Lecture 5: Molecular Dynamics

Junmei Wang

Department of Pharmacology, University of Texas
Southwestern Medical Center at Dallas

Junmei.wang@utsouthwestern.edu

Potential Functions

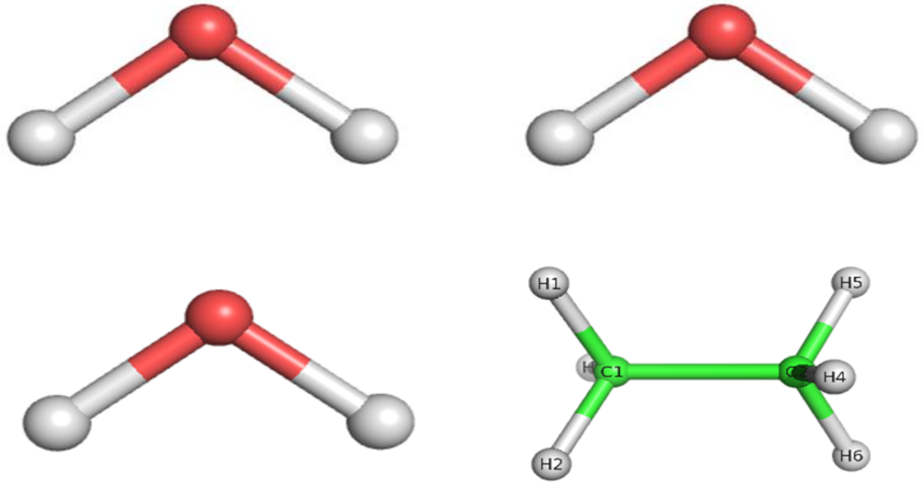
Potential Functions

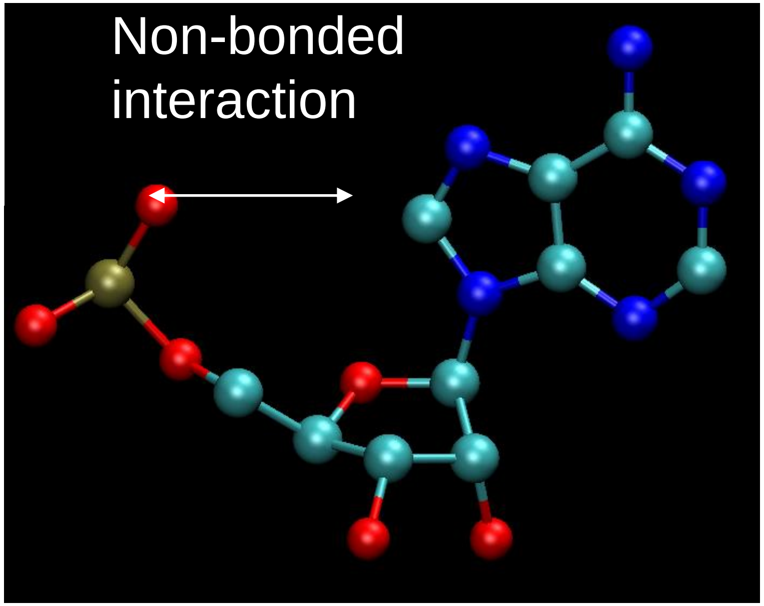
- **Molecular Mechanics**

$$\boxed{V_{potential}} = \sum_{\text{bonds}} K_r\left(r - r_{eq}\right)^2 + \sum_{\text{angles}} K_\theta\left(\theta - \theta_{eq}\right)^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}\left[1 + \cos\left(n\varphi - \gamma\right)\right] +$$

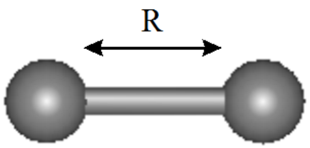
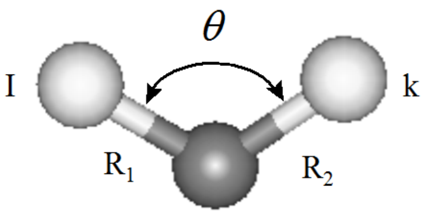
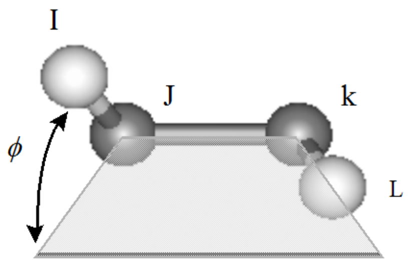
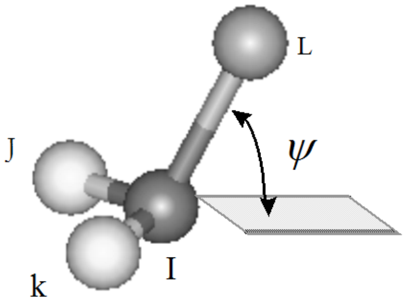
$$\sum_{\substack{i<j \\ \text{van der Waals}}} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right) + \sum_{\substack{i<j \\ \text{electrostatic}}} \frac{q_i q_j}{R_{ij}}$$

Additive
Harmonic
Effective

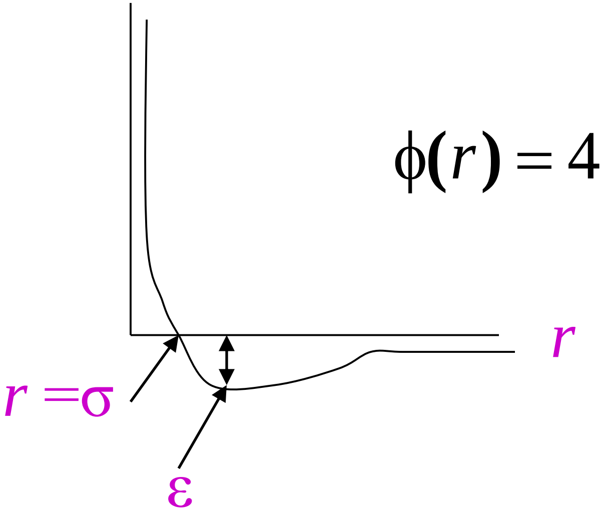




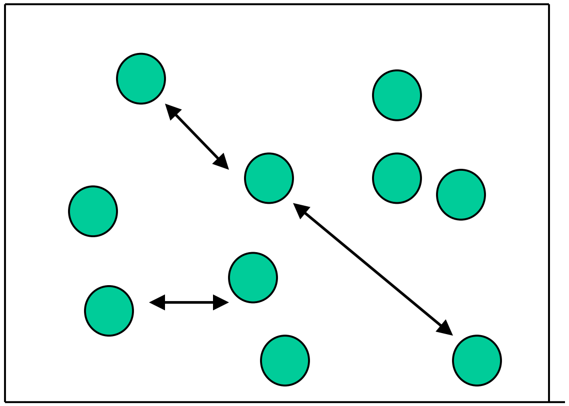
Valence Force Field

Description	Ilustration	Typical Expressions	
		Potential	Forces
Bond Stretch		$U_r = \frac{1}{2} K_b (R - R_0)^2$	$-\frac{\partial U}{\partial R} = K_b (R - R_0)$
Bond Angle bend		$U_\theta = \frac{1}{2} C(\cos\theta - \cos\theta_0)^2$	$-\frac{\partial U}{\partial \theta} = C(\cos(\theta) - \cos(\theta_0))\sin(\theta)$
Torsion		$U_\phi = \sum_{n=1}^{p} \frac{1}{2} K_{\theta,n} [1 - d\cos(n\phi)]$	$-\frac{\partial U}{\partial R} = \frac{1}{2} K_b \cdot d \cdot n \cdot \sin(n\theta)$
Inversion		$U_\psi = \frac{1}{2} C(\cos\psi - \cos\psi_0)^2$	$-\frac{\partial U}{\partial \psi} = C(\cos(\psi) - \cos(\psi_0))\sin(\psi)$

Lenard-Jones Potential



$$\phi(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$



The force in x_1 direction between (x_1, y_1, z_1) and (x_2, y_2, z_2)

$$F_{x_1} = -\frac{\partial\phi}{\partial x_1} = -\frac{\partial\phi}{\partial r}\frac{\partial r}{\partial x_1} = -\frac{\partial\phi}{\partial r}\frac{\partial[(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2]^{1/2}}{\partial x_1} =$$

$$= -\frac{\partial\phi}{\partial r}\frac{x_1}{r} = 24\varepsilon\left[\frac{2\sigma^{12}}{r^{13}} - \frac{\sigma^6}{r^7}\right]\frac{x_1}{r}$$

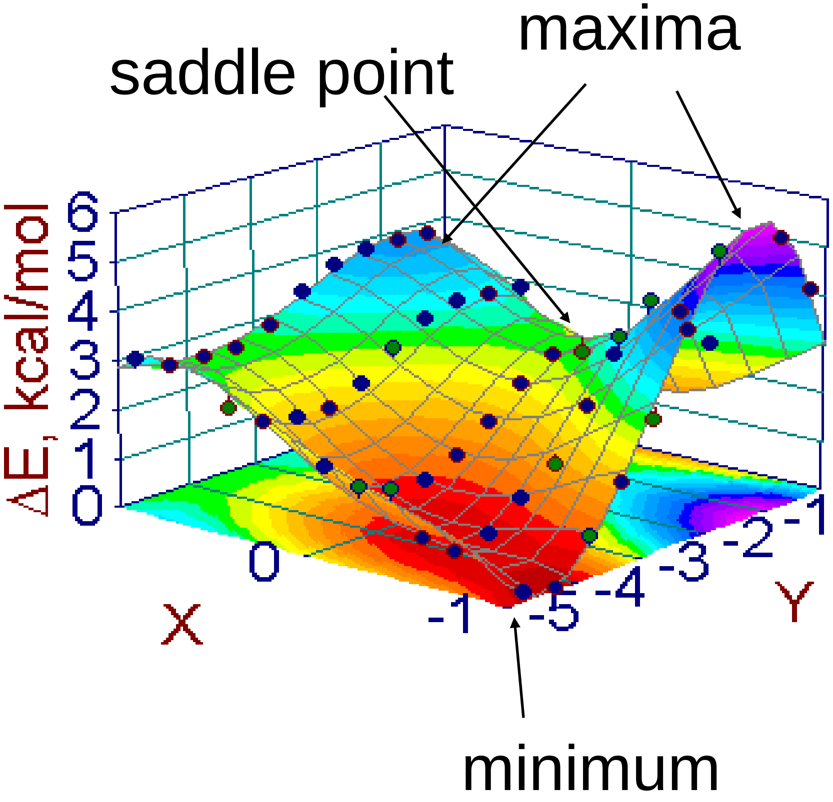
$$r_{ij} = \sqrt{(x_i-x_j)^2+(y_i-y_j)^2+(z_i-z_j)^2}$$

$$\frac{\partial v}{\partial x_i} = \frac{\partial v}{\partial r_{ij}}\frac{\partial r_{ij}}{\partial x_i} \qquad \frac{\partial r_{ij}}{\partial x_i} = \frac{(x_i-x_j)}{r_{ij}}$$

Minimization

Potential Energy Surface

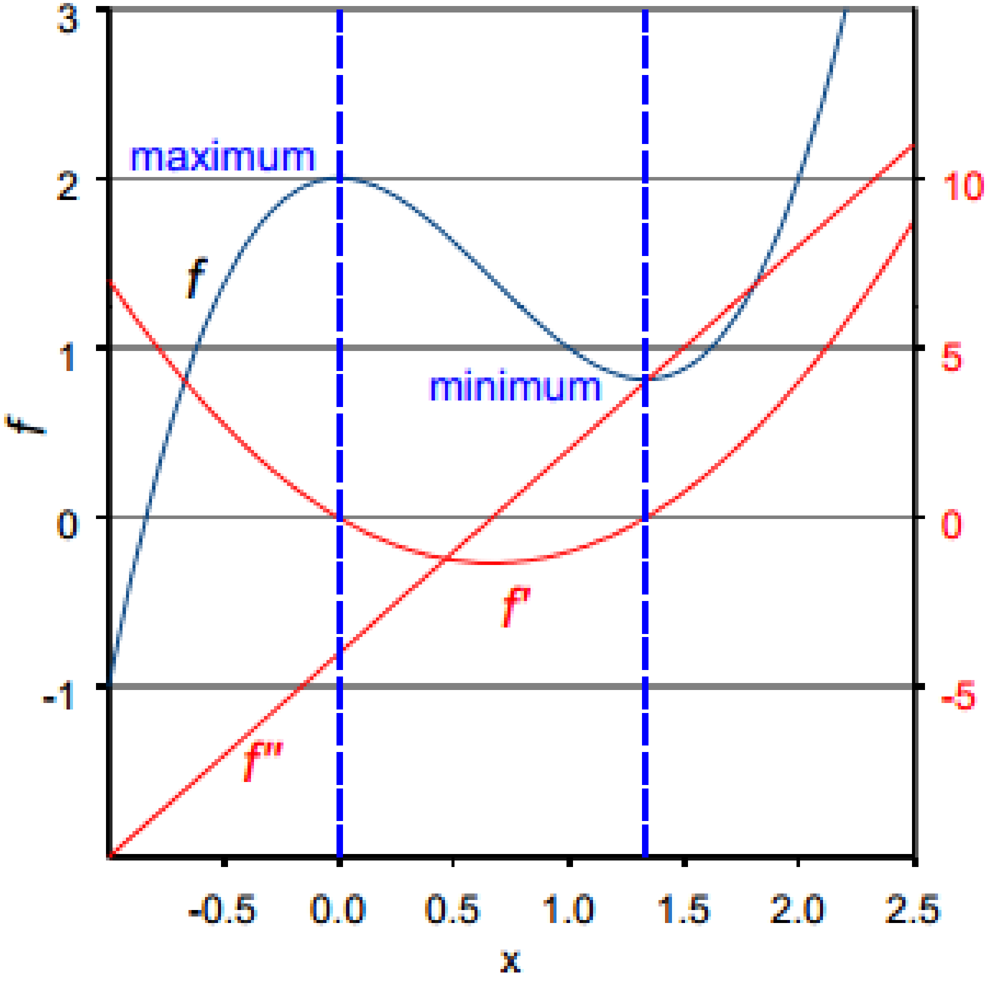
Extrema (stationary points, where the gradient is zero):



Minimization Methods

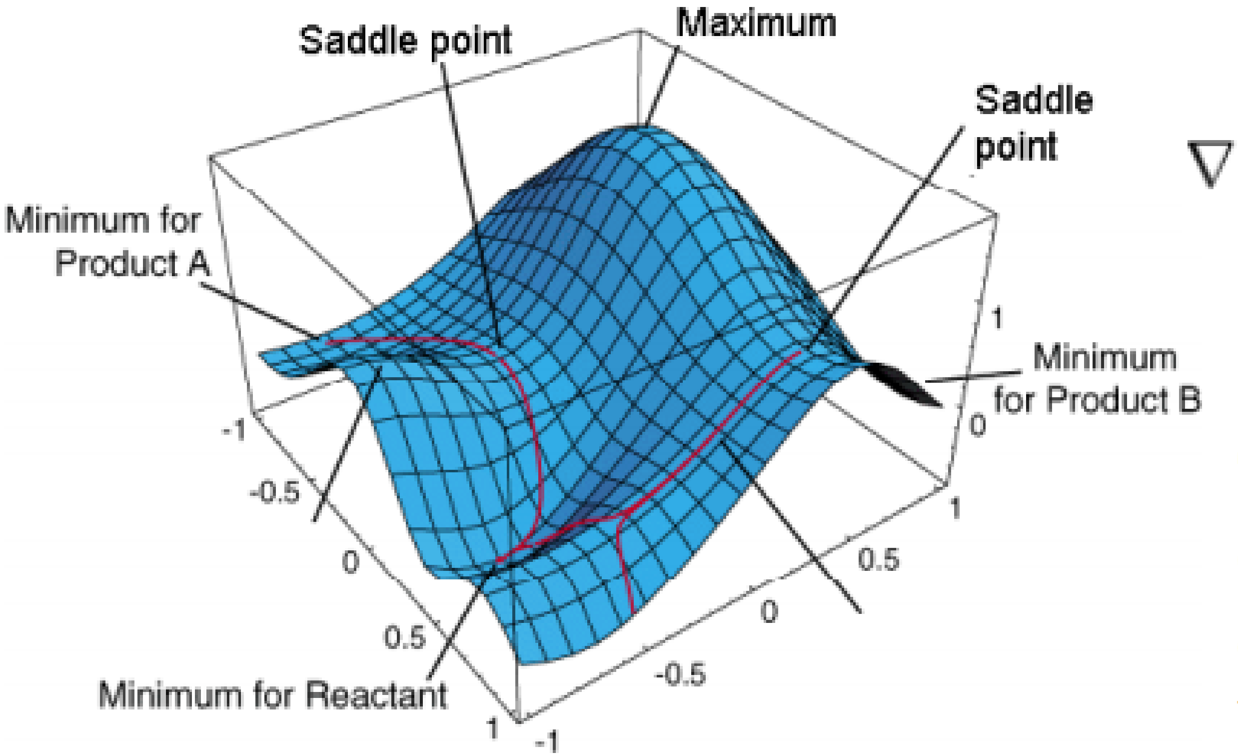
1. Simplex
2. Steepest descent
3. Conjugated gradient
4. Newton Raphson

One-dimensional optimization



- *f'(x)=0* – stationary point
- *f''(x)<0* – maximum
- *f''(x)>0* – minimum

Multidimensional Optimization



$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right)$$

- gradient

The coordinates can be Cartesian (x,y,z..), or internal, such as bondlength and angle displacements)

Force: $F = -\nabla E$ (Potential energy)

$\nabla f = 0$ - Stationary Point (minimum, maximum, or saddle point)

Hessian matrix

A matrix of second-order derivatives of the energy with respect to atomic coordinates (e.g., Cartesian or internal coordinates)

Sometimes called force matrix – matrix size of (3N-6)x(3N-6)

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

$$H_{ij}(f) = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

Quadratic approximation

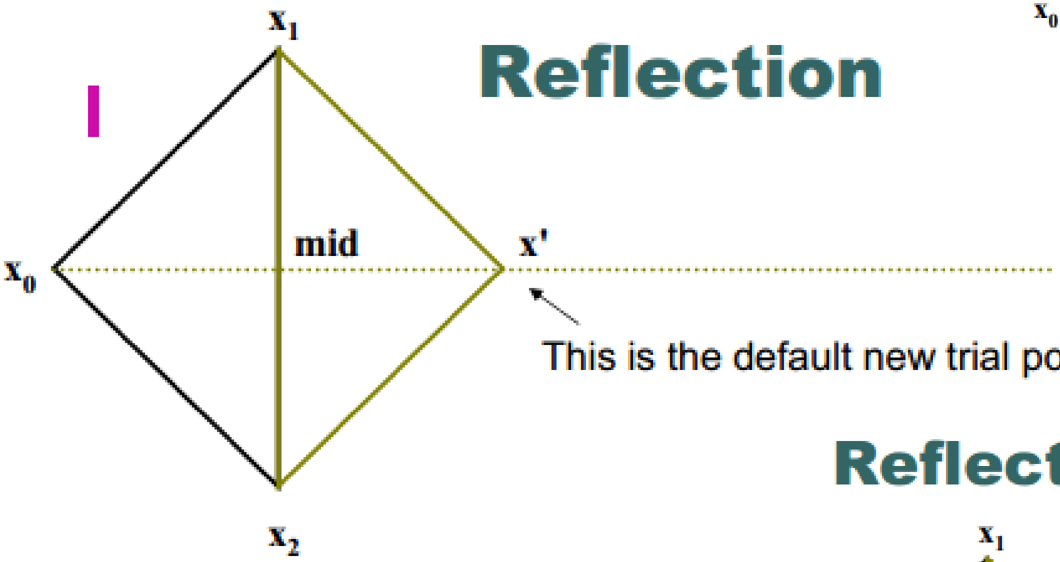
Approximate the complex energy landscape by harmonic potentials around a stationary point $(x_1^{(st)},\dots x_n^{(st)})$ $[\nabla E(x_1^{(st)},\dots x_n^{(st)}) = 0]$

$$E(x_1,\dots x_n) = E(x_1^{(st)},\dots x_n^{(st)}) + \tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} H_{ij}(E)(x_i - x_i^{(st)})(x_j - x_j^{(st)})$$

Simplex Method

I

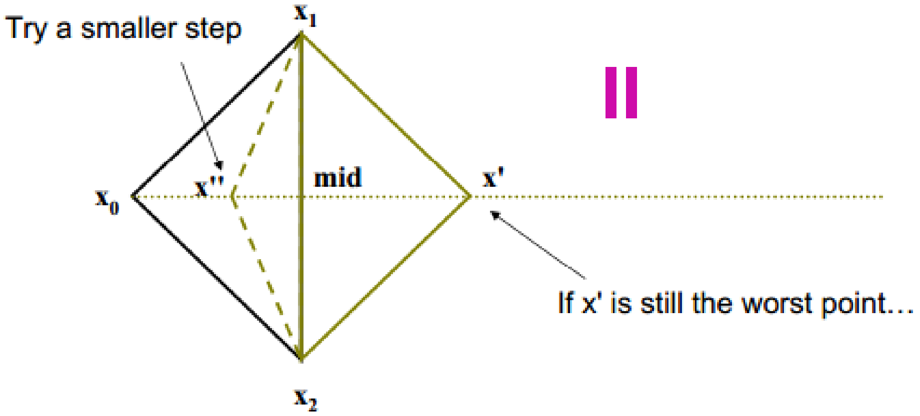
Reflection



This is the default new trial point

II

Contraction (One Dimension)

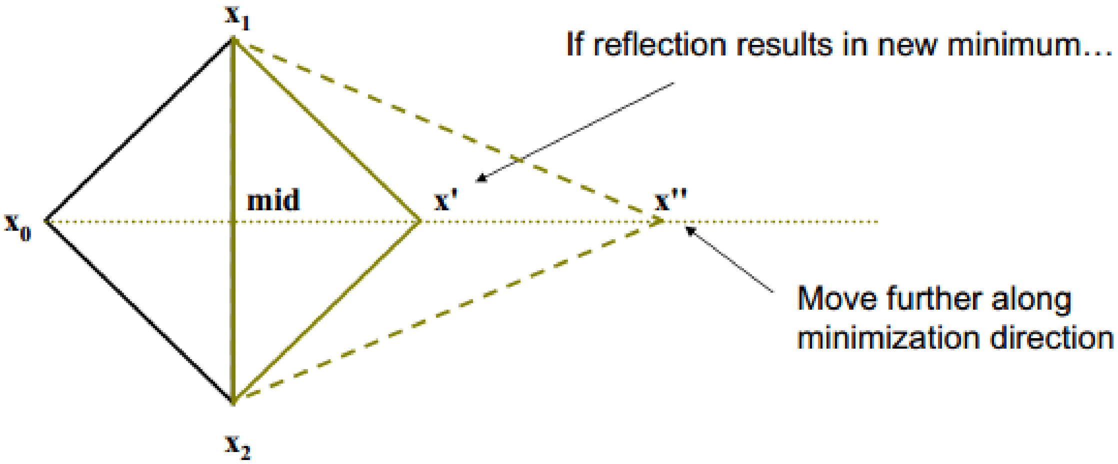


Try a smaller step

If x' is still the worst point…

III

Reflection and Expansion



If reflection results in new minimum…

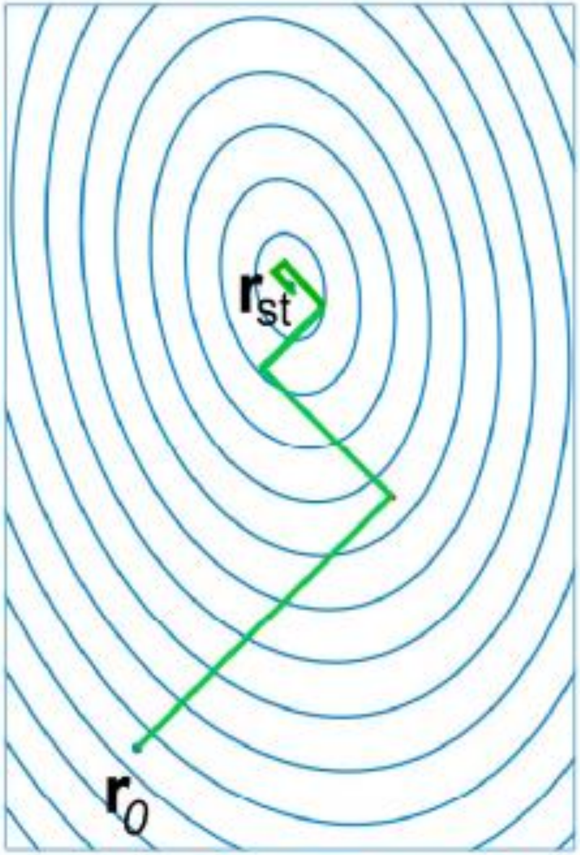
Move further along
minimization direction

Steepest Descent

Steepest descent direction $\mathbf{g}=-\nabla E$

$\mathbf{r}=(x_1,\ldots x_n)$, $\mathbf{g}$ - (3N-6) -dimensional vectors

$$\mathbf{r}_{i+1}=\mathbf{r}_i+\lambda_i\mathbf{g}_i/|\mathbf{g}_i|$$



Advantage:

- local minimization is guaranteed
- fast minimization far from the minimum

Disadvantage:

- slow descent along narrow valleys
- slow convergence near minima

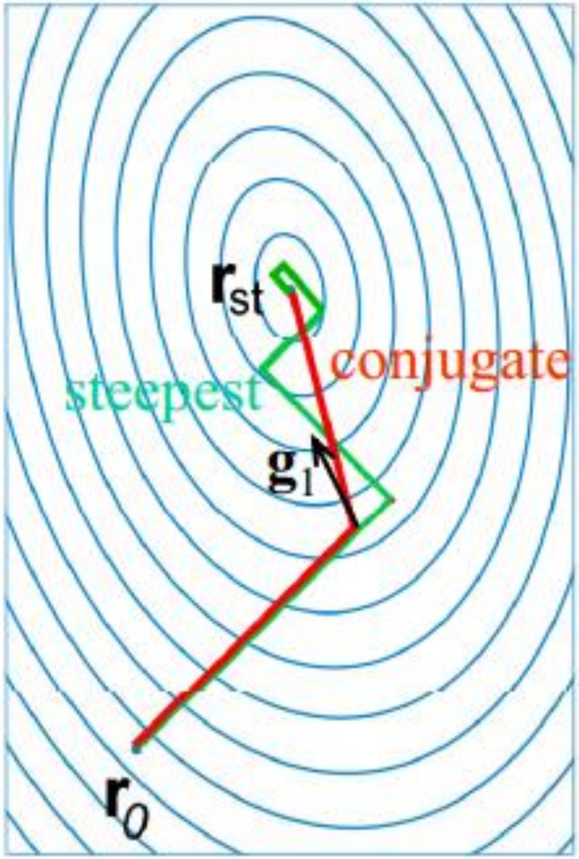
History gradient information is not kept

Conjugate Gradient Methods

Search direction is chosen using history gradient information

Initial direction – the steepest descent direction, $\mathbf{h}_0 = \mathbf{g}_0 = -\nabla E$

$$\mathbf{r}_{i+1} = \mathbf{r}_i + \xi \mathbf{h}_i$$



ξ is defined by a one-dimensional minimization along the search direction.

$$\mathbf{h}_{i+1} = \mathbf{g}_{i+1} + \gamma_{i+1} \mathbf{h}_i$$

Fletcher-Reeves method: $\gamma_{i+1} = \dfrac{|\mathbf{g}_{i+1}|^2}{|\mathbf{g}_i|^2}$

Polak-Ribiere method: $\gamma_{i+1} = \dfrac{(\mathbf{g}_{i+1} - \mathbf{g}_i)\mathbf{g}_{i+1}}{|\mathbf{g}_i|^2}$

Polak-Ribiere may be superior for non-quadratic functions.

Gradient history is equivalent to implicit use of the Hessian matrix.

Newton-Raphson Methods

Explicit use of Hessian matrix.

Quadratic approximation : $E(\mathbf{r}) = E(\mathbf{r}_0) - \sum_{k=1}^{n} (F_k q_k - \frac{1}{2} \varepsilon_k q_k^2)$

q_k- normal coordinates, $\mathbf{r} - \mathbf{r}_0 = \sum_{k=1}^{n} q_k \mathbf{l}_k$ $\quad \nabla E(\mathbf{r}_0) = -\sum_{k=1}^{n} F_k \mathbf{l}_k$

ε_k and $\mathbf{l}_k$ - eigenvalues and eigenvectors of the Hessian matrix at $\mathbf{r}_0$

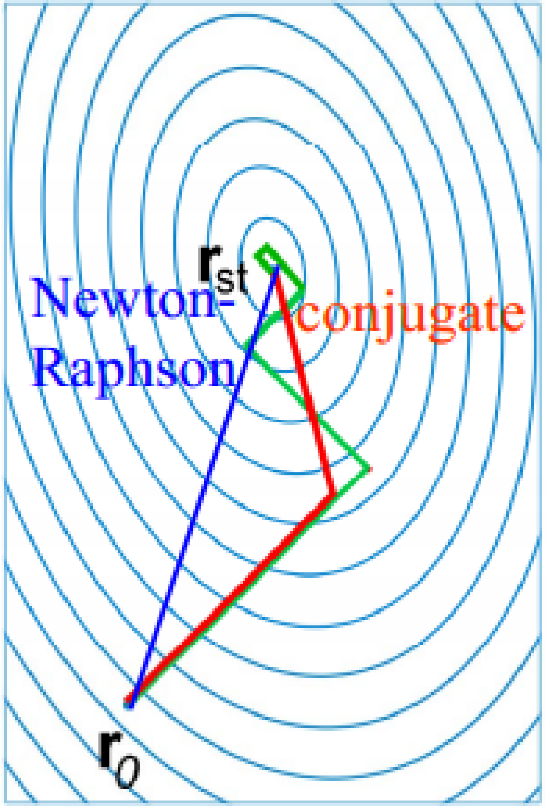
One-step optimization of quadratic functions $q_k = \frac{F_k}{\varepsilon_k} \Rightarrow \nabla E(\mathbf{r}) = 0$

For arbitrary functions

$$\mathbf{r}_{i+1} = \mathbf{r}_i + \sum_{k=1}^{n} \mathbf{l}_k(\mathbf{r}_i) F_k(\mathbf{r}_i) / \varepsilon_k(\mathbf{r}_i)$$

Descent direction for $\varepsilon_k(\mathbf{r}_i) > 0$

Finds the closest stationary point (either minimum, maximum, or saddle point).



Convergence Criteria

1. Maximum steps √
2. RMS displacement
3. RMS force √
4. Maximum Displacement
5. Maximum Force

√: used by AMBER

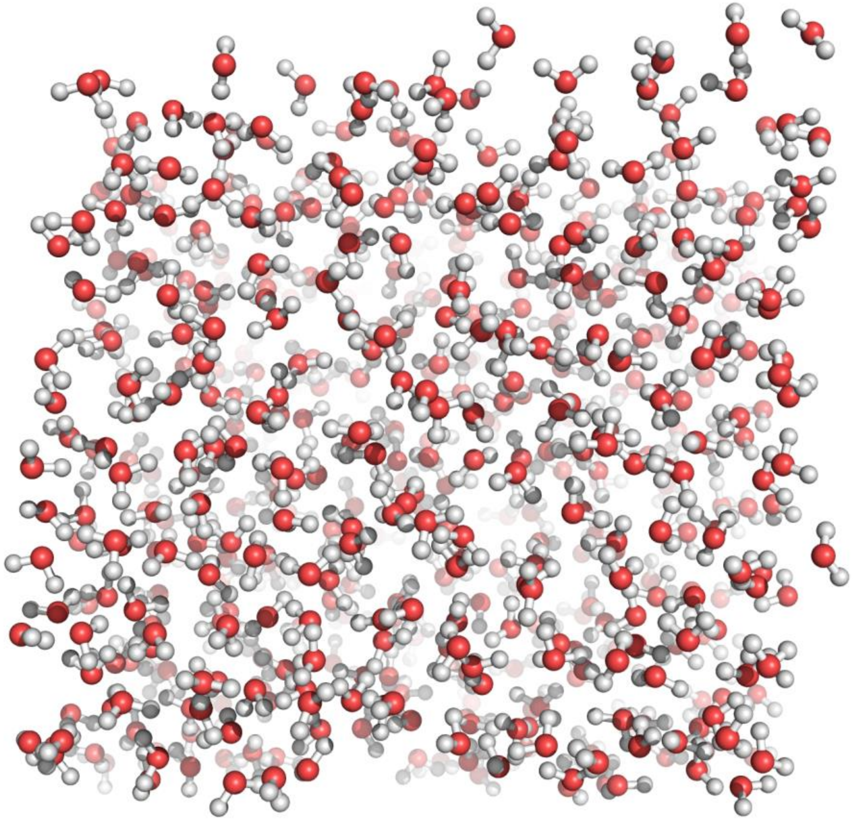
Molecular Dynamics Simulations

Molecular Dynamics (MD) Simulations

Why?

Static → Dynamic

562 water molecules



Molecular Dynamics

- MD is our approximation to how molecules explore their potential energy surface in the real world
- The atoms are “heated” by giving them a distribution of velocities corresponding to temperature we wish to simulate
- The wiggling and jiggling of the atoms is then obtained by integrating the Newtonian laws of motion
- This gives us the energy of all states occupied at that Condition (temperature, pressure) **as long as we simulate long enough (Ergodic hypothesis)**

Molecular Dynamics Simulations

How?

Newton’s Law of Motions

$$F(X) = -\nabla \boxed{V_{potential}} = Ma(t) \quad (1)$$

$$V(t=0) = V_0 \quad (2)$$

$$\frac{m_i}{2}\left\langle V_{0i,a}^{2} \right\rangle = \frac{k_B T}{2} \quad (a = x, y, z) \quad (3)$$ **(Equipartition Theorem)**

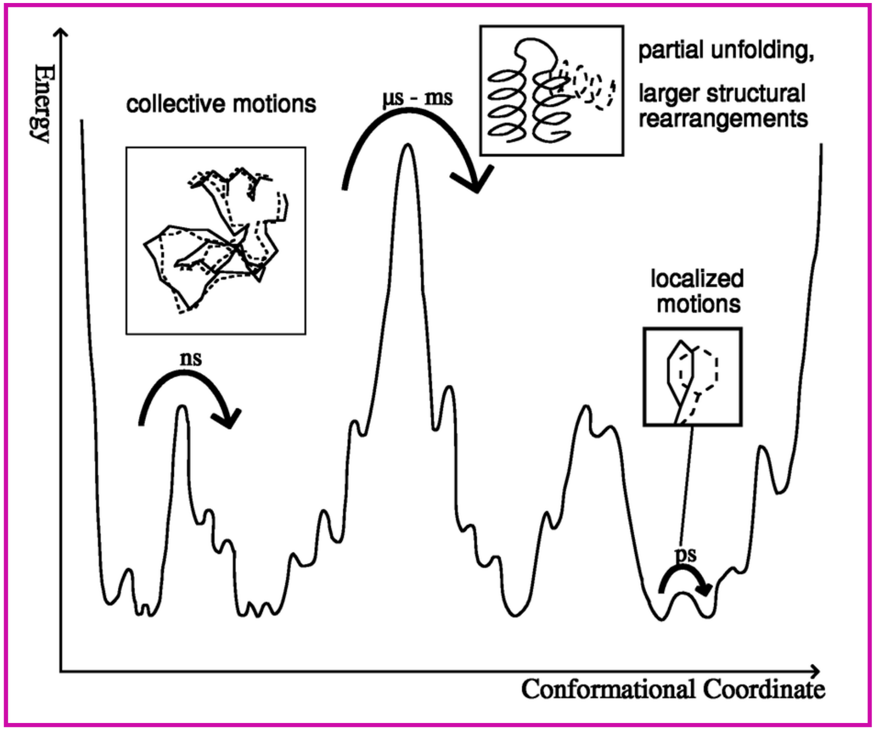
$$V(t+\Delta t) = V(t) + a(t)\Delta t \quad (4)$$

$$X(t+\boxed{\Delta t}) = X(t) + \Delta t V(t) \quad (5)$$

Two Biggest challenges in MD simulations:

1. Values of inter-atomic Newtonian forces must be accurate.
2. Time step Δt must be very small to represent fastest vibrations (2 fs).

Timescales



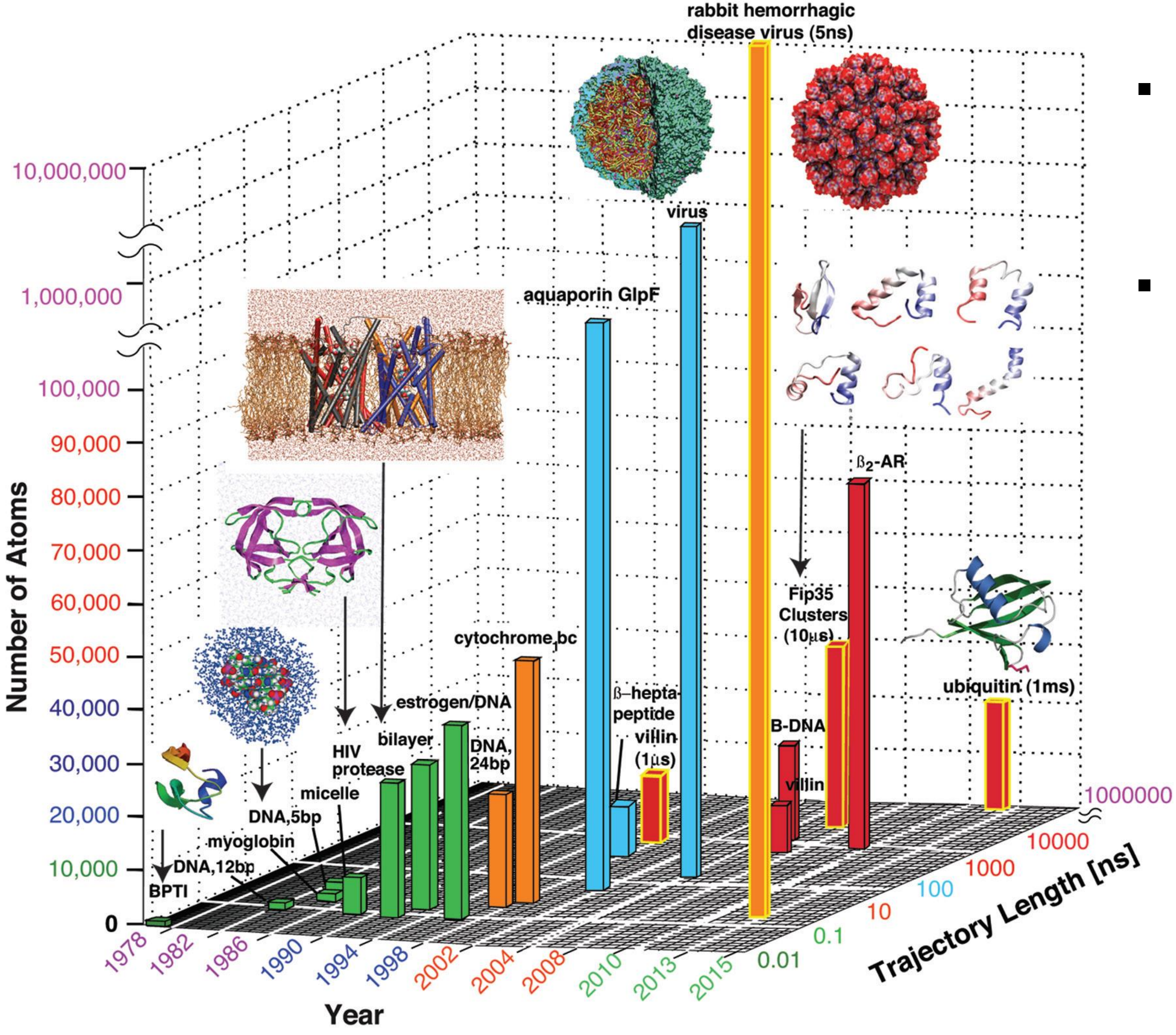
1. Bond vibrations - 1 fs
2. Collective vibrations - 1 ps
3. Conformational transitions - ps or longer
4. Enzyme catalysis - microsecond/millisecond
5. Ligand Binding - micro/millisecond
6. Protein Folding - millisecond/second

Molecular dynamics:

Integration time step – 1 or 2 femtosecond

Accessible timescale: 10 nanoseconds to 1µ seconds.

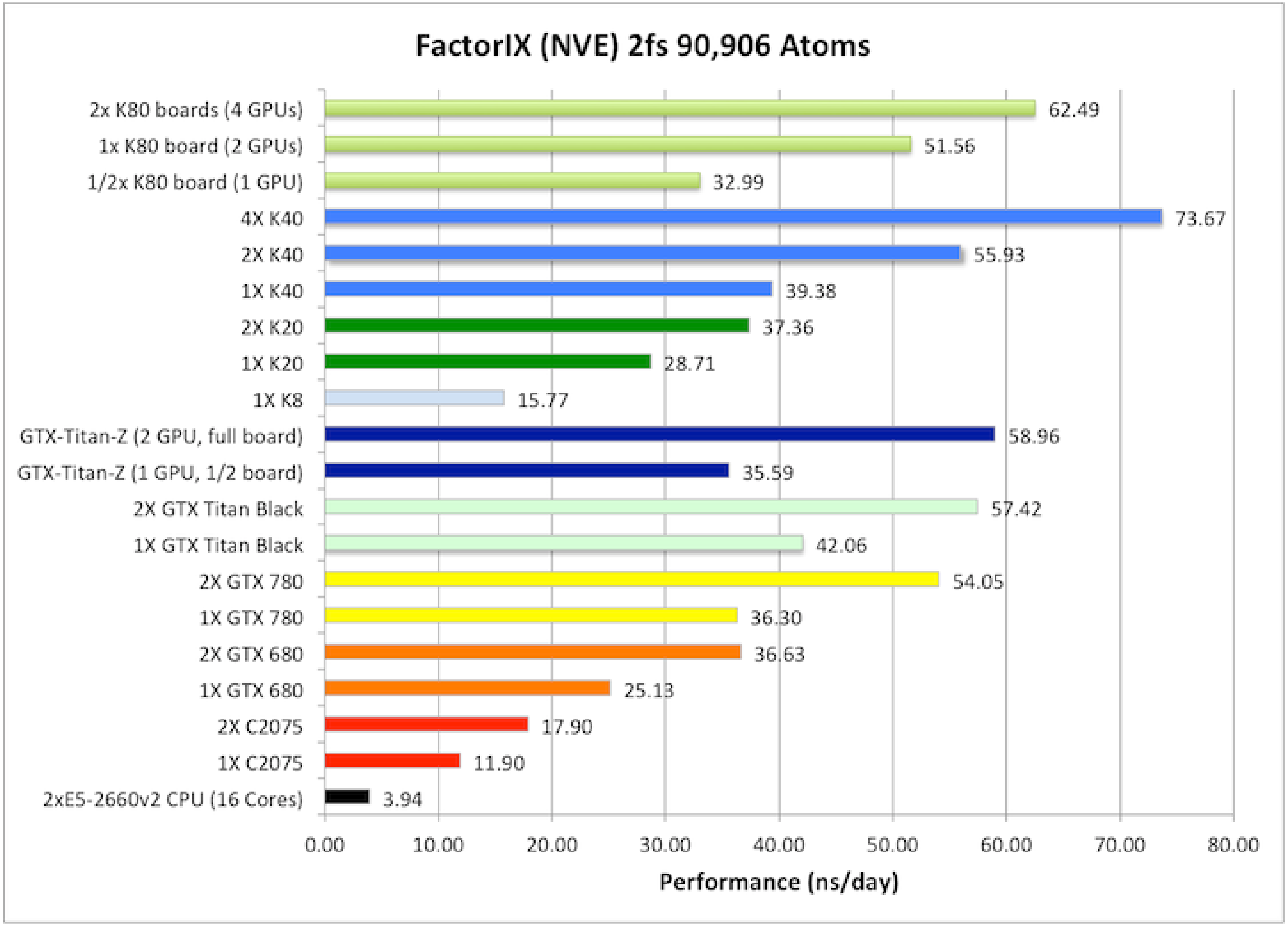
MD Simulations in Studying Biological Systems



- 1-ms simulations of ubiquitin by D. E. Shaw
- 5-ns simulations of HIV-1 capsid consisting of 64 million atoms

Courtesy of Tamar Schlick

Benchmark of AMBER GPU-MD for FactorIX



Courtesy of ambermd.org

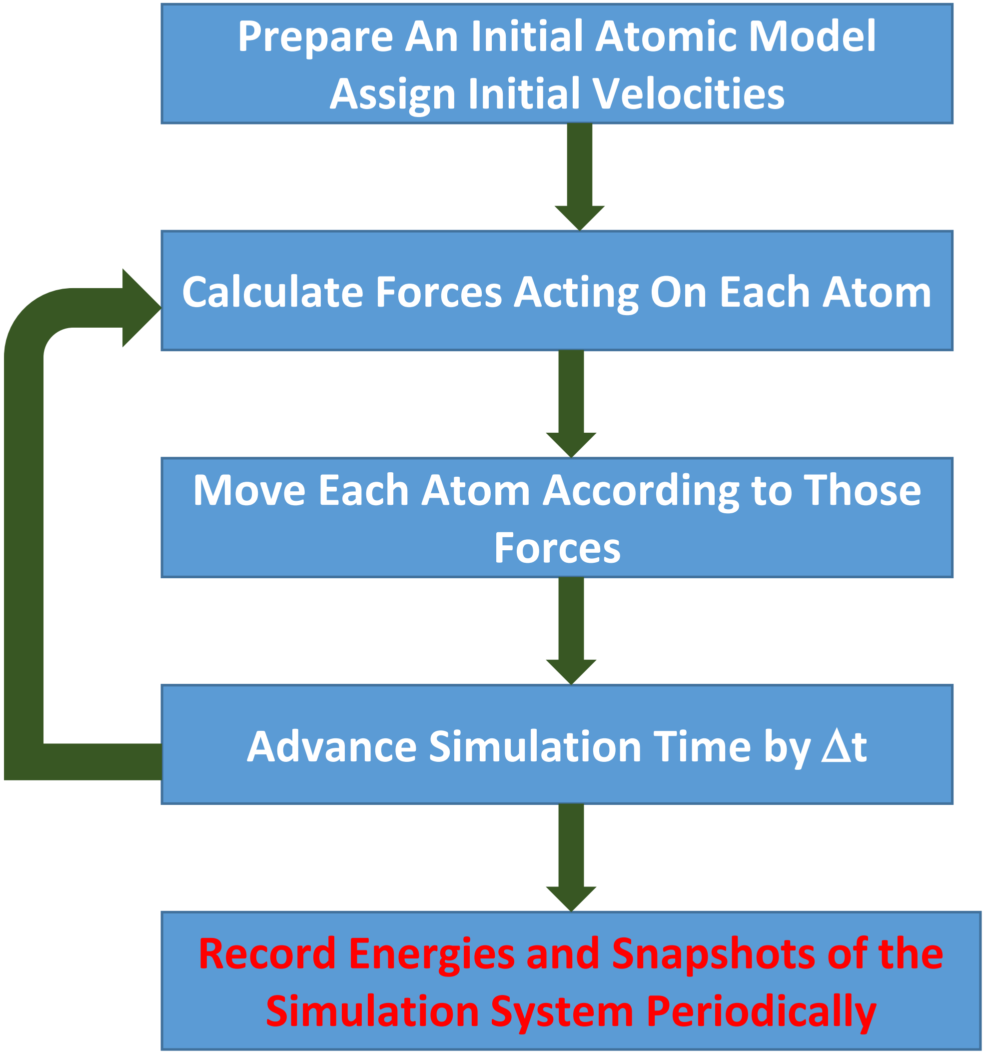
Steps in Molecular Simulations Studies

1. Build realistic atomistic model of the system

2. Simulate the behavior of your system over time using specific conditions (temperature, pressure, volume, etc)

3. Analyze the results obtained from MD and relate to macroscopic level properties

Basic Procedure of Performing MD Simulations



Prepare An Initial Atomic Model
Assign Initial Velocities

Calculate Forces Acting On Each Atom

Move Each Atom According to Those Forces

Advance Simulation Time by Δt

Record Energies and Snapshots of the Simulation System Periodically

Molecular Dynamics

- **Solve Newton's equation for a molecular system:**

Potential Function → Force

$$F_i = -\frac{\partial V_i}{\partial r_i}$$

Newton's Law:

$$F_i = m_i a_i$$

F_i=force on *i*th atom
m_i = mass of *ith* atom
a_i=acceleration of *i*th atom

How do you run a MD simulation?

- **Get the initial configuration**
- **Assign initial velocities**

At thermal equilibrium, the expected value of the kinetic energy of the system at temperature T is:

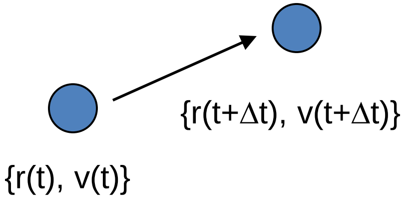
$$\langle E_{kin} \rangle = \frac{1}{2} \sum_{i=1}^{3N} m_i v_i^2 = \frac{1}{2}(3N) k_B T$$

This can be obtained by assigning the velocity components v_i from a random Gaussian distribution
with mean 0 and standard deviation ($k_B T/m_i$):

$$\langle v_i^2 \rangle = \frac{k_B T}{m_i}$$

Integrator: Verlet Algorithm

Start with {r(t), v(t)}, integrate it to {r(t+Δt), v(t+Δt)}:



{r(t+Δt), v(t+Δt)}

{r(t), v(t)}

The new position at t+Δt:

$$r(t+\Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}\Delta t^2 a(t) + O(\Delta t^3) \tag{1}$$

Similarly, the old position at t-Δt:

$$r(t-\Delta t) = r(t) - v(t)\Delta t + \frac{1}{2}\Delta t^2 a(t) - O(\Delta t^3) \tag{2}$$

Add (1) and (2):

$$r(t+\Delta t) = 2r(t) - r(t-\Delta t) + \Delta t^2 a(t) + O(\Delta t^4) \tag{3}$$

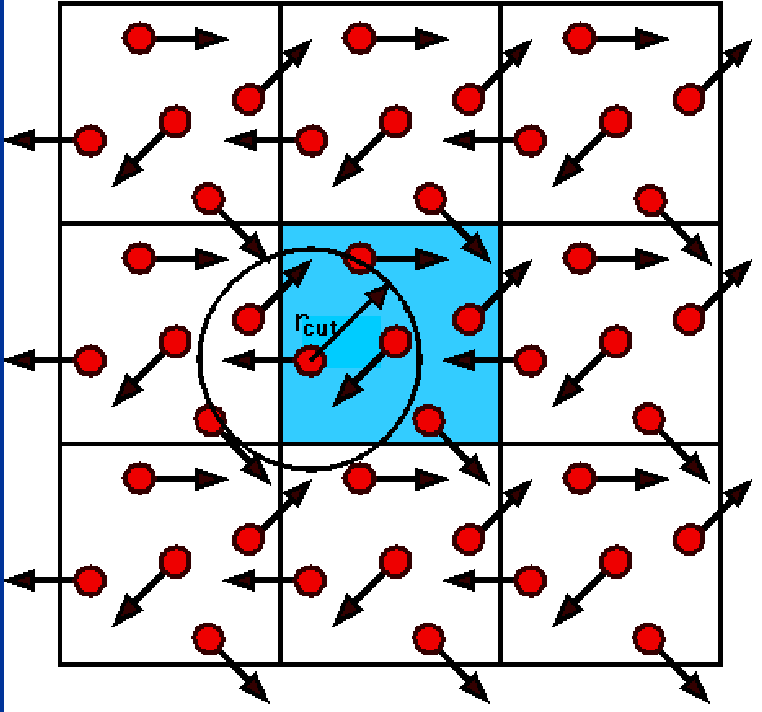
Thus the velocity at t is:

$$v(t) = \dot{r}(t) = \frac{1}{2\Delta t}(r(t+\Delta t) - r(t-\Delta t)) + O(\Delta t^2) \tag{4}$$

Periodic Boundary Condition

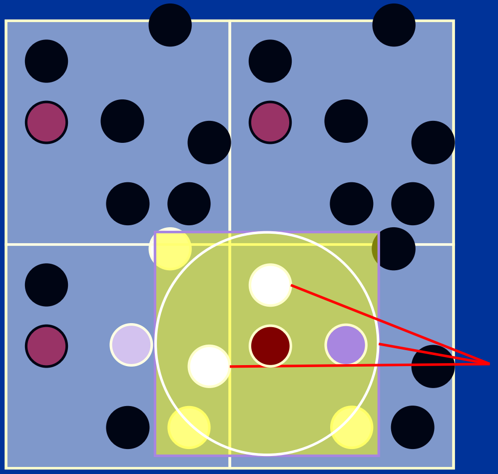
Periodic Boundary Conditions

- infinite system with small number of particles
- remove surface effects
- shaded box represents the system we are simulating, while the surrounding boxes are exact copies in every detail
- whenever an atom leaves the simulation cell, it is replaced by another with exactly the same velocity, entering from the opposite cell face (number of atoms in the cell is conserved)
- r_{cut} is the cutoff radius when calculating the force between two atoms

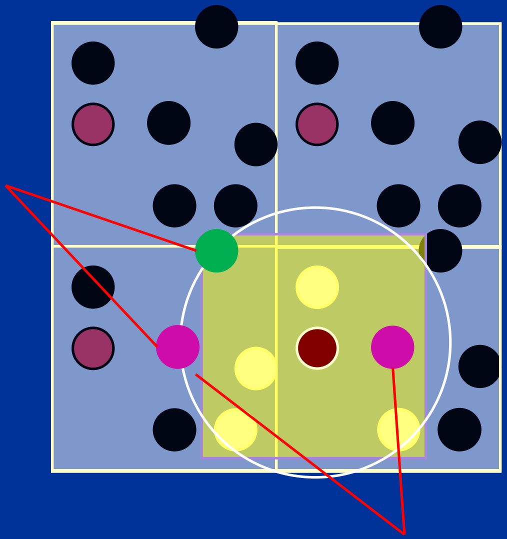


Minimum Image

1. Bulk system modeled via periodic boundary condition
   - not feasible to include interactions with all images
   - must truncate potential at half the box length (at most) to have all separations treated consistently
2. Contributions from distant separations may be important



Only interactions considered

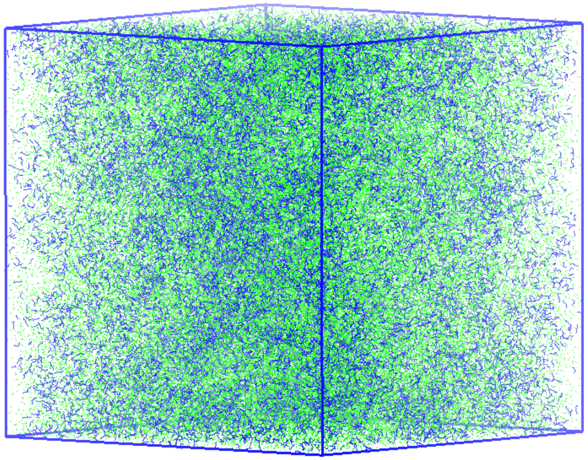


These two are same distance from central atom

Same atoms

System Configurations

Statistical Ensembles



$$Q(N,V,\beta) = \sum_i e^{-\beta E_i}$$

Q is called the
Canonical Partition Function.

$A(N,V,T) = -kTlnQ$ canonical

$S(N,V,E) = -kln\Omega$ $\text{micro} - \text{canonical}$

$pV(V,T,\mu) = -kTln\Xi$ grand canonical

$G(N,P,T) = -kTln\Delta$ $\text{thermal} - \text{isobaric}$

In Equilibrium MD, we want to sample the ensemble as best as possible!

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)

Simulating at constant T: the Berendsen scheme

- Bath supplies or removes heat from the system as appropriate
- Exponentially scale the velocities at each time step by the factor λ:

$$\lambda = \sqrt{1 - \frac{\Delta t}{\tau}\left(1 - \frac{T_{bath}}{T}\right)}$$

T: “kinetic” temperature

where τ determines how strong the bath influences the system

Berendsen et al. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684 (1984)

Simulating at constant P: Berendsen scheme

- Couple the system to a pressure bath
- Exponentially scale the volume of the simulation box at each time step by a factor λ:

$$\lambda = 1 - \kappa \frac{\Delta t}{\tau_P}\left(P - P_{bath}\right)$$

where

$$P = \frac{2}{3v}\left(E_{kin} + \sum_{i=1}^{N} x_i \bullet F_i\right)$$

υ : volume
x_i: position of particle i
F_i : force on particle i

where κ : isothermal compressibility
τ_P : coupling constant

Berendsen et al. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684 (1984)

Time averages and ensemble averages

macroscopic numbers of atoms or molecules (*of the order of* 10^{23}*, Avogadro's number is 6.02214199 ×* 10^{23}): impossible to handle for MD

statistical mechanics (Boltzmann, Gibbs): a single system evolving in time is replaced by a large number of replications of the same system that are considered simultaneously

time average is replaced by an *ensemble average*:

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time}$$

Ergodic hypothesis

$$\langle A \rangle_{ensemble} = \int\int dp^N dr^N A(p^N, r^N)\rho(p^N, r^N)$$

$$\langle A \rangle_{time} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A\left(p^N(t), r^N(t)\right) dt \approx \frac{1}{M} \sum_{t=1}^{M} A\left(p^N, r^N\right)$$

Ergodic hypothesis

- Classical statistical mechanics integrates over all of *phase space* {r,p}.
- The *ergodic hypothesis* assumes that for sufficiently long time the *phase trajectory* of a closed system passes arbitrarily close to every point in phase space.
- Thus the two averages are equal

Analyzation MD

- Averages
- Fluctuations
- Time Correlations

Time variation of energies

- kinetic energies
- potential energies

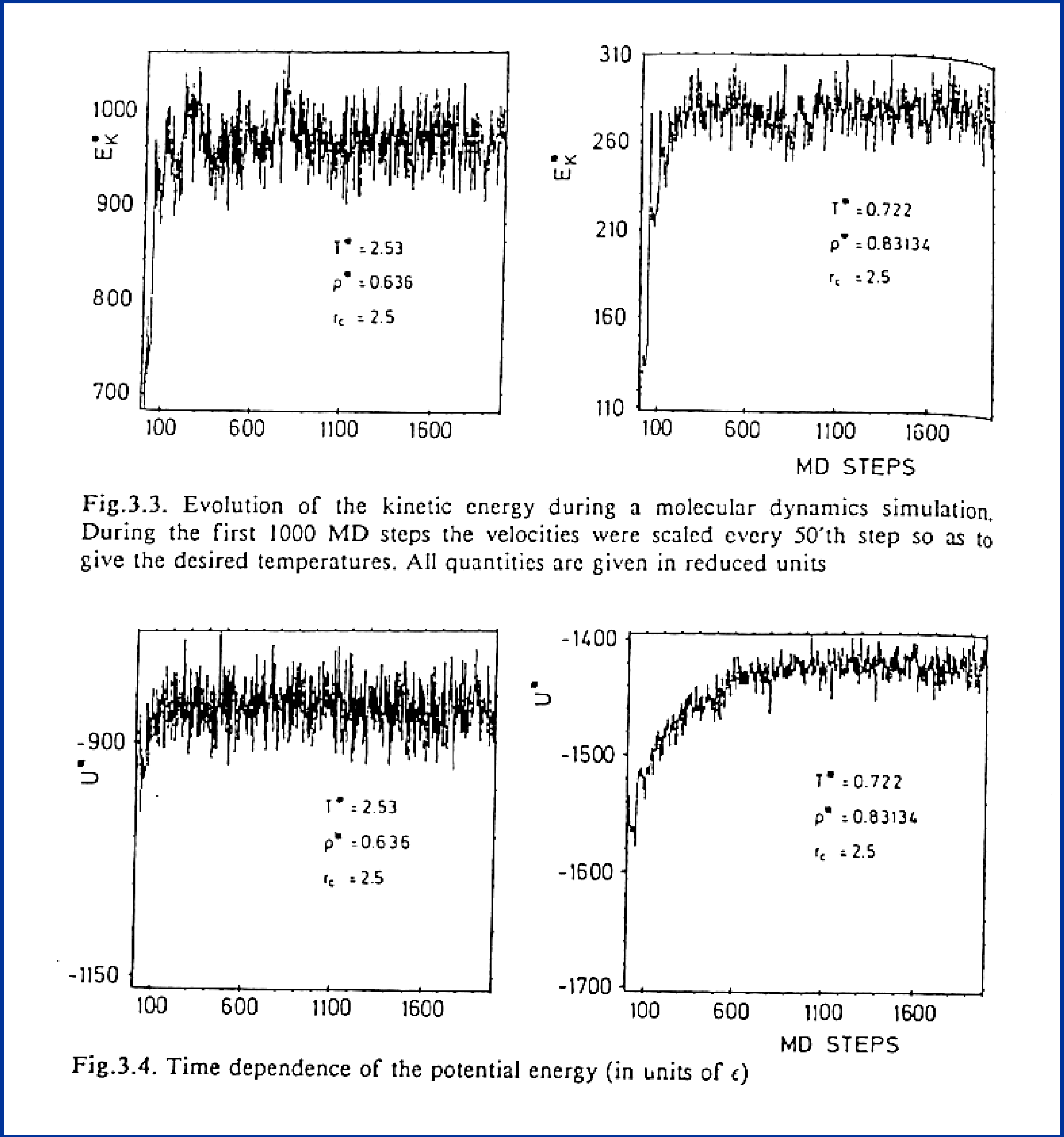
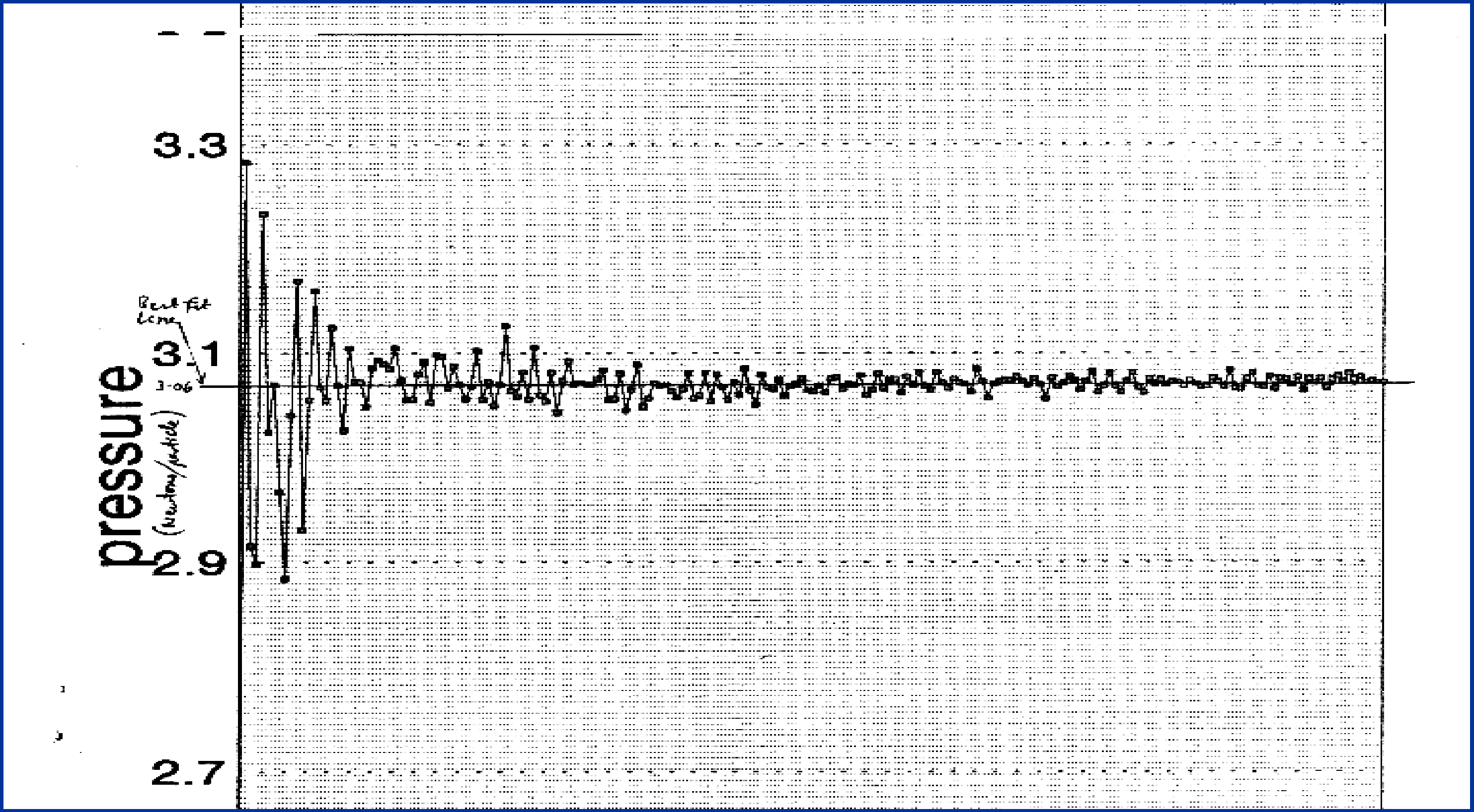


Fig.3.3. Evolution of the kinetic energy during a molecular dynamics simulation. During the first 1000 MD steps the velocities were scaled every 50'th step so as to give the desired temperatures. All quantities are given in reduced units

Fig.3.4. Time dependence of the potential energy (in units of ϵ)

Time variation of pressure

- Equilibration of pressure with time



Simple Average

- Mean energy

$$\langle E \rangle = \frac{1}{N}\sum_{i=1}^{N} E_i$$

- Mean Structure

- RMS difference between two structures

$$RMSD_i = \left\langle \left(r_i^{\alpha} - r_i^{\beta}\right)^2 \right\rangle^{\frac{1}{2}} = \sqrt{\frac{1}{N_i}\sum_i \left(r_i^{\alpha} - r_i^{\beta}\right)^2}$$

- B-factor

$$B_i = \frac{8}{3}\pi^2 (RMSD_i)^2$$

Bulk-Density

- Density is one of a major molecular properties for which van der Waals parameterization intends to reproduce.

$$\langle \rho \rangle = \frac{NM}{N_A \langle V \rangle}$$

- Perform NTP simulations

Temperature

Temperature is related to the ensemble averaged kinetic energy by

$$\tfrac{1}{2}\,(3N - N_c)\,k_B\,T = \langle\, E_k \,\rangle_M$$

where N is the number of atoms and N_c is the number of constraints. Typically we require the total linear momentum of the system is constrained to zero (the center of mass of the system does not move), and N_c is 3.

Pressure

Pressure is related to the product of the positions and forces (for pairwise interactions):

Virial Theorem

$$PV = \underbrace{Nk_\mathrm{B}T}_{\text{ideal gas contribution}} + (1/3)\,\langle\, \textstyle\sum_{i<j} r_{ij}\, f_{ij}\, \rangle_M$$

where N is the number of atoms, r_{ij} is the distance between a pair of interacting atoms, f_{ij} is the corresponding force, and the sum is over *all pairwise* interactions.

Compressibility Factor

Compressibility measures the deviation from the ideal gas law $PV = Nk_BT$

$$Z = \frac{PV}{Nk_BT} = 1 - \frac{1}{3Nk_BT}\left\langle \sum_{i<j} r_{ij}.F_{ij} \right\rangle$$

where N is the number of atoms, r_{ij} is the distance between a pair of interacting atoms, f_{ij} is the corresponding force, and the sum is over *all pairwise* interactions.

Isothermal Compressibility And Thermal Expension Coefficient

Compressibility is a measure of the relative volume change of a fluid or solid as a response to a pressure change

$$\beta_T = -\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_T = \frac{\langle V^2\rangle - \langle V\rangle^2}{k_B\langle T\rangle\langle V\rangle}$$

Thermal expansion coefficient, α_p is defined as

$$\alpha_p = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_p = \frac{\langle VH\rangle - \langle V\rangle\langle H\rangle}{k_B\langle T\rangle^2\langle V\rangle} = -\frac{d\ln\langle\rho(T)\rangle}{dT} \approx \frac{\ln\langle\rho_2\rangle - \ln\langle\rho_1\rangle}{T_2 - T_1}$$

Heat Capacity At Constant Volume

- Calculate internal energies at different temperatures and take the partial derivatives:

$$C_V = (\partial U / \partial T)_V$$

$$= (U2 - U1) / (T2 - T1) \text{ at constant } V$$

- Calculate the fluctuation of internal energy around its mean value:

$$Nk_B T^2 C_V = \langle (U - \langle U \rangle_M)^2 \rangle = \langle U^2 \rangle_M - \langle U \rangle_M^2$$

It requires a longer simulation time for one simulation at one temperature. (A trade-off!)

Isobaric Heat Capacity

- Heat capacity is a measure of the heat energy required to increase the temperature of a unity quality of a substance by a unit of temperature.

$$c_p = \left(\frac{\partial H}{\partial T}\right)_p = \frac{\langle H^2 \rangle - \langle H \rangle^2}{N k_B \langle T \rangle^2} \approx \frac{\langle H_2 \rangle - \langle H_1 \rangle}{T_2 - T_1}$$

$$\delta c_p \approx \frac{\sqrt{\delta \langle H_1 \rangle^2 + \delta \langle H_2 \rangle^2}}{T_2 - T_1}$$

Enthalpy of Vaporization

- Enthalpy of vaporization is the enthalpy change that occurs during the transition of one mol of substance from the liquid to the gas phase, where each of the phases is under the equilibrium pressure.
- Another important molecular property to be used in van der Waals parameterization

$$\Delta H_{vap}(T) = H(p,T)_{gas} - H(p,T)_{liquid} = E_{gas} - E_{liquid} + p(V_{gas} - V_{liquid})$$

$$\Delta H_{vap}(T) = E_{gas} - E_{liquid} + RT - p\left\langle V_{liquid} \right\rangle + C$$

$$C = C_{vib} + C_{pol} + C_{ni} + C_{x}$$

$$E_{vib}^{CM} = n_{vib} k_B T$$

$$E_{vib}^{QM} = \sum_{i=1}^{n_{vib}} \left(\frac{h\nu_i}{2} + \frac{h\nu_i}{e^{h\nu_i/kT} - 1} \right)$$

$$C_{vib} = E_{vib,g,intra}^{QM} - E_{vib,l,intra}^{QM} + E_{vib,l,inter}^{QM} - E_{vib,l,inter}^{CM}$$

Static Dielectric Constant

- The static dieletric constant ε(0) of a medium is determined by the magnitude and density of the molecular dipole moments and the extent to which the directions of the dipole moments are correlated.

$$\varepsilon(0) = 1 + 4\pi \frac{\langle M^2 \rangle}{3\langle V \rangle k_B \langle T \rangle}$$

$$\langle M^2 \rangle = \langle M.M \rangle - \langle M \rangle . \langle M \rangle = \langle M_x^2 + M_y^2 + M_z^2 \rangle - \left(\langle M_x^2 \rangle + \langle M_y^2 \rangle + \langle M_z^2 \rangle \right)$$

M is the total system dipole moment

- Polarizability correction

$$\varepsilon_{pol}^{corr} = 4\pi N \alpha / \langle V \rangle$$

Diffusion Coefficient

- Einstein–Smoluchowski Relation

$$D = \lim_{t\to\infty} \frac{1}{6t} \left\langle \left| r(t) - r(0) \right|^2 \right\rangle$$

- Self-diffusion

  TIP3P - 50

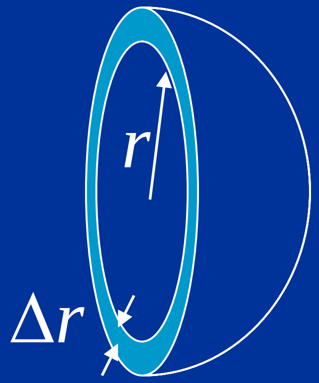
  CH3OH – NMA -

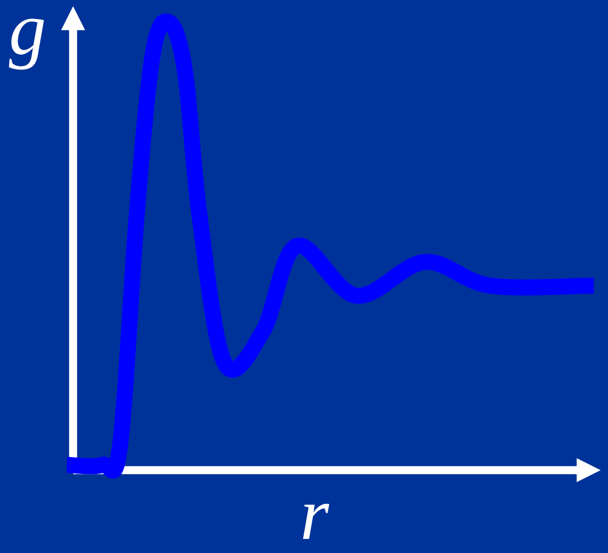
- Diffusion of binary liquid mixtures

Radial Distribution Function

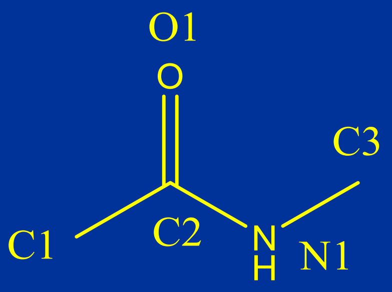
A radial distribution function measures the probability of finding a particle as a function of distance from a given particle.

$$g(r, \Delta r) \propto \langle \underbrace{N(r, \Delta r)}_{\text{Number of particles between r and $r + \Delta r$ from the given particle}} \rangle_M / \underbrace{4\pi r^2 \Delta r}_{\text{The volume of a spherical shell with thickness Δr}}$$



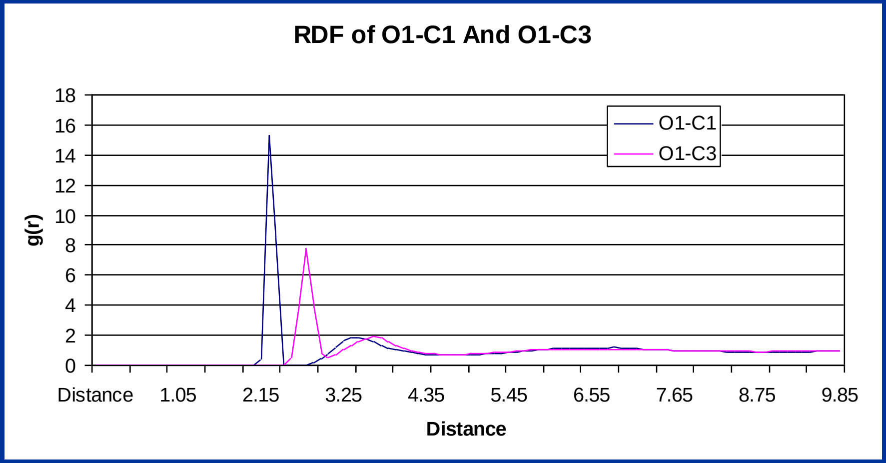


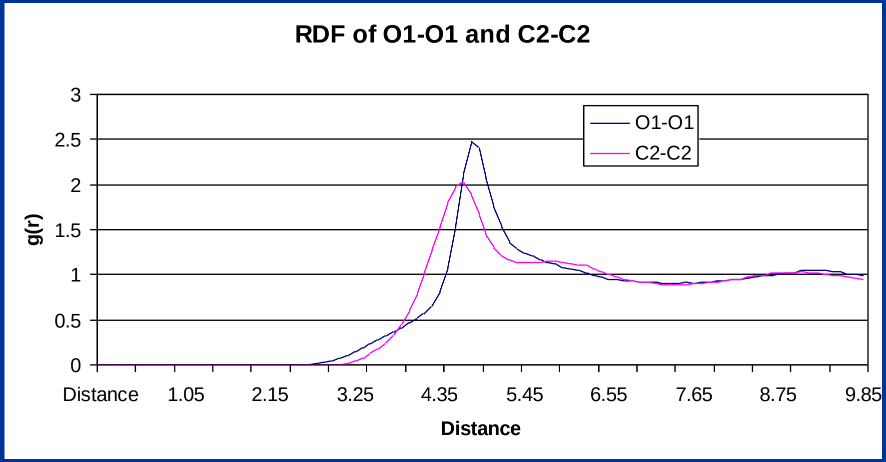
RDF of NMA



N-methyl acetic amide

```
#
trajin j14_md8.trj.gz
#
radial rad1 0.1 10 @O1 @O1
radial rad2 0.1 10 @C2 @C2
radial rad3 0.1 10 @O1 @C1
radial rad4 0.1 10 @O1 @C3
```





Correlation Function

A correlation function measures the relationship between two variables:

$$C_{xy} = \langle\, x\, y \,\rangle_M \,/\, (\langle\, x^2 \,\rangle_M \,\langle\, y^2 \,\rangle_M)^{\,½}$$

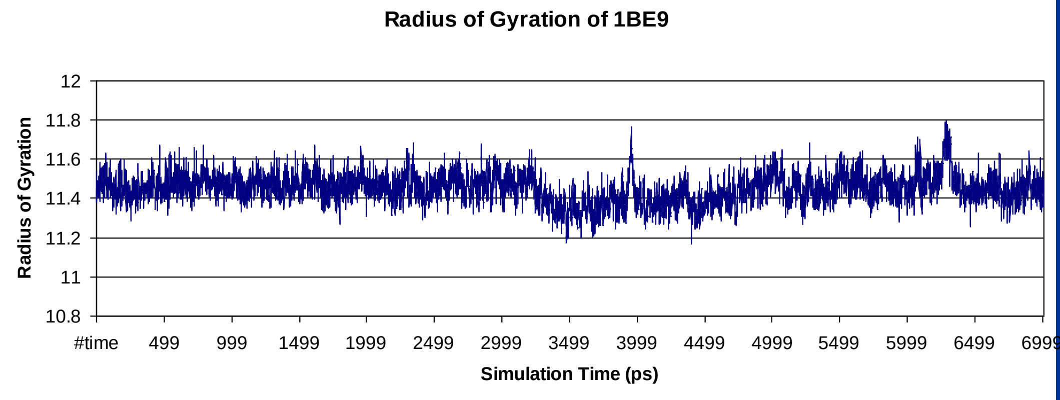
- If x (or y) fluctuate about a non-zero mean value, replace x (or y) in the above equation by $x - \langle\, x \,\rangle_M$ (or $y - \langle\, y \,\rangle_M$).
- If $x = y$, C_{xx} is called an auto-correlation function.

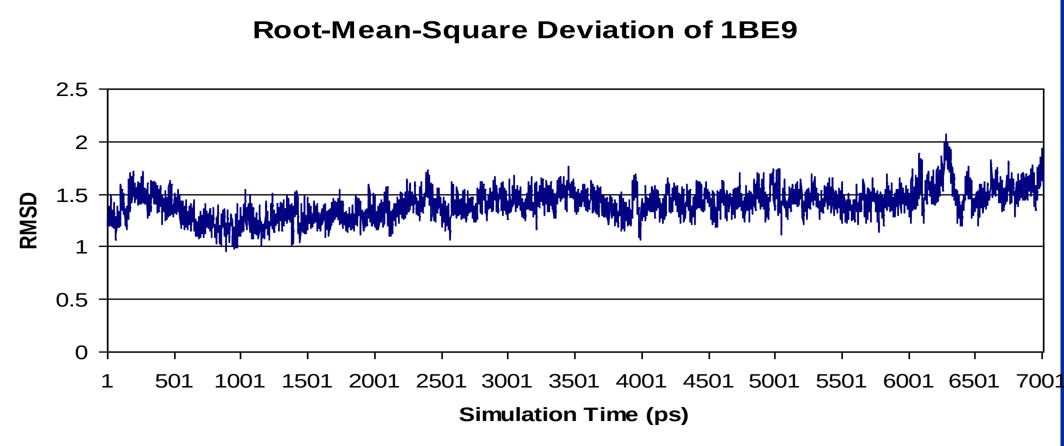
$$C_{xy} = \begin{cases} 1 & \text{completely correlated} \\ 0 & \text{independent} \\ -1 & \text{completely (anti-)correlated} \end{cases}$$

Radius of Gyration

Radius of gyration is used to describe the dimension of a molecule

$$R_g^2 = \frac{1}{N}\sum_{i=1}^{N}(r_i - r_{mean})^2$$





Advanced Topics

Potential cut-offs

Bonded interactions: local, therefore O(N),
where N is the number of atoms in the molecule considered)

Non-bonded interactions: involve all pairs of
Atoms, therefore O(N^2)

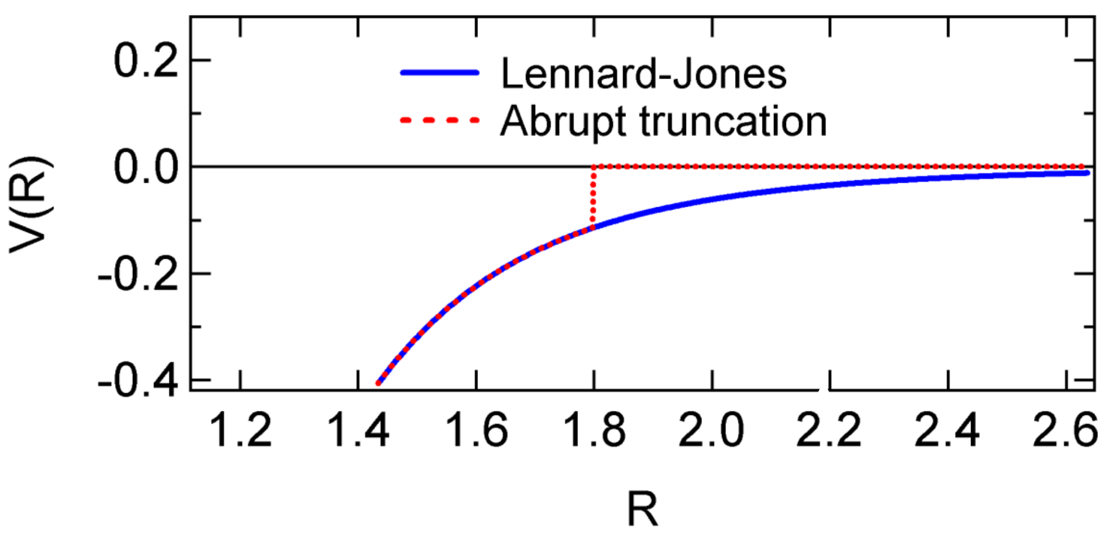
$$U_{NB} = \sum_{i,j\,nonbonded} \varepsilon_{ij}\left[\left(\frac{R_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{i,j\,nonbonded} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}}$$

Reducing the computing time: use of cut-off in U_{NB}
The cutoff distance may be no greater than ½ L (L= box length)

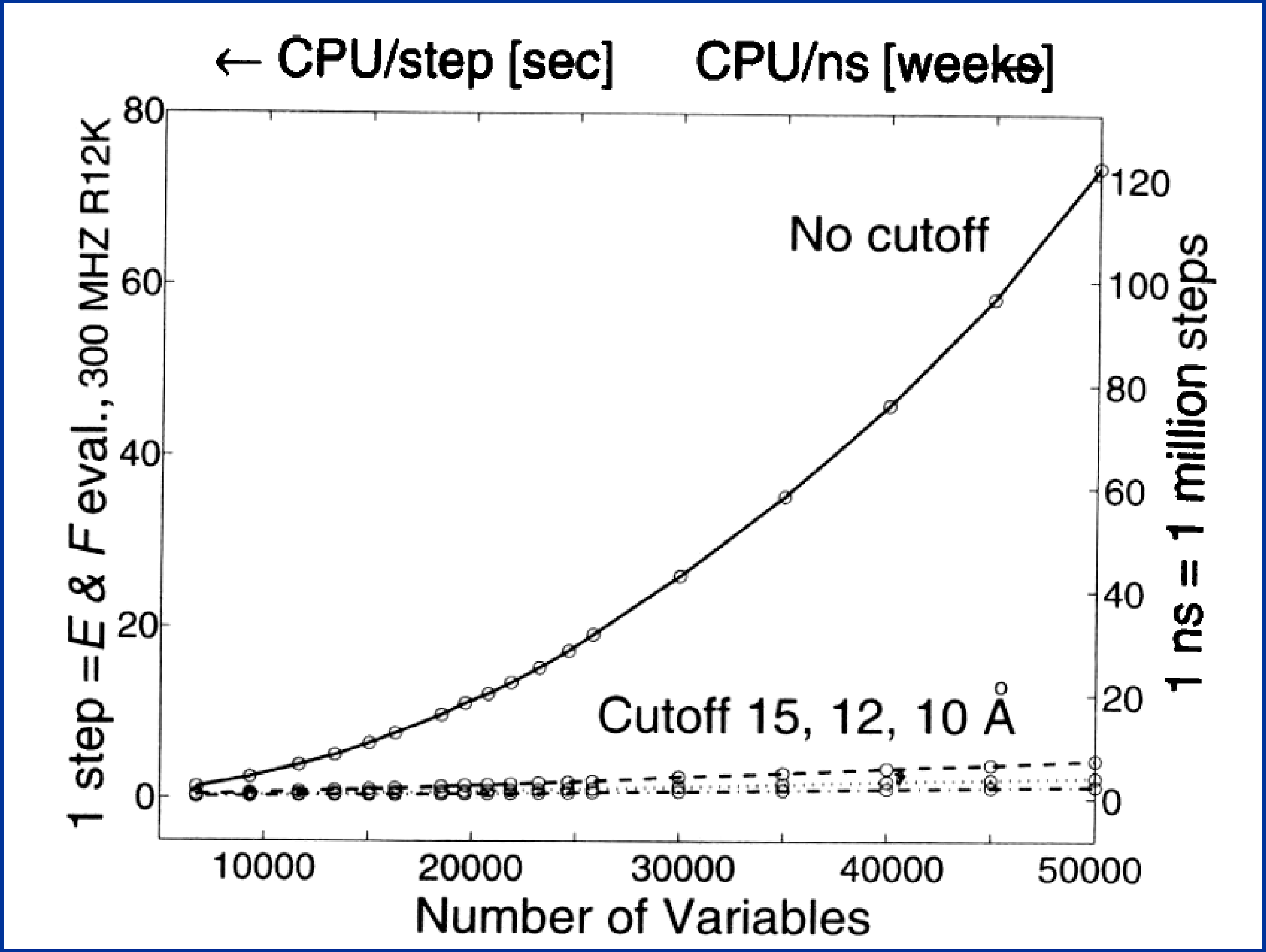
Potential truncation

common approach:
cut-off the at a fixed value R_{cut}

problem: discontinuity in energy and force
possibility of large errors



Speed-up



Tamar Schlick, "Molecular Modeling and Simulation", Springer

Cutoff schemes for faster energy computation

$$U_{NB}=\sum_{i,j}\omega_{ij}S\left(r_{ij}\right)\varepsilon_{ij}\left[\left(\frac{R_{ij}}{r_{ij}}\right)^{12}-2\left(\frac{R_{ij}}{r_{ij}}\right)^{6}\right]+\sum_{i,j}\omega_{ij}S\left(r_{ij}\right)\frac{q_iq_j}{4\pi\varepsilon_0\varepsilon r_{ij}}$$

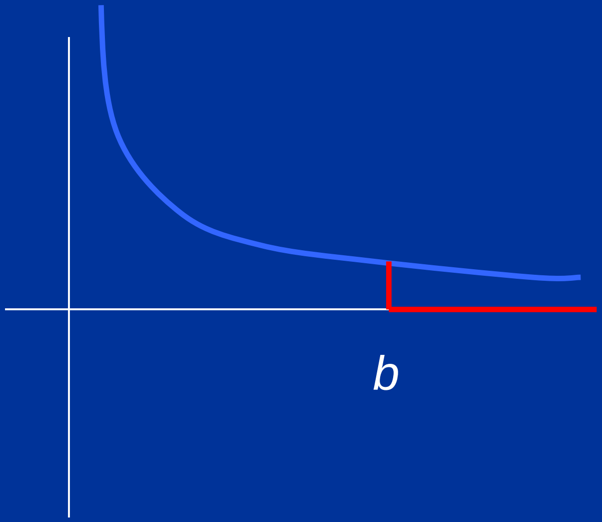
- ω_{ij} : weights *(0< ω_{ij} <1). Can be used to exclude bonded terms, or to scale some interactions (usually 1-4)*

- *S(r) : cut-off function.*

Three types:

1) Truncation:

$$S(r)=\begin{cases}1 & r<b\\ 0 & r\geq b\end{cases}$$

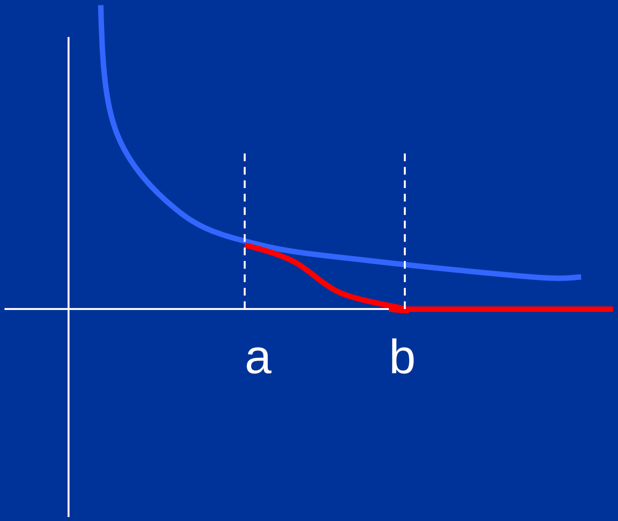


Cutoff schemes for faster energy computation

2. Switching

$$S(r) = \begin{cases} 1 & r < a \\ 1 + y(r)^2\left[2y(r) - 3\right] & a \le r \le b \\ 0 & r > b \end{cases}$$

with $$y(r) = \frac{r^2 - a^2}{b^2 - a^2}$$

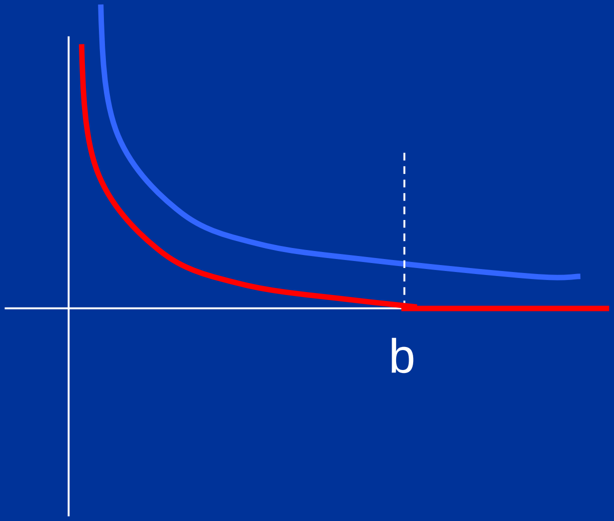


3. Shifting

$$S_1(r) = \left[1 - \left(\frac{r}{b}\right)^2\right]^2 \qquad r \le b$$

or

$$S_2(r) = \left[1 - \frac{r}{b}\right]^2 \qquad r \le b$$



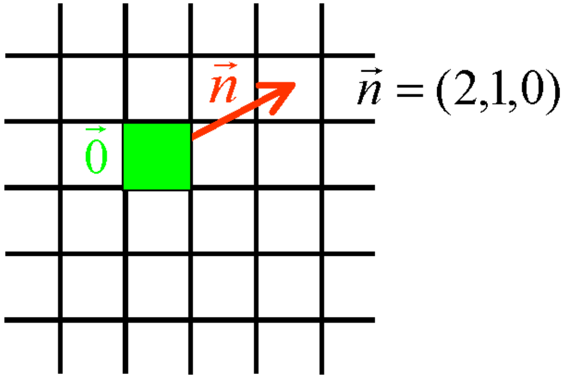
Ewald Sum Method

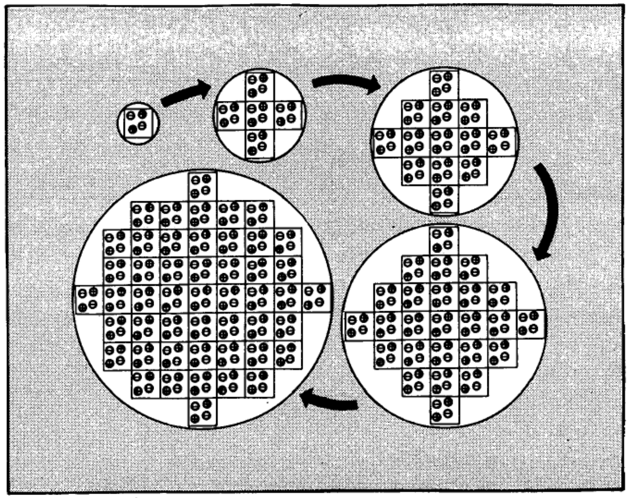
- calculate all interactions with infinite lattice implied by periodic BCs using ideas developed for treating ionic crystals
- consider charge version (also versions for dipoles & other multipoles)
- within primary box $V_{elect} = \frac{1}{4\pi\varepsilon_0}\sum_{j>i}^{N}\sum_{i=1}^{N}\frac{q_i q_j}{|\vec{r}_i - \vec{r}_j|}$
- for periodic lattice

$$V_{elect} = \frac{1}{4\pi\varepsilon_0}\frac{1}{2}\sum_{\vec{n}=\vec{0}}^{\infty}{}'\sum_{j=1}^{N}\sum_{i=1}^{N}\frac{q_i q_j}{|\vec{r}_i - \vec{r}_j + \underline{\vec{n}}|}$$

(exclude $j_0=i_0$)

lattice vector





- Σ is only conditionally convergent (must group terms to get sensible results); consider spherical arrays of cells

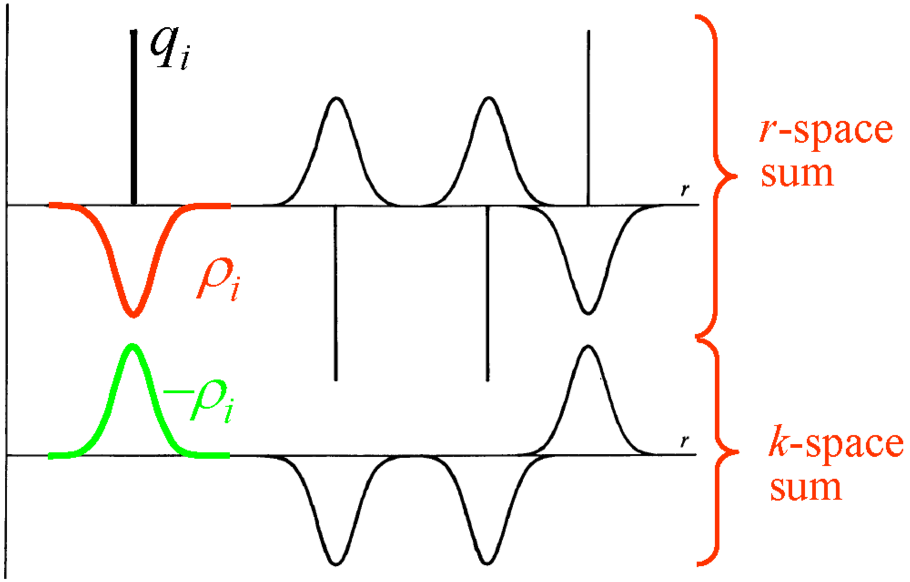
Ewald Sum Method

- in the Ewald method the direct sum is made short ranged by adding a neutralizing charge distribution around each charge:

$$\rho_i(\vec{r}) = -q_i \frac{\alpha^3}{\pi^{3/2}} \exp(-\alpha^2 r^2)$$

- $\rho_i(r)$ is then subtracted in a 2nd summation performed in reciprocal "*k*" space

$$\rho_i(\vec{k}) = \int d\vec{r} \exp(-\vec{k}\cdot\vec{r})\rho_i(\vec{r})$$



- the result is:

$$\mathcal{V} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\begin{cases} \displaystyle\sum_{|\mathbf{n}|=0}^{\infty}{}' \frac{q_i q_j}{4\pi\varepsilon_0} \frac{\operatorname{erfc}(\alpha|\mathbf{r}_{ij}+\mathbf{n}|)}{|\mathbf{r}_{ij}+\mathbf{n}|} \\ \displaystyle + \sum_{k\neq 0} \frac{1}{\pi L^3} \frac{q_i q_j}{4\pi\varepsilon_0} \frac{4\pi^2}{k^2} \exp\left(-\frac{k^2}{4\alpha^2}\right) \cos(\mathbf{k}\cdot\mathbf{r}_{ij}) \\ \displaystyle - \frac{\alpha}{\sqrt{\pi}} \sum_{k=1}^{N} \frac{q_k^2}{4\pi\varepsilon_0} + \frac{2\pi}{3L^3} \left| \sum_{k=1}^{N} \frac{q_k}{4\pi\varepsilon_0} r_k \right|^2 \end{cases}$$

r-space sum

k-space sum

self term and surface term (for array in vacuum)

L = cell length

Ewald Sum Method

Additional corrections:

- arises from a gaussian acting on its own site
- (self-energy correction)

- or from a surface in vacuum

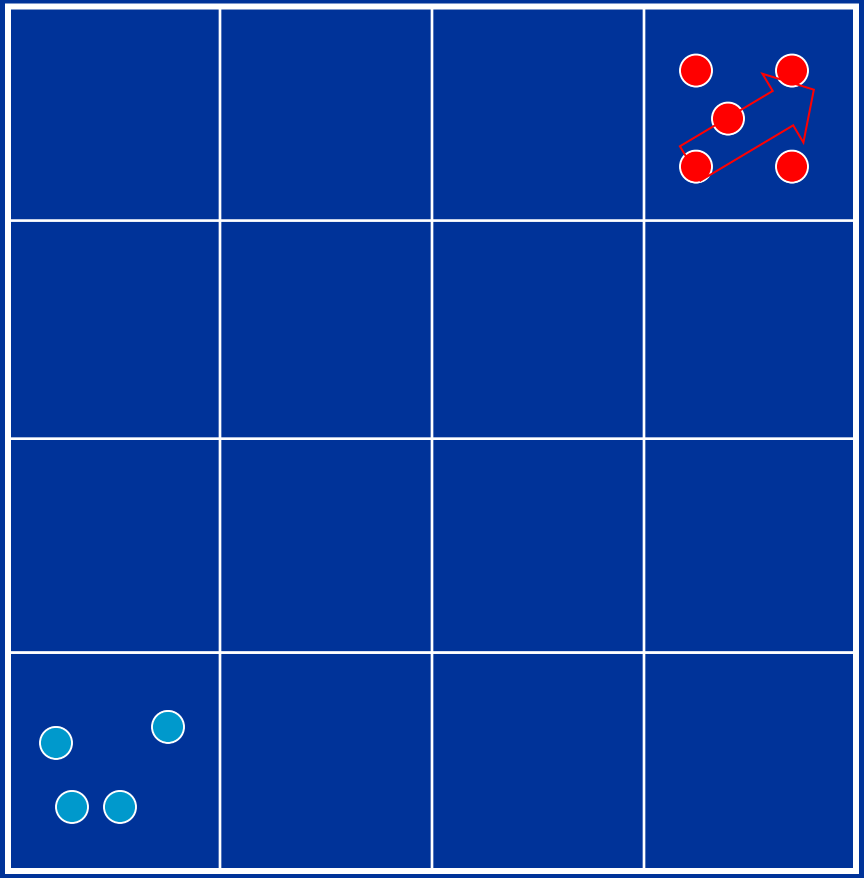
Particle Mesh Ewald

- Similar to Ewald method except that it uses FFT
- P3ME method has a very similar spirit with PME

1. Assigning charges onto grids
2. Use Fast Fourier Transform to speed up the k-space evaluation
3. Differentiation to determine forces on the grids
4. Interpolating the forces on the grid back to particles
5. Calculating the real-space potential as normal Ewald

Fast Multipole Method

- Represent charge distributions in a hierarchically structured multipole expansion
- Translate distant multipoles into local electric field
- Particles interact with local fields to count for the interactions from distant charges
- Short-range interactions are evaluated pairwise directly



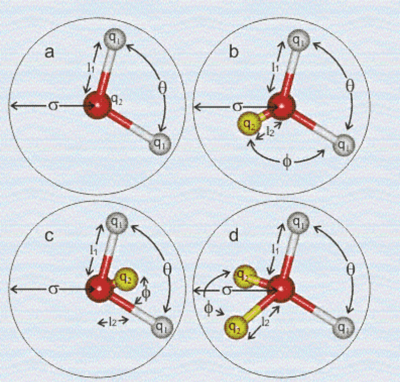
CPU: O(N)

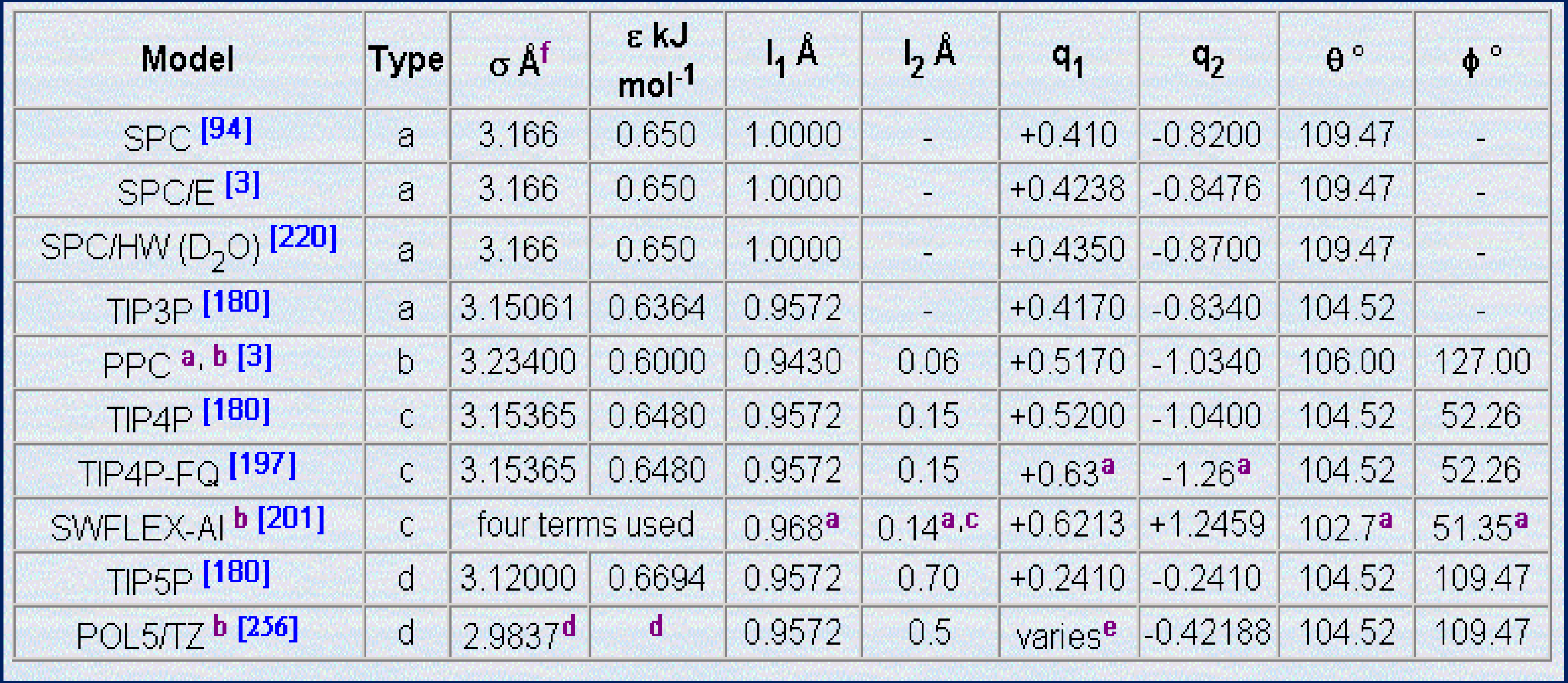
Multiple time step dynamics

- Reversible reference system propagation algorithm (r-RESPA)
  - Forces within a system classified into a number of groups according to how rapidly the force changes
  - Each group has its own time step, while maintaining accuracy and numerical stability

Water Model Parameters

- SPC, SPC/E (Berendsen)
- TIP3P, TIP4P, TIP5P (Jorgensen)
- TIP4P/FQ, POL5 (Berne)





Model	Type	σ Å[f]	ε kJ mol^{-1}	l_1 Å	l_2 Å	q_1	q_2	θ °	φ °
SPC [94]	a	3.166	0.650	1.0000	-	+0.410	-0.8200	109.47	-
SPC/E [3]	a	3.166	0.650	1.0000	-	+0.4238	-0.8476	109.47	-
SPC/HW (D_2O) [220]	a	3.166	0.650	1.0000	-	+0.4350	-0.8700	109.47	-
TIP3P [180]	a	3.15061	0.6364	0.9572	-	+0.4170	-0.8340	104.52	-
PPC [a], [b] [3]	b	3.23400	0.6000	0.9430	0.06	+0.5170	-1.0340	106.00	127.00
TIP4P [180]	c	3.15365	0.6480	0.9572	0.15	+0.5200	-1.0400	104.52	52.26
TIP4P-FQ [197]	c	3.15365	0.6480	0.9572	0.15	+0.63[a]	-1.26[a]	104.52	52.26
SWFLEX-AI [b] [201]	c	four terms used		0.968[a]	0.14[a,c]	+0.6213	+1.2459	102.7[a]	51.35[a]
TIP5P [180]	d	3.12000	0.6694	0.9572	0.70	+0.2410	-0.2410	104.52	109.47
POL5/TZ [b] [256]	d	2.9837[d]	[d]	0.9572	0.5	varies[e]	-0.42188	104.52	109.47

MD –Limitation

1. Quality of the force field

2. Size and Time – atomistic simulations can be performed only for systems of a few tenths of angstroms on the length scale and for a few nanoseconds on the time scale

3. Conformational freedom of the molecule – the number of possible conformations a molecule can adopt is enormous, growing exponentially with the number or rotatable bonds.

4. Only applicable to systems that have been parameterized

5. Connectivity of atoms cannot change during dynamics – no chemical reactions

AMBER

AMBER Force Field Development

1. A Brief Introduction on AMBER
2. General AMBER Force Field (GAFF)
3. AMBER Polarizable force field
4. Molecular Mechanical Toolkits

ambermd.org

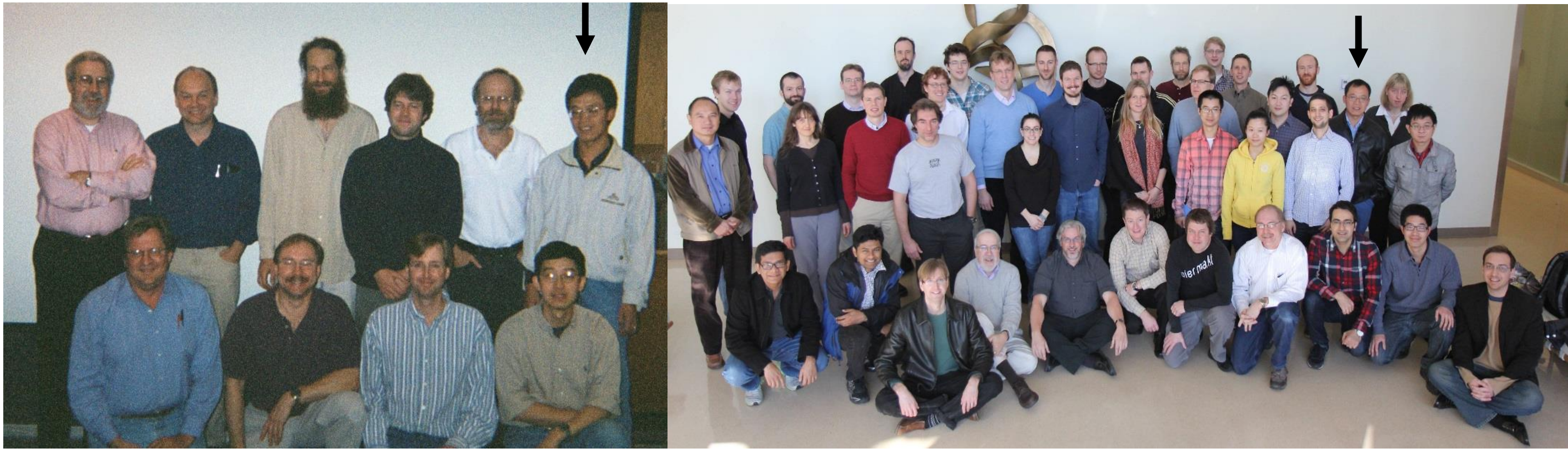
AMBER
(<u>A</u>ssisted <u>M</u>odel-<u>B</u>uilding with <u>E</u>nergy <u>R</u>efinement)

Founder: Peter A. Kollman at UCSF (1944 - 2001)

Current Leader: David A. Case at Rutgers University

AMBER Facts:

- One of the two major molecular simulation packages
- The world's fastest MD package using GPU
- Extensively applied both in academia and industry



2002

2014

Molecular Mechanical Toolkits

1. **Antechamber**

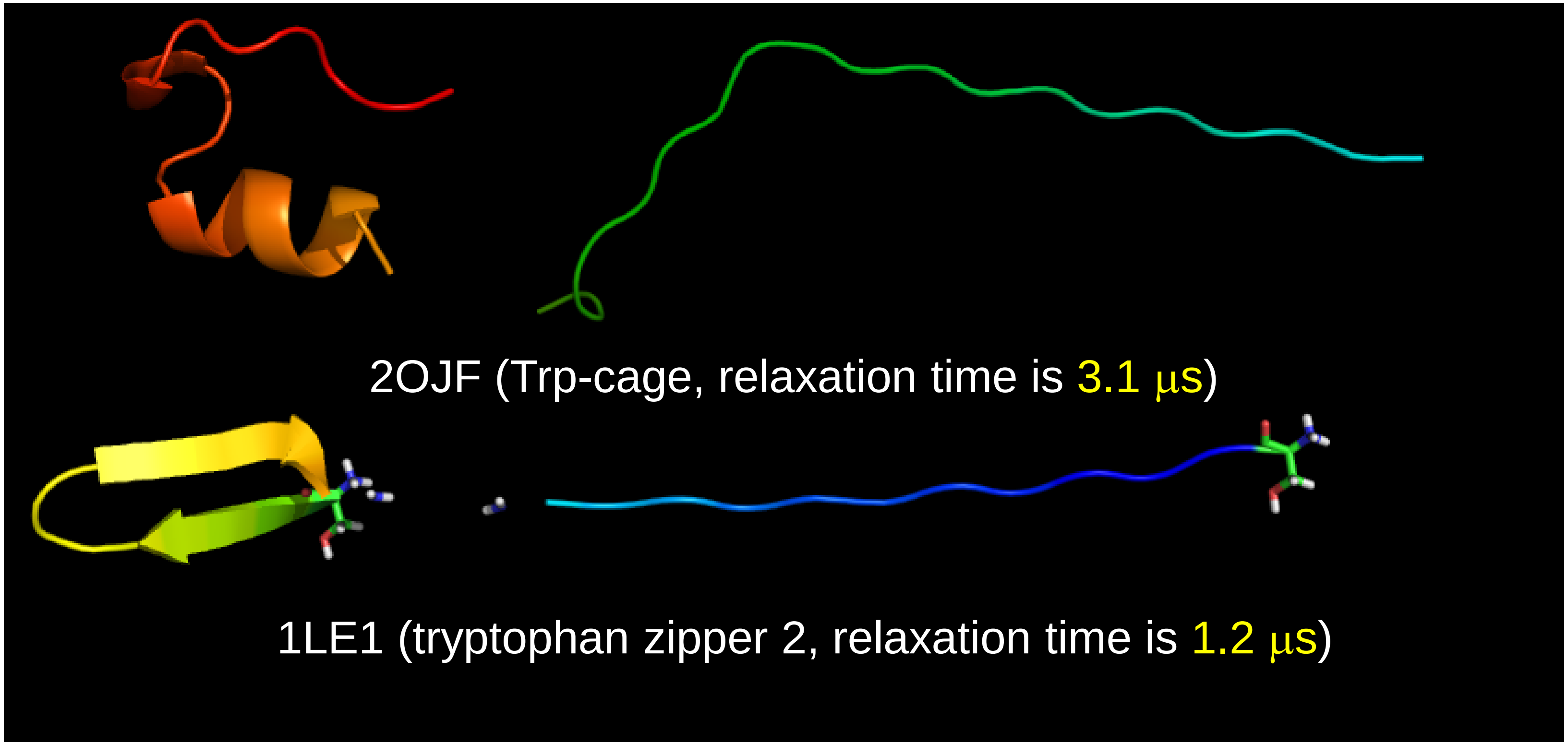
   Wang et al. J. Mol. Graphics & Modeling, 25, 247-260

   - First tool in the field to automatically generate models for arbitrary organic molecules
   - Extensively applied in structure-based drug design
   - Extensively used among and outside the AMBER community (it is able to generate MMFF models in CHARMM/NAMD formats)

2. **Online Tools**
   - Generate force field parameters for arbitrary organic molecules from various inputs
   - Database search
   - Calculate binding free energy with MM-GBSA

https://mulan.swmed.edu/mmfft

Protein Folding From Extended Conformations



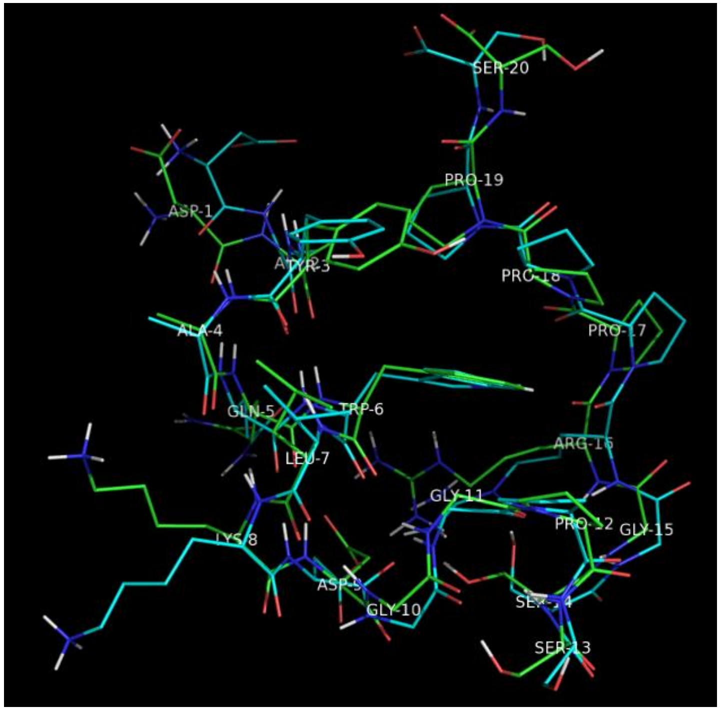
2OJF (Trp-cage, relaxation time is 3.1 μs)

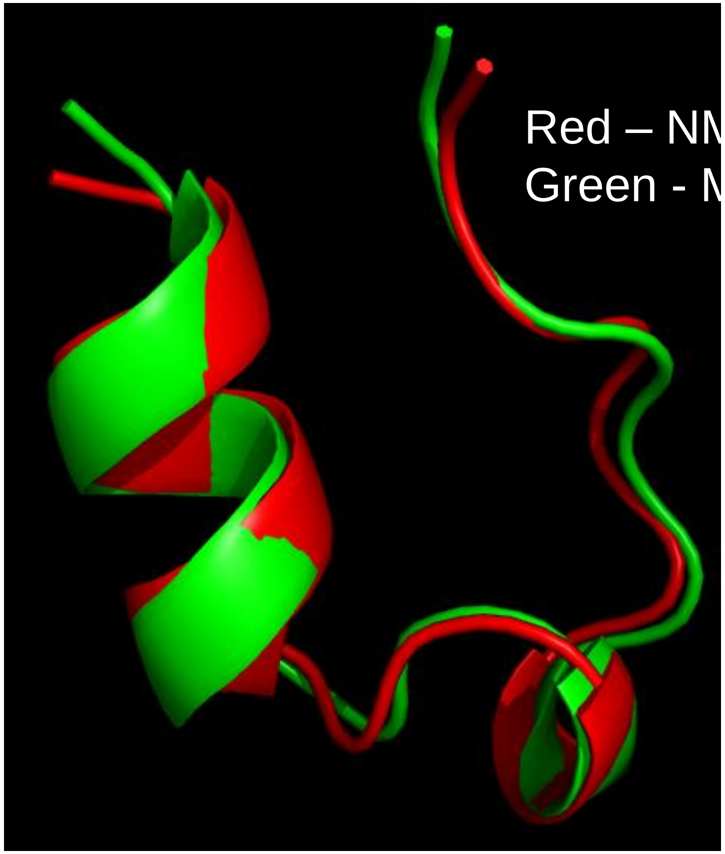
1LE1 (tryptophan zipper 2, relaxation time is 1.2 μs)

tryptophan zipper 2

2. Parm99 fails to fold Trp-cage at room temperature (Trp-cage can be folded with Parm99 at 325 K)
3. Starting from fully extended conformation
4. GBMD at 298 K

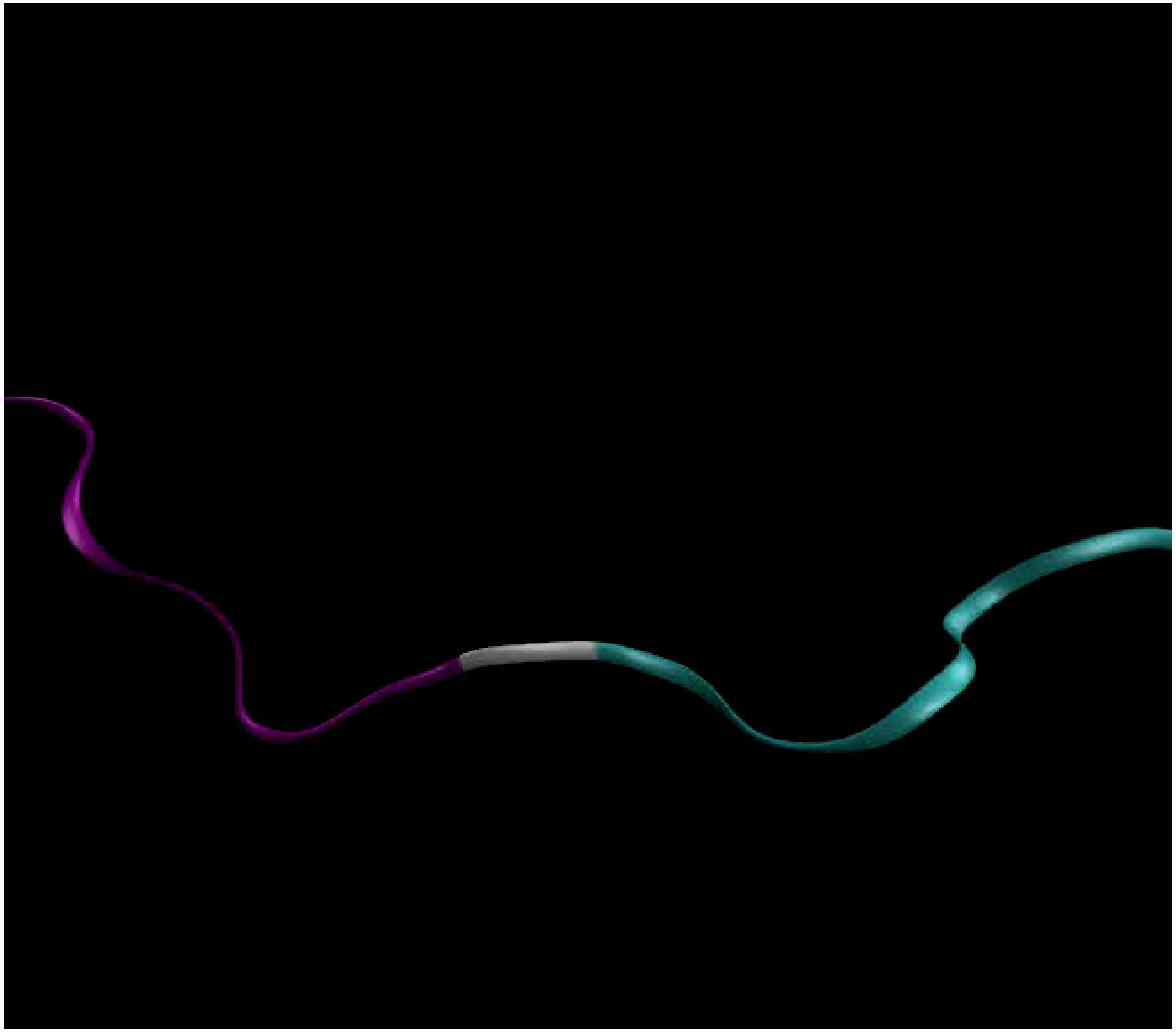
Trp-cage Is Folded within 200 Nanoseconds with Parm99-Mod



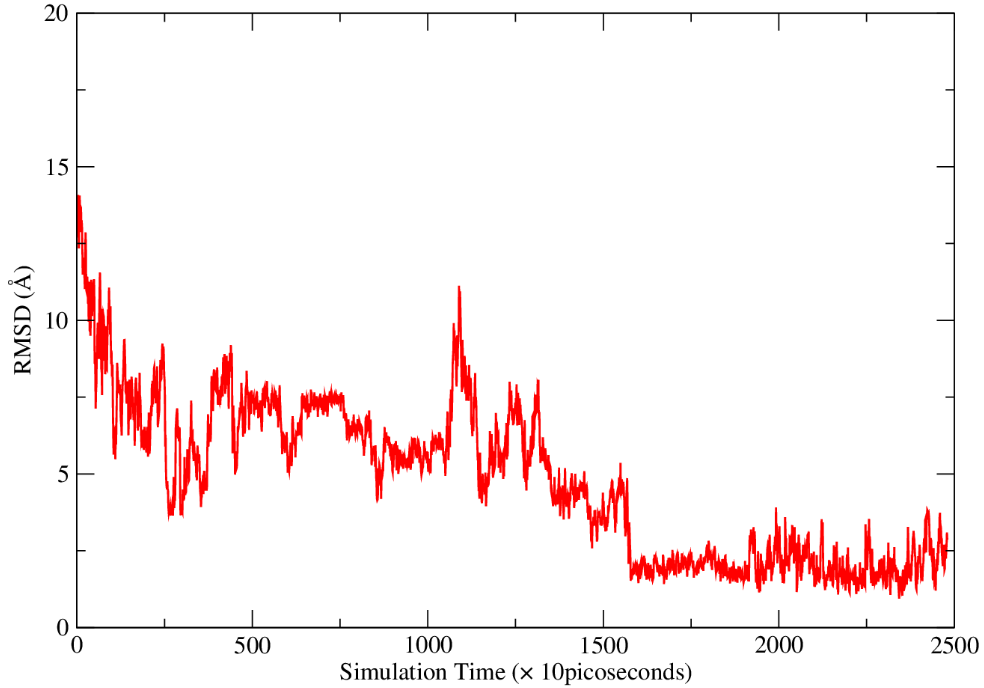

Red – NM
Green - M

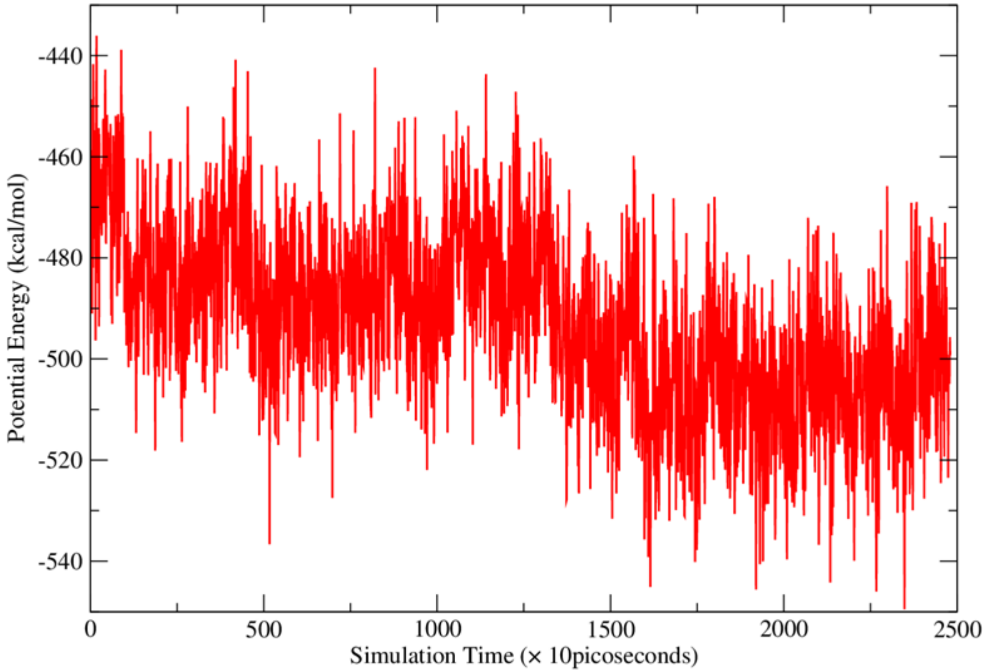
Main chain RMSD is 0.80 Å
Simmerling et al’s folding achieved an RMSD of 1.1 Å using Parm99 at 325 K (JACS, 124, 11258-11259)

In silico Folding of Trp-cage

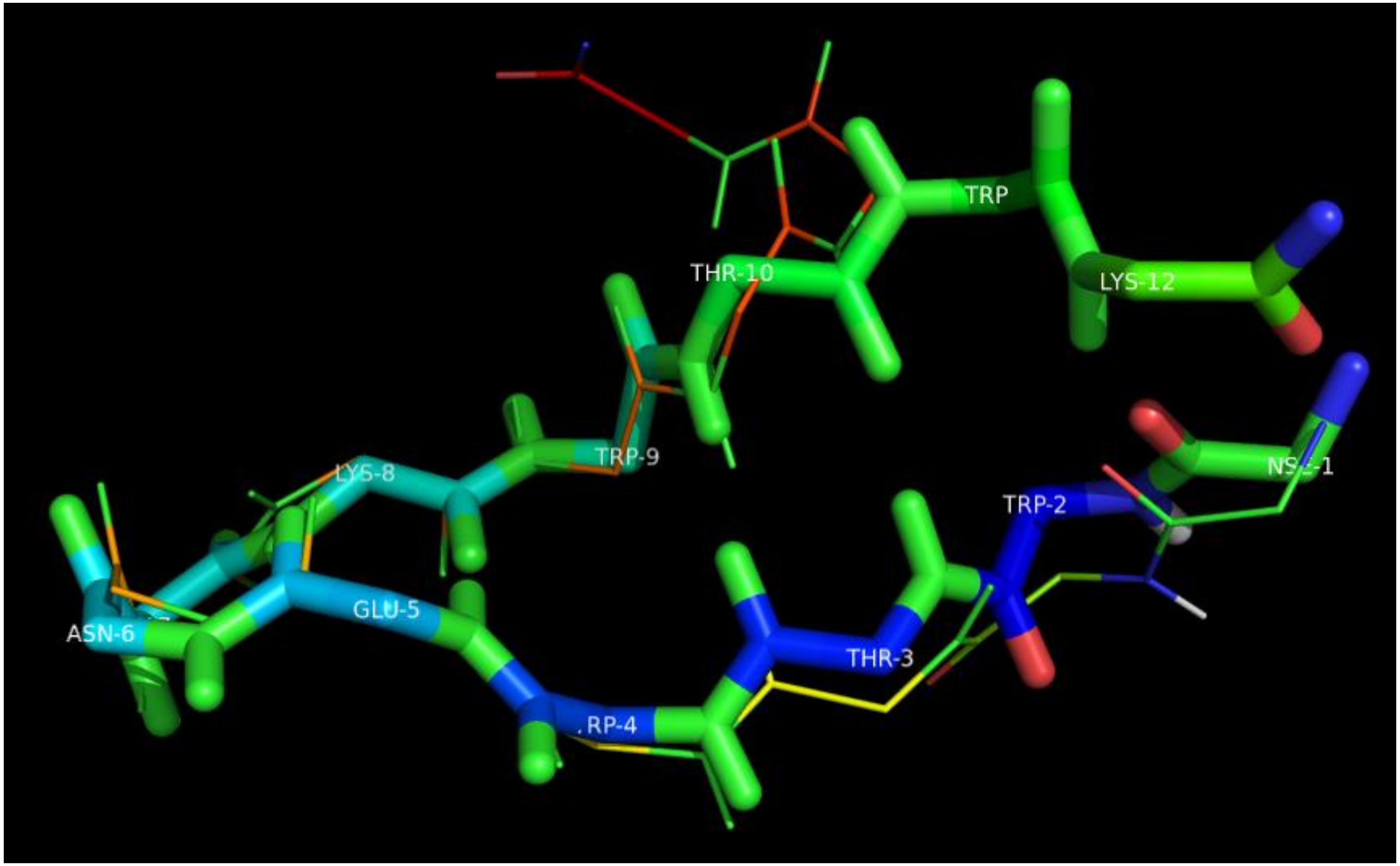


Convergence of MD Trp-cage Structures to a NMR Structure





Tryptophan Zipper 2 Is Folded Within 100 Nanoseconds with Parm99-Mod

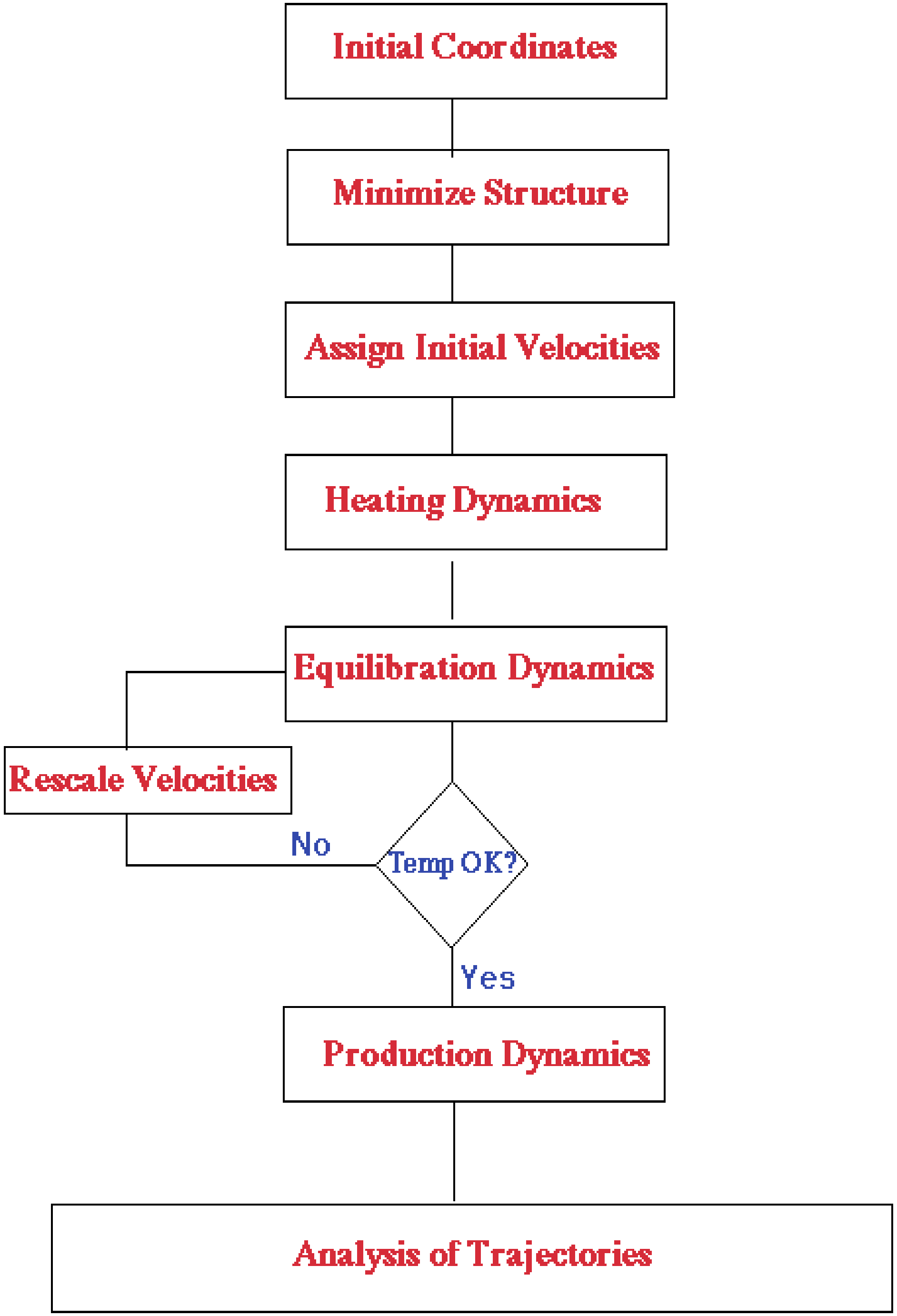


Main chain RMSD is 0.79 Å.

How To Generate System Topologies

1. **Residue topology files**
2. **Force field parameter files**
3. **Additional force field parameter files**
4. **Topology files**
5. **LEAP (xleap and tleap)**
6. **Basic procedure**
    1) Source leaprc.ffXX
    2) Load/generate structure file
    3) Add counter ions
    4) Add water box
    5) Save topology files

Detailed MD Scheme



Graphics Antechamber

https://Mulan.swmed.edu/mmfft

Running MD Simulations

1. **Programs**

   sander pmemd pmemd.MPI pmemd.cuda

2. **Input file for MD simulations**
3. **Running programs**

Analyze MD Snapshots

1. **Programs**

   ptraj, cpptraj

2. **Input file for ptraj and cpptraj**

Replay MD Trajectories With VMD