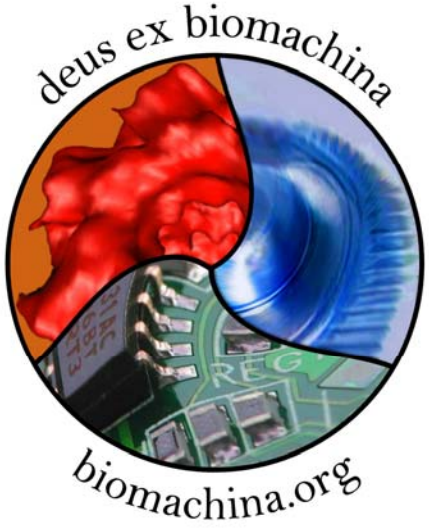


Molecular Modeling

John M. Robinson

School of Health Information Sciences
University of Texas – Houston

Why do simulations?

1. *What are the functional motions?*
   - Molecular (deterministic) dynamics
   - Stochastic dynamics
   - Normal modes

Dynamics

2. *What are the most probable conformations?*
   - Molecular dynamics
   - Monte Carlo methods
   - Hybrid MD-MC methods

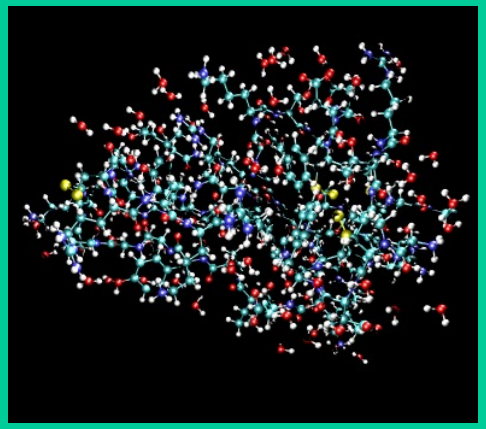
Sampling

3. *What is the most stable (probable) structure?*
   - Energy minimization
   - Simulated annealing

Optimization

Protein Structure and Energy

Model System

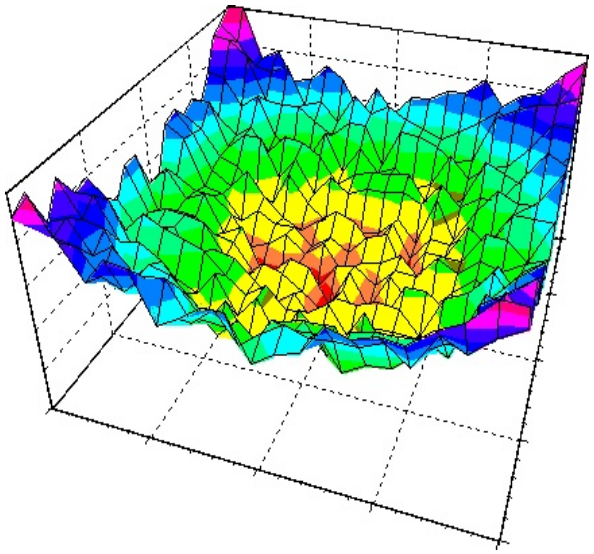


Molecular Mechanics Potential

$$V = \sum_{bonds} k_b (b - b_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 +$$

$$+ \sum_{dihedrals} \sum_{n=1}^{N} K_\phi^{(n)} [1 + \cos(n\phi - \delta)] + \sum_{impropers} K_\omega (\omega - \omega_0)^2$$

$$+ \sum_{i,j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \sum_{i,j} \left(\frac{q_i q_j}{D r_{ij}} \right)$$



Energy Surface $\rightarrow$
Exploration by Simulation..

Molecular Dynamics

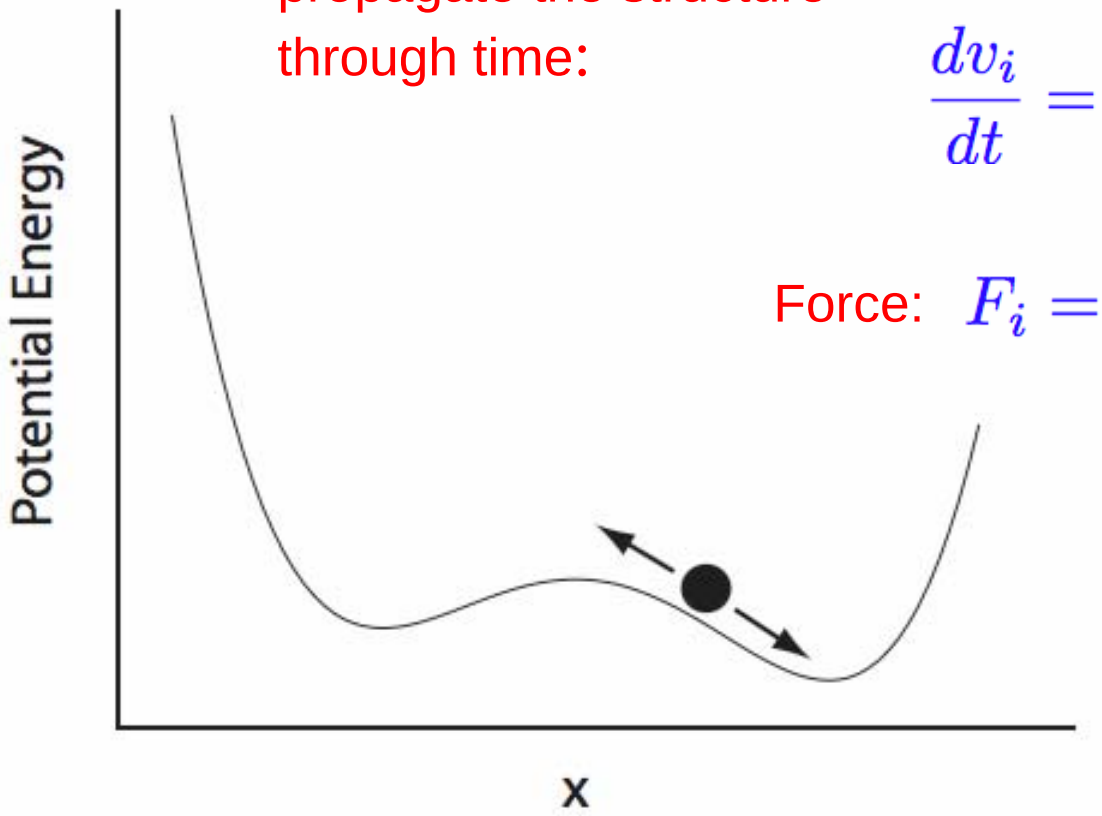
Movement on the potential energy surface

Newton's Equations propagate the structure through time:

$$\frac{dx_i}{dt} = v_i$$

$$\frac{dv_i}{dt} = m^{-1} F_i$$

Force: $F_i = \nabla_i E$



Brownian Dynamics

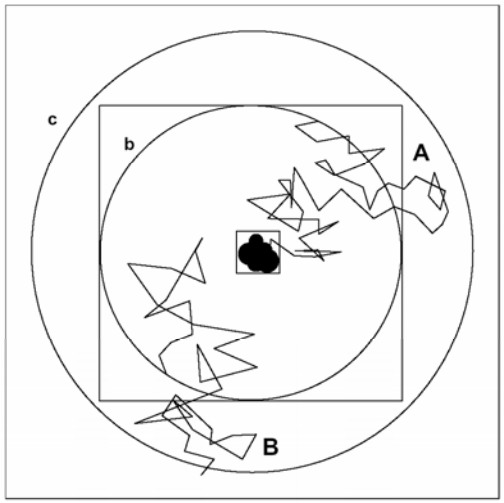
Brownian Dynamics = Newtonian Dynamics + Random Terms

$$\frac{d^2\vec{r}_i(t)}{dt^2} = m_i^{-1}\vec{F}_i + m_i^{-1}\vec{R}_i - \beta_i\frac{d\vec{r}(t)}{dt}$$

In Biomolecular Simulations:

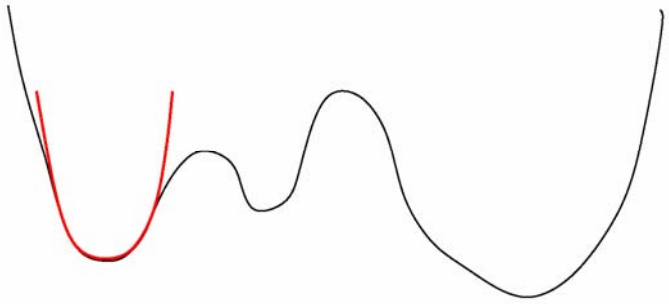
- Diffusion of Macromolecules
- Simulation of Association Processes

Molecules are treated as rigid or only semirigid macroscopic objects.



Normal Mode Analysis

Approximate the complex energy landscape by harmonic potentials.



Pros:

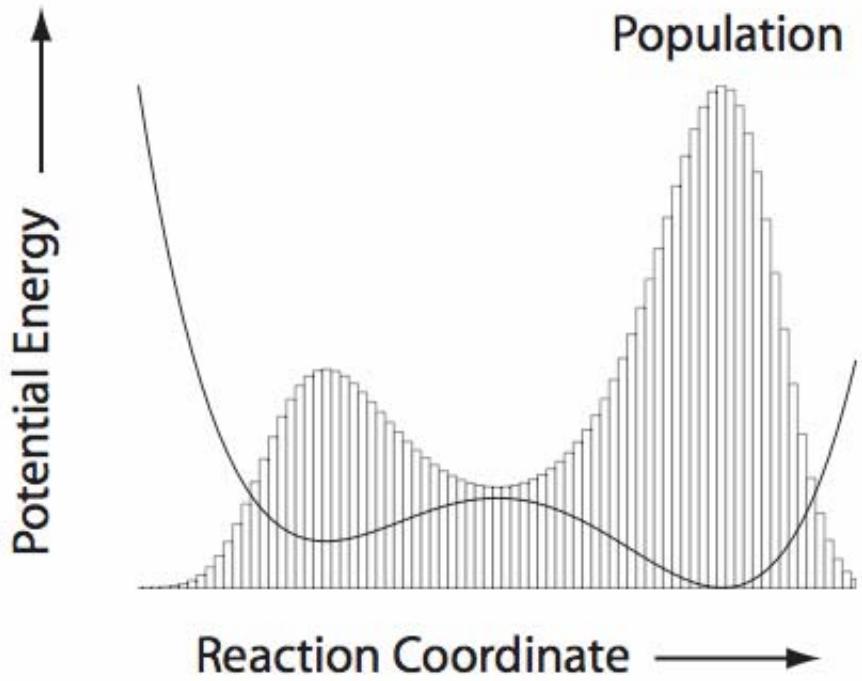
- good local approximation
- yields large scale, correlated motions
- thermodynamic theory (entropy, enthalpy, free energy)

Cons:

- no transitions between states
- forced orthogonalization of modes
- what do modes mean (are all functionally relevant)?

Monte Carlo

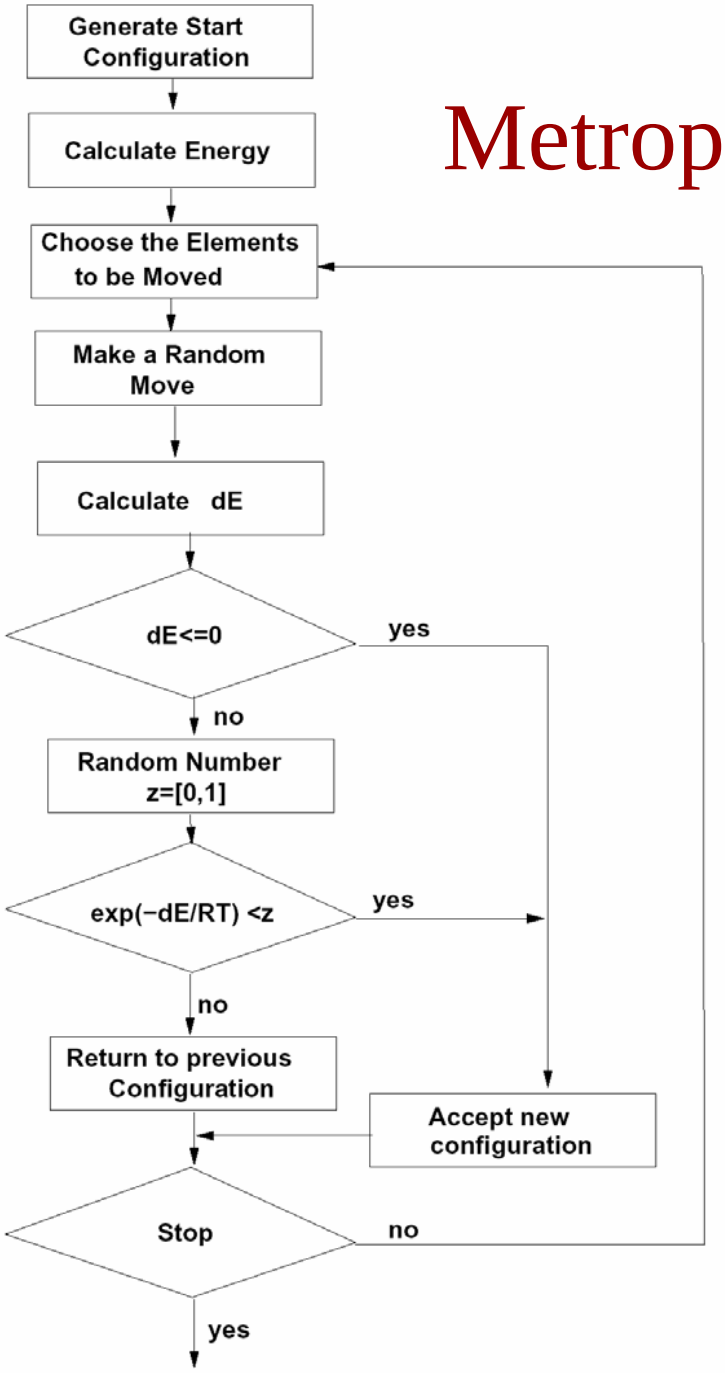
Quantity of interest: $\langle A(x) \rangle_{NVT} = \int \rho_{NVT}(x) A(x) dx$



System will preferentially populate the lowest energy states

$$\rho_{NVT}(x) = \frac{1}{Z} \exp(-\beta \Delta E(x))$$

Metropolis Monte Carlo



- Rigorous canonical sampling
- Produces Boltzmann weighted populations:

$$\rho_{NVT}(x) = \frac{1}{Z} \exp(-\beta \Delta E(x))$$

- information about equilibrium states
- easy to implement

Monte Carlo Pros/Cons

Simulation involves random trial steps. (Analogy with gambling, hence the name Monte Carlo)

Pros:

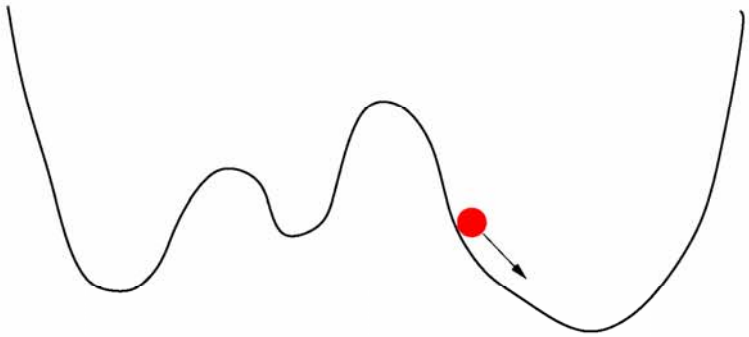
- does not require a continuous energy function (as in MD)
- number of particles can easily vary (very hard in MD)

Cons:

- highly correlated movements are hard to simulate, leads to a poor sampling of large-scale changes

Optimization by Energy Minimization

Goal: finding low energy conformations, i.e., most probable conformations



Various numerical optimization procedures are applied to find these minima.

But:

- only local minima are found
- only minimum potential energy, not minimum free energy

Physics Based Modeling

1. *Quantum Mechanics - focus on the electrons*
   - Ab initio methods (MO = LCAO)
     HF, DFT
   - Semi-empirical methods

2. *Molecular Mechanics - focus on the nuclei*
   - Empirical pairwise potential
   - Continuum potential field

3. *Hybrid QM/MM*

Quantum Chemistry

Physical model: Born-Oppenheimer approximation: electron waves in a field of the nuclei.

Stationary 1D Schrodinger equation: $-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) = E\psi(x)$

Quantum chemical calculations yield:

– Structure: ground states ,excited states, transition states

– Energy: reaction energy (equilibrium constant), activation energy, ionization energy, torsional energy, vibrational energy

– Spectra: electronic, vibrational (normal modes), NMR

– Charge: partial charges, dipoles

– Chemical reactions: bond breakage

– Quantum effects: tunneling, spin

Quantum Chemistry Pros/Cons

Pros:

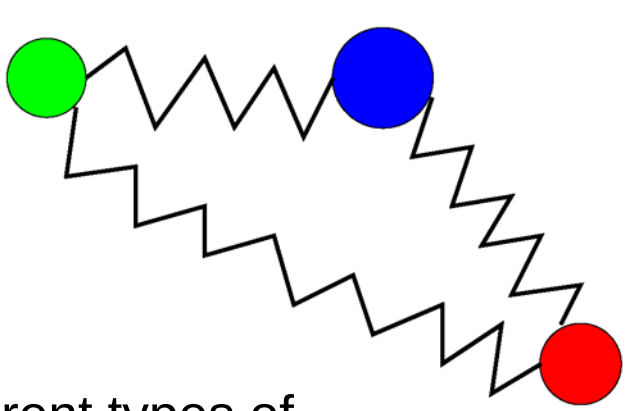
- detailed picture of a molecule (electrons etc.)
- electronic description, chemical reactions, excited states
- limited conformational flexibility (relaxation)
- often used to parametrize MD force fields

Cons:

- computationally very demanding (ab initio > DFT > semiempirical)
- static model (no time)
- only small systems: 10-1000 atoms

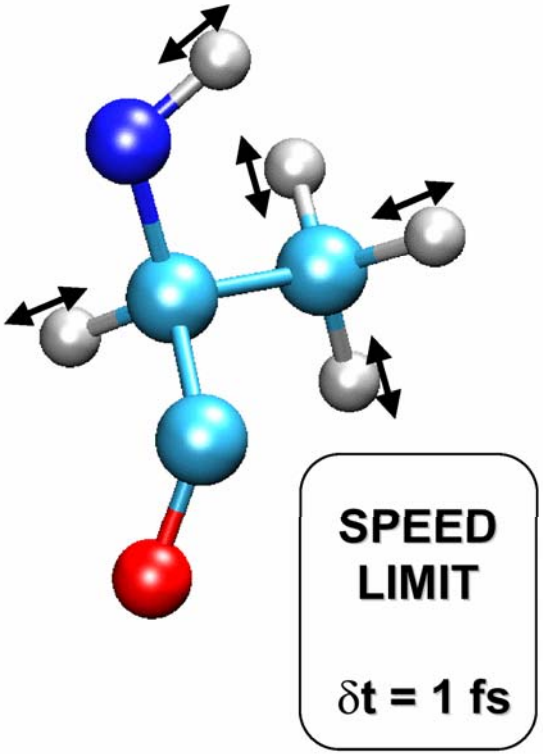
Molecular Mechanics

•A molecule is described by interacting (soft) spheres.



- Different types of spheres describe different types of atoms.
- The interaction between chemically bound atoms is described by special bonding interaction terms.
- The interaction of not chemically bound atoms is described by non-bonding interaction terms.
- The motion of all the atoms in the molecule is described by Newtonian classical mechanics.

Timescale Limitations



Molecular dynamics:

Integration timestep - 1 fs, set by fastest varying force.

Accessible timescale: about 10 nanoseconds.

Molecular Mechanics Pros/Cons

Pros:

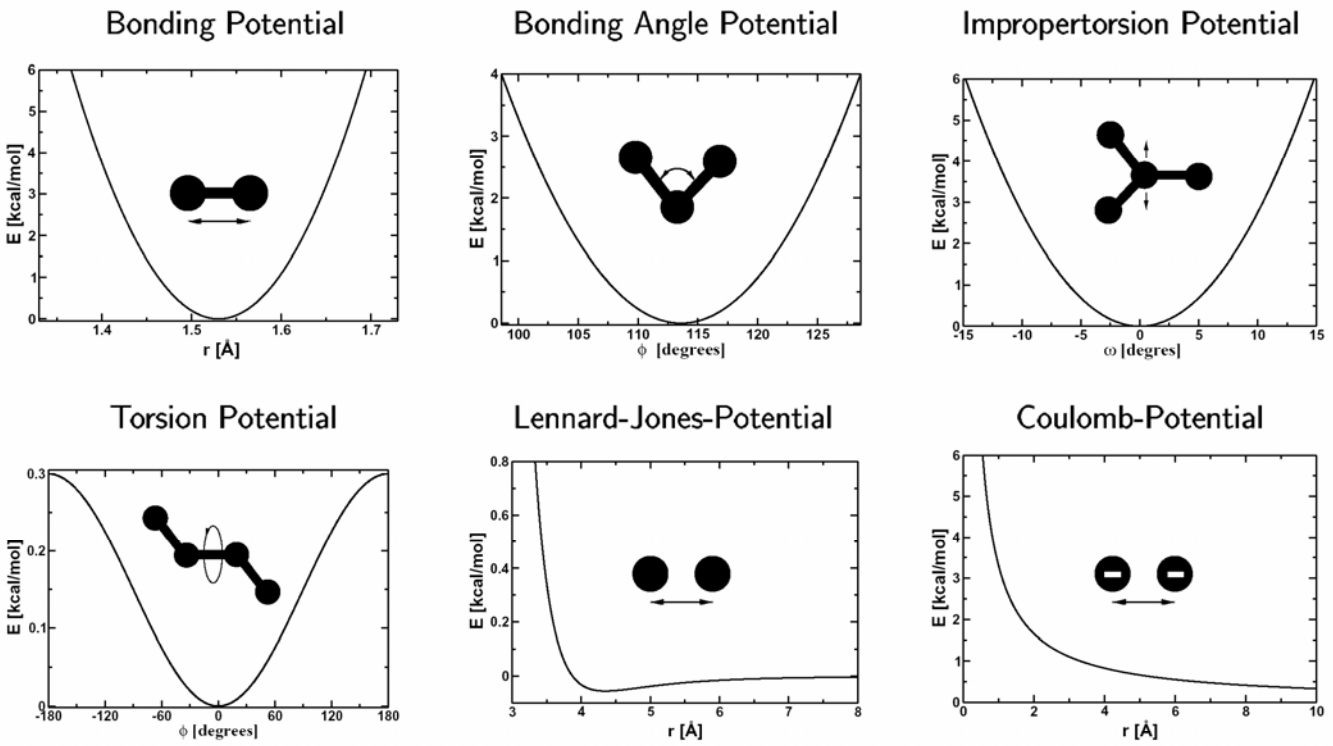
- detailed stereochemical model that describes certain aspects of biomolecules very well
- conformational flexibility
- dynamic model (time dependence) is possible
- large systems (> 10^4 atoms) can be modeled

Cons:

- computationally demanding
- large scale conformational changes are hard to model
- no electronic (quantum) desciption, no chemical reaction (bond breaking/forming), no excited states, …
- limited run times

Molecular Dynamics Implementation

Molecular Dynamics: Pairwise Potential



$$E = \sum_b k_b(r - r_0)^2 + \sum_\theta k_\theta(\theta - \theta_0)^2 + \sum_\omega k_\omega(\omega - \omega_0)^2$$

$$+ \sum_\phi k_\phi(1 - \cos(n\phi - \delta)) + \sum_{i<j} 4\,\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right] + \sum_{i<j}\frac{q_i q_j}{r}$$

Verlet's Numeric Integration Method

Taylor expansion:

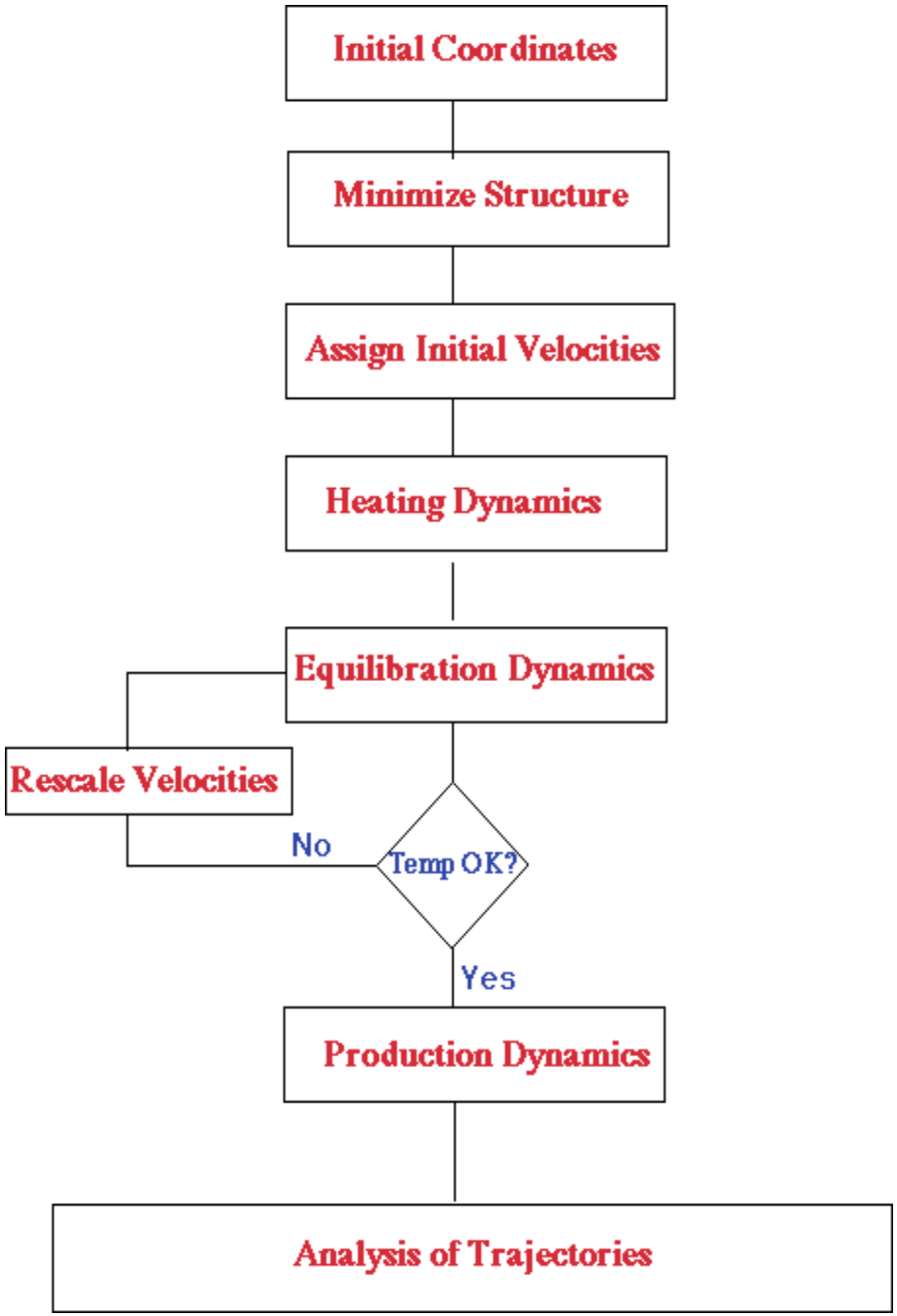
$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2$$

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

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

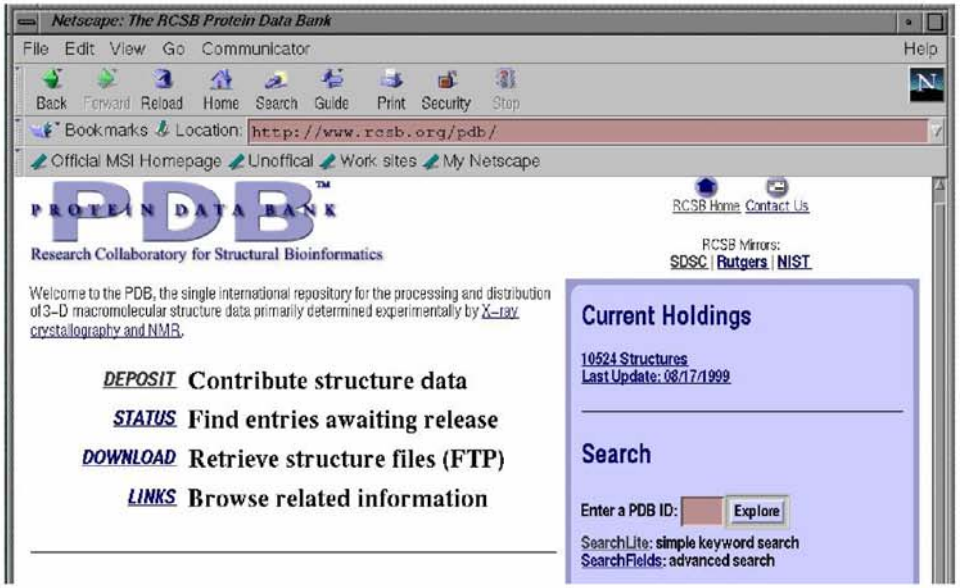
Verlet's Method

Production Run Protocol



Input Files: PDB

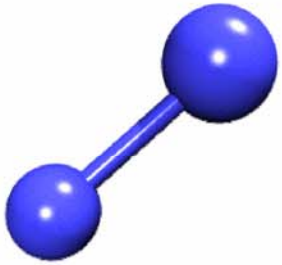
- Simulations start with atomic structures from the Protein Data Bank, in the standard PDB file format.
- PDB files contain standard records for species, tissue, authorship, citations, sequence, secondary structure, etc.
- We only care about the atom records…
  - atom name (N, C, CA)
  - residue name (ALA, HIS)
  - residue id (integer)
  - coordinates (x, y, z)
  - occupancy (0.0 to 1.0)
  - temp. factor (a.k.a. beta)
  - segment id (6PTI)
- No hydrogen atoms!
  (We must add them ourselves.)



Input Files: PSF

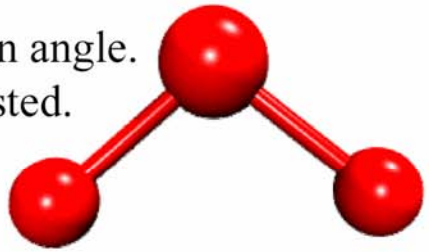
- Every atom in the simulation is listed.
- Provides all static atom-specific values:
  - atom name (N, C, CA)
  - atom type (NH1, C, CT1)
  - residue name (ALA, HIS)
  - residue id (integer)
  - segment id (6PTI)
  - atomic mass (in atomic mass units)
  - partial charge (in electronic charge units)
- What is not in the PSF file?
  - coordinates (dynamic data, initially read from PDB file)
  - velocities (dynamic data, initially from Boltzmann distribution)
  - force field parameters (non-specific, used for many molecules)

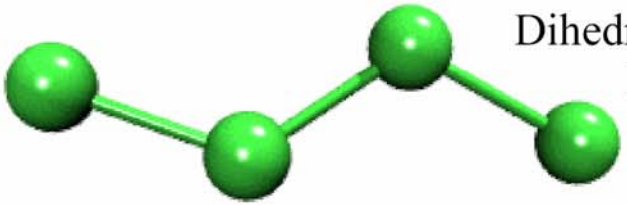
Input Files: PSF



Bonds: Every pair of covalently bonded atoms is listed.

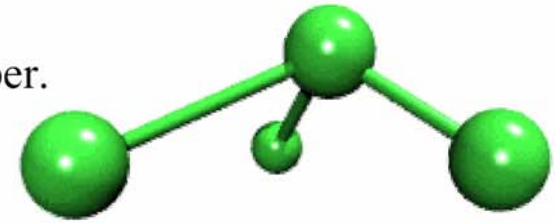
Angles: Two bonds that share a common atom form an angle.
Every such set of three atoms in the molecule is listed.





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

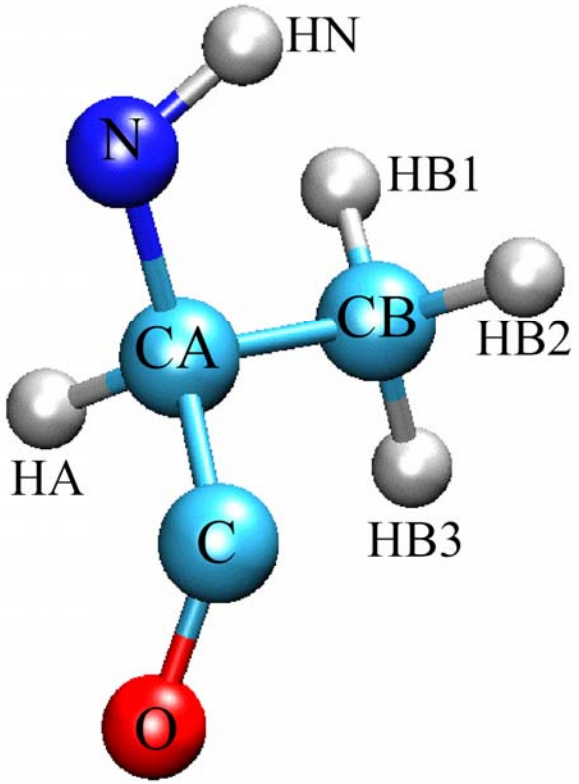
Impropers: Any *planar* group of four atoms forms an improper.
Every such set of four atoms in the molecule is listed.



Input Files: Topology Files

blueprints for building a PSF file

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



Input Files: Parameter Files

defining the MM energy terms

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

Practical Tips for Setting up MD

1. *Decide what you want to simulate*
   (protein, DNA, sugars, water, ions, lipids, etc)

2. *Build Individual Components*
   - add missing atoms
   - add hydrogens
   - modify ionization states
   - graft functional groups onto residues
   - compute missing energy parameters with QM

3. *Solvate Structure*

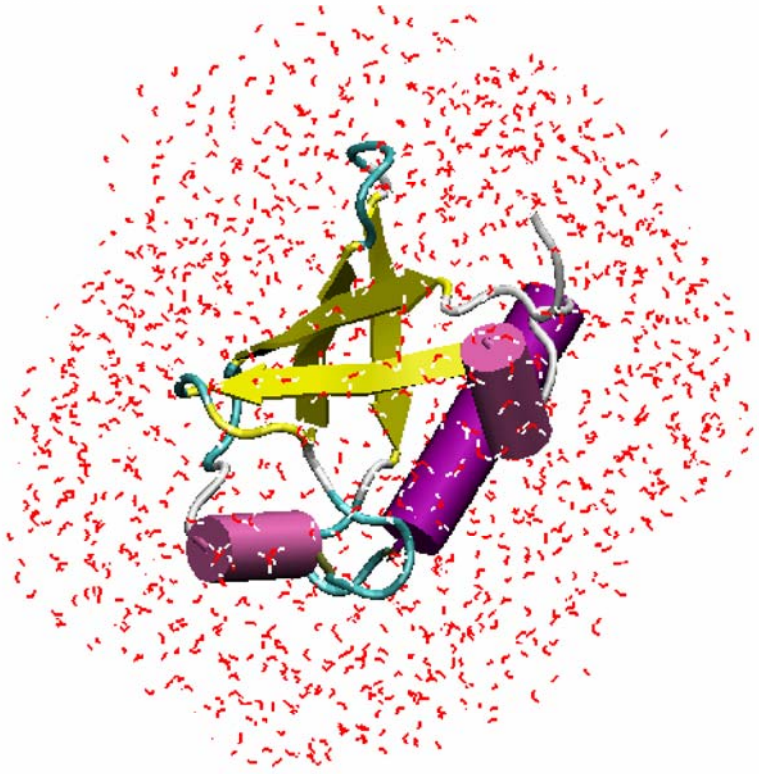
4. *Combine Molecular Components*
   (lipid bilayer, water, ions, polymeric chains)

5. *Minimize Energy / Equilibrate*

Setting up a MD Simulation

Example: Building Ubiquitin

- Obtain file from PDB (1ubq);
- Add missing hydrogen atoms;
- Determine protonation state of HIS residues;
- Add a water box;
- Trim the water box down to a sphere.

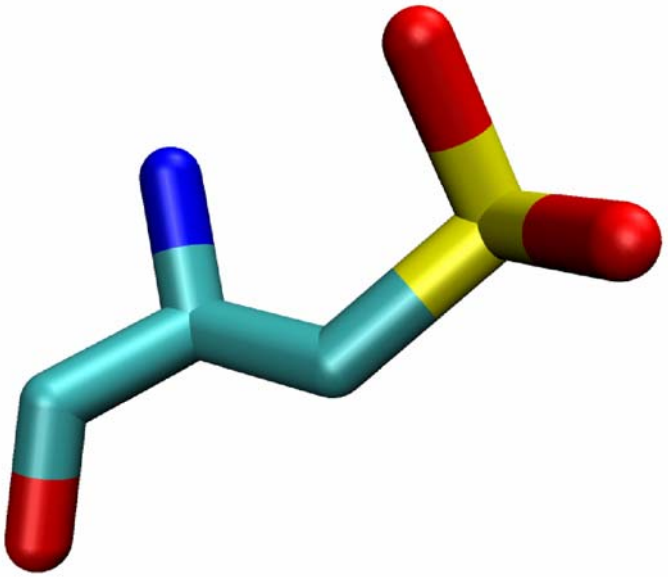


Build Individual Components

- Split the structure into individual, connected segments
- Delete all hydrogens (avoids atom name conflicts later, they are mobile and will be placed by MD program anyway)
- Correct atom names (compare PDB to topology file, edit PDB)

Deal with Unknown Residues

- Your system may contain residues that aren't in your topology file.
- In many cases the residue can be built as a chimera out of existing topology groups.
- Exotic new groups may require quantum chemistry to parameterize accurately.



Solvate the Structure

Implicit vs. Explicit Water Molecules?

Implicit: Distance-dependent dielectric (X-PLOR)

`parameter nbonds RDIE SWITCH end end`

+ Inexpensive

- Conformation artifacts (“quick and dirty”)

$$f_{ELEC}(R) = Q_i Q_j \frac{C}{\varepsilon_o R^2}$$

Implicit: Poisson-Boltzmann

+ More accurate than distance-dep. epsilon

- Experimental, research in progress (CHARMM, AMBER)

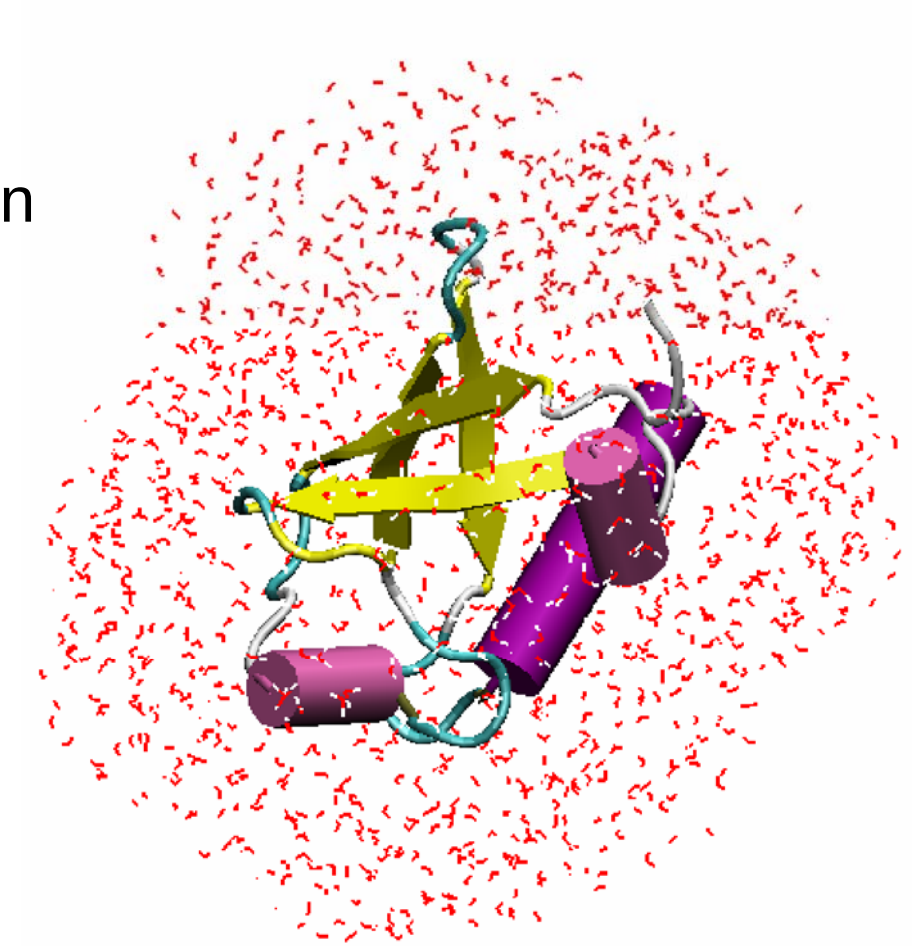
Explicit: Solvent

+ Best modeling of solvation effects

- Expensive, slow dynamics of water molecules (displacement difficult)

Explicit Solvation Scripts

- VMD *solvate* (Tcl script in library).
- The basic building block is an equilibrated cube of water
- Replicates the water box as many times as necessary, renaming segments and removing overlapping atoms.
- The VMD *solvate* package uses VMD's atom selection capabilities.
- *solvate* can deal with periodic boundary conditions

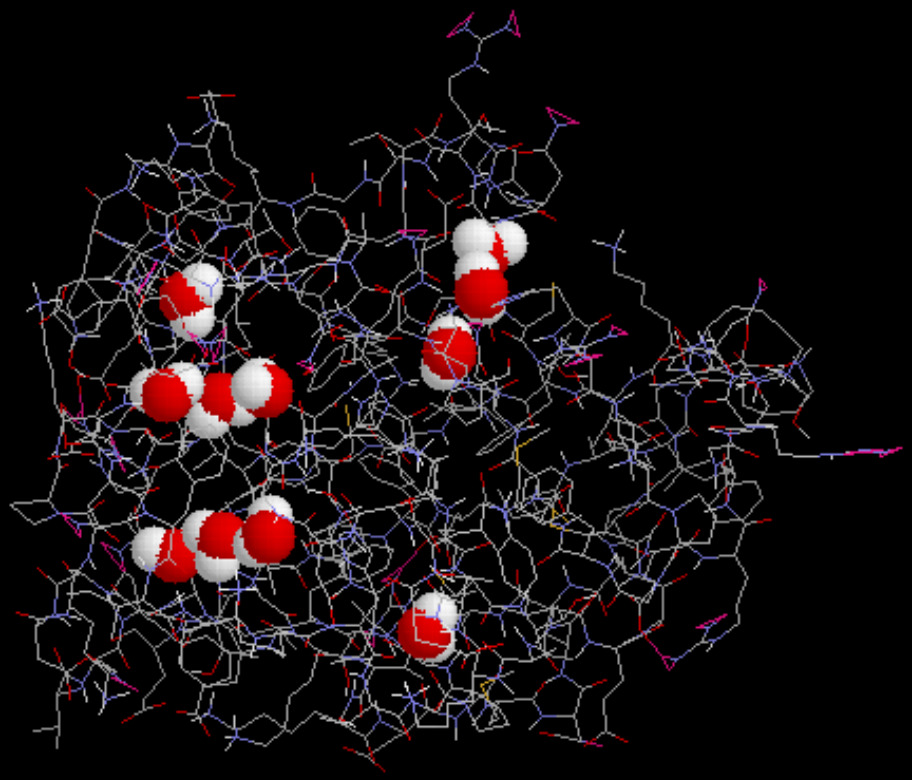


Predicting Buried Water

To prevent collapse of any cavities, we need to fill them with water molecules

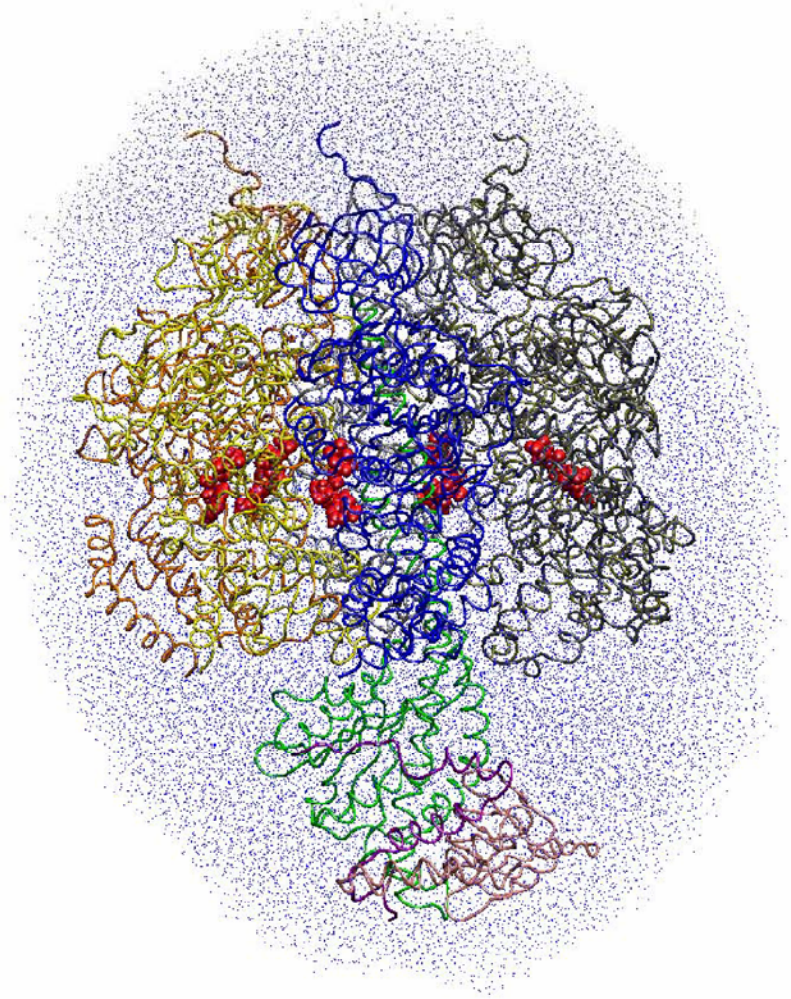
DOWSER program (Jan Hermans, UNC Chapel Hill)

URL:
http://femto.med.unc.edu/DOWSER/Dowser.htm



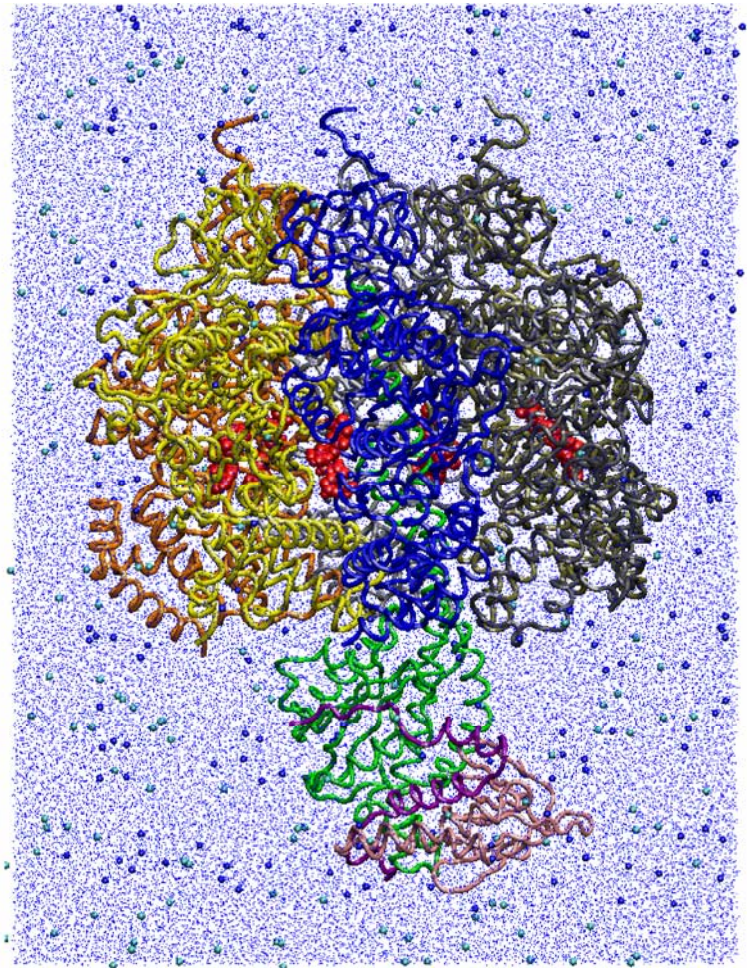
Water Layer vs. Periodic Boundaries

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



Periodic Boundary Conditions

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

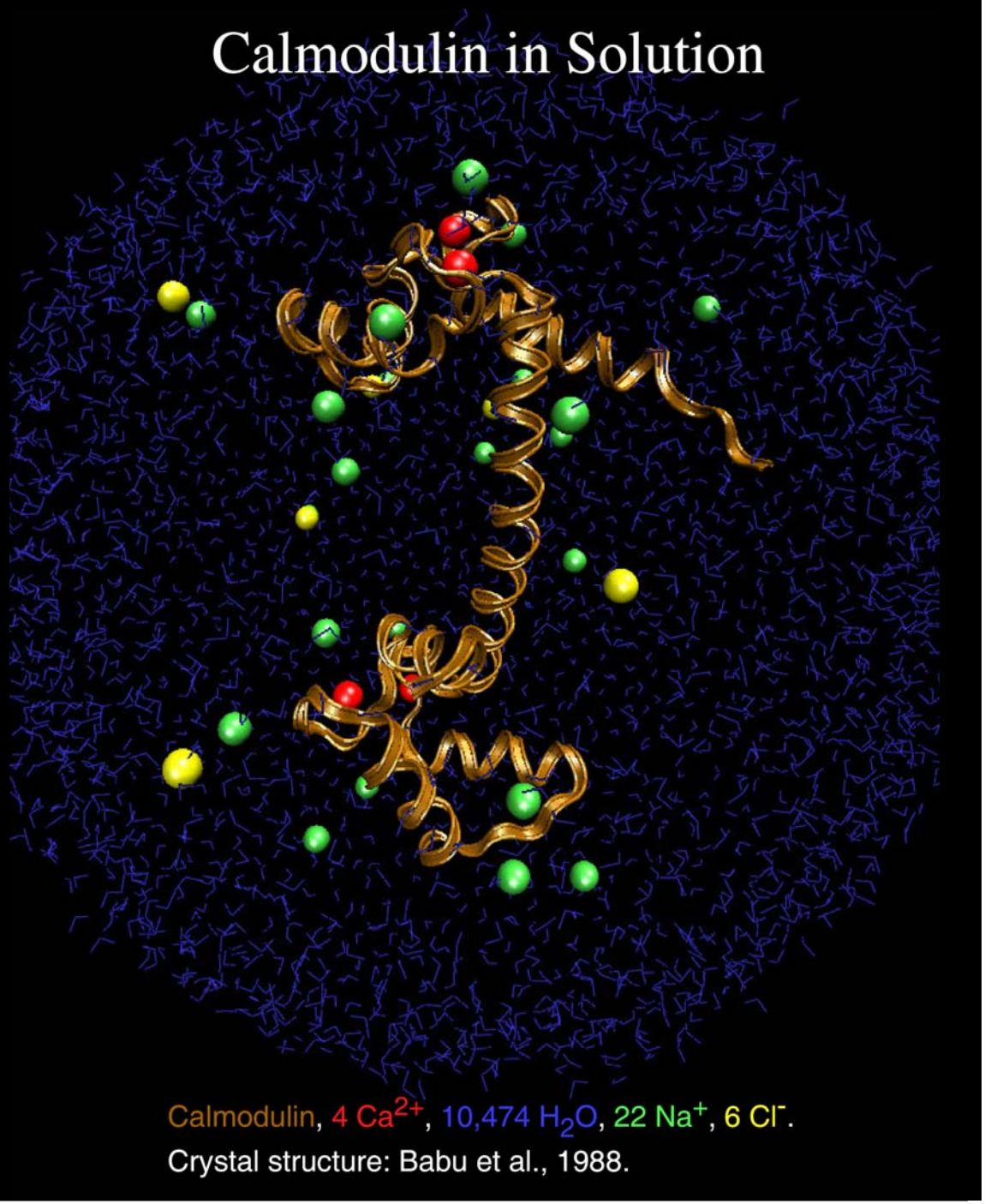


Predicting Ion Positions

Explicit ions neutralize the system and mimic physiological ionic strength.

Placed sequentially at minimum of electrostatic stat energy with X-PLOR script.

Wriggers et al., *Biophys.Journal* 1998, 74:1622-1639.



Combine Molecular Components

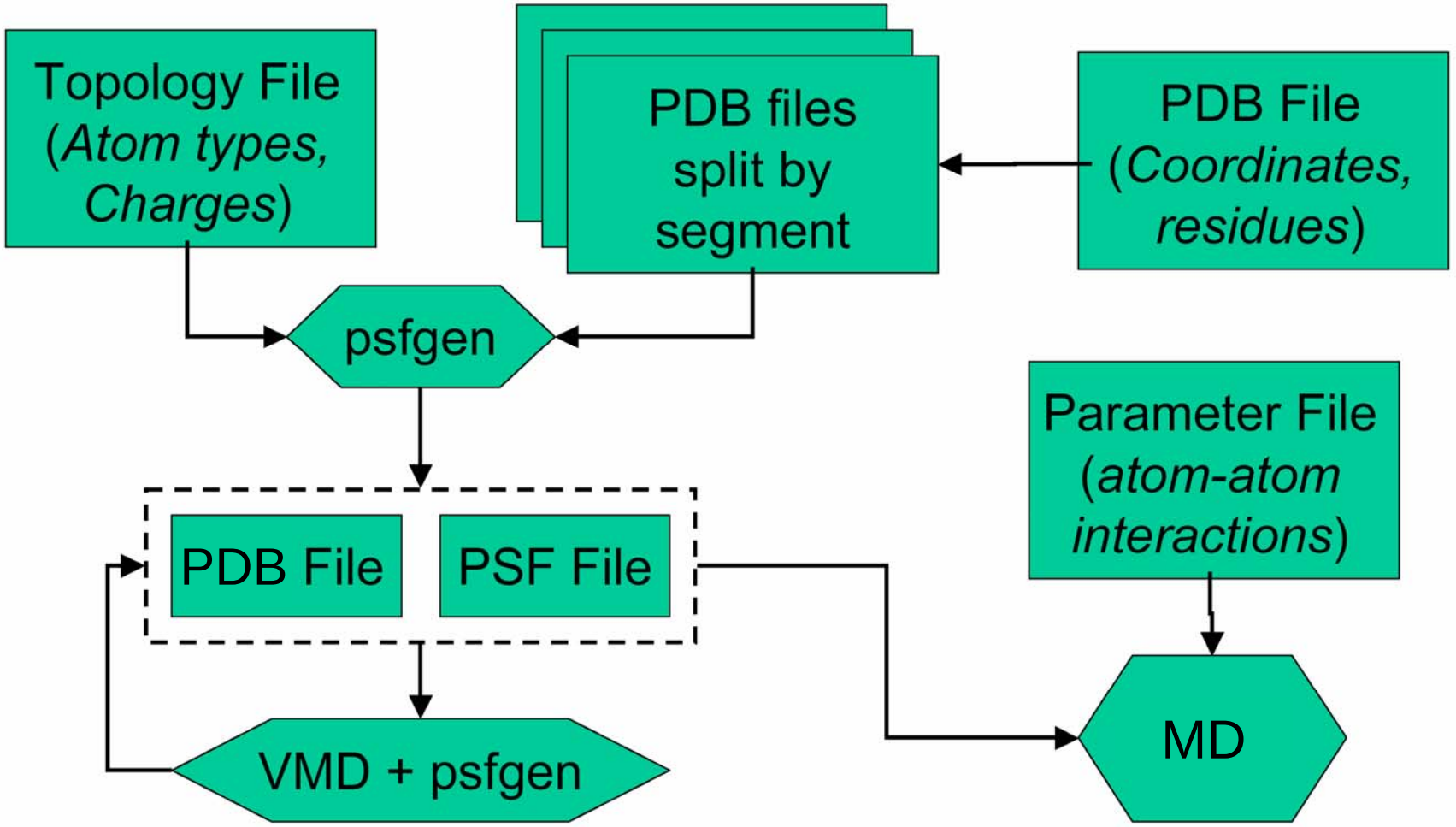
•Once you have all the components (protein water, membrane, etc) combine them into one structure (PDB + PSF), e.g. with X-PLOR script as shown in earlier sessions.

•Alternatively, use VMD/psfgen to assemble the PDB files

Structure Building in VMD: psfgen

- Tcl script in VMD script library.
- Maps residues to entries in a CHARMM topology file.
- Links residues to form connected segments.
- Combines segments to form a complete structure file.
- Patches residues to form new covalent bonds or modify charge states.
- Guesses coordinates for missing atoms.
- Writes PSF and PDB files.

psfgen Flow Chart



Take Home Message

- Setting up MD simulations is very tricky.
- May take expert hours to weeks of preparation.
- In this workshop we have already prepared scripts and structures for the demos.
- We use relatively ancient X-PLOR (1990s) in flexible fitting tutorials because of ease of implementation of constraints and ubiquitous distribution/documentation. CNS, NAMD, etc are alternatives that we are currently exploring.

References

•*Molecular Modeling and Simulation: An Interdisciplinary Guide.* Tamar Schlick, Springer Verlag 2002.

•*NAMD/VMD workshop tutorials.* www.ks.uiuc.edu/Training/Tutorials

•*Molecular Thermodynamics*. McQuarrie and Simon, University Science Books, 1999.

•*Computer Simulations of Liquids*. Allen and Tildesley, Oxford University Press, 1987.

Acknowledgements/Credits

Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign

G. Matthias Ullmann, Uni Heidelberg

Jeremy Smith, Biocomputing Unit, U Heidelberg, Germany

Nurit Haspel, David Zanuy, Ruth Nussinov, Tel Aviv U., Israel

Loredana Vaccaro and Mark Samson, U. Oxford

http://xplor.csb.yale.edu