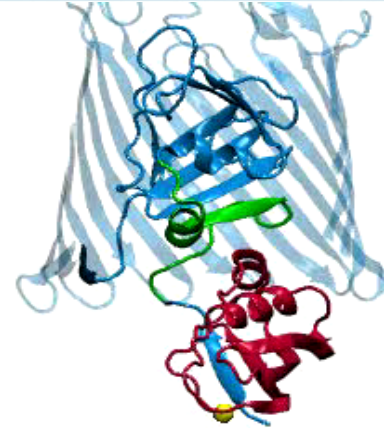


National Institute of Allergy and Infectious Diseases

Molecular Modeling
ACE-Mali/USTTB

Advanced molecular dynamics

January 5, 2016



NIAID



National Institute of
Allergy and
Infectious Diseases

Michael Dolan, Ph.D.

Bioinformatics and Computational Biosciences Branch

Office of Cyber Infrastructure and Computational Biology

Schedule



14h00 – 16h00, Tuesdays & Thursdays

Week	Date	Course	Instructor
2	Nov. 10	Molecular Properties	Darrell Hurt
	Nov. 12	Molecular Surfaces & Solvation: Part 1	Darrell Hurt
3	Nov. 17	Electrostatic Potential	Darrell Hurt
	Nov. 19	Molecular Surfaces & Solvation: Part 2	Mike Dolan
4	Nov. 24	Electron Microscopy & Volumetric Data	Amit Roy
	Nov. 26	No NIAID instruction – U.S. holiday	---
5	Dec. 1	Crystallography	Darrell Hurt
	Dec. 3	NMR	Phil Cruz
6	Dec. 8	Molecular Mechanics (force fields, charges, protonation, minimization): Part 1	Amit Roy
	Dec. 10	Molecular Mechanics (force fields, charges, protonation, minimization): Part 2	Amit Roy
7	Dec. 15	Sequence to Structure: Part 1	Conrad Shyu
	Dec. 17	Sequence to Structure: Part 2	Amit Roy
8	Dec. 21	Sequence to Structure: Part 3	Amit Roy
	Dec. 24	VMD Visualization	Conrad Shyu
9	Dec. 29	Movies in VMD	Conrad Shyu
	Dec. 31	Molecular Dynamics: Part 1	Mike Dolan
10	Jan. 5	Molecular Dynamics: Part 2	Mike Dolan
	Jan. 7	Ligand Parameterization	Phil Cruz
11	Jan. 12	Drug Design: Part 1	Phil Cruz
	Jan. 14	Drug Design: Part 2	Phil Cruz

NIAID

Important dates for molecular modeling course



Molecular modeling course homework deadline: **Friday, Jan. 15**

Molecular modeling final exam: **Friday, Jan. 22**

Outline



Classical forcefields: Topologies and parameters defined

- Creating novel forcefield parameters

Mechanics of classical MD

Solvation models

Electrostatics models

Thermostats and barostats

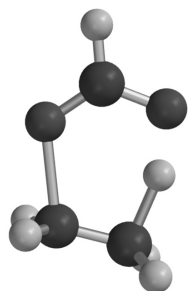
Hands-on exercise

Homework assignment

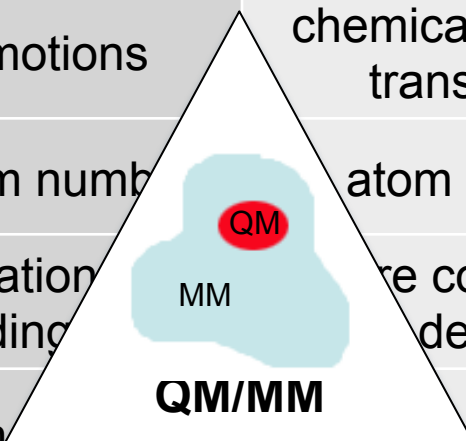
Molecular dynamics (MD)



A computer technique where atoms or molecules are approximated by physics and allowed to interact over a period of time giving a *view of motion*



classical (molecular mechanical) MD	quantum MD
ball and springs	electron density
$F = ma$	$i\hbar \frac{\partial}{\partial t} \Psi(\mathbf{r}, t) = -\frac{\hbar^2}{2m} \nabla^2 \Psi(\mathbf{r}, t) + V(\mathbf{r}) \Psi(\mathbf{r}, t)$
molecular motions	chemical reactions and transition states
unlimited atom number	atom number limit
less computation demanding	more computationally demanding
approximation of reality	closer to reality



MD



A word about quantum mechanical MD...

Why is there an atom limit? Limit is due to computational power.

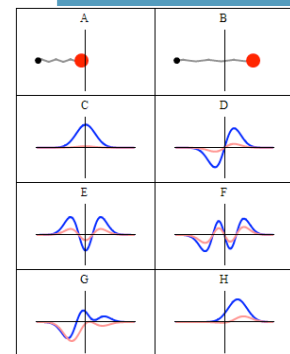
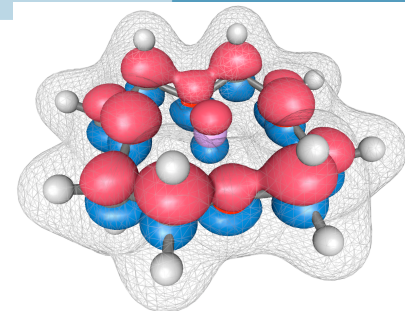
Why so computationally demanding? Takes time to solve the electronic Schrödinger equation.

Schrödinger equation – linear, partial differential equation that describes how the quantum state of some physical system changes with time – analog to Newton's 2nd Law. (The Hartree-Fock method is one way to solve this equation.) The “solutions” are functions which describe wave-like motion for a group of particles.

$$E\Psi(\mathbf{r}) = \frac{-\hbar^2}{2m}\nabla^2\Psi(\mathbf{r}) + V(\mathbf{r})\Psi(\mathbf{r})$$

wavefunction - probability (density) amplitude describing the quantum state of a particle (electrons, nuclei) and how it behaves as a function of space and time

basis sets - a set of functions which are combined to build molecular orbitals (STO-3G, 3-21G, 6-31G*, 6-311G*, cc-pVDZ ...)



MPG



semi-empirical models - simplest of the quantum chemical schemes,
– use Hartree-Fock method, but make approximations and obtain some parameters from use empirical data (like pre-calculated orbitals). (PM3, MNDO, AM1, RM1, PM3, and PM6)

Why is it “more real” than molecular mechanics?

- bond breaking and formation
- using molecular orbitals instead of balls and springs
- In the real world, electron clouds surrounding atoms are constantly shifting according to their environments; charges of atoms can be represented as dynamic and responsive (electronic polarization) instead of fixed

Spartan (Wavefunction), **Gaussian**, **GAMESS** (Gordon group, Iowa State Univ.)



“Classical” or “molecular mechanical” MD

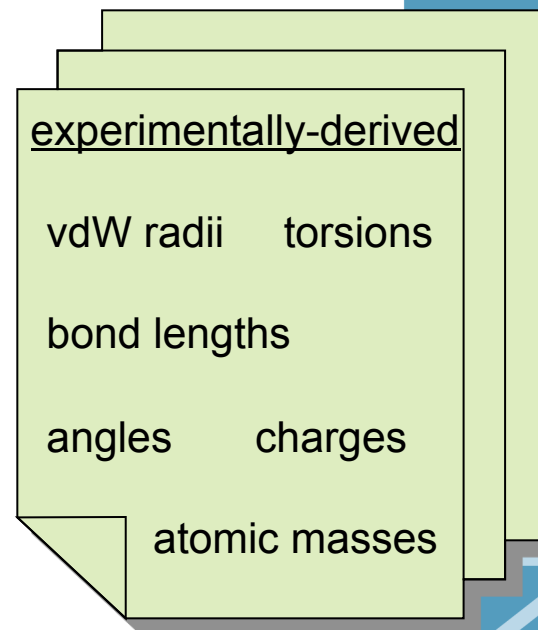
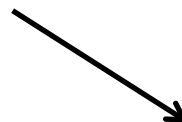
- represents a molecule mechanically using a “force field”
- force field – describes the energy of a configuration of atoms using an *equation* and *set of parameters*



$$U = \sum_{\text{bonds}} k_r (r - r_0)^2$$
$$+ \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2$$
$$+ \sum_{\text{dihedrals}} k_\phi [1 + \cos(n\phi + \phi_0)]$$
$$+ \sum_{\text{atom } i} \sum_{j \neq i} 4\epsilon_{i,j} \left[\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^6 \right]$$
$$+ \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon_0 r_{i,j}}$$

bonded

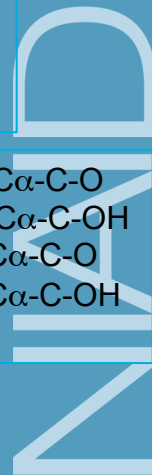
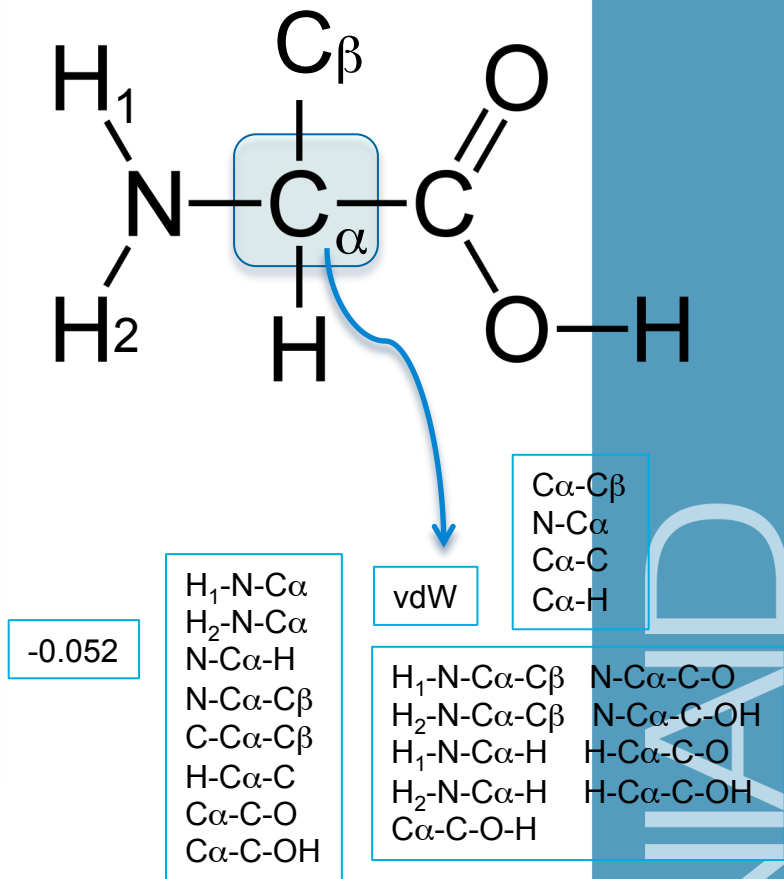
non-bonded





Force field parameters

- Geometries from real-world experiments (X-ray crystallography, NMR, spectroscopy)
- Charges from rigorous quantum-mechanical *ab initio* methods
- Algorithm assigns a type to each atom
- Nucleic acids, amino acids, common small molecules (ATP, heme, modified aa) already solved

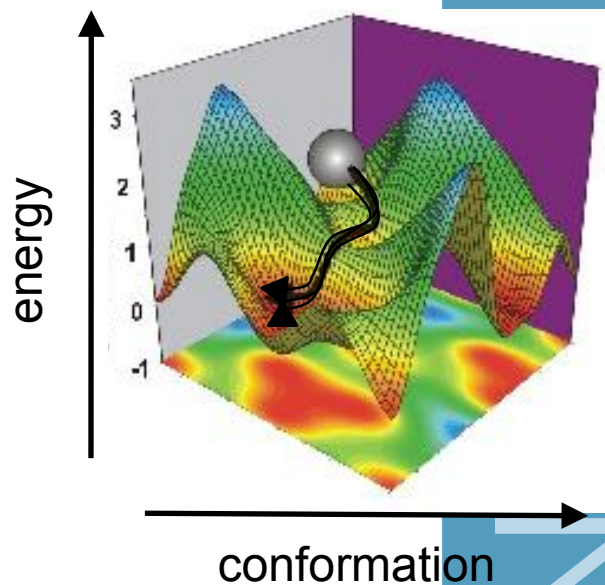




Potential energy surface

$$E = \sum_{\text{bonds}} K_b (r - r_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\phi [1 + \cos(n\phi - \delta)] + \sum_{\text{impropers}} K_\psi (\psi - \psi_0)^2$$
$$+ \sum_{i>j} \varepsilon \left[\left(\frac{r_m}{r} \right)^{12} - 2 \left(\frac{r_m}{r} \right)^6 \right] + \sum_{i>j} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r}$$

- all possible conformations of a molecule describes a [potential energy surface](#)
- represented as a curve (or multidimensional surface) with atomic positions as variables.



Some forcefields

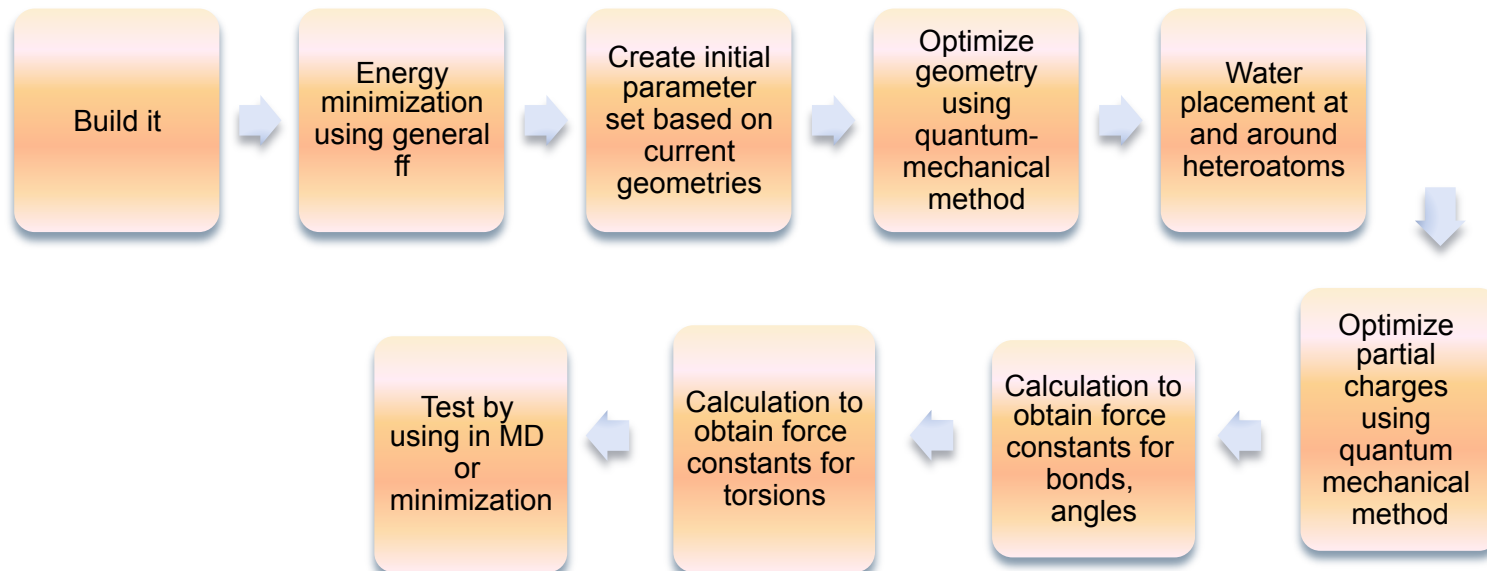


Name	Type	Application	Notes
Amber	First generation	proteins and nucleic acid	
CHARMM	First generation	general	
CVFF	First generation	general	
COSMOS-NMR	First generation	Inorganic and organic compounds	Bond Polarization Thoery (BPT) ; QM/MM
GROMACS	First generation		
GROMOS	First generation	general	
OPLS (OPLS-AA, OPLS-UA, OPLS-2001, OPLS-2005)	First generation		
ENZYMIK	First generation; polarizable ff based on induced dipoles	general	Empirical valence bond (EVB) method; semi-macroscopic PDL in MOLARIS
ECEPP/2	First generation	polypeptides	
QCFF/PI	First generation	General for conjugated molecules	
CFF	Second generation	organic compounds, includes force fields for polymers, metals, etc.	
MMFF	Second generation	general	
MM2, MM3, MM4	Second generation	general	
CFF/ind	Polarizable ff based on induced dipoles	general	

Name	Type	Application	Notes
PFF	Polarizable (based on point charges)		
DRF90	Polarizable (based on point charges)		
SP-basis Chem Potential Equal.	Polarizable (based on point charges)		
CHARMM	Polarizable (based on point charges)		
AMBER	Polarizable (based on point charges)		
SIBFA	Polarizable Force Fields based on distributed multipoles	Small molecules and flexible proteins; transition metals	procedure formulated and calibrated on the basis of ab initio supermolecule computations
AMOEBA	Polarizable Force Fields based on distributed multipoles		
ORIENT	Polarizable Force Fields based on distributed multipoles		
NEMO	Polarizable Force Fields based on distributed multipoles		
Gaussian Electrostatic Model (GEM)	Polarizable Force Fields based on density		
Polarizable procedure based on the Kim-Gordon approach	Polarizable Force Fields based on density		
ReaxFF	Reactive forcefields	atomistic-scale dynamical simulations of chemical reactions.	
EVB (empirical valence bond)	Reactive forcefields	Chemical reactions	EVB facilitates calculations of actual activation free energies in condensed phases and in enzymes.
RWFF	Reactive forcefields	simulate the bond formation/ breaking of water and acids.	reproduces the experimental data of neutron scattering experiments
COSMOS-NMR	Polarizable Force Fields based on Bond Polarization Theory (BPT)		Hybrid QM/MM force field



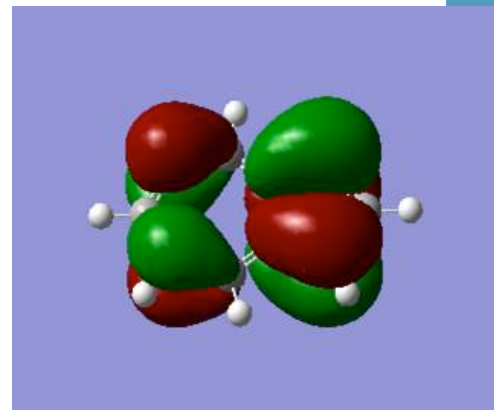
Creating novel ff parameters



QM programs: Gaussian, GAMESS, Spartan

basis sets – algorithms describing molecular orbitals

Example: 6-31G*



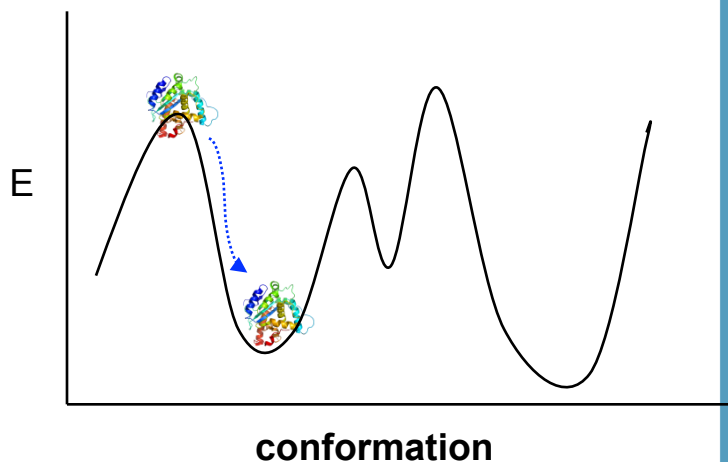
NIAAD



Minimization vs. MD

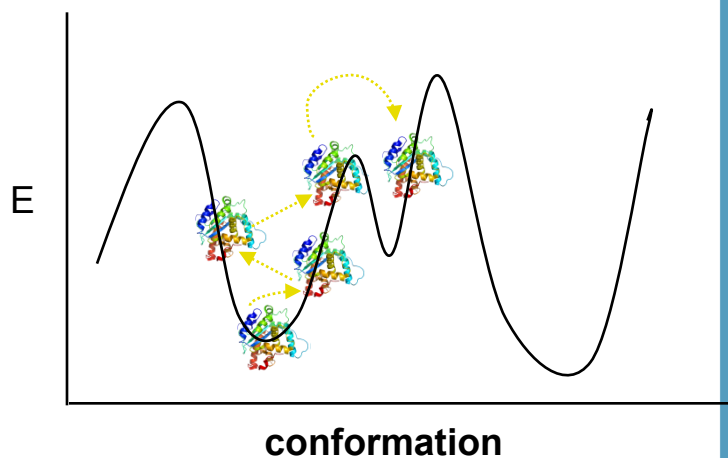
Minimization

Attempts to find the nearest, low energy conformation

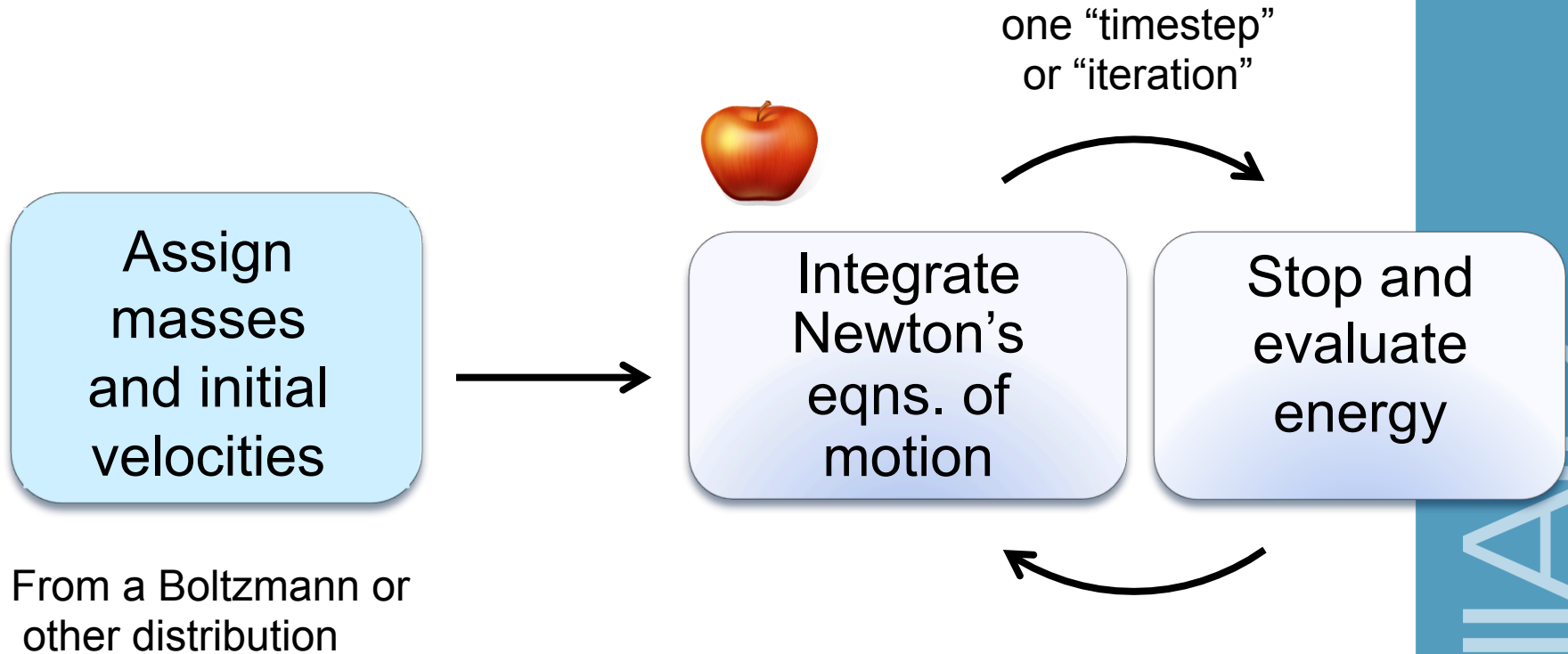


Molecular dynamics

Seeks to explore conformational space; can access multiple low energy conformations



Running a classical MD simulation





Mechanics of classical MD

Newton's 2nd Law of motion:



$$F_i = m_i a_i$$

$$v(t) = \frac{dx}{dt} \quad a(t) = \frac{dv}{dt}$$

Basic physics reminder

The force can also be expressed as the gradient of the potential energy:

$$F_i = -\nabla_i E \quad (\text{less force} = \text{less energy})$$

Combining these two equations gives:

$$-\frac{dE}{dx_i} = m_i \frac{d^2 x_i}{dt^2}$$

Newton's equation of motion can then relate the derivative of the potential energy to the changes in position as a function of time.

How do we calculate Δ in position?



- knowledge of atomic forces and masses is used to solve for the positions of each atom along a series of small time steps
- resulting series of snapshots of structural changes over time is called a *trajectory*
- atoms are assigned *initial velocities* that conform to the total kinetic energy of the system dictated by a desired simulation temperature
- force on an atom can be calculated from the change in energy between its current position and its position a small distance away

“time step” usually 1 fsec (10^{-15} sec) = fastest bond vibration

SHAKE – algorithm that fixes H-bonds allowing the use of a 2 fsec time step



Newton's 2nd Law: A simple example

$$F = ma = m \frac{dv}{dt} = m \frac{d^2x}{dt^2}$$

Take the simple case where acceleration is constant:

$$a = \frac{dv}{dt}$$

After integrating, an expression for velocity (v):

$$v = at + v_0$$

Integrate again to get position (x): $x = vt + x_0$

Substitute for velocity in above:

$$x = at^2 + v_0t + x_0$$

integration

We get the value of x at time t as a function of the acceleration, a , the initial position, x_0 , and the initial velocity, v_0

Summary



To calculate a trajectory, one only needs the initial positions of the atoms, an initial distribution of velocities and the acceleration, which is determined by the gradient of the potential energy function.

$$x = at^2 + v_0t + x_0$$



$$a = -\frac{1}{m} \frac{dE}{dx}$$

The acceleration is given as the derivative of the potential energy with respect to the position, x

The equations of motion are deterministic ... the positions and the velocities at time zero determine the positions and velocities at all other times.



Initial velocity assignment

- initial assignment of velocities are pulled from a Maxwell-Boltzmann or Gaussian distribution for a given temperature
- gives the probability that an atom i has a velocity $v(x)$ in the x direction at a given temperature
- velocities can be calculated for an input T using the relation:

$$T = \frac{1}{(3N)} \sum_{i=1}^N \frac{|p_i|^2}{2m_i}$$

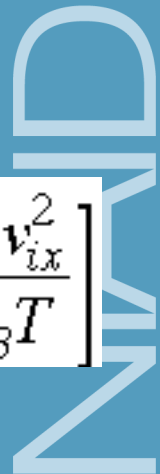
where N is the number of atoms

- velocities chosen randomly resulting in *no overall momentum*:

$$P = \sum_{i=1}^N m_i v_i = 0$$



$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T} \right)^{1/2} \exp \left[-\frac{1}{2} \frac{m_i v_{ix}^2}{k_B T} \right]$$



Integrators

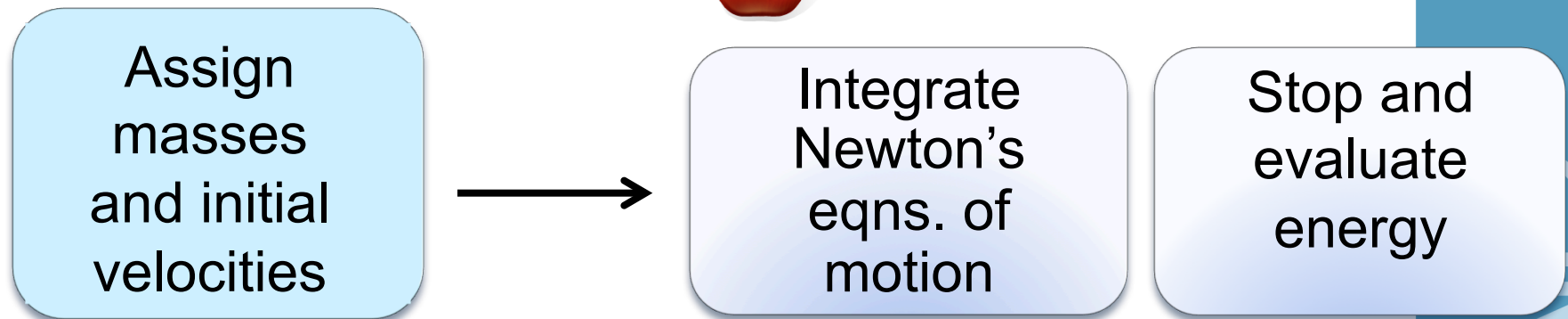


- potential energy is a function of the atomic positions ($3N$) of all the atoms in the system \longrightarrow a complicated function!
- no analytical solution to the equations of motion; must be solved numerically
- numerical algorithms have been developed for integrating the equations of motion (Taylor series expansions)

- * Verlet
- * Leap-frog
- * Velocity Verlet
- * Beeman's

1. The algorithm should conserve energy and momentum.
2. It should be computationally efficient
3. It should permit a long time step for integration.

Running a classical MD simulation



From a Boltzmann or other distribution



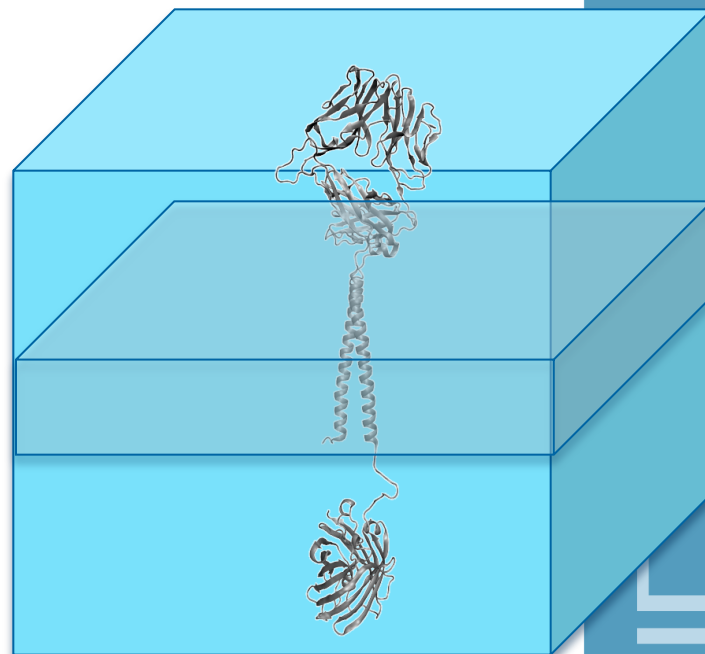
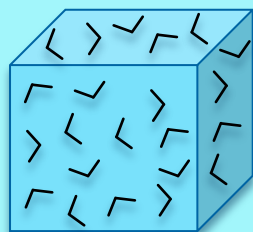
A typical classical MD experiment

Two approaches to solvation:

1. Implicit solvent



2. Explicit solvent



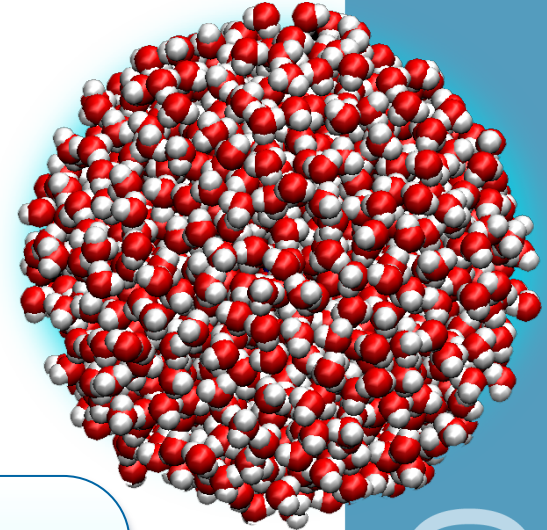
Production run

steps



Implicit solvation

- *approximates* solvent effects in MD systems by treating solvent as a “continuum”
- interactions between solvent and solute are described as a function of solute coordinates alone
- computationally more efficient than explicit solvent



Does not account for: *solvent degrees of freedom*
hydrophobic effect (entropy component)
viscosity
H-bonds with the solute

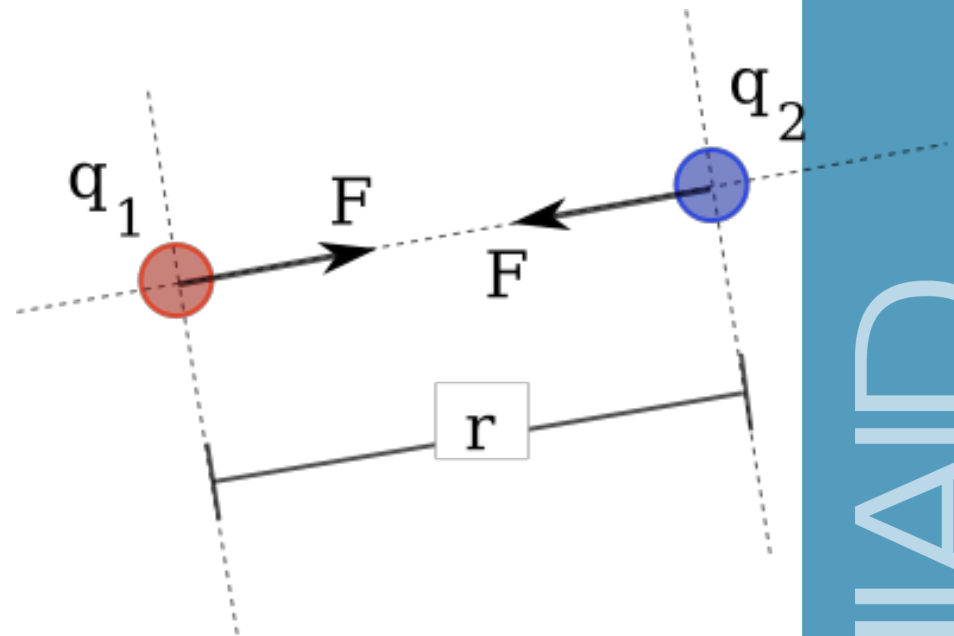


Distance-dependent dielectric

- simplest form of implicit solvation
- describes the magnitude of the electrostatic force between two charges

$$F = k_e \frac{q_1 q_2}{r^2}$$
$$k_e = \frac{1}{4\pi\epsilon_0}$$

Coulomb's constant



Generalized Born/Surface Area (GB/SA)



- most popular
- based on experimental linear relations between Gibbs free energy of transfer and the surface area of a solute molecule



Born

$$G_{\text{solv}} = G_{\text{cav}} + G_{\text{vdW}} + G_{\text{pol}}$$

linear function of the solvent-accessible surface areas of each atom

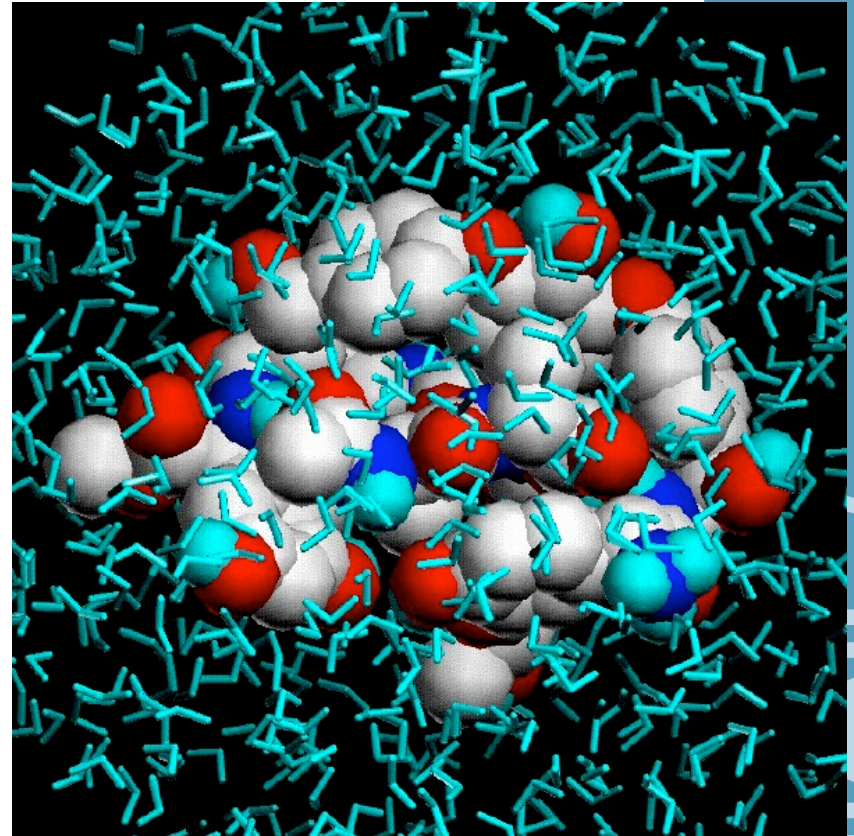
$$G_{\text{pol}} = -166.0 \left(1 - \frac{1}{\epsilon} \right) \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{\sqrt{r_{ij}^2 + b_i b_j} \exp\left(-\frac{r_{ij}^2}{4b_i b_j}\right)}$$

Generalized Born eqn.

Explicit solvation



- solvent molecules modeled using all-atoms
- counterions modeled explicitly
- potentially doubles system size, more computation
- includes H-bonding, hydrophobic effect, viscosity, solvent degrees of freedom



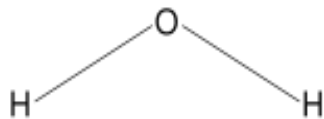
Explicit water models



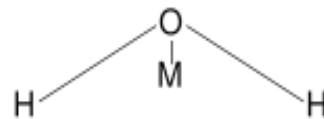
Types of models:

- rigid models (fixed atom positions)
- flexible models (atoms on “springs”)
- polarizable models (include explicit polarization term)

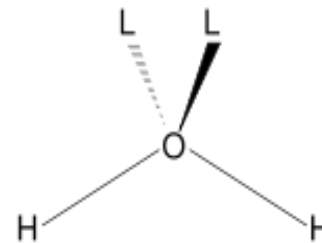
SPC	(1981)
SPC/E	(1987)
TIP3P	(1983)
TIP4P	(1983)
TIP5P	(2000)
TIP6P	(2003)



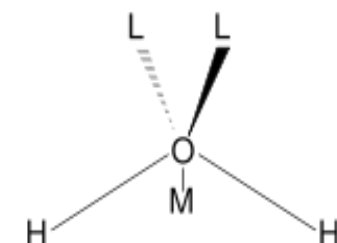
3-site



4-site

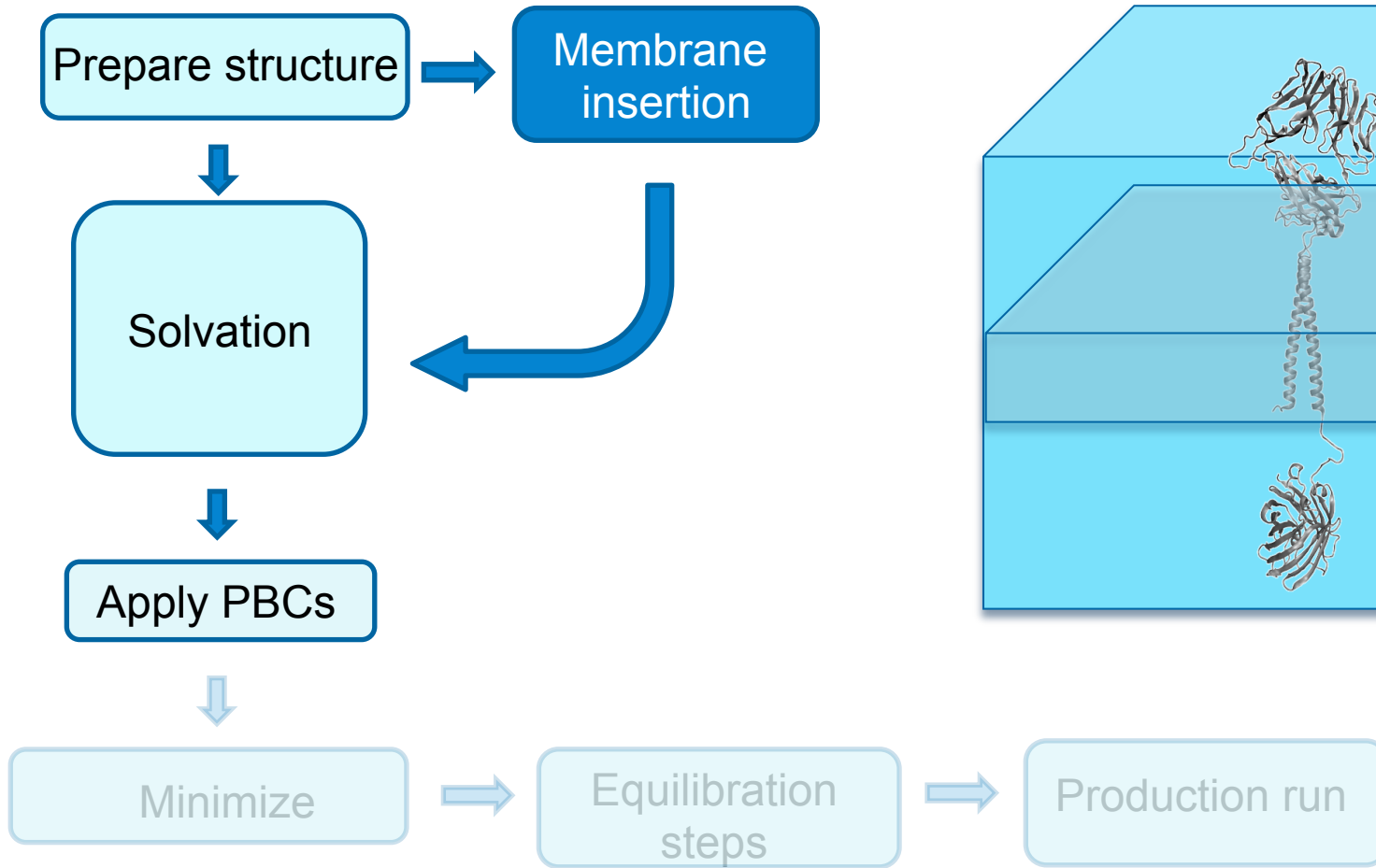


5-site



6-site

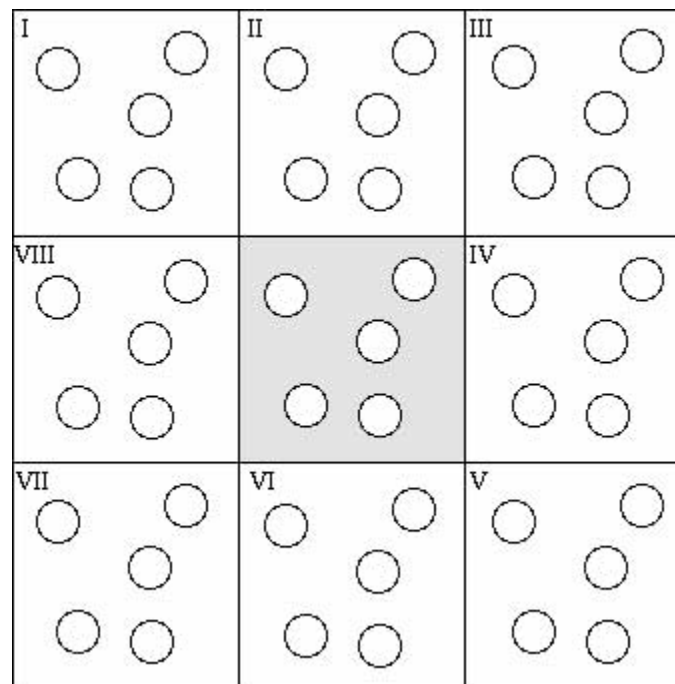
A typical classical MD experiment



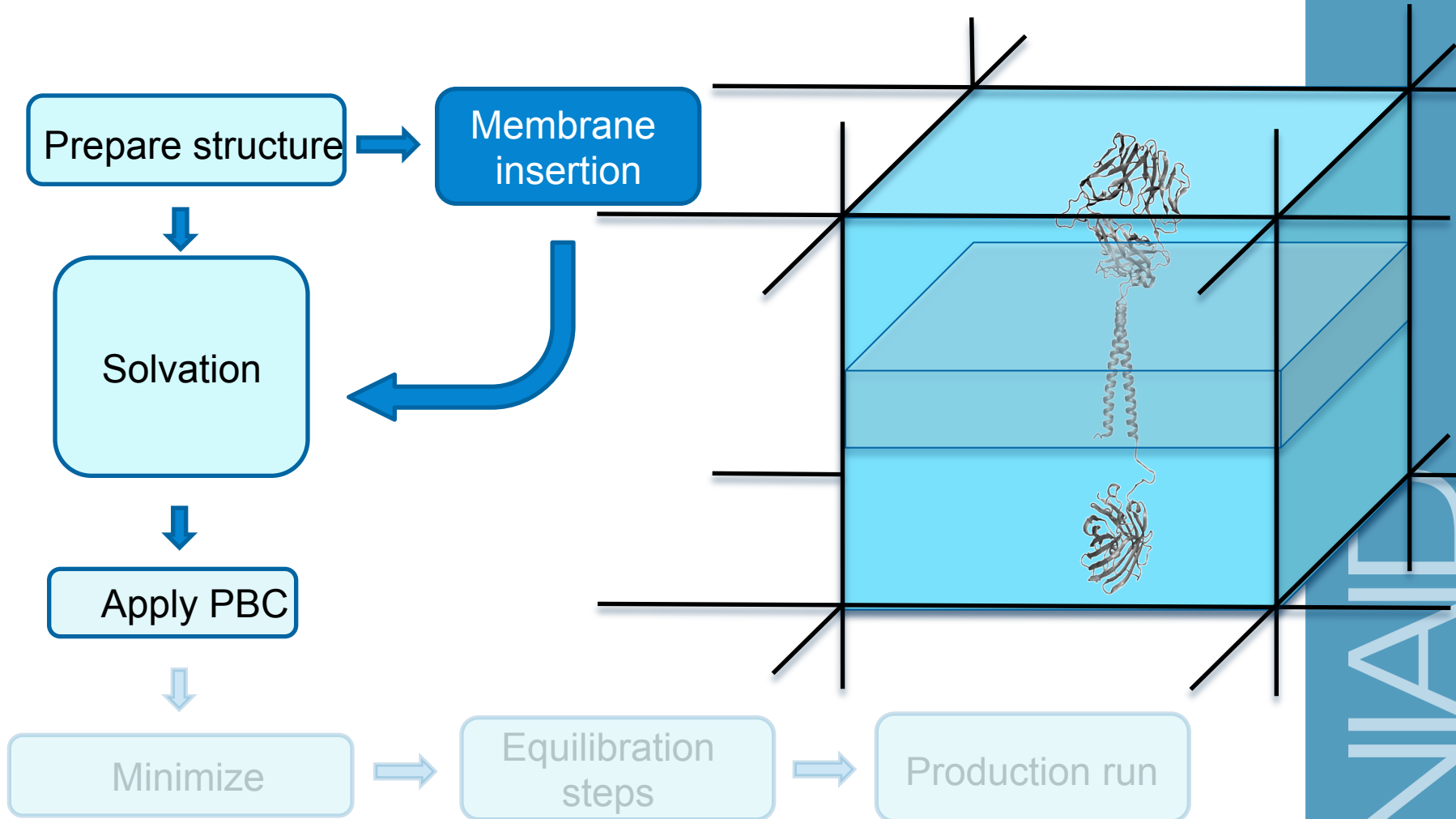


Periodic Boundary Conditions (PBC)

The fully solvated central cell is simulated, in the environment produced by the repetition of this cell in all directions.



A typical classical MD experiment

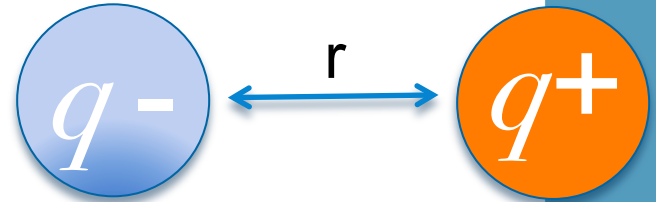


Electrostatics



Coulomb's law

$$E_{\text{elec}} = \frac{q_i q_j}{4\pi\epsilon_0\epsilon r_{ij}}$$



- Coulomb energy decreases only as $1/r$
- long range interaction
- Special techniques to reduce combinatorial problem:

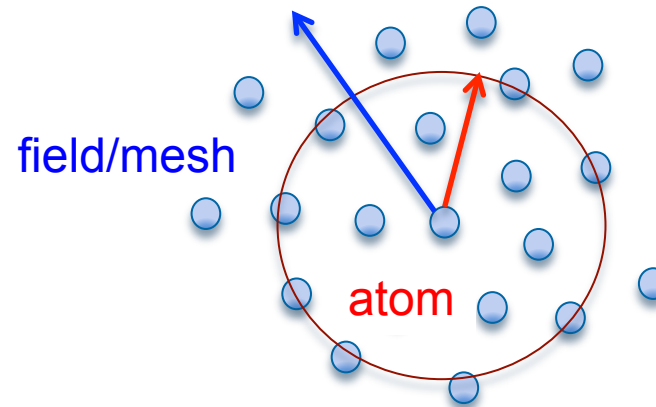
Particle Mesh Ewald (PME)

Reaction Field



Particle/Mesh Ewald method

- method for computing the electrostatic interaction energies of periodic systems (e.g. crystals)
- “speed trick” - interaction potential decomposed into a **short-range component** (summed in real space) and a **long-range component** (summed in Fourier space)

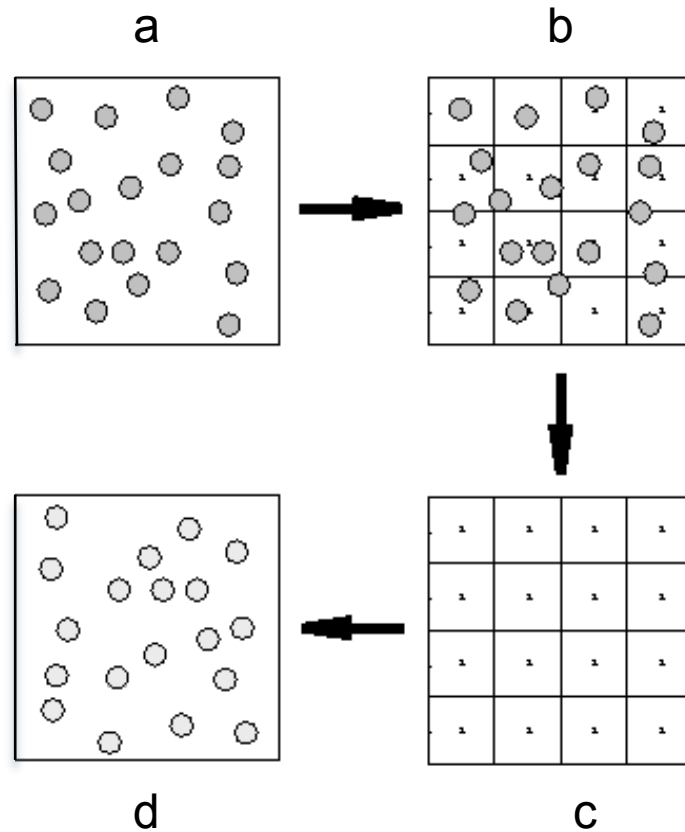


- Fourier-space summation of longer range interactions rapidly converges compared to its real-space equivalent (*needed for PBC where solvent goes to ∞*)

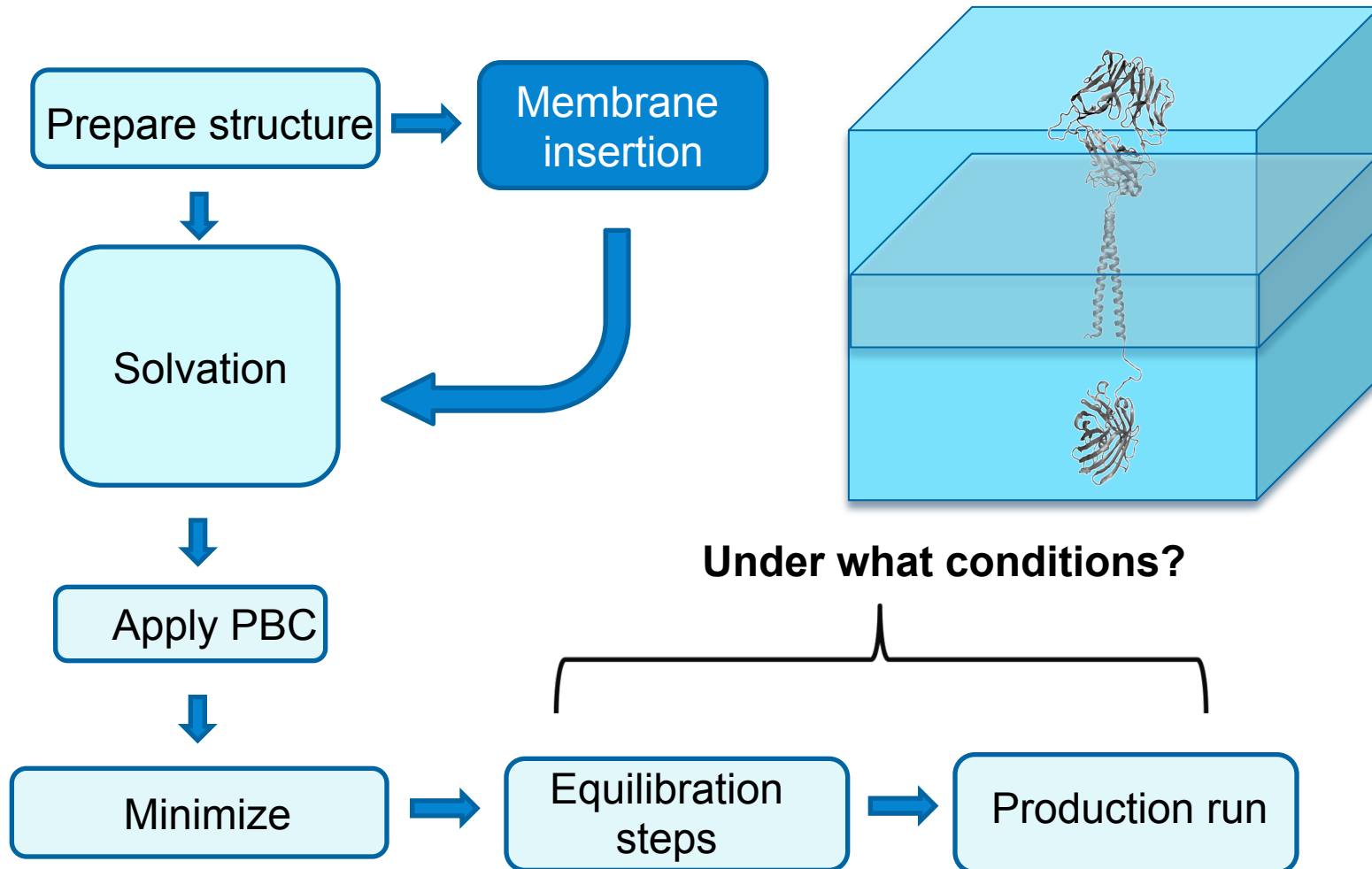
Particle/Mesh Ewald: *In actu*



- a. System of charged particles
- b. Charges interpolated onto a grid
- c. Calculate forces and potentials at grid points using a FFT
- d. Interpolate forces back to particles and update coordinates



A typical classical MD experiment



Ensembles



- s
 - e
 - r
- For isothermal-isobaric ensemble, how are fluctuations in P and T minimized and kept on target?

thermostats and barostats



Thermostats: Temperature Control



- system is coupled to a large heat bath that imposes a desired temperature
- coupling represented by stochastic impulsive forces that act occasionally on randomly selected particles
- ensures that all accessible constant-energy shells are visited according to their Boltzmann weight



Anderson
Berendsen
Langevin ←
Nosé-Hoover
DCD

NAMD default

Langevin Thermostat



In the Langevin equation of motion, a frictional force added to the conservative force is proportional to the velocity, and it adjusts the kinetic energy of the particle so that the temperature matches the set temperature.

$$M_i \frac{dv_i}{dt} = -\gamma v_i + F_i + G_i(T)$$

Mass of particle

Viscosity of hypothetical medium

interatomic force

Random force

Temperature of medium

Barostats: Pressure control



- system is coupled to a large pressure bath that imposes a desired pressure
- pressure controlled by dynamically changing the size of the simulation cell
- change can be isotropic (shape unchanged) or anisotropic (shape changes)

Anderson
Berendsen
Langevin

Nosé-Hoover

Parrinello-Rahman

← NAMD default





Langevin piston/Nosé-Hoover barostat

- modified Nose-Hoover method in which Langevin dynamics are used to control fluctuations in the barostat

$$r' = p/m + e' r$$

$$p' = F - e' p - gp + R$$

$$V' = 3V e'$$

$$e'' = 3V/W(P - P_0) - g_e e' + R_e/W$$

$$W = 3N\tau^2 kT$$

$$\langle R^2 \rangle = 2mgkT/h$$

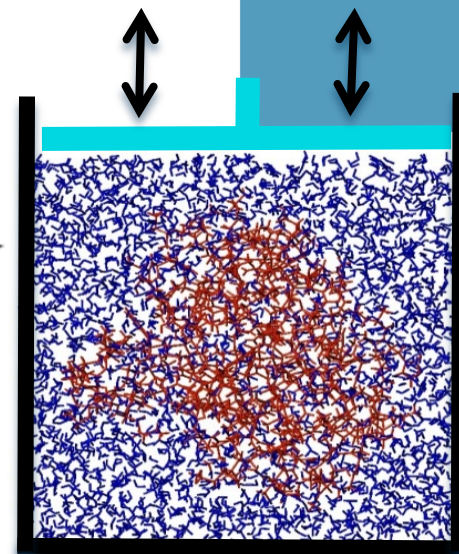
$$\tau = \text{oscillation period}$$

$$\langle R_e^2 \rangle = 2W g_e kT/h$$

Piston mass

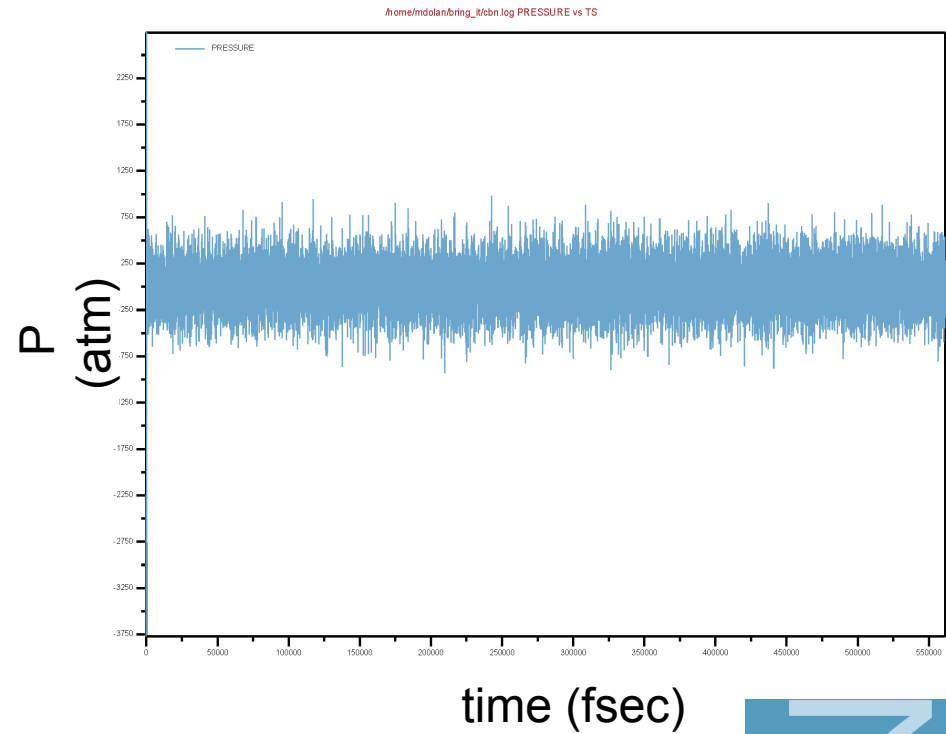
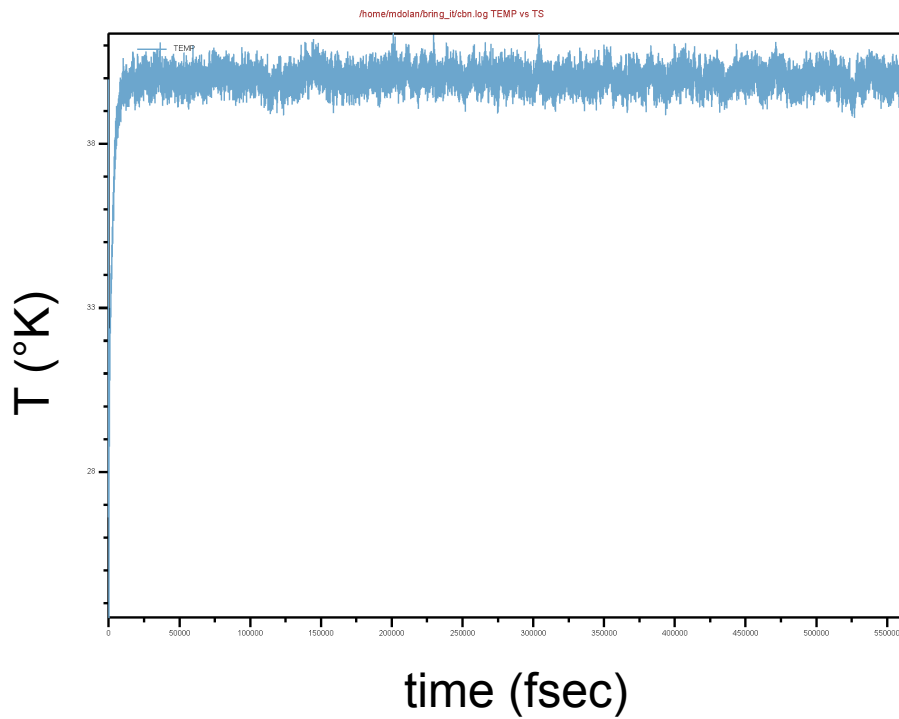
Atom noise

Piston noise



User defines desired P , oscillation and decay times of piston, temperature of piston, damping coefficients, and temperature of atoms for Langevin dynamics.

An isobaric-isothermal MD run





Time Scales of Biological Processes

Large-Scale Motions (μsec to sec)

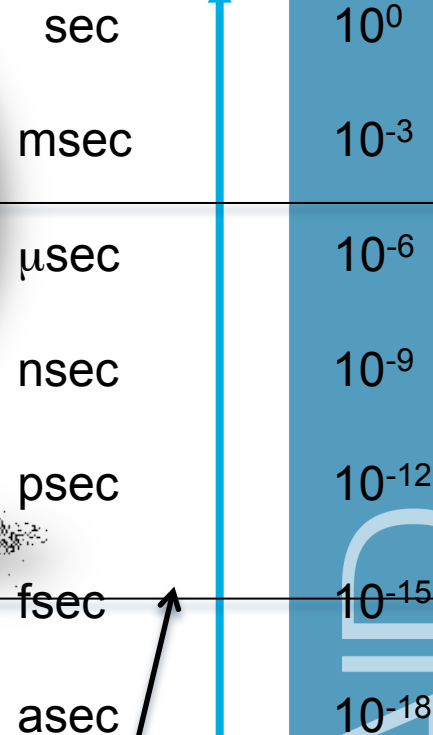
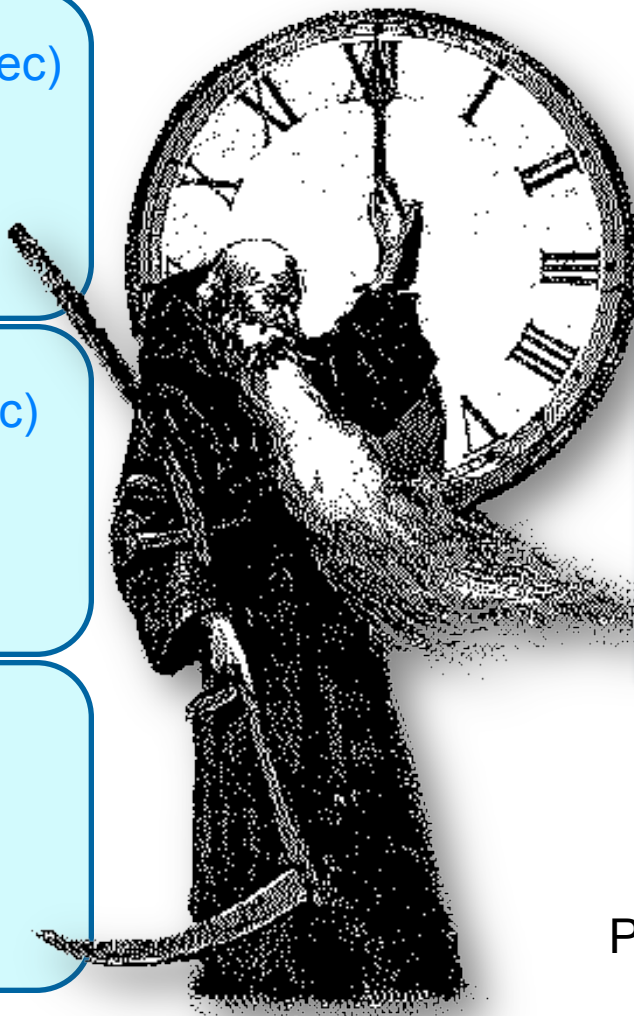
- helix-coil transitions
- dissociation/association
- folding and unfolding

Rigid Body Motions (nsec to sec)

- helix motions
- domain (hinge) bending
- subunit motions

Local Motions (fsec)

- bond fluctuations
- sidechain motions
- loop motions



Possible at NIH

NAMD



- ***N*ot *A*nother *M*olecular *D*ynamics** program
- parallel molecular dynamics code for large biomolecular systems
- available for free from the Univ. Ill, Urbana-Champaign (Google “namd”)
- installed on the NIH Biowulf cluster

Forcefields: CHARMM (default), Amber

Barostat: modified Nosé-Hoover constant pressure method with piston fluctuation control implemented using Langevin dynamics

Thermostat: Langevin (default), Berendsen

NAMD



Hands-on exercise

Goal: Run a short MD simulation of explicitly solvated crambin using the CHARMM ff with CHARMM atom types and charges

1. Examine the protein (VMD)
2. Create NAMD input files (VMD)
3. Examine and modify a configuration file (.conf)
4. Run simulation (NAMD)
5. View trajectory (VMD)
6. Analyze run (VMD)

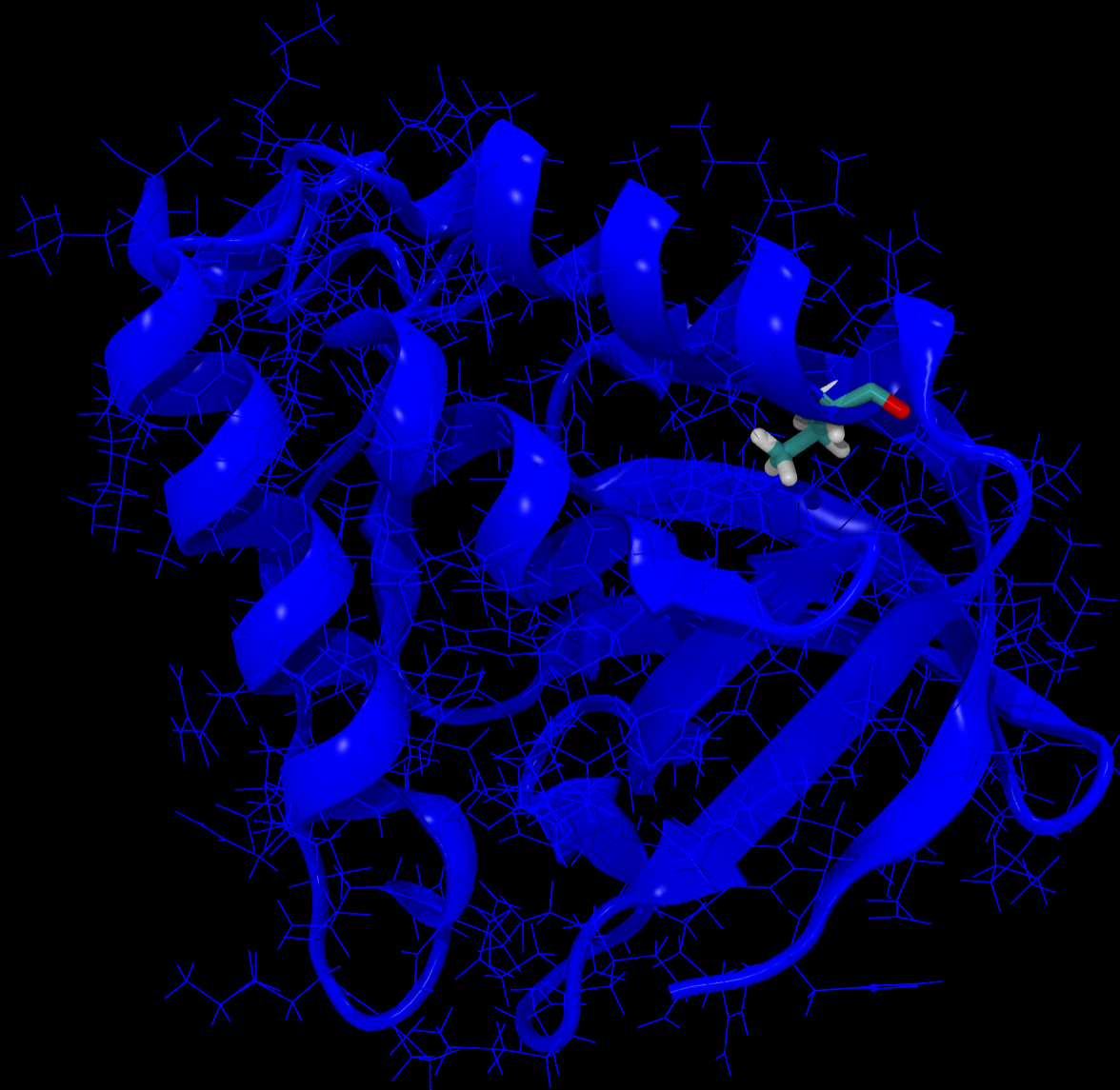


Steered and interactive MD

- External forces are applied reducing energy barriers
- Increases the probability of unlikely events on the time scale of molecular dynamics
- Corresponds closely to micromanipulation through atomic force microscopy or “optical tweezers”



Free Energy Perturbation (FEP)





Flexible fitting to a cryo-EM map

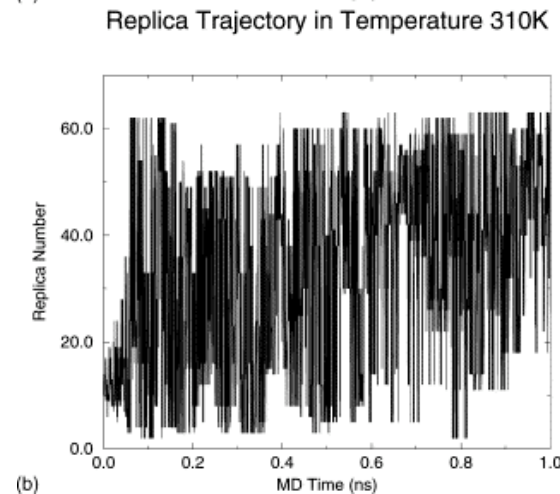
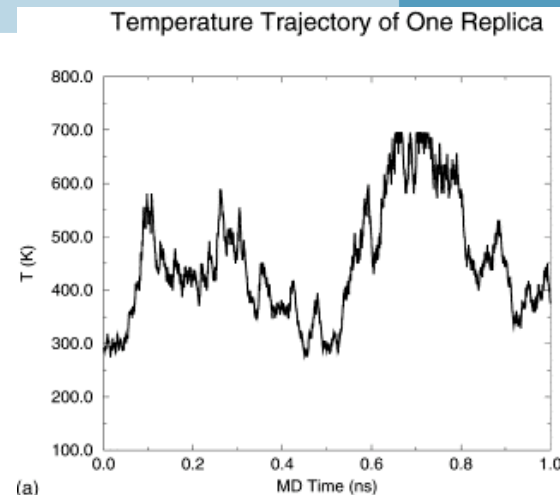


Acetyl-CoA Synthase Trabuco, L.G. *et al.* (2008) Structure

Replica Exchange MD (REMD)



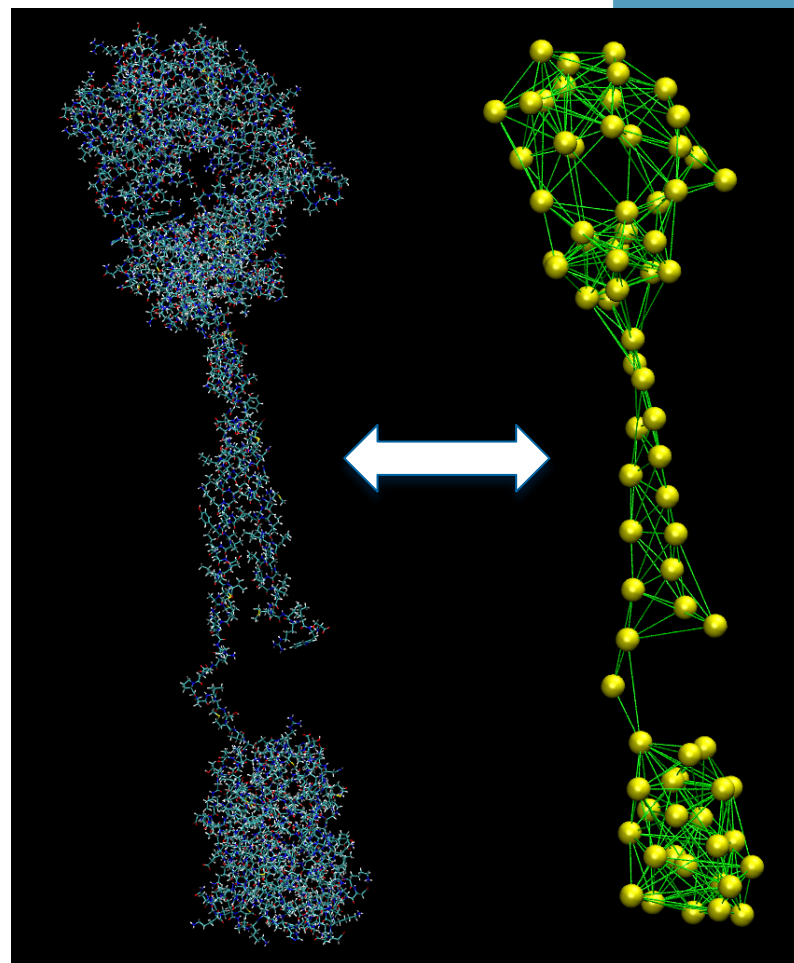
- couples MD trajectories with a temperature exchange Monte Carlo process to efficiently sample conformational space
- series of replicas (<65) are run in parallel between desired T and a high T , where a replica can surmount energy barriers
- periodically, configurations of neighboring replicas are exchanged
- exchange accepted or rejected based on criterion to maintain balance
- combined trajectories in T and conformational space allow a replica to overcome free energy barriers present at low T



Coarse-grained MD



- Groups of atoms are represented by a single particle (“bead”) reducing degrees of freedom
- Shaped based and residue based
- Requires less resources and is faster compared to all-atom representation
- Results in an orders of magnitude increase in simulation time



Summary



- Forcefields, parameters
- Molecular mechanics of MD (don't forget about quantum MD)
- Implicit vs. explicit solvent models
- Electrostatics models (PME)
- Control: Thermostats, barostats
- Other methods: Steered, coarse-grained, replica exchange, gravity

Contact



Questions or comments about molecular dynamics?

dolanmi@niaid.nih.gov

General questions about what the BCBB group has to offer?

scienceapps@niaid.nih.gov



Momentum

$$P = mv$$

Force

$$F = ma$$

Kinetic energy

$$E_k = \frac{1}{2}mv^2$$

Boltzmann's constant

$$k_b = \frac{R}{N_A}$$

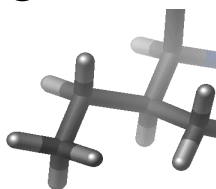
Ideal Gas Constant

$$R = 8.314\,472\text{ J K}^{-1}\text{ mol}^{-1}$$



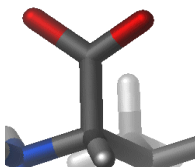
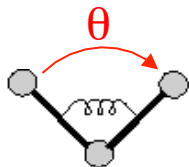
Covalent bonds and angles

Bonds

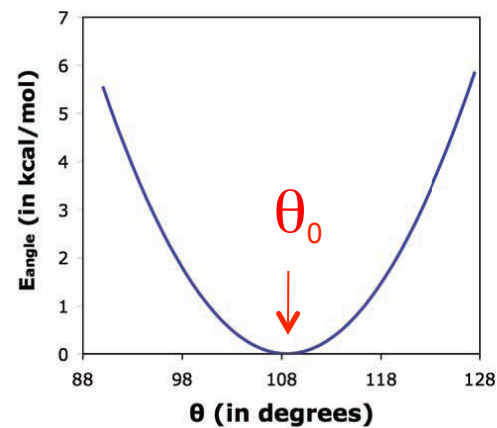
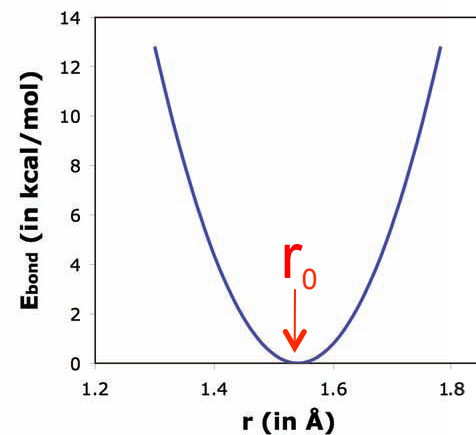


$$E_{\text{bond}} = K_b (r - r_0)^2$$

Angles



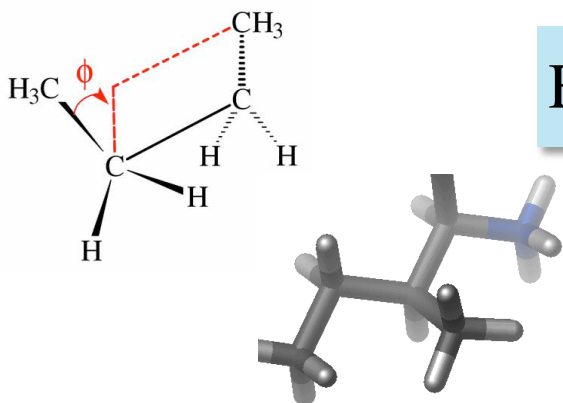
$$E_{\text{angle}} = K_{\theta} (\theta - \theta_0)^2$$



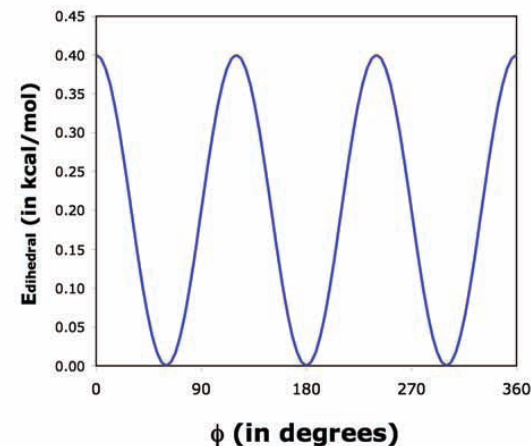


Dihedrals and improper angles

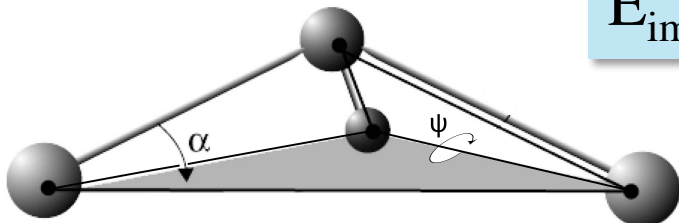
Dihedral angle



$$E_{\text{dihedral}} = K_{\phi} [1 + \cos(n\phi - \delta)]$$

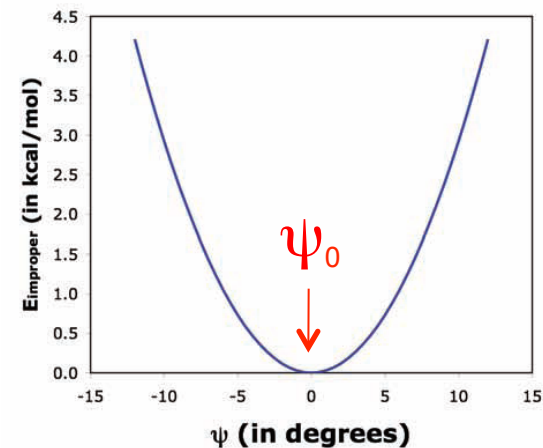


Improper angles



$$E_{\text{improper}} = K_{\psi} (\psi - \psi_0)^2$$

“out-of-plane” (oop)





van der Waals interactions

Lennard -Jones potential

$$E_{\text{vdW}} = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] = \varepsilon \left[\left(\frac{r_m}{r} \right)^{12} - 2 \left(\frac{r_m}{r} \right)^6 \right]$$

Repulsive : Pauli exclusion principle

$$\propto \frac{1}{r^{12}}$$

Attractive: induced dipole / induced dipole

$$\propto -\frac{1}{r^6}$$

