Molecular Dynamics Method 2

Justin Gullingsrud

Theoretical and Computational
Biophysics Group

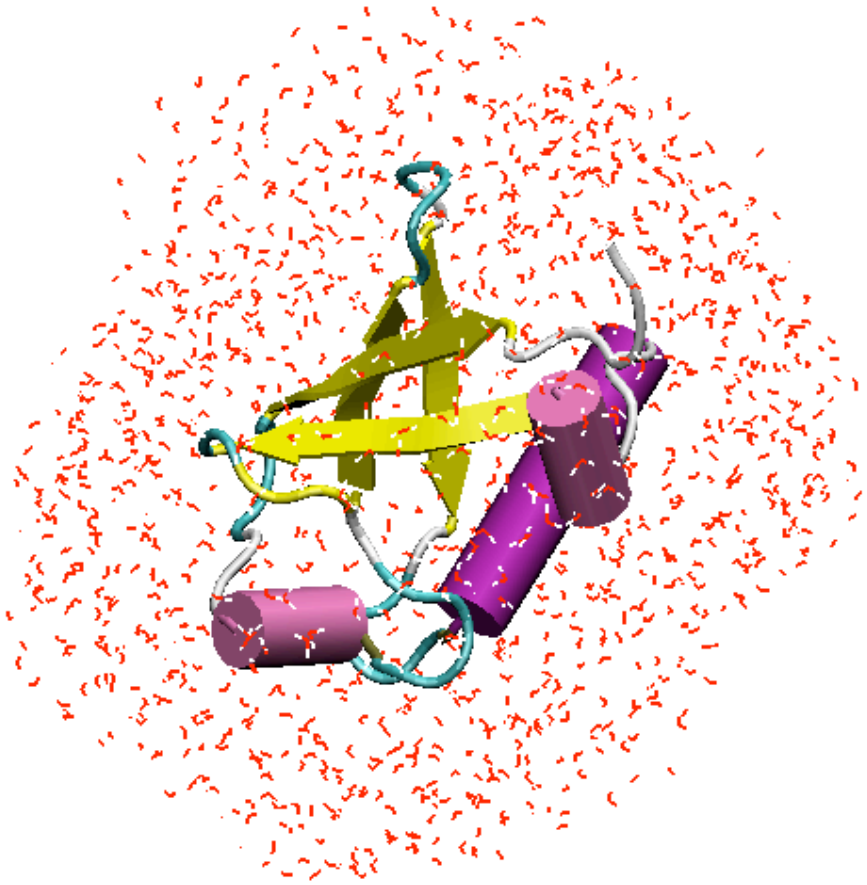
Beckman Institute, UIUC

Why do we need to build molecular structures?

- "I thought PDB files contained structure information already."
- Biomolecules can be represented in a variety of ways; many different force fields can be used to describe their interactions.
- Structure building maps the *abstract representation* of a molecule in a PDB file to a *concrete representation* needed for an MD simulation.
- "If we knew what we were doing, it wouldn't be research."

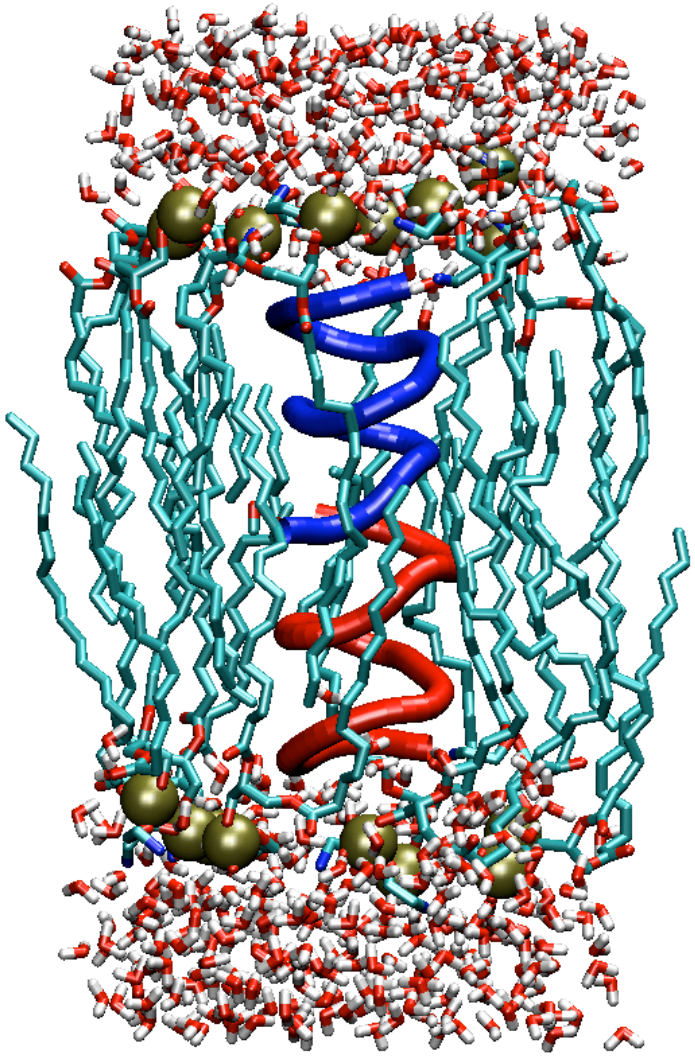
Example: Building Ubiquitin

- 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.



Example: Building 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



General Strategy

- 0. **Decide** what you want to simulate! Determine the components of the simulation (protein, dna, water, ions, lipids, etc.)
- 1. **Build** individual components.
  - Add missing atoms, modify ionization states, graft functional groups onto particular residues, etc.
- 2. **Combine** molecular components.
  - Lipid bilayer
  - Water
  - Ions
- 3. **Minimize.**

Structure building in VMD: psfgen

- 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 for NAMD.

The power of scripting

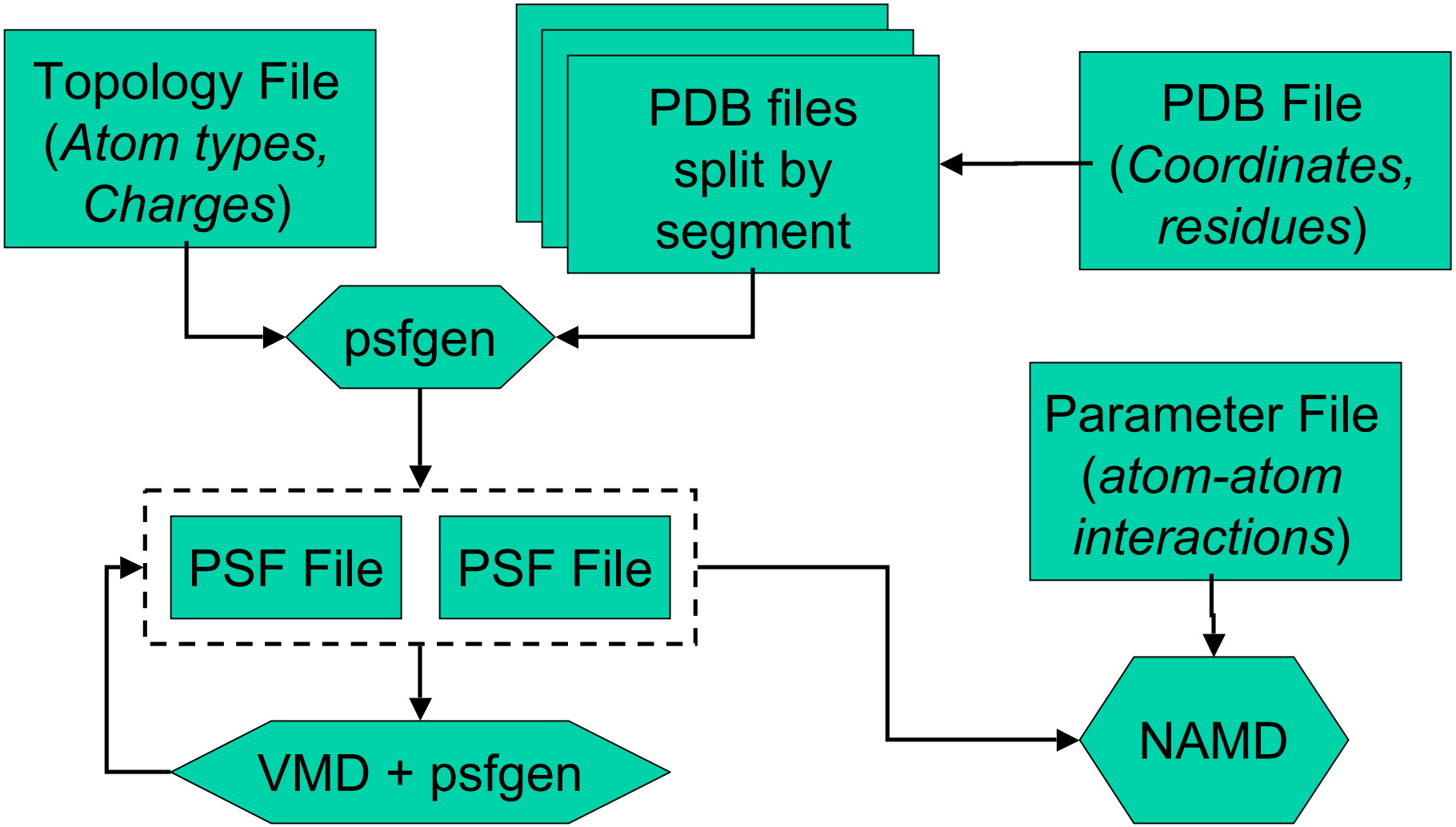
- Tcl is a full-featured scripting language, and psfgen extends Tcl with structure-building commands.
- Running psfgen from within VMD gives you access to VMD's powerful atom selection capabilities.
- You can write Tcl scripts that generate lipid bilayers or automatically solvate proteins.

Running a structure building script

- The name of the structure building package within VMD is `psfgen.`
- To access the structure building commands, your script must contain the line `package require psfgen` as its first command.
- Structure building commands can be freely intermingled with other VMD commands:

```
vmd> set badwat [atomselect top "water and
within 2.4 of protein"]
vmd> foreach segid [$badwat get segid] resid
[$badwat get resid] {
?  delatom $segid $resid]
}
```

Structure building flowchart

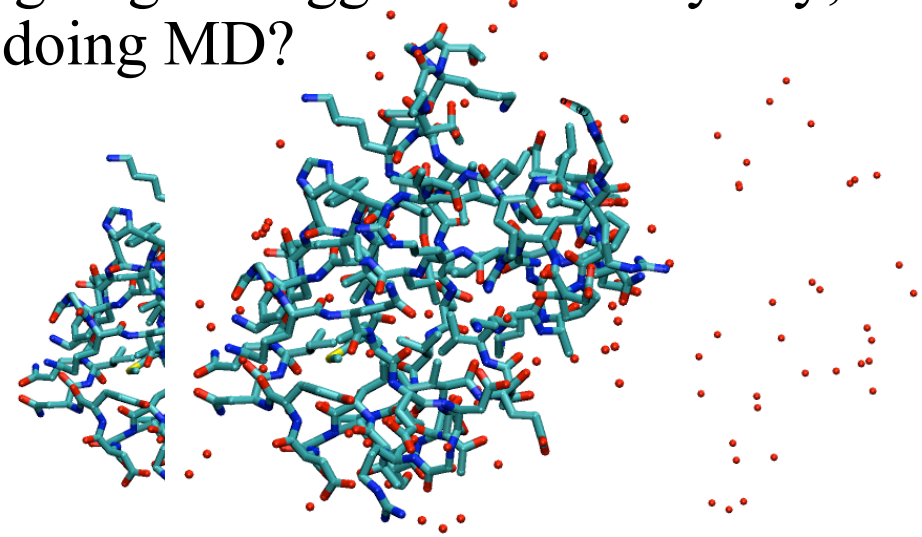


Structure building data files

- Topology files:
  - Atom definitions (just the mass)
  - Residue definitions:
    - atom names, types, and charges;
    - bonds and impropers (but not angles and dihedrals)
  - Patches for initial, terminal and other residues
- PDB file: sequence and coordinate data
- PSF file: Every interaction in the simulation (bonds, angles, dihedrals, etc.)

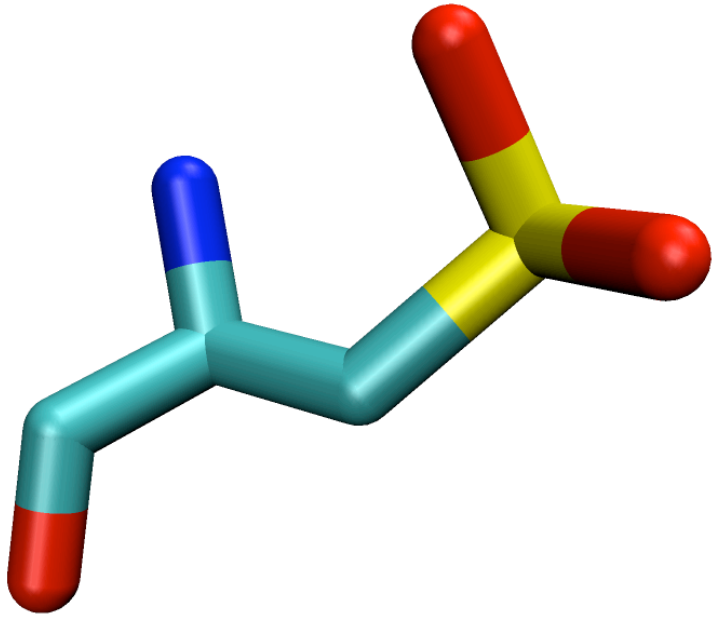
1. **Building** the Protein Structure

- Split the structure into connected segments
- If your structure contains hydrogens… delete them!
  - Positions can be obtained from the topology file
  - Avoid tedious atom name conflicts.
  - They're going to wiggle around anyway; otherwise why are you doing MD?



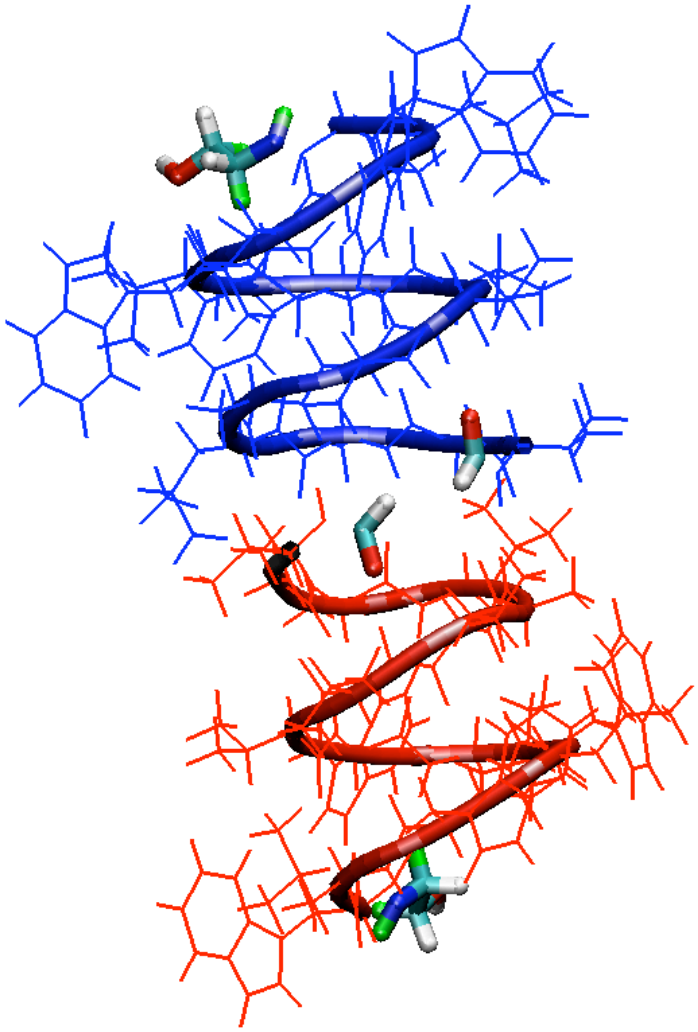
Dealing with Unknown Residues

- 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.



Example: Gramicidin A Peptide

- D-Val and D-Leu residues
- Formyl group at N-terminus, ethanolamide group at C-terminus
- Created new topology, parameter entries by analogy with existing structures and terms.



Correcting atom names

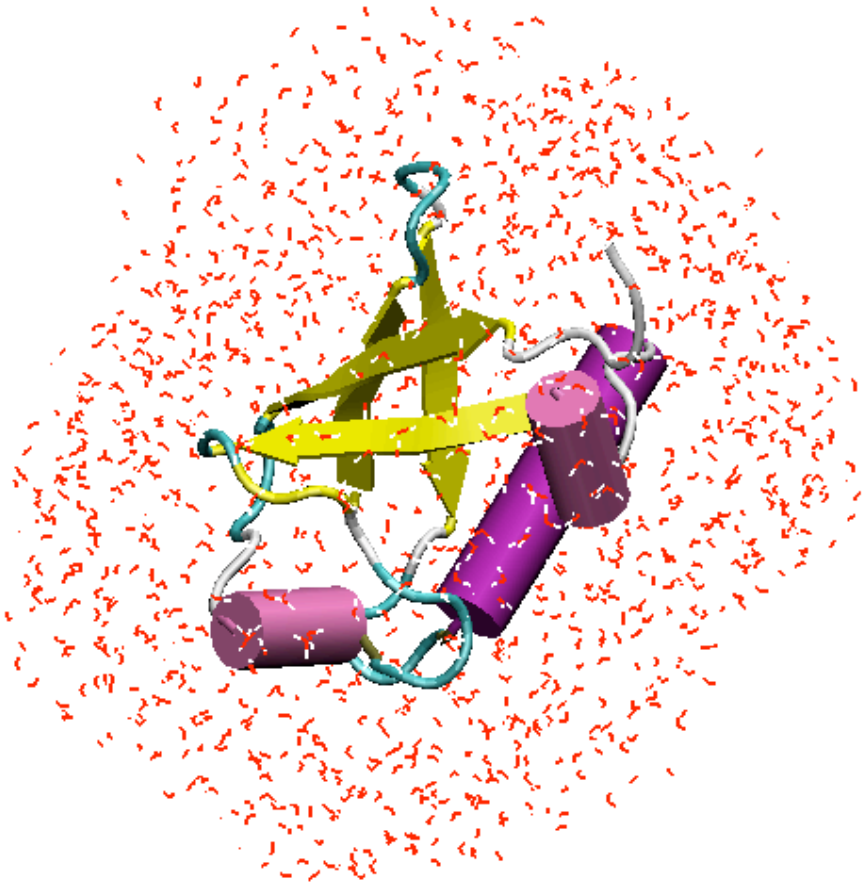
- If errors occur when reading coordinates:
  - Look at source pdb in VMD w/o psf file.
  - Compare guessed structure to topology file.
  - Alias atom names to match.
- Reversed atom names will slip through:
  - Look for strange guessed coordinates.
  - Use two atom aliases to reverse this.

2. **Combining** Simulation Components

- Once you have all the components (protein, water, membrane, etc.), combine them into one structure.
- Load the structure into VMD, and use atom selections to create PDB files containing the atoms you want to keep.
- Use VMD/psfgen to assemble the new PDB files into a reasonable starting configuration.

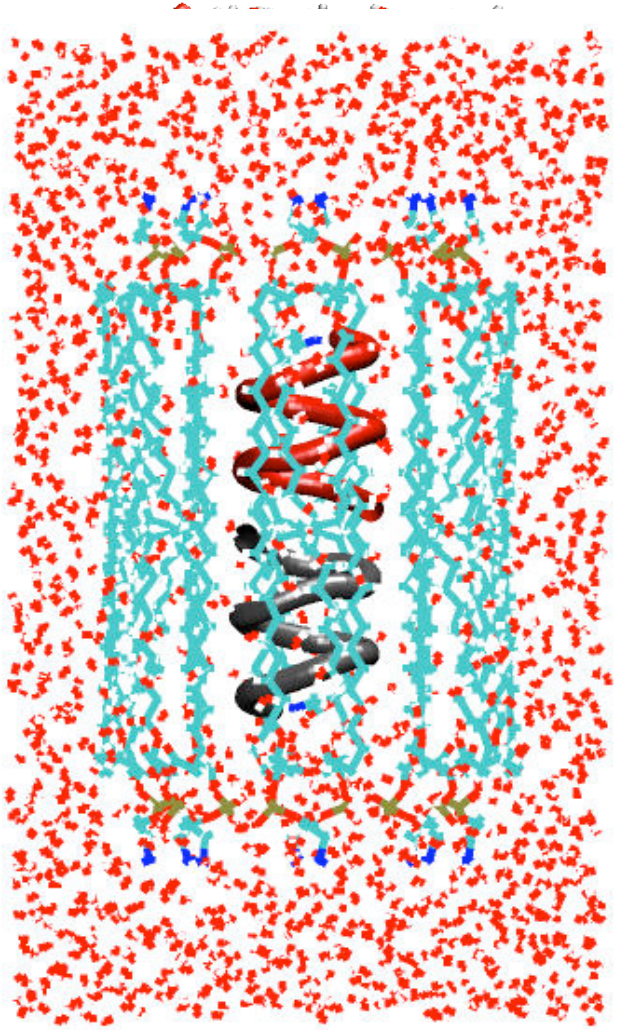
VMD's solvate package

- The *solvate* package uses psfgen commands and VMD's atom selection capabilities.
- The basic building block is a cube of water equilibrated in an NpT ensemble.
- *Solvate* replicates the water box as many times as necessary, renaming segments and removing overlapping atoms.

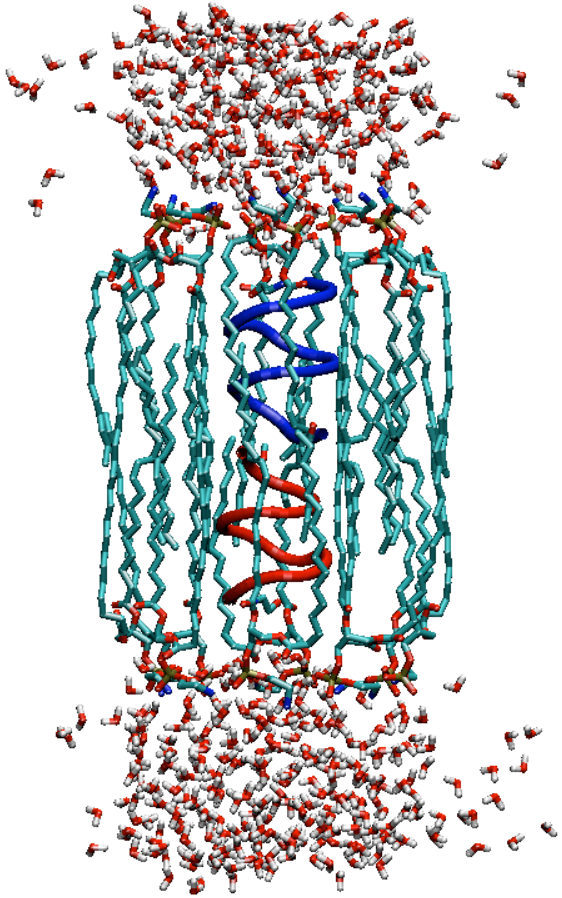


Example: Solvating Gramicidin

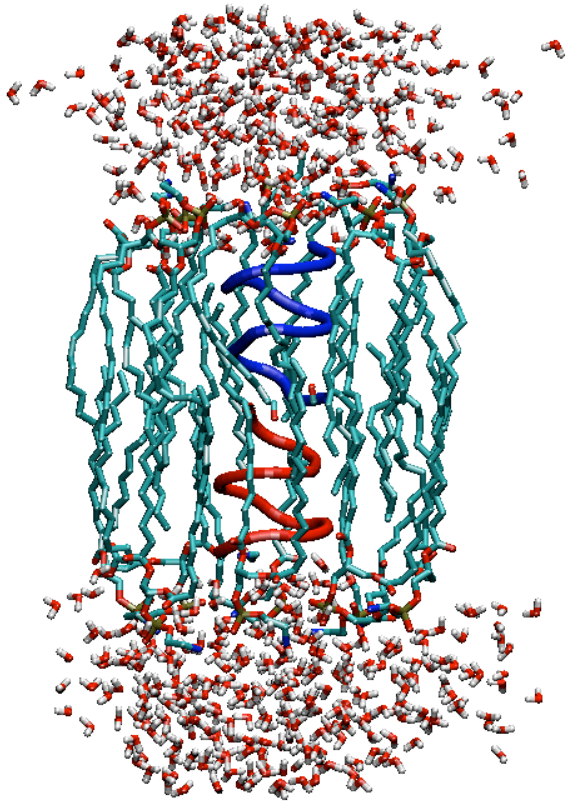
- Begin with a block of equilibrated water.
- Overlay the entire system with the water.
- Chop water outside the desired periodic cell, inside the membrane, and too close to protein or membrane.



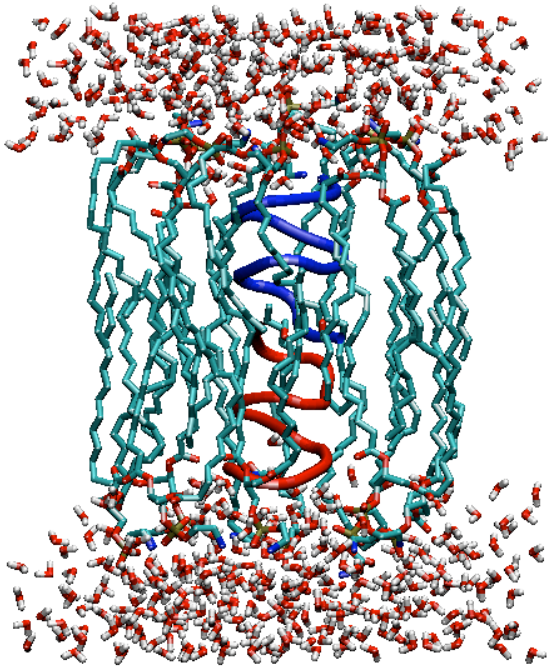
3. **Minimizing** and Equilibrating Gramicidin A



Minimization



Restrained equilibration



Free equilibration

Checking results

- Minimize guessed atoms:
  - Large motions indicate bad guesses.
  - May indicate indicate switched atom names.
- Minimize entire system:
  - Look for strange conformations.
  - May indicate errors in topology file.

Let the production run commence!