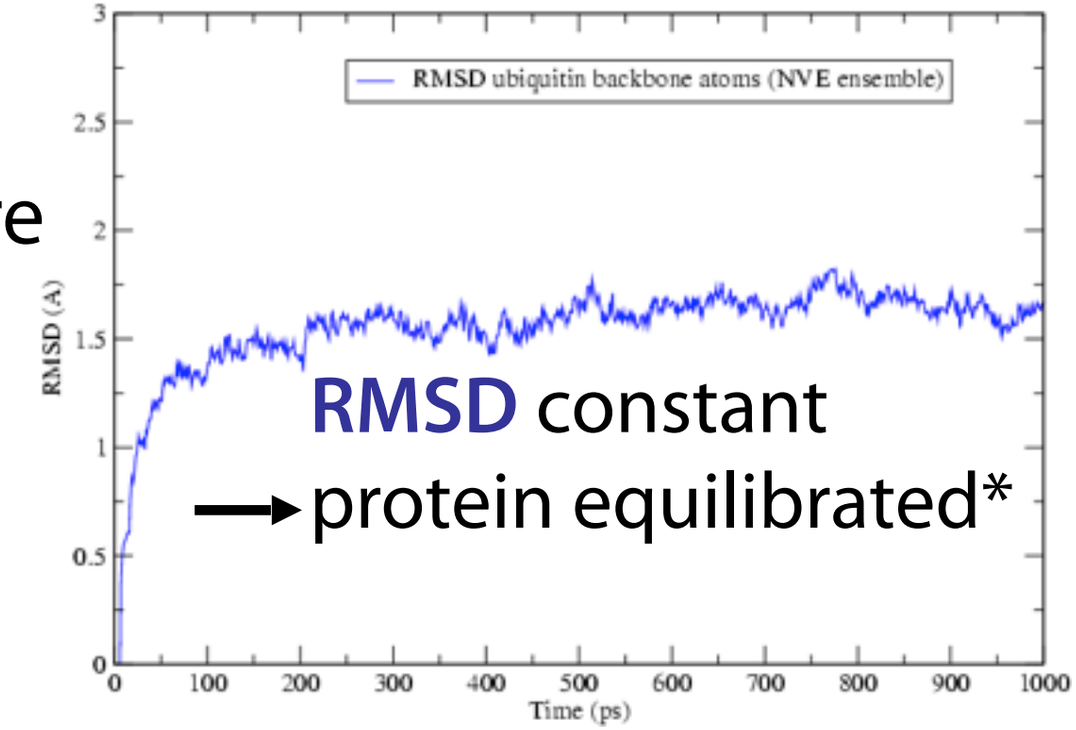
The atomic velocities of a protein establish a thermometer.


Normal Distribution of Temperatures
NVE Simulation (N=11000)
Normal Distribution
μ = 297.7 K
2σ = 2.4 K
ubiquitin in water bath
$\sigma^2 = \frac{2T^2}{3N}$
Normalized frequency
Temperature (K)
0.12
0.1
0.08
0.06
0.04
0.02
0
294
295
296
297
298
299
300
301
302
303

Equilibrium Properties of Proteins

Root Mean Squared Deviation: measure for equilibration and protein flexibility

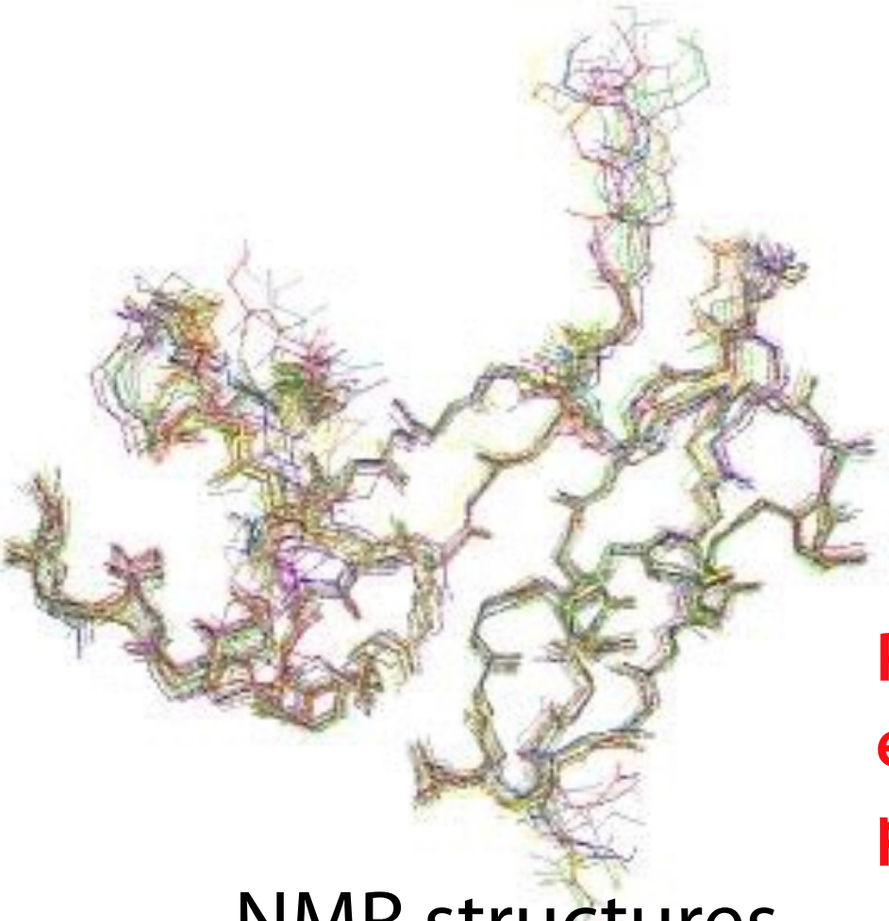
$$RMSD(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(R_i(t) - R_i(0)\right)^2}$$



RMSD constant

→ protein equilibrated*

**you never really know for sure*

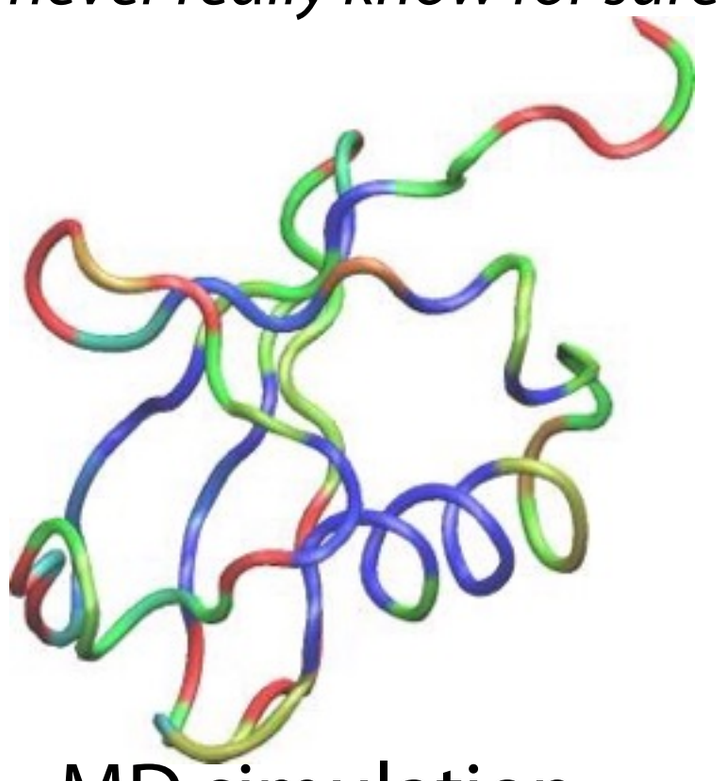


NMR structures

aligned together to see flexibility

Ubiquitin

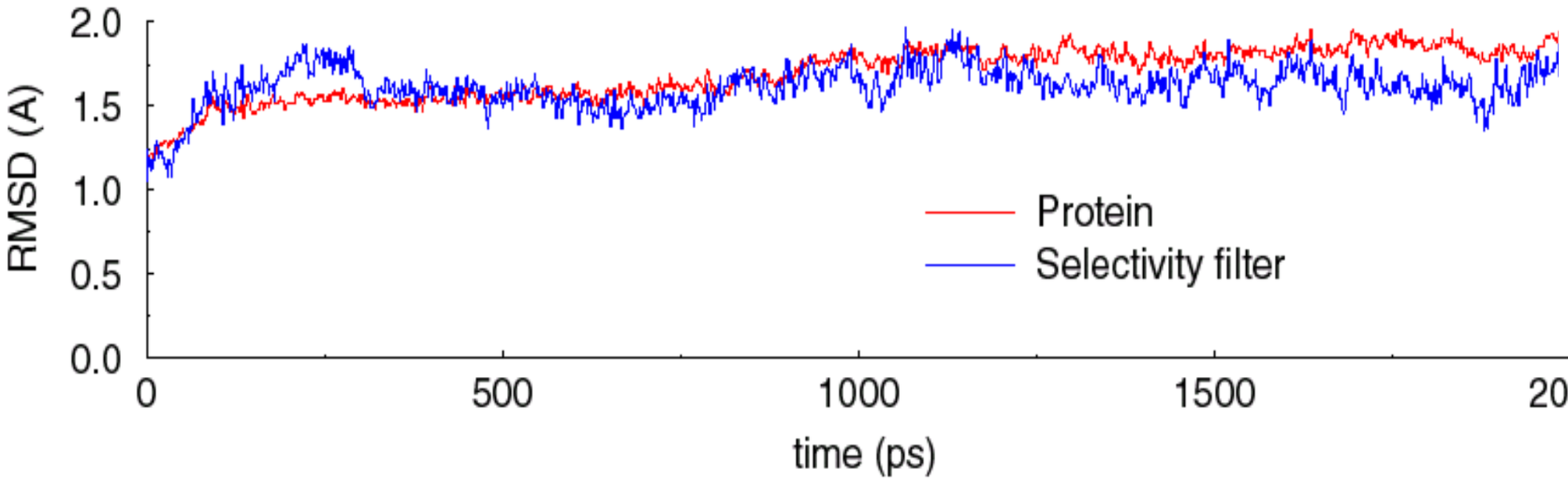
Protein sequence exhibits characteristic permanent flexibility!

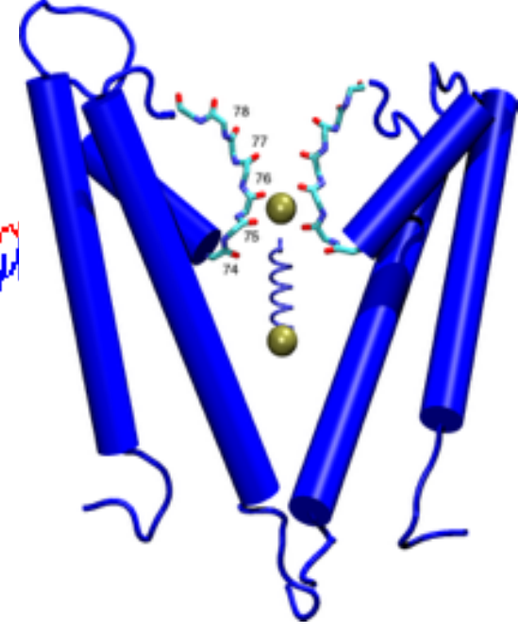


MD simulation

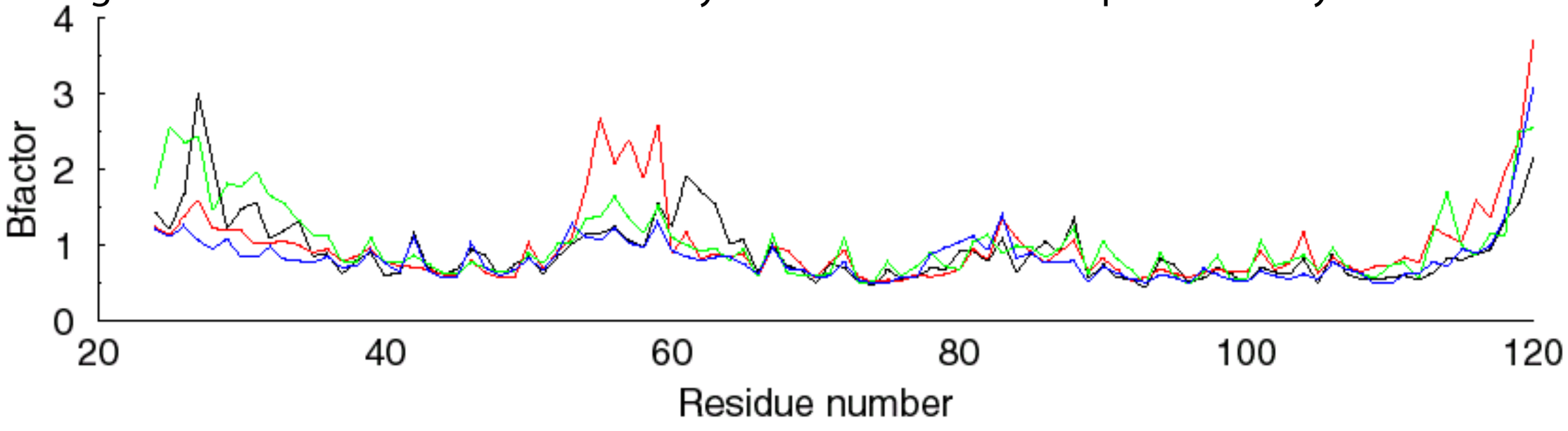
The color represents mobility of the protein through simulation (red = more flexible)

MD Results





RMS deviations for the KcsA protein and its selectivity filer 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.