Building Molecular Structures for NAMD

Jim Phillips
Theoretical Biophysics Group
Beckman Institute, UIUC

What does psfgen do?

- Read Charmm topology files.
- Read sequences from PDB files.
- Generate atoms, bonds, angles, etc.
- Apply patches to link or modify residues.
- Read coordinates from PDB files.
- Guess coordinates which are missing.
- Write PSF and PDB files for NAMD.

What does psfgen not do?

- Arbitrary manipulation of strucure.
- Translation or rotation of coordinates.
- Selection and deletion of atoms or residues.
- Automatic hydration of molecules.
- Interpretation of non-atom PDB records.
- Determination of protonation states.
- Force and energy evaluation.

Data files for psfgen

- Topology files:
  - list of atom types and masses
  - residue types and definitions
    - atom names, types, and charges
    - bonds and impropers (but not angles and dihedrals)
  - patches for initial, terminal and other residues
- PDB: sequence and coordinate data
- PSF: all but coordinates and parameters

Running psfgen

- Typically run in batch mode:
  psfgen < mkmol.inp >& mkmol.log
- Can be used interactively, but why?
- Generally, building a molecule is iterative:
  psfgen < mkmol.inp |& less
  vmd -psf mol.psf -pdb mol.pdb
  vi mkmol.inp

Tcl in psfgen

- psfgen is just an extended Tcl interpreter.
- Full scripting for input, but…
  - Tcl can't access results or errors.
  - Tcl can't access internal structures.

BPTI walkthrough

- Builds BPTI from Protein Data Bank files.
- Illustrates multiple segments and patches.
- Input PDB files created from 6PTI.pdb via:

  grep -v '^HETATM' 6PTI.pdb >6PTI_protein.pdb

  grep 'HOH' 6PTI.pdb > 6PTI_water.pdb

Topology files

- Input:

  topology toppar/top_all22_prot.inp

- Output:

```
reading topology file toppar/top_all22_prot.inp

>>>>>>>>CHARMM22 All-Hydrogen Topology File for Proteins <<<<<
>>>>>>>>>>>>>>>>>>>>>>> July 1997 <<<<<<<<<<<<<<<<<<<<<<<<<<<<
>>>>>>> Direct comments to Alexander D. MacKerell Jr. <<<<<<<<<
>>>>>> 410-706-7442 or email: alex,mmiris.ab.umd.edu  <<<<<<<<<

Created by CHARMM version 22 1
```

Reading sequences

- Input:

  segment BPTI { pdb output/6PTI_protein.pdb }

- Output:

```
reading residues from pdb file output/6PTI_protein.pdb
extracted 57 residues from pdb file
generating structure at end of segment
no residue 1 before ARG:1 of segment BPTI
add improper failed in residue ARG:1
no residue 1 past GLY:57 of segment BPTI
add bond C(0) N(1) failed in residue GLY:57
```

Applying patches

- Input:

  patch DISU BPTI:5 BPTI:55
  patch DISU BPTI:14 BPTI:38
  patch DISU BPTI:30 BPTI:51

- Output:

```
applying patch DISU to 2 residues
applying patch DISU to 2 residues
applying patch DISU to 2 residues
```

Reading coordinates

- Input:

  alias atom ILE CD1 CD

  alias atom LEU CD1 CD2

  alias atom LEU CD2 CD1

  coordpdb output/6PTI_protein.pdb BPTI

- Output:

```
aliasing residue ILE atom CD1 to CD
aliasing residue LEU atom CD1 to CD2
aliasing residue LEU atom CD2 to CD1
reading coordinates from pdb file output/6PTI_protein.pdb for
   segment BPTI
```

Using aliases

- Alternative to editing the input files.
  - Only affects reading from input files.
  - Output names match topology files.
- Aliasing residue names (for sequence):

  alias residue HIS HSD
- Aliasing atom names (for coordinates):

  alias atom ILE CD1 CD

A segment of water

- Input:

```
alias residue HOH TIP3
segment SOLV {
  auto none
  pdb output/6PTI_water.pdb
}
```

- Output:

```
aliasing residue HOH to TIP3
building segment SOLV
disabling angle autogeneration
disabling dihedral autogeneration
reading residues from pdb file output/6PTI_water.pdb
extracted 73 residues from pdb file
generating structure at end of segment
```

Water coordinates

- Input:

```
alias atom HOH O OH2
coordpdb output/6PTI_water.pdb SOLV
```

- Output:

```
aliasing residue HOH atom O to OH2
reading coordinates from pdb file
  output/6PTI_water.pdb for segment SOLV
```

Guessing coordinates

- Input:

  guesscoord

- Output:

```
guessing coordinates based on topology file
Warning: guessing coordinates for 583 atoms
Warning: poorly guessed coordinates for 151 atoms
```

Writing output

- Input:

  writepsf output/bpti.psf

  writepdb output/bpti.pdb

- Output:

```
writing psf file output/bpti.psf
total of 1101 atoms
total of 1115 bonds
total of 1681 angles
total of 2366 dihedrals
total of 139 impropers
writing pdb file output/bpti.pdb
```

Minimizing guesses

- Atomic coordinates should always be minimized before starting a simulation:

  "minimization on"
- For initial minimization, can minimize only atoms with guessed coordinates:

  "fixedAtoms on"

Understanding IC records

```
IC -C   CA   *N   HN    1.3551 126.4900  180.0000 115.4200  0.9996
IC -C   N    CA   C     1.3551 126.4900  180.0000 114.4400  1.5390
```

- Four atom names (A,B,C,D):
  - optional -/+/# for prev, next, next-of-next
  - number of residue (1,2,3,4) for patches
  - * on third indicates improper version
- Five numbers:

  d(AB), ang(ABC), dihe(ABCD), ang(BCD), d(CD), or
  d(AC), ang(ACB), impr(ABCD), ang(BCD), d(CD)

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.
- Bad IC records for can often be removed.

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.

What is the future of psfgen?

- Better integration with Tcl.
- Automated methods for solvation.
- Incorporation into NAMD front end.
- Incorporation into VMD.