THE UNIVERSITY of TEXAS
HEALTH SCIENCE CENTER AT HOUSTON
SCHOOL of HEALTH INFORMATION SCIENCES

Molecular Dynamics Simulation: Preparation

For students of HI 6001-100 “Biomolecular Modeling”

Willy Wriggers, Ph.D.
School of Health Information Sciences

http://biomachina.org/courses/modeling/06.html

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

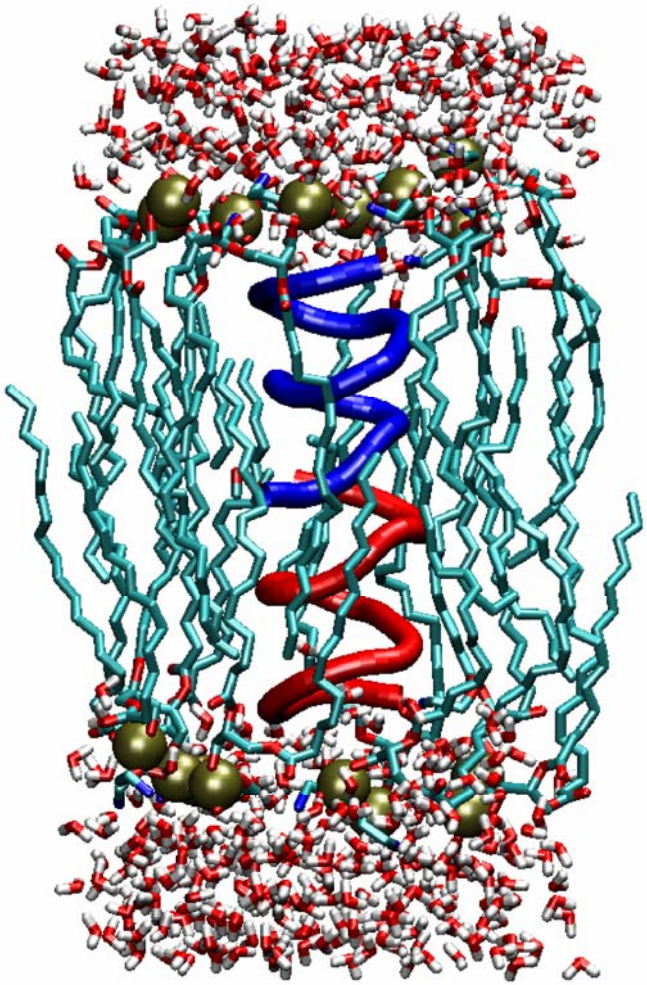
4. Combine Molecular Components
(lipid bilayer, water, ions, polymeric chains)

5. Minimize Energy / Equilibrate

Decide What You Want to Simulate

Example: Gramicidin A

- Obtain GA structure from the PDB databank (www.rcsb.org)
- Deal with non-standard N-terminal and C-terminal residues
- Build a lipid membrane around the peptide
- Add water
- Equilibrate

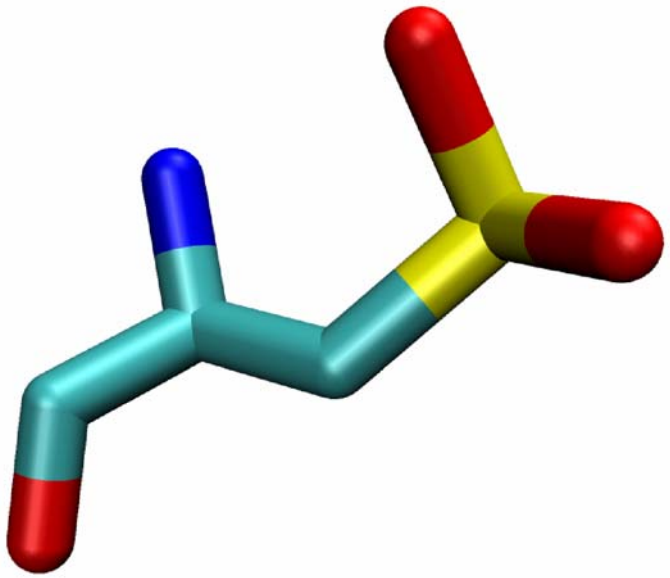


Build Individual Components

- 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)

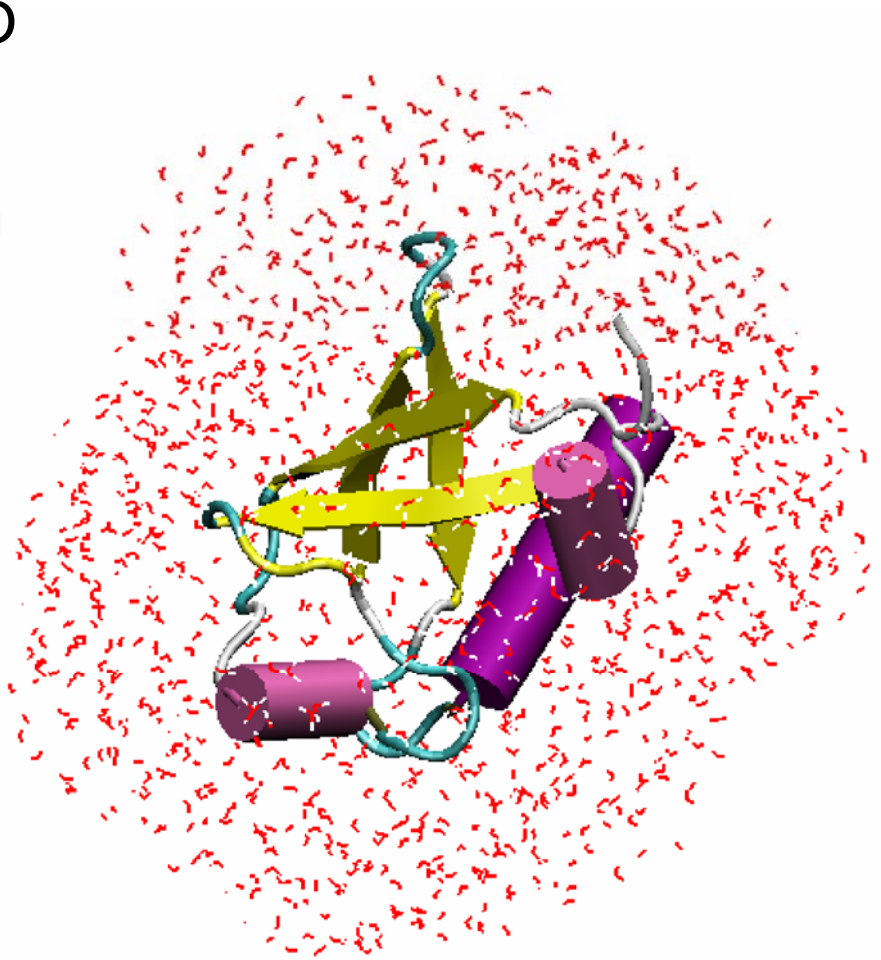
Explicit: Solvent

+ Best modeling of solvation effects

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

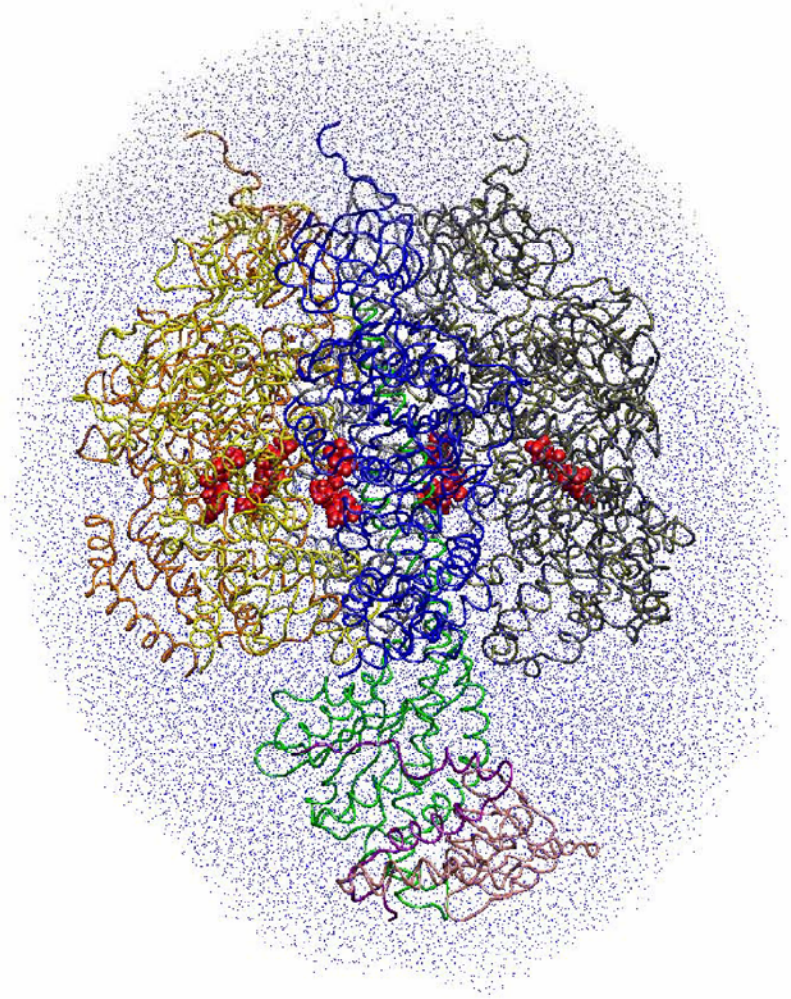
Explicit Solvation Scripts

- X-PLOR (Solvate.inp) or 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



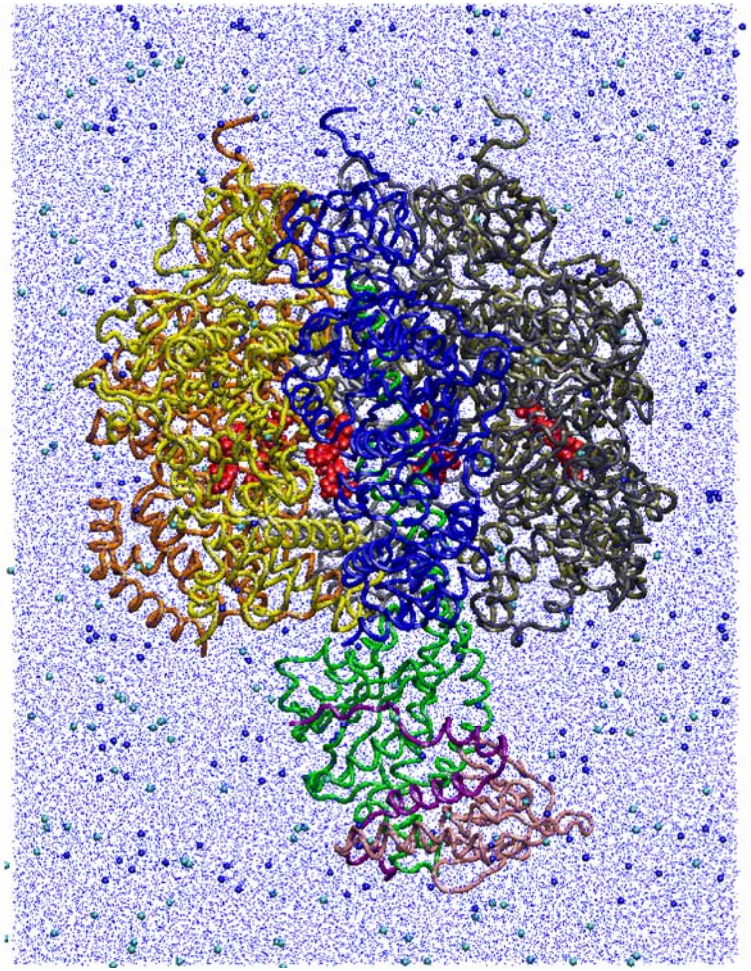
Water Layer vs. Periodic Boundaries

- 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.
-



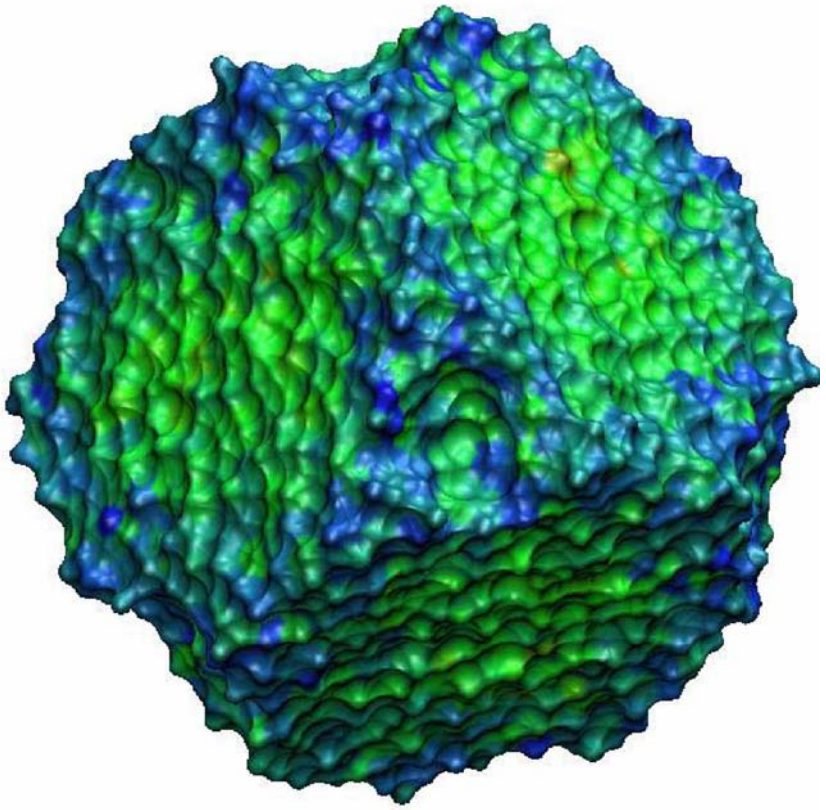
Periodic Boundary Conditions

- 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.



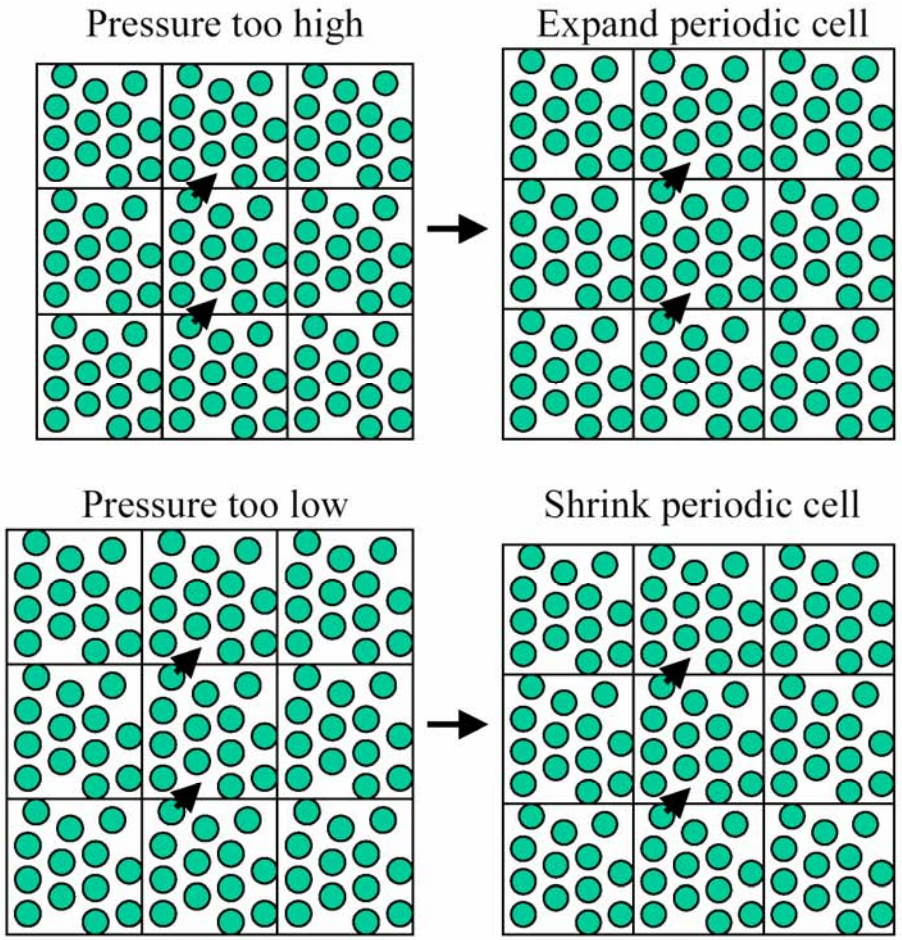
Need for Constant Pressure

- Periodic boundaries are used to eliminate surface effects.
- This assumes that the simulation completely fills the periodic cell.
- A gas can expand to fill any container, but water has a narrow range of densities.
- What happens if the volume we choose for the periodic cell is too large?



Constant Pressure Simulation

- The pressure of a molecular system depends on its volume and temperature.
- Non-periodic systems can adjust themselves in infinite volume and are at zero pressure.
- Periodic systems must use a barostat to vary cell volume and maintain constant pressure.
- Atomic coordinates are rescaled along with cell.



Explicit Solvent: Layer or Periodic?

Water Layer:

+ 1-2 Layers (5-8 Å) often sufficient to hydrate surface, water molecules remain attached to surface (no boiling off)

+ Fewer steric restrictions on global molecular shape

- Thicker layers introduce surface tension

- Boiling off of some water molecules

Periodic Boundary Conditions:

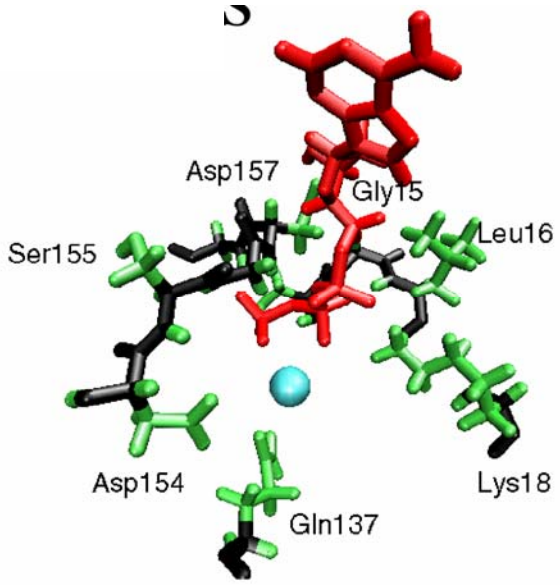
+ Excellent simulation of “infinite” solvent effects

- Restricts conformational dynamics (difficult to sample large changes)

- Requires constant pressure simulations (NAMD, CHARMM, Amber)

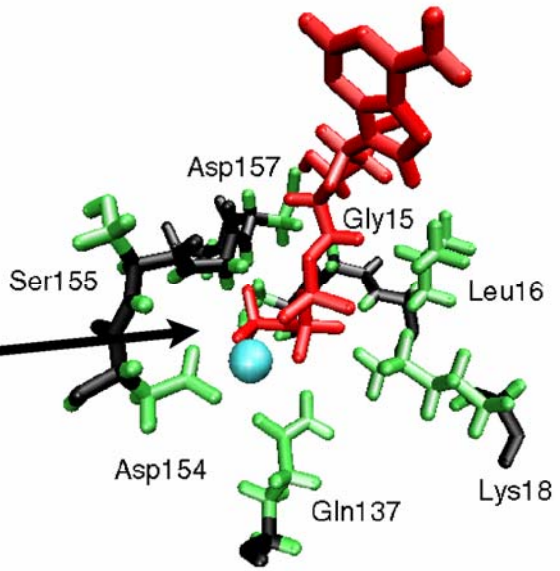
Importance of Buried Solvent Molecules

(a) Actin: Kabsch crystal structure (2.5 Å resolution).



(b) Actin: 500ps simulation.

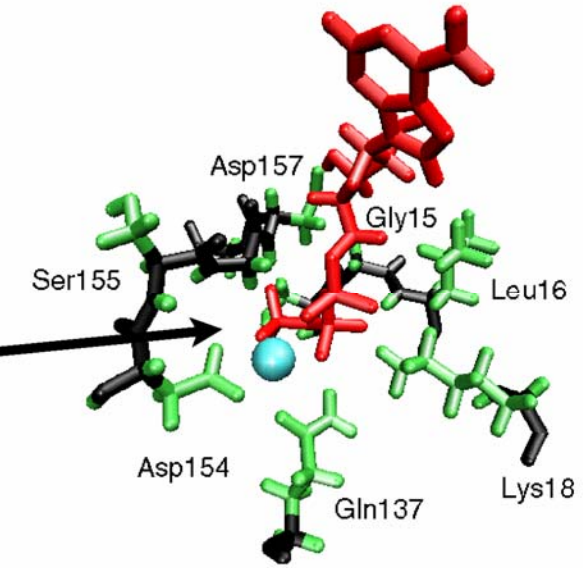
Compaction of enzymatic site



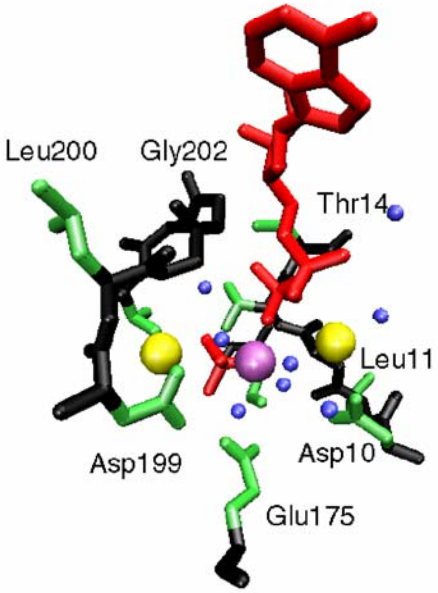
Importance of Buried Solvent Molecules

(b) Actin:
500ps simulation.

Compaction of enzymatic site



(c) Hsc70 crystal structure
(1.7 Å resolution).



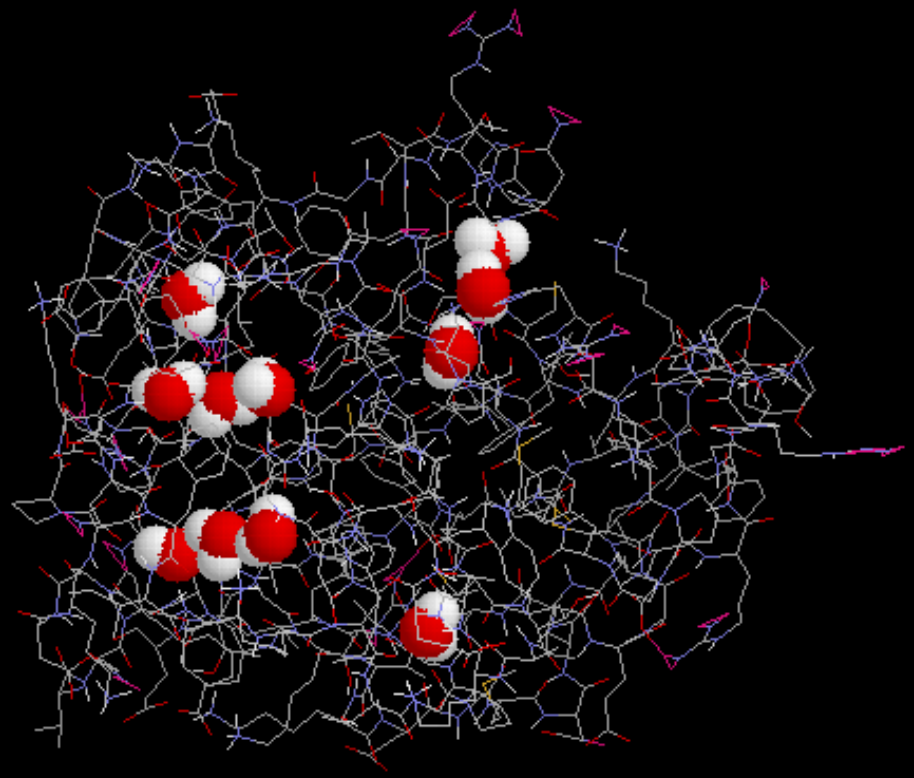
water molecules and ions stabilize enzymatic site

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

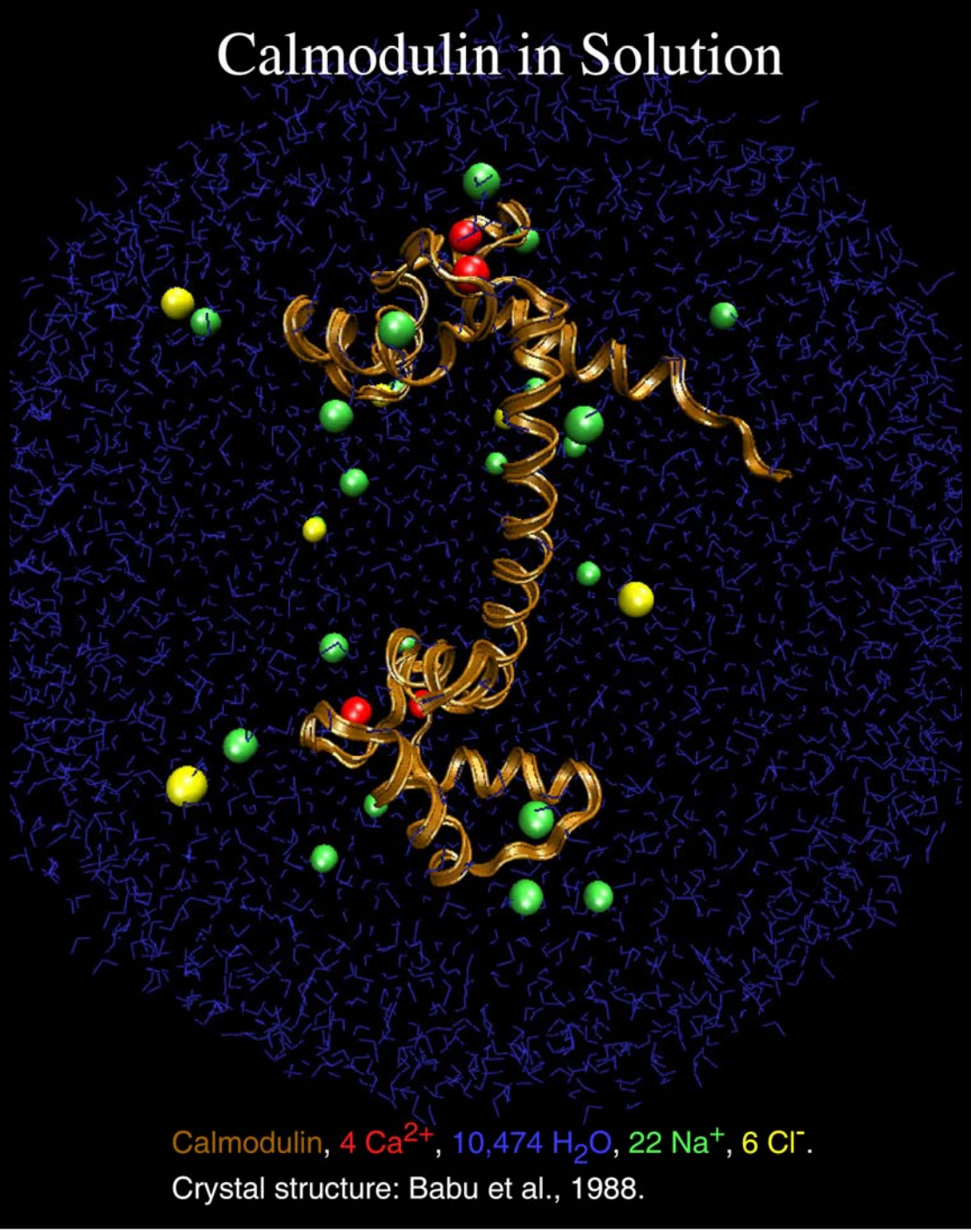


Predicting Ions

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

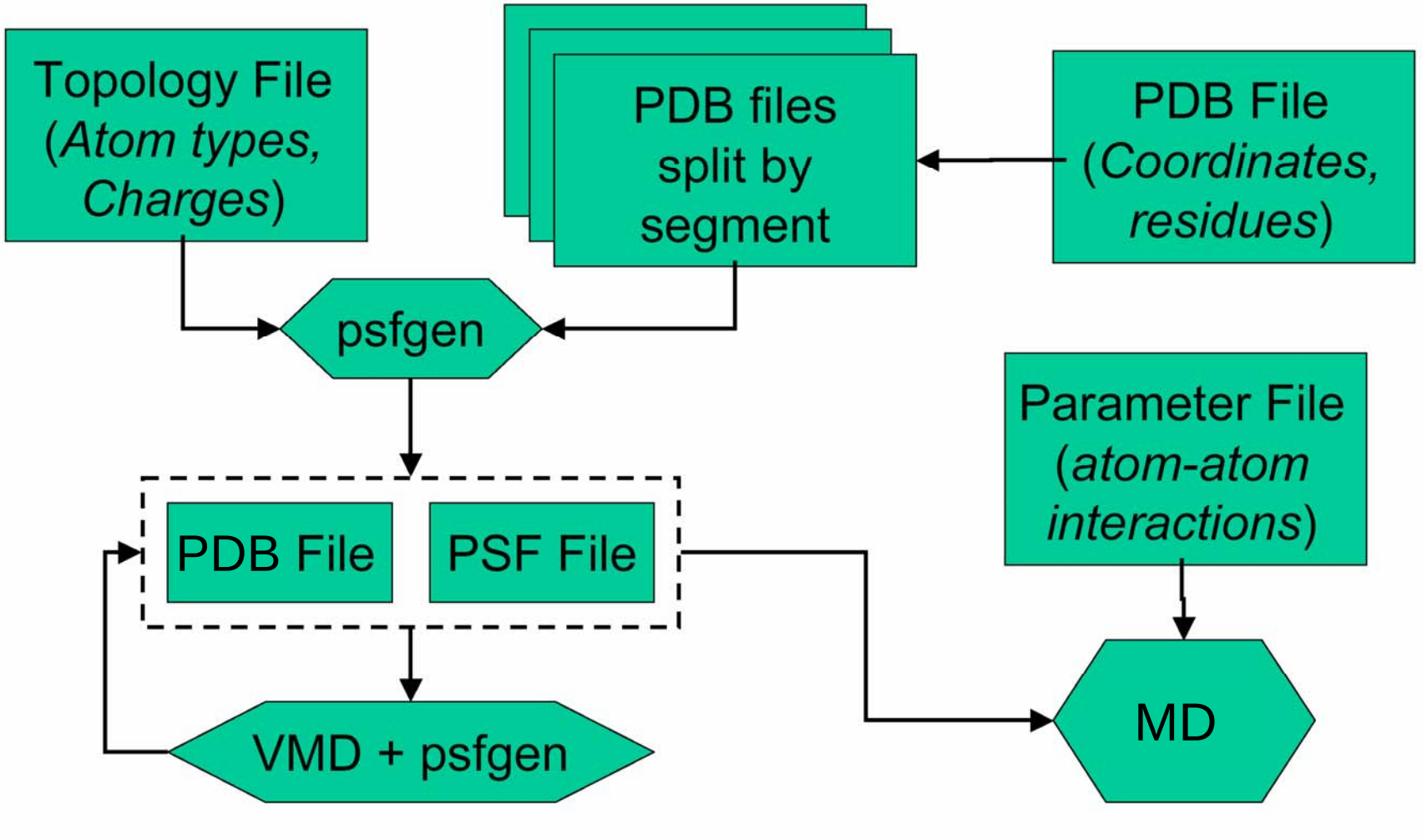
•Once you have all the components (protein water, membrane, etc) combine them into one structure (PDB + PSF), e.g. with X-PLOR script as shown in earlier sessions.

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

Structure Building in VMD: psfgen

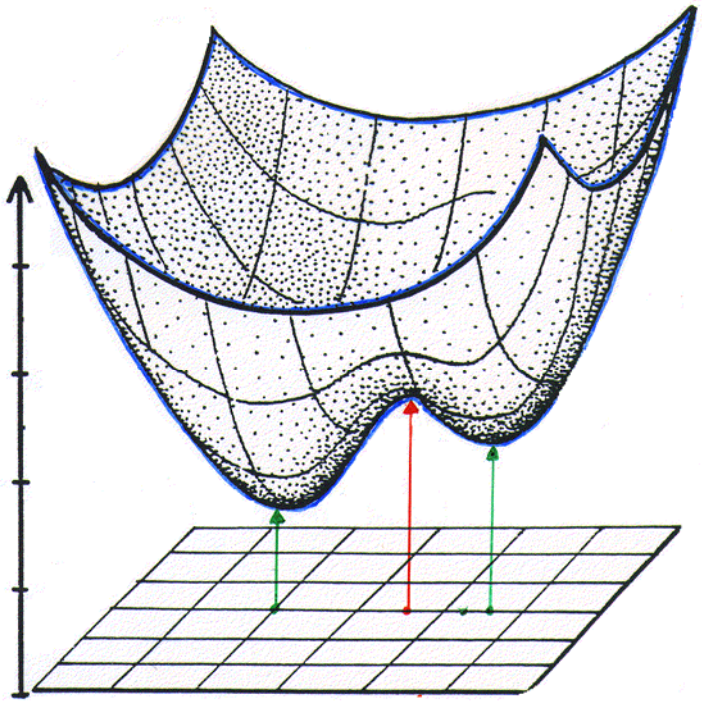
- Tcl script in VMD script library.
- Maps residues to entries in a CHARMM topology file.
- Links residues to form connected segments.
- Combines segments to form a complete structure file.
- Patches residues to form new covalent bonds or modify charge states.
- Guesses coordinates for missing atoms.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

psfgen Flow Chart



Energy Minimization

- First order algorithms:
  - Steepest descent
  - Conjugated gradient
- Second order algorithms:
  - Newton-Raphson
  - Adopted basis Newton Raphson (ABNR)



Why is Minimization Required?

- Most of the coordinates are obtained using X-ray diffraction or NMR.
- The methods do not resolve all atoms of the system (e.g. hydrogens).
- Missing parts are added later using modeling programs, which are not 100% accurate.
- Structures have small deviations from the idealized stereochemistry of the MM energy function.
- Minimization is therefore required to resolve the clashes that may “blow up” the energy function.

Steepest Descent

This is the simplest minimization method:

- The first directional derivative (gradient) of the potential is calculated and displacement is added to every coordinate in the opposite direction (the direction of the force).
- The step is increased if the new conformation has a lower energy.
- Advantages: Simple and fast.
- Disadvantages: Inaccurate, usually does not converge.

Conjugate Gradient

- Uses first derivative information + information from previous steps – the weighted average of the current gradient and the previous step direction.
- The weight factor is calculated from the ratio of the previous and current steps.
- This method converges much better than SD.

Check Results

Minimize guessed atoms:

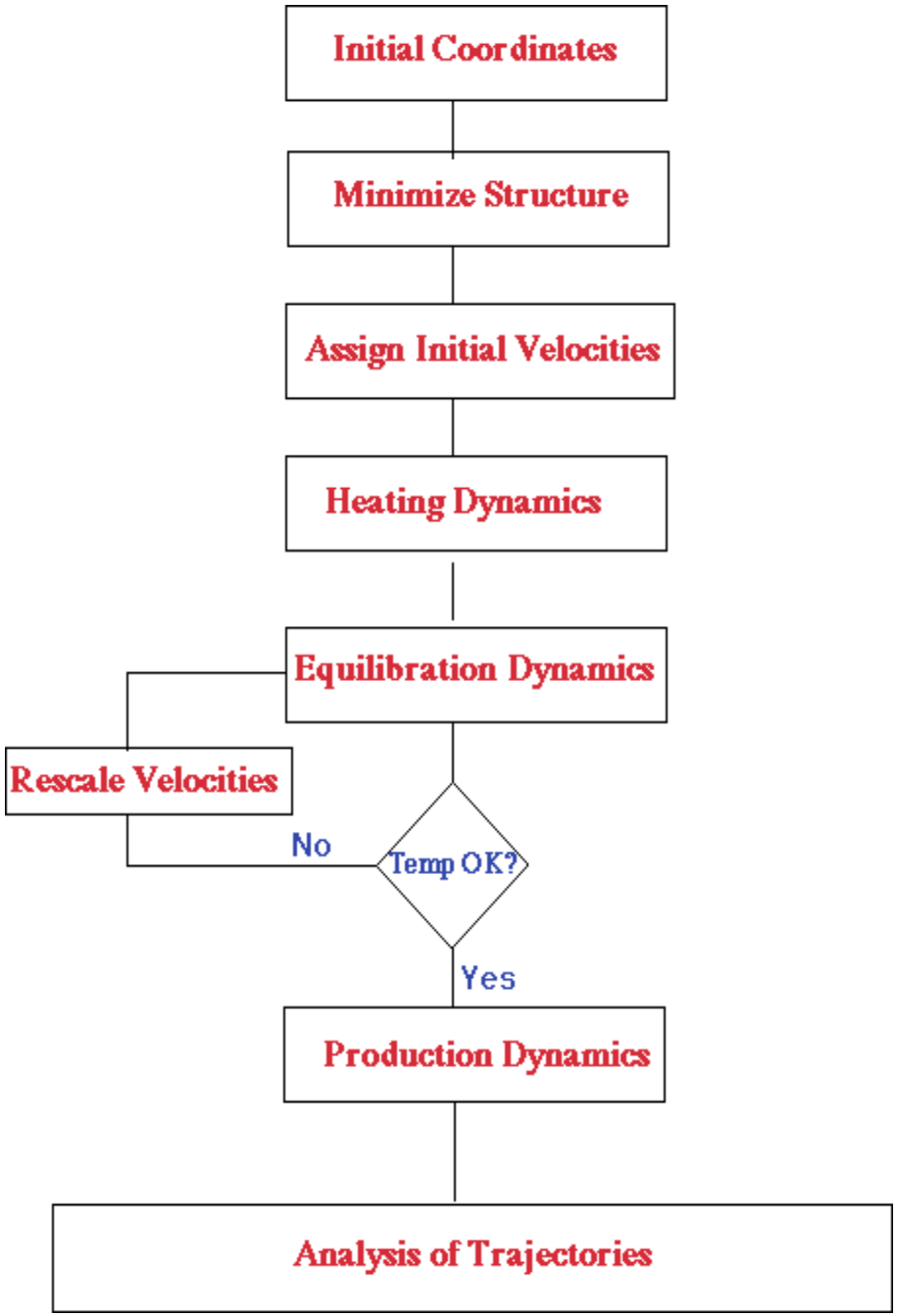
– Large motions indicate bad guesses.

– May indicate confused atom names.

Minimize entire system:

– Look for strange conformations.

– May indicate errors in topology file.



Minimize Energy / Equilibrate

Equilibrating the System

Velocity distribution may change during simulation, especially if the system is far from equilibrium.

- Perform a simulation, scaling the velocities occasionally to reach the desired temperature.

- The system is at equilibrium if:
  - Quantities fluctuate around an average value.
  - The average remains constant over time.

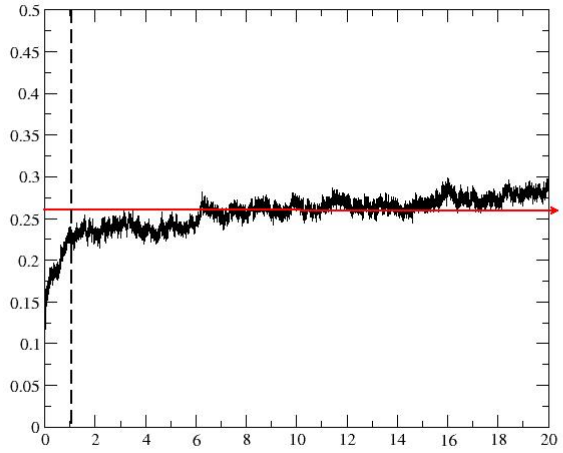
- Variables to monitor:
  - Structural properties (RMSD, order parameters…)
  - Thermodynamics quantities (Potential Energy…)

Root Mean Square Deviation

$$RMSD(t_1, t_2) = \left[\frac{1}{M} \sum_{i=1}^{N} m_i \|\mathbf{r}_i(t_1) - \mathbf{r}_i(t_2)\|^2 \right]^{\frac{1}{2}}$$

Equilibrium:

- No drift
- Oscillations around an equilibrium value

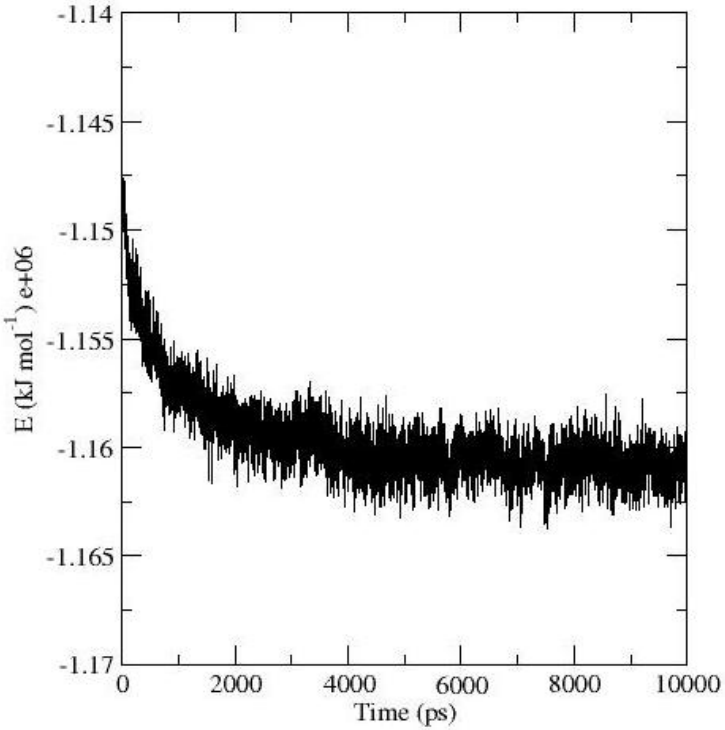


0.259nm±0.016

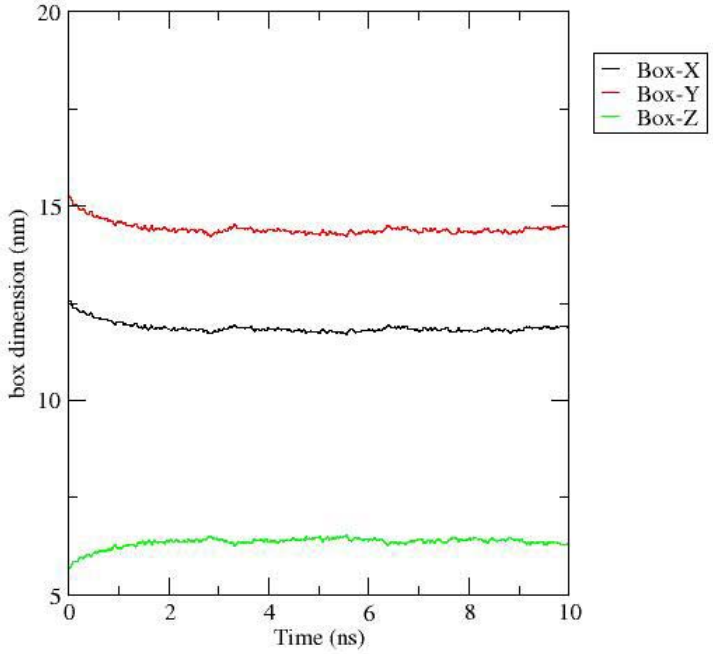
AcrB (*E. coli*) TM region in POPC

Energy and Box Size

Potential energy



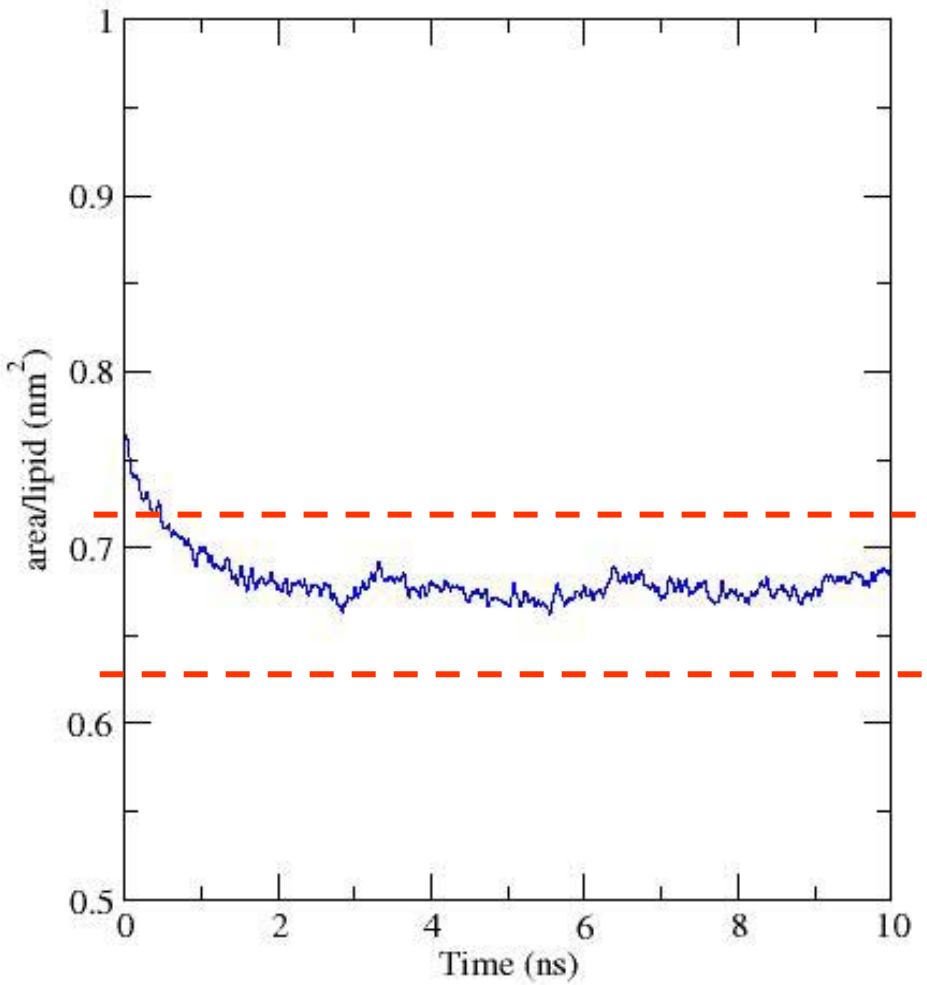
Box dimensions



POPC 502lipids

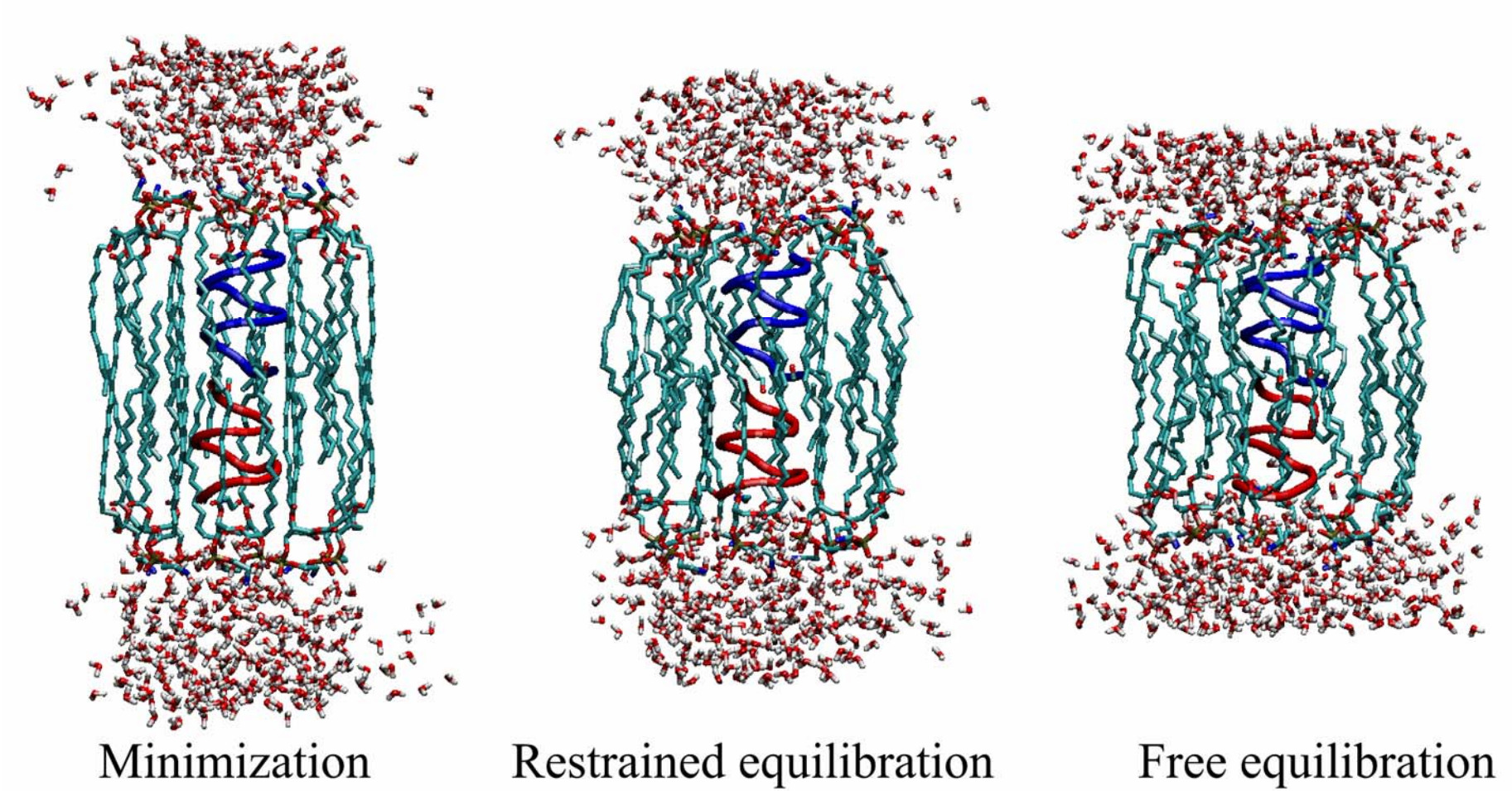
Order Parameters

Area/lipid



POPC 502lipids

Example: Gramicidin A

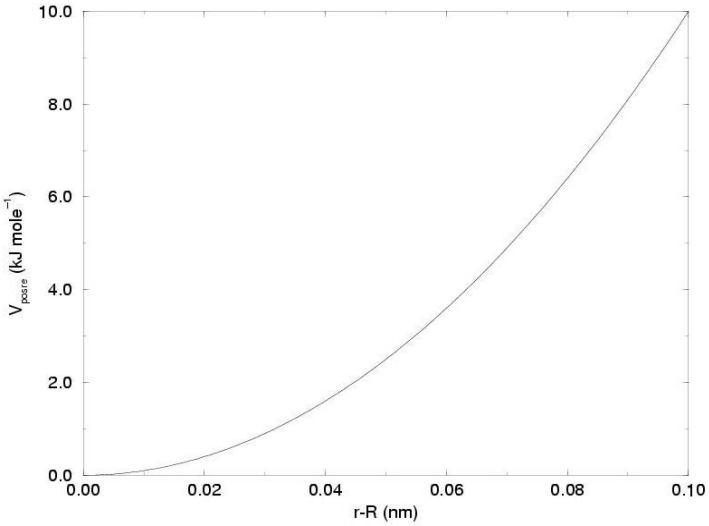


Position Restraints

To a fixed reference position $\boldsymbol{R}_i$: used during equilibration to avoid dramatic rearrangements of some parts of the system

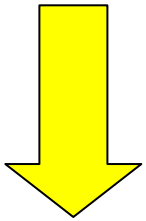
e.g. position restraints to protein after insertion in bilayer to re-equilibrate the lipids

$$V_{pr}(\boldsymbol{r}_i) = \frac{1}{2} k_{pr} |\boldsymbol{r}_i - \boldsymbol{R}_i|^2$$



How Long Should We Simulate?

- The simulation runs are of finite length
- Is the conformational space fully sampled?



Convergence analysis
(depends on what we are looking for)

Resources and Further Reading

WWW:
http://cmm.info.nih.gov/intro_simulation
http://xplor.csb.yale.edu/

Books:
Schlick, Chapters 8, 9, 12, 13
Brunger, X-PLOR Version 3.1, Chapters 1-11
online free at http://alpha2.bmc.uu.se/local_html/xplor_mirror.html

Acknowledgements/Credits

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

Jeremy Smith, Biocomputing Unit, U Heidelberg, Germany

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

Loredana Vaccaro and Mark Samson, U. Oxford