A PDB Files

The term PDB can refer to the Protein Data Bank (http://www.rcsb.org/pdb/), to a data file provided there, or to any file following the PDB format. Files in the PDB include information such as the name of the compound, the species and tissue from which is was obtained, authorship, revision history, journal citation, references, amino acid sequence, stoichiometry, secondary structure locations, crystal lattice and symmetry group, and finally the ATOM and HETATM records containing the coordinates of the protein and any waters, ions, or other heterogeneous atoms in the crystal. Some PDB files include multiple sets of coordinates for some or all atoms. Due to the limits of x-ray crystallography and NMR structure analysis, the coordinates of hydrogen atoms are not included in the PDB.

NAMD and VMD ignore everything in a PDB file except for the ATOM and HETATM records, and when writing PDB files the ATOM record type is used for all atoms in the system, including solvent and ions. Here are the ATOM records for the first two residues of ubiquitin from the 1UBQ entry in the PDB:

```
ATOM      1  N   MET     1      27.340  24.430   2.614  1.00  9.67      1UBQ  71
ATOM      2  CA  MET     1      26.266  25.413   2.842  1.00 10.38      1UBQ  72
ATOM      3  C   MET     1      26.913  26.639   3.531  1.00  9.62      1UBQ  73
ATOM      4  O   MET     1      27.886  26.463   4.263  1.00  9.62      1UBQ  74
ATOM      5  CB  MET     1      25.112  24.880   3.649  1.00 13.77      1UBQ  75
ATOM      6  CG  MET     1      25.353  24.860   5.134  1.00 16.29      1UBQ  76
ATOM      7  SD  MET     1      23.930  23.959   5.904  1.00 17.17      1UBQ  77
ATOM      8  CE  MET     1      24.447  23.984   7.620  1.00 16.11      1UBQ  78
ATOM      9  N   GLN     2      26.335  27.770   3.258  1.00  9.27      1UBQ  79
ATOM     10  CA  GLN     2      26.850  29.021   3.898  1.00  9.07      1UBQ  80
ATOM     11  C   GLN     2      26.100  29.253   5.202  1.00  8.72      1UBQ  81
ATOM     12  O   GLN     2      24.865  29.024   5.330  1.00  8.22      1UBQ  82
ATOM     13  CB  GLN     2      26.733  30.148   2.905  1.00 14.46      1UBQ  83
ATOM     14  CG  GLN     2      26.882  31.546   3.409  1.00 17.01      1UBQ  84
ATOM     15  CD  GLN     2      26.786  32.562   2.270  1.00 20.10      1UBQ  85
ATOM     16  OE1 GLN     2      27.783  33.160   1.870  1.00 21.89      1UBQ  86
ATOM     17  NE2 GLN     2      25.562  32.733   1.806  1.00 19.49      1UBQ  87
```

The fields seen here in order from left to right are the record type, atom ID, atom name, residue name, residue ID, x, y, and z coordinates, occupancy, temperature factor (called beta), segment name, and line number.

If this file is loaded into VMD and then written out as a new file, most of the extra information will be removed, the HETATM records will become ATOM records, and the previously empty chain ID field (between residue name and residue ID) will be set to X (unless present in the original file), and the line number will be omitted, as seen here:

```
ATOM      1  N   MET X   1      27.340  24.430   2.614  1.00  9.67      1UBQ
ATOM      2  CA  MET X   1      26.266  25.413   2.842  1.00 10.38      1UBQ
ATOM      3  C   MET X   1      26.913  26.639   3.531  1.00  9.62      1UBQ
```

B PSF Files

A PSF file, also called a protein structure file, contains all of the molecule-specific information needed to apply a particular force field to a molecular system. The CHARMM force field is divided into a topology file, which is needed to generate the PSF file, and a parameter file, which supplies specific numerical values for the generic CHARMM potential function. The topology file defines the atom types used in the force field; the atom names, types, bonds, and partial charges of each residue type; and any patches necessary to link or otherwise mutate these basic residues. The parameter file provides a mapping between bonded and nonbonded interactions involving the various combinations of atom types found in the topology file and specific spring constants and similar parameters for all of the bond, angle, dihedral, improper, and van der Waals terms in the CHARMM potential function.

The PSF file contains six main sections of interest: atoms, bonds, angles, dihedrals, impropers (dihedral force terms used to maintain planarity), and cross-terms. The following is taken from a PSF file for ubiquitin. First is the title and atom records:

```
PSF CMAP

       6 !NTITLE
 REMARKS original generated structure x-plor psf file
 REMARKS 2 patches were applied to the molecule.
 REMARKS topology top_all27_prot_lipid.inp
 REMARKS segment U { first NTER; last CTER; auto angles dihedrals }
 REMARKS defaultpatch NTER U:1
 REMARKS defaultpatch CTER U:76

    1231 !NATOM
       1 U    1    MET  N    NH3   -0.300000       14.0070           0
       2 U    1    MET  HT1  HC     0.330000        1.0080           0
       3 U    1    MET  HT2  HC     0.330000        1.0080           0
       4 U    1    MET  HT3  HC     0.330000        1.0080           0
       5 U    1    MET  CA   CT1    0.210000       12.0110           0
       6 U    1    MET  HA   HB     0.100000        1.0080           0
       7 U    1    MET  CB   CT2   -0.180000       12.0110           0
```

The fields in the atom section are atom ID, segment name, residue ID, residue name, atom name, atom type, charge, mass, and an unused 0. PSF files may be in either CHARMM or X-PLOR format, with the CHARMM format using an integer rather than a name for the atom type. NAMD requires the X-PLOR format, which is also more flexible since it is not tied to the specific order of atom types in a single parameter file. NAMD and VMD require that the order of atoms in any PDB, DCD, or other atomic data file exactly matches the order found in the PSF file.

Notice that hydrogen atoms are included, that multiple atoms in a residue may share the same type (e.g., HT1, HT2, and HT3 are of type HC), and that

atoms of the same element may have different types (e.g., CA and CB, HA, and HT1). Much of the information in the PDB is also included in the PSF file. Often, the only information from the PDB file used by NAMD is the atomic coordinates. In some cases the contents of the occupancy and beta fields of the PDB file are used as parameters to optional simulation methods.

The covalent bond section lists four pairs of atoms per line:

```
 1237 !NBOND: bonds
    1       5       2       1       3       1       4       1
    5       6       7       5       7       8       7       9
   10       7      10      11      10      12      13      10
```

The angle section lists three triples of atoms per line:

```
 2257 !NTHETA: angles
    1       5       6       1       5      18       2       1       5
    2       1       4       2       1       3       3       1       5
    3       1       4       4       1       5       5      18      19
```

The dihedral and improper sections list two quadruples of atoms per line:

```
 3293 !NPHI: dihedrals
    1       5       7      10       1       5       7       8
    1       5       7       9       1       5      18      20
    1       5      18      19       2       1       5       7

  204 !NIMPHI: impropers
   18       5      20      19      20      18      22      21
   30      32      27      31      30      27      32      31
   32      30      33      34      32      30      34      33
```

The cross-term sections list two quadruples of atoms per line:

```
74 !NCRTERM: cross-terms
     18      20      22      35      20      22      35      37
     35      37      39      54      37      39      54      56
     54      56      58      74      56      58      74      76
```

The order of the atoms within each bond is not significant. The order of atoms within angles, dihedrals, and impropers is significant, but the order can be reversed without affecting the resulting potential. In no case is the relative order of different bonds etc. significant. Since every bond and improper is explicitly listed in the topology file, the atom order within these terms tend to match the original order given. Angles and dihedrals are typically generated automatically and therefore appear in a sorted order.

There is a difference between X-PLOR formatted CHARMM PSF files, which VMD generates with the psfgen module and NAMD uses in conjunction with CHARMM parameter files, and PSF files generated by X-PLOR, which NAMD uses in conjunction with X-PLOR parameter files. Dihedral terms sometimes require multiple sinusoids to represent a torsional potential and therefore

multiple parameter sets appear in the parameter file. In PSF files generated by X-PLOR multiple dihedrals would be indicated by duplicate dihedrals in the PSF file. When using CHARMM PSF and parameter files NAMD extracts any multiple dihedral terms directly from the parameter file and each dihedral appears only once in the PSF file.

C Topology Files

A CHARMM forcefield topology file contains all of the information needed to convert a list of residue names into a complete PSF structure file. It also contains internal coordinates that allow the automatic assignment of coordinates to hydrogens and other atoms missing from a crystal PDB file.

The current versions of the CHARMM forcefield are CHARMM22 for proteins and CHARMM27 for lipids and nucleic acids including CMAP correction to proteins. The individual topology files are named, respectively, `top_all22_prot_cmap.inp`, `top_all27_lipid.rtf`, and `top_all27_na.rtf`. To enable computation on hybrid systems, combinations are also provided, named `top_all27_na_lipid.rtf`, `top_all27_prot_lipid.rtf`, and `top_all27_prot_na.rtf` which can all be found in the CHARMM31 release. While the tools used with NAMD allow multiple topology and parameter files to be used simultaneously, it is preferable to use these pre-combined files. The CHARMM31 release is available for download from the MacKerell web site:

http://mackerell.umaryland.edu/charmm_ff.shtml

We will examine `top_all27_prot_lipid.rtf`; the other files are similar. At the beginning of the file is the header, indicated by lines beginning with *s, followed in this case by "31 1" to indicate the version of CHARMM that generated the file:

```
*> CHARMM22 All-Hydrogen Topology File for Proteins and Lipids <<
*>>>>>> Includes phi, psi cross term map (CMAP) correction <<<<<<
*>>>>>>>>>>>>>>>>>>>>>   July 2004    <<<<<<<<<<<<<<<<<<<<<<<<<
* All comments to ADM jr. via the CHARMM web site: www.charmm.org
*               parameter set discussion forum
*
31  1
```

Comments in the topology file are indicated by ! anywhere on a line. Important usage information is often contained in these comments, so it is always a good idea to inspect the topology file in a text editor when using it to build a structure. The next part of the file is a long set of comments containing references to the papers and other sources of the parameters in the file:

```
! references
!
!PROTEINS
!
!MacKerell, A.D., Jr,. Feig, M., Brooks, C.L., III, Extending the
!treatment of backbone energetics in protein force fields: limitations
!of gas-phase quantum mechanics in reproducing protein conformational
!distributions in molecular dynamics simulations, Journal of
!Computational Chemistry, 25: 1400-1415, 2004.
!
!MacKerell, Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr., R.L.;
```

```
!Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.;
!Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.T.K.; Mattos,
!C.; Michnick, S.; Ngo, T.; Nguyen, D.T.; Prodhom, B.; Reiher, III,
!W.E.; Roux, B.; Schlenkrich, M.; Smith, J.C.; Stote, R.; Straub, J.;
!Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M.  All-atom
!empirical potential for molecular modeling and dynamics Studies of
!proteins.  Journal of Physical Chemistry B, 1998, 102, 3586-3616.
!
```

The topology file must define the type, mass, and charge of every atom in every residue, so that a PSF file can be constructed. While the partial charges assigned to atoms of the same type vary between residues, their masses do not. Therefore, the mass of every atom type is declared once at the beginning of the file in a MASS statement. This statement also pairs an integer with each type name, which is used in CHARMM formatted PSF files, but *not* in the X-PLOR formatted PSF files used by NAMD. The type indices are unique but not necessarily consecutive. Notice in the following except that there are many types of hydrogen and carbon atoms defined, but the atomic masses are the same:

```
MASS     1 H      1.00800 H ! polar H
MASS     2 HC     1.00800 H ! N-ter H
MASS     3 HA     1.00800 H ! nonpolar H
MASS     4 HT     1.00800 H ! TIPS3P WATER HYDROGEN
MASS     5 HP     1.00800 H ! aromatic H
MASS     6 HB     1.00800 H ! backbone H
MASS     7 HR1    1.00800 H ! his he1, (+) his HG,HD2
MASS     8 HR2    1.00800 H ! (+) his HE1
MASS     9 HR3    1.00800 H ! neutral his HG, HD2
MASS    10 HS     1.00800 H ! thiol hydrogen
MASS    11 HE1    1.00800 H ! for alkene; RHC=CR
MASS    12 HE2    1.00800 H ! for alkene; H2C=CR
MASS    13 HA1    1.00800 H ! alkane, CH, new LJ params (see toppar_all22_prot_aliphatic_c27.str)
MASS    14 HA2    1.00800 H ! alkane, CH2, new LJ params (see toppar_all22_prot_aliphatic_c27.str)
MASS    15 HA3    1.00800 H ! alkane, CH3, new LJ params (see toppar_all22_prot_aliphatic_c27.str)
MASS    16 HF1    1.00800 H ! Aliphatic H on fluorinated C (see toppar_all22_prot_fluoro_alkanes.str)
MASS    17 HF2    1.00800 H ! Aliphatic H on fluorinated C (see toppar_all22_prot_fluoro_alkanes.str)
MASS    20 C     12.01100 C ! carbonyl C, peptide backbone
MASS    21 CA    12.01100 C ! aromatic C
MASS    22 CT1   12.01100 C ! aliphatic sp3 C for CH
MASS    23 CT2   12.01100 C ! aliphatic sp3 C for CH2
MASS    24 CT3   12.01100 C ! aliphatic sp3 C for CH3
```

When specifying the connectivity of a chain of residues in a protein, it is necessary to refer to atoms in the previous or succeeding residue. The CHARMM topology file declares those atom types that will be referenced in adjoining residues as such:

```
DECL -CA
```

```
DECL -C
DECL -O
DECL +N
DECL +HN
DECL +CA
```

The first and last residues of a chain obviously have different connectivity from those in the center, since they have one fewer neighbor. This is handled by applying *patch residues*, normally referred to as *patches*, to the terminal residues. As will be seen, any residue definition can specify the patch to be applied when it is the first or last residue in a segment. However, a default set is declared for the entire file, as in the following where the default patch is NTER for the first residue of a segment and CTER for the last:

```
DEFA FIRS NTER LAST CTER
```

While the covalent bond connectivity between atoms must necessarily be provided by the topology file, enumerating all of the required angles and dihedrals would be tedious and error-prone, as well as enormously complicated since every combination of residues joined by a peptide bond would require a different set. Therefore, angles and dihedrals are automatically generated for every pair or triple of connected bonds when a segment is built. This autogeneration may be enabled or disabled on a per-segment basis as it should never be used on segments of water, but the default is defined in the topology file:

```
AUTO ANGLES DIHE
```

We are now ready for the actual residue definitions, beginning with alanine, as shown below. A residue is indicated by the RESI statement with the residue name (ALA) and total charge (0.00). Next are listed all of the atoms in the residue in ATOM statements with the atom name (N, HN, CA), type (NH1, H, CT1), and partial charge (-0.47, 0.31, 0.07). The GROUP statements, dividing the atoms into integer-charge groups, are not used by NAMD and should not be confused with the *hydrogen groups*, each a non-hydrogen atom and all hydrogens bonded to it, that NAMD uses to accelerate distance-testing for nonbonded calculations.

```
RESI ALA          0.00
GROUP
ATOM N    NH1    -0.47  !     |
ATOM HN   H       0.31  !  HN-N
ATOM CA   CT1     0.07  !     |     HB1
ATOM HA   HB      0.09  !     |    /
GROUP                   !  HA-CA--CB-HB2
ATOM CB   CT3    -0.27  !     |    \
ATOM HB1  HA      0.09  !     |     HB3
ATOM HB2  HA      0.09  !   O=C
ATOM HB3  HA      0.09  !     |
GROUP                   !
```

```
ATOM C    C       0.51
ATOM O    O      -0.51
```

The ALA residue continues by defining connectivity, with each BOND statement followed by a list of pairs of atoms to be connected with bonds. The DOUBLE statement is a synonym for BOND and does not affect the resulting PSF file. Observe that the atom C is bonded to +N, the N of the following residue. A bond between N and -C will be provided by the preceding residue. The order of bonds, or of the atoms within a bond, is not significant.

```
BOND CB CA  N  HN  N  CA
BOND C  CA  C  +N  CA HA  CB HB1  CB HB2  CB HB3
DOUBLE O  C
```

As noted above, the angle and dihedral terms will be autogenerated and are therefore not listed for this residue. The less common improper dihedrals (normally just called impropers), however, must be listed explicitly. In this case there are two impropers, which maintain the planarity of the peptide bonds. As with dihedrals, the order of atoms within an improper may be reversed. As shown below, impropers are specified by the IMPR statement followed by sets of four atoms, with the central atom to which the other three are bonded typically listed first.

```
IMPR N -C CA HN  C CA +N O
```

The CMAP correction terms should also be listed explicitly since they are only applied to the backbone dihedrals, indicated following the CMAP statement.

```
CMAP -C  N  CA  C   N  CA  C  +N
```

Explicit hydrogen bond terms are no longer present in the CHARMM force field and are therefore not calculated by NAMD. The DONOR and ACCEPTOR statements, shown below, specify pairs of atoms eligible to form hydrogen bonds. The psfgen module in VMD ignores these statements and does not incorporate hydrogen bonding information into the PSF file.

```
DONOR HN N
ACCEPTOR O C
```

Finally in the residue definition are the internal coordinate IC statements. For each set of four atoms 1 2 3 4, the IC specifies in order the bond length 1-2, the angle 1-2-3, the dihedral 1-2-3-4, the angle 2-3-4, and the bond length 3-4. With this set of data, the position of atom 1 may be determined based on the positions of atoms 2–4, and the position of atom 4 may be determined from the positions of atoms 1–3, allowing the recursive generation of coordinates for all atoms in the structure based on a three-atom seed. Improper IC statements are indicated by a * preceding the third atom, the atom to which the other three are bonded, as in 1 2 *3 4. The order of atoms in an IC statement is different from that of an IMPR statement, and values provide the length 1-3, the angle 1-3-2, the dihedral 1-2-3-4, the angle 2-3-4, and the length 3-4.

```
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
IC N    CA   C    +N    1.4592 114.4400  180.0000 116.8400  1.3558
IC +N   CA   *C   O     1.3558 116.8400  180.0000 122.5200  1.2297
IC CA   C    +N   +CA   1.5390 116.8400  180.0000 126.7700  1.4613
IC N    C    *CA  CB    1.4592 114.4400  123.2300 111.0900  1.5461
IC N    C    *CA  HA    1.4592 114.4400 -120.4500 106.3900  1.0840
IC C    CA   CB   HB1   1.5390 111.0900  177.2500 109.6000  1.1109
IC HB1  CA   *CB  HB2   1.1109 109.6000  119.1300 111.0500  1.1119
IC HB1  CA   *CB  HB3   1.1109 109.6000 -119.5800 111.6100  1.1114
```

It was mentioned above that special treatment is required for the first or last residue in a chain, and that this is implemented as a patch. The following is the default first-residue patch for proteins, NTER. The syntax is almost identical to a normal residue, but the ATOM statements may refer to either add a new atom (HT1, HT2, HT3) or modify the type and charge of an existing atom of the given name (N, CA, HA). The DELETE statement operates as one would expect, removing the atom (HN) and any bonds that include it.

```
PRES NTER         1.00 ! standard N-terminus
GROUP                  ! use in generate statement
ATOM N    NH3    -0.30 !
ATOM HT1  HC      0.33 !         HT1
ATOM HT2  HC      0.33 !     (+)/
ATOM HT3  HC      0.33 ! --CA--N--HT2
ATOM CA   CT1     0.21 !   |    \
ATOM HA   HB      0.10 !   HA    HT3
DELETE ATOM HN
BOND HT1 N HT2 N HT3 N
DONOR HT1 N
DONOR HT2 N
DONOR HT3 N
IC HT1  N    CA   C     0.0000  0.0000  180.0000  0.0000  0.0000
IC HT2  CA   *N   HT1   0.0000  0.0000  120.0000  0.0000  0.0000
IC HT3  CA   *N   HT2   0.0000  0.0000  120.0000  0.0000  0.0000
```

The comment "use in generate statement" indicates that the NTER patch is used during segment generation, and is applied before angles and dihedrals are generated. An example of the other type of patch, those applied after segment generation, is the LINK patch given below. The patch statement requires, in this case, two residues as arguments, the first being the last residue of a segment generated without the CTER patch and the second being the first residue of a segment generated without the NTER patch (the comments have the N and C termini reversed). The numbers 1 and 2 preceding atom names indicate which argument residue the named atom belongs to. Since angles and dihedrals will not be regenerated, they are enumerated.

```
PRES LINK         0.00 ! linkage for IMAGES or for joining segments
                       ! 1 refers to previous (N terminal)
```

```
                    ! 2 refers to next (C terminal)
                    ! use in a patch statement
                    ! follow with AUTOgenerate ANGLes DIHEdrals command
BOND 1C 2N
!the need for the explicit specification of angles and dihedrals in
!patches linking images has not been tested
!ANGLE 1C 2N 2CA  1CA 1C 2N
!ANGLE 1O 1C 2N   1C  2N 2HN
!DIHE 1C  2N  2CA 2C   1C  2N  2CA 2HA  1C  2N  2CA 2CB
!DIHE 1HA 1CA 1C  2N   1N  1CA 1C  2N   1CB 1CA 1C  2N
!DIHE 1CA 1C  2N  2HN  1CA 1C  2N  2CA
!DIHE 1O  1C  2N  2HN  1O  1C  2N  2CA
IMPR 2N 1C 2CA 2HN  1C 1CA 2N 1O
IC 1N   1CA  1C   2N    0.0000  0.0000  180.0000  0.0000  0.0000
IC 2N   1CA  *1C  1O    0.0000  0.0000  180.0000  0.0000  0.0000
IC 1CA  1C   2N   2CA   0.0000  0.0000  180.0000  0.0000  0.0000
IC 1C   2N   2CA  2C    0.0000  0.0000  180.0000  0.0000  0.0000
IC 1C   2CA  *2N  2HN   0.0000  0.0000  180.0000  0.0000  0.0000
```

These types of patches are used to alter protonation states (ASPP, GLUP, HS2), create disulphide bonds (DISU), attach HEME groups (PHEM) and their ligands (PLO2, PLIG), and even remove unwanted autogenerated angles (FHEM).

The following is the complete residue definition for GLY, the smallest of the amino acids. Compare GLY to the ALA residue dissected above.

```
RESI GLY          0.00
GROUP
ATOM N    NH1    -0.47  !     |
ATOM HN   H       0.31  !     N-H
ATOM CA   CT2    -0.02  !     |
ATOM HA1  HB      0.09  !     |
ATOM HA2  HB      0.09  ! HA1-CA-HA2
GROUP                   !     |
ATOM C    C       0.51  !     |
ATOM O    O      -0.51  !     C=O
                        !     |
BOND N HN  N  CA  C CA
BOND C +N  CA HA1 CA HA2
DOUBLE O  C
IMPR N -C  CA HN  C CA   +N O
CMAP -C  N  CA  C   N  CA  C  +N
DONOR HN N
ACCEPTOR O C
IC -C   CA   *N   HN    1.3475 122.8200  180.0000 115.6200  0.9992
IC -C   N    CA   C     1.3475 122.8200  180.0000 108.9400  1.4971
IC N    CA   C    +N    1.4553 108.9400  180.0000 117.6000  1.3479
IC +N   CA   *C   O     1.3479 117.6000  180.0000 120.8500  1.2289
IC CA   C    +N   +CA   1.4971 117.6000  180.0000 124.0800  1.4560
IC N    C    *CA  HA1   1.4553 108.9400  117.8600 108.0300  1.0814
```

```
IC N    C    *CA  HA2   1.4553 108.9400 -118.1200 107.9500  1.0817
PATCHING FIRS GLYP
```

Unlike ALA and most protein residues, in which CA is bonded to HA and CB, in GLY it is bonded to a pair of hydrogens, HA1 and HA2. Also, the hydrogen bonded to N is named H rather than HN. For these reasons, the default NTER patch cannot be applied to GLY and the PATCHING statement is used to change the default first-residue patch for GLY to GLYP. Similarly, the LINK patch above cannot be used for GLY residues, so the additional patches LIG1, LIG2, and LIG3 are provided to link GLY to non-GLY, non-GLY to GLY, and GLY to GLY.

Water is also defined as a residue, as shown below, but some care is required for its use. A third bond, not required by NAMD, is included to allow CHARMM to make the water molecule rigid. This ring structure would confuse angle and dihedral autogeneration, so segments of water must be generated with autogeneration disabled, and therefore an explicit angle is also included. If this third bond is removed from the water topology, then autogeneration must still be disabled to avoid duplicating the angle, unless the angle is also removed.

```
RESI TIP3         0.000 ! tip3p water model, generate using noangle nodihedral
GROUP
ATOM OH2  OT     -0.834
ATOM H1   HT      0.417
ATOM H2   HT      0.417
BOND OH2 H1 OH2 H2 H1 H2    ! the last bond is needed for shake
ANGLE H1 OH2 H2             ! required
ACCEPTOR OH2
PATCHING FIRS NONE LAST NONE
```

The residue definitions for the ions below are exceedingly simple.

```
RESI SOD          1.00 ! Sodium Ion
GROUP
ATOM SOD  SOD     1.00
PATCHING FIRST NONE LAST NONE

RESI CLA         -1.00 ! Chloride Anion
GROUP
ATOM CLA  CLA    -1.00
PATCHING FIRST NONE LAST NONE
```

We have discussed only those parts of the topology files associated with proteins and solvent. There is much additional information regarding proteins, not to mention lipids and nucleic acids, included in the comments in the topology files themselves.

D Parameter Files

A CHARMM forcefield parameter file contains all of the numerical constants needed to evaluate forces and energies, given a PSF structure file and atomic coordinates. The parameter file is closely tied to the topology file that was used to generate the PSF file, and the two are typically distributed together and given matching names.

The current versions of the CHARMM forcefield are CHARMM22 for proteins and CHARMM27 for lipids and nucleic acids including CMAP correction to proteins. The individual parameter files are named, respectively, `par_all22_prot_cmap.inp`, `par_all27_lipid.prm`, and `par_all27_na.prm`. To enable hybrid systems, combinations are also provided, named `par_all27_na_lipid.prm`, `par_all27_prot_lipid.prm`, and `par_all27_prot_na.prm` which can all be found in the CHARMM31 release. While the tools used with NAMD allow multiple topology and parameter files to be used simultaneously, it is preferable to use these pre-combined files. The CHARMM parameters are available for download from the MacKerell web site:

http://mackerell.umaryland.edu/charmm_ff.shtml

We will examine `par_all27_prot_lipid.prm`; the other files are similar. At the beginning of the file is the header, indicated by lines beginning with *s:

```
*>CHARMM22 All-Hydrogen Parameter File for Proteins and Lipids <<
*>>>>> Includes phi, psi cross term map (CMAP) correction <<<<<<<
*>>>>>>>>>>>>>>>>>>>>>   July, 2003   <<<<<<<<<<<<<<<<<<<<<<<<
* All comments to ADM jr. via the CHARMM web site: www.charmm.org
*               parameter set discussion forum
*
```

Comments in the parameter file are indicated by ! anywhere on a line. The next part of the file is a long set of comments containing references to the papers and other sources of the parameters in the file:

```
! references
!
!PROTEINS
!
!MacKerell, A.D., Jr,. Feig, M., Brooks, C.L., III, Extending the
!treatment of backbone energetics in protein force fields: limitations
!of gas-phase quantum mechanics in reproducing protein conformational
!distributions in molecular dynamics simulations, Journal of
!Computational Chemistry, 25: 1400-1415, 2004.
!
!MacKerell, Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr., R.L.;
!Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.;
!Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.T.K.; Mattos,
!C.; Michnick, S.; Ngo, T.; Nguyen, D.T.; Prodhom, B.; Reiher, III,
!W.E.; Roux, B.; Schlenkrich, M.; Smith, J.C.; Stote, R.; Straub, J.;
```

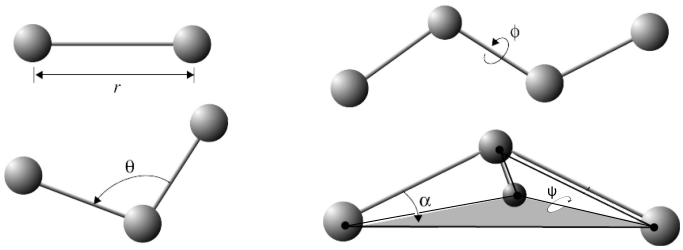

Figure 24: Internal coordinates for bonded interactons: r governs bond stretching; θ represents the bond angle term; ϕ gives the dihedral angle; the small out-of-plane angle α is governed by the so-called "improper" dihedral angle ψ.

```
!Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M.  All-atom
!empirical potential for molecular modeling and dynamics Studies of
!proteins.  Journal of Physical Chemistry B, 1998, 102, 3586-3616.
!
```

The first set of entries in the parameter file are those for bonds, indicated by the BONDS keyword. Each entry consists of a pair of atom types, a spring constant, and an equilibrium length. The bond potential function is $K_b(b-b_0)^2$, where b is the bond length in Angstroms. Bonds are a stiff degree of freedom in biomolecules, so the energy function is only accurate for values near the equilibrium length. Entries are present for every type of bond present in the topology file. Fig. 24 illustrates how bond length, bond angle, dihedral angle and improper angle are defined. The beginning of the bonds section is shown below:

```
BONDS
!
!V(bond) = Kb(b - b0)**2
!
!Kb: kcal/mole/A**2
!b0: A
!
!atom type Kb          b0
!
!Carbon Dioxide
CST  OST   937.96      1.1600 ! JES
!Heme to Sulfate (PSUL) link
SS   FE    250.0       2.3200 !force constant a guess
           !equilbrium bond length optimized to reproduce
           !CSD survey values of
           !2.341pm0.01 (mean, standard error)
           !adm jr., 7/01
C    C     600.000     1.3350 ! ALLOW ARO HEM
```

```
                ! Heme vinyl substituent (KK, from propene (JCS))
CA   CA    305.000     1.3750 ! ALLOW   ARO
                ! benzene, JES 8/25/89
CE1  CE1   440.000     1.3400   !
! for butene; from propene, yin/adm jr., 12/95
CE1  CE2   500.000     1.3420   !
! for propene, yin/adm jr., 12/95
CE1  CT2   365.000     1.5020   !
! for butene; from propene, yin/adm jr., 12/95
CE1  CT3   383.000     1.5040   !
! for butene, yin/adm jr., 12/95
CE2  CE2   510.000     1.3300   !
! for ethene, yin/adm jr., 12/95
```

The CHARMM potential function is designed to confuse physicists, as the form is $K_b(b - b_0)^2$, rather than the traditional $\frac{1}{2}k_b(b - b_0)^2$, and therefore the K_b given in the parameter files is half the value of a traditional spring constant. Also, while the units kcal/mol for energy, Angstroms for length, atomic masses, electron charges, and either fs or ps for time may be convenient for input and output, tortuous unit conversions are required to express the equations of motion.

The next section gives parameters for every type of angle present in the topology file, indicated by the ANGLES keyword. The angle potential function is $K_\theta(\theta - \theta_0)^2$, where θ is the measure of the angle in degrees. Angles are a stiff degree of freedom in biomolecules as well, so the energy function is only accurate for values near the equilibrium angle. Since angles are formed from combinations of bonds, there are many more types of angles than types of bonds. Each entry consists of three atom types, a spring constant, and an equilibrium angle. A small minority of entries also contant Urey-Bradley parameters, which are a spring constant and equilibrium length for a bond-like term between the first and third atoms in the angle. The beginning of the angles section is shown below, with a Urey-Bradley term in the first entry only:

```
ANGLES
!
!V(angle) = Ktheta(Theta - Theta0)**2
!
!V(Urey-Bradley) = Kub(S - S0)**2
!
!Ktheta: kcal/mole/rad**2
!Theta0: degrees
!Kub: kcal/mole/A**2 (Urey-Bradley)
!S0: A
!
!atom types     Ktheta    Theta0   Kub     S0
!
!Carbon Dioxide, JES
OST  CST  OST    3000.00  180.0000 ! CO2, JES
!Heme to Sulfate (PSUL) link
```

```
CS   SS   FE    50.0      100.6    !force constant a guess
                !equilibrium angle optimized to reproduce
                !CSD survey values
                !107.5pm0.6 (mean, standard error)
                !adm jr., 7/01
SS   FE   NPH   100.0       90.0    !force constant a guess
                !adm jr., 7/01
!
CA   CA   CA    40.000    120.00   35.00   2.41620 ! ALLOW   ARO
                ! JES 8/25/89
CE1  CE1  CT2    48.00    123.50   !
                ! for 2-butene, yin/adm jr., 12/95
CE1  CE1  CT3    48.00    123.50   !
! for 2-butene, yin/adm jr., 12/95
CE1  CT2  CT3    32.00    112.20   !
! for 1-butene; from propene, yin/adm jr., 12/95
CE2  CE1  CT2    48.00    126.00   !
! for 1-butene; from propene, yin/adm jr., 12/95
CE2  CE1  CT3    47.00    125.20   !
! for propene, yin/adm jr., 12/95
```

The next section gives parameters for every type of dihedral present in the topology file; there are even more dihedrals than there are angles. Since dihedrals represent the energy of rotation around a covalent bond, which is the source of most conformational flexibility in biomolecules, they must provide a smooth energy for 360 degrees. This is done in most cases with a single sinusoid, $K_\chi(1 + \cos(n(\chi - \delta))$ where χ is the angle between the plane containing the first three atoms in the dihedral and the plane containing the last three. The "multiplicity" n is typically 1, 2, or 3, although for a small number of cases two or three terms with different values of n are provided for the same atom types. You may can observe in the excerpts below that the dihedral spring constants are one to two orders of magnitude lower than for angles, with an order of magnitude difference between flexible and inflexible dihedrals.

```
DIHEDRALS
!
!V(dihedral) = Kchi(1 + cos(n(chi) - delta))
!
!Kchi: kcal/mole
!n: multiplicity
!delta: degrees
!
!atom types             Kchi    n   delta
!
!Heme to Sulfate (PSUL) link
X    FE   SS   X        0.0000  4     0.00 ! guess
                !adm jr., 7/01
X    CS   SS   X        0.0000  3     0.20 ! guess
                !from methanethiol, HS S CT3 HA
```

```
               !adm jr., 7/01

C    CT1  NH1  C        0.2000  1   180.00 ! ALLOW PEP
               ! ala dipeptide update for new C VDW Rmin, adm jr., 3/3/93c
C    CT2  NH1  C        0.2000  1   180.00 ! ALLOW PEP
               ! ala dipeptide update for new C VDW Rmin, adm jr., 3/3/93c
C    N    CP1  C        0.8000  3     0.00 ! ALLOW PRO PEP
               ! 6-31g* AcProNH2, ProNH2, 6-31g*//3-21g AcProNHCH3 RLD 4/23/93
CA   CA   CA   CA       3.1000  2   180.00 ! ALLOW   ARO
               ! JES 8/25/89
CA   CPT  CPT  CA       3.1000  2   180.00 ! ALLOW   ARO
               ! JWK 05/14/91 fit to indole
```

Because of the large numbers of dihedral terms required to describe a complete protein, the wildcard atom type X is occasionally used. These parameters will be used in NAMD if a more specific match is not found elsewhere in the parameter file.

```
!X    C    C    X        4.0000  2   180.00 ! ALLOW HEM
               ! Heme (6-liganded): substituents (KK 05/13/91)
X    C    NC2  X        2.2500  2   180.00 ! ALLOW   PEP POL ARO
               ! 9.0->2.25 GUANIDINIUM (KK)
X    CD   OH1  X        2.0500  2   180.00 ! ALLOW   PEP POL ARO ALC
               ! adm jr, 10/17/90, acetic acid C-Oh rotation barrier
X    CD   OS   X        2.0500  2   180.00 ! ALLOW   PEP POL
               ! adm jr. 3/19/92, from lipid methyl acetate
X    CE1  CE1  X        0.1500  1     0.00
               ! 2-butene, adm jr., 2/00 update
X    CE1  CE1  X        8.5000  2   180.00
               ! 2-butene, adm jr., 2/00 update
X    CE2  CE2  X        4.9000  2   180.00 !
! for ethene, yin/adm jr., 12/95
```

The final bond-like terms in the parameter file are impropers, which are used exclusively and explicitly in the molecular topology to maintain planarity. As such, the harmonic form $K_\psi(\psi - \psi_0)^2$ with a large spring constant and ψ_0 typically zero is used to restrain deformations among an atom and three atoms bonded to it. As with dihedrals, ψ is angle between the plane containing the first three atoms and the plane containing the last three. Notice below that wildcard atom types occur in the second and third positions, rather than the first and fourth as in dihedrals.

```
IMPROPER
!
!V(improper) = Kpsi(psi - psi0)**2
!
!Kpsi: kcal/mole/rad**2
!psi0: degrees
!note that the second column of numbers (0) is ignored
```

```
!
!atom types           Kpsi                   psi0
!
CPB  CPA  NPH  CPA    20.8000         0      0.0000 ! ALLOW HEM
                ! Heme (6-liganded): porphyrin macrocycle (KK 05/13/91)
CPB  X    X    CE1    90.0000         0      0.0000 ! ALLOW HEM
                ! Heme (6-liganded): substituents (KK 05/13/91)
CT2  X    X    CPB    90.0000         0      0.0000 ! ALLOW HEM
                ! Heme (6-liganded): substituents (KK 05/13/91)
CT3  X    X    CPB    90.0000         0      0.0000 ! ALLOW HEM
                ! Heme (6-liganded): substituents (KK 05/13/91)
!HA   C    C    HA     20.0000         0      0.0000 ! ALLOW   PEP POL ARO
```

The next section gives the interpolation values for CMAP, which is a correction map to the backbone dihedral energy. Each ϕ and ψ dihedral angle, ranging from -180 to 180, is divided into 24 grid points; the energy correction at each point is given. A continuous function can be constructed using these values along with an interpolation formula, which can be found in the references below.

```
CMAP
! 2D grid correction data.  The following surfaces are the correction
! to the CHARMM22 phi, psi alanine, proline and glycine dipeptide surfaces.
! Use of CMAP requires generation with the topology file containing the
! CMAP specifications along with version 31 or later of CHARMM.  Note that
! use of "skip CMAP" yields the charmm22 energy surfaces.
!
! references
!MacKerell, A.D., Jr., Feig, M., Brooks, C.L., III, Accurate Treatment of
!Protein Backbone Conformational Energetics in Empirical Force Fields, Submitted
!for publication.

!MacKerell, A.D., Jr., Feig, M., Brooks, C.L., III, Improved Treatment of the
!Protein Backbone in Empirical Force Fields, Journal of the American Chemical
!Society, In Press.


! alanine map
C    NH1  CT1  C    NH1  CT1  C    NH1   24

!-180
0.126790 0.768700 0.971260 1.250970 2.121010
2.695430 2.064440 1.764790 0.755870 -0.713470
0.976130 -2.475520 -5.455650 -5.096450 -5.305850
-3.975630 -3.088580 -2.784200 -2.677120 -2.646060
-2.335350 -2.010440 -1.608040 -0.482250

!-165
-0.802290 1.377090 1.577020 1.872290 2.398990
2.461630 2.333840 1.904070 1.061460 0.518400
```

```
-0.116320 -3.575440 -5.284480 -5.160310 -4.196010
-3.276210 -2.715340 -1.806200 -1.101780 -1.210320
-1.008810 -0.637100 -1.603360 -1.776870
```

Finally we come to the nonbonded interaction parameters. The NONBONDED statement includes a list of parameters which are used as defaults by the program CHARMM, but are ignored by NAMD. Those shown below correspond to the NAMD settings exclude scaled1-4, switching on, pairlistdist 14.0, cutoff 12.0, switchdist 10.0, dielectric 1.0, and 1-4scaling 1.0:

```
NONBONDED nbxmod  5 atom cdiel shift vatom vdistance vswitch -
cutnb 14.0 ctofnb 12.0 ctonnb 10.0 eps 1.0 e14fac 1.0 wmin 1.5
               !adm jr., 5/08/91, suggested cutoff scheme
```

Recall that the partial charge of each atom is specified in the topology and PSF files and is independent of the atom type. Therefore the only type-based parameters are for the van der Waals interactions, which are represented by the classic Lennard-Jones potential (expressed in the somewhat unconventional form) $\epsilon[(r_{min}/r)^{12} - 2(r_{min}/r)^6]$. Observe that at $r = r_{min}$ the force is zero and the energy is $-\epsilon$. Rather than providing a different value of epsilon could be provided for every possible combination of atom types, only one value is provided per type and inter-type interactions are calculated using the sum of the radii $r_{min}/2$ and the geometric mean of the well-depths ϵ. By convention, the ϵ values are negative in the parameter file, as seen here:

```
!
!V(Lennard-Jones) = Eps,i,j[(Rmin,i,j/ri,j)**12 - 2(Rmin,i,j/ri,j)**6]
!
!epsilon: kcal/mole, Eps,i,j = sqrt(eps,i * eps,j)
!Rmin/2: A, Rmin,i,j = Rmin/2,i + Rmin/2,j
!
!atom  ignored    epsilon      Rmin/2   ignored   eps,1-4       Rmin/2,1-4
!
C      0.000000  -0.110000     2.000000 ! ALLOW   PEP POL ARO
               ! NMA pure solvent, adm jr., 3/3/93
CA     0.000000  -0.070000     1.992400 ! ALLOW   ARO
               ! benzene (JES)
CC     0.000000  -0.070000     2.000000 ! ALLOW   PEP POL ARO
               ! adm jr. 3/3/92, acetic acid heat of solvation
CD     0.000000  -0.070000     2.000000 ! ALLOW  POL
               ! adm jr. 3/19/92, acetate a.i. and dH of solvation
CE1    0.000000  -0.068000     2.090000 !
               ! for propene, yin/adm jr., 12/95
CE2    0.000000  -0.064000     2.080000 !
               ! for ethene, yin/adm jr., 12/95
CM     0.000000  -0.110000     2.100000 ! ALLOW HEM
               ! Heme (6-liganded): CO ligand carbon (KK 05/13/91)
```

When the scaled1-4 exclusion policy is used (as it should with the CHARMM force field) nonbonded interactions of atoms separated by three bonds (i.e.,

atoms 1 and 4 in the chain 1-2-3-4) are modified. Even if the scaling factor for electrostatics is 1.0 (as it should be for modern CHARMM force fields), special modified van der Waals parameters are used for 1-4 pairs of atoms for which they are specified, as in the examples below.

```
!atom  ignored    epsilon      Rmin/2   ignored   eps,1-4       Rmin/2,1-4
!
CP1    0.000000  -0.020000     2.275000   0.000000  -0.010000     1.900000 ! ALLOW   ALI
                ! alkane update, adm jr., 3/2/92
CP2    0.000000  -0.055000     2.175000   0.000000  -0.010000     1.900000 ! ALLOW   ALI
                ! alkane update, adm jr., 3/2/92
CP3    0.000000  -0.055000     2.175000   0.000000  -0.010000     1.900000 ! ALLOW   ALI
                ! alkane update, adm jr., 3/2/92
```

The parameter file ends with a reference to parameters for explicit hydrogen bond energy terms. These are obsolete, no longer present in the CHARMM force field, and therefore not implemented by NAMD.

```
HBOND CUTHB 0.5  ! If you want to do hbond analysis (only), then use
                 ! READ PARAM APPEND CARD
                 ! to append hbond parameters from the file: par_hbond.inp

END
```

For information on how the parameters have been derived, you must consult the corresponding publications referenced in the parameter files themselves or listed on the MacKerell web site http://mackerell.umaryland.edu/charmm_ff.shtml.

E NAMD Configuration Files

The NAMD configuration file (also called a config file, .conf file, or .namd file) is given to NAMD on the command line and specifies virtually everything about the simulation to be done. The only exceptions are details relating to the parallel execution environment, which vary between platforms. Therefore, the config file should be portable between machines, platforms, or numbers of processors in a run, as long as the referenced input files are available.

As a convenience, on startup NAMD will switch to the directory that contains the config file, so that all file paths are relative to that directory. NAMD also accepts multiple config files on the command line, but changes directories before parsing each file, so it is best to keep everything in the same directory when using multiple config files.

NAMD parses its configuration file using the Tcl scripting language, which you should be familiar with if you have done any serious work using VMD. However, to remain compatible with config files from earlier version of NAMD which did not include Tcl, several compromises have been made. Tcl is a case sensitive language, so the `$temp` and `$TEMP` are different variables. NAMD's config file options are implemented as case-insensitive Tcl commands, however, so `temperature 300`, `Temperature 300`, or even `tEmPerATURe 300` will all have the same effect.

NAMD config files were historically order-independent, and the whole file was parsed before any files were read or calculations done. Now that NAMD incorporates more advanced scripting, the config file is parsed up to the point where one of the case-sensitive "action" commands such as `minimize` or `run` are encountered; at this point the input files are read and real calculation begins. Beyond this point, the config file is order-dependent and only these action commands and a subset of configuration options may be given. Also, at this point any comments must follow the rules of Tcl, using a semicolon to end a command before using # to start a comment on the same line; it is good practice to follow these rules everywhere in the file.

We will now walk through a NAMD configuration file template that contains options for common simulation methods in NAMD. A version of this file should be available in the tutorial materials as `sample.conf`.

First we specify the files that contain the molecular structure and initial conditions. A `coordinates` PDB file is always required, even if the actual coordinates will come from a binary coordinates file.

```
structure          mypsf.psf
coordinates        mypdb.pdb
```

Setting the Tcl variable `$temperature` makes it easy to change the target temperature for many options.

```
set temperature    310  ;# target temperature used several times below
```

If we are starting from scratch, we'll use the coordinates from the PDB file and take velocities randomly from a Boltzmann distribution, using the `$temperature` variable.

```
# starting from scratch
temperature         $temperature    ;# initialize velocities randomly
```

Otherwise we'll read in the binary output files of the previous run. We use the Tcl variable `$inputname` to avoid errors in typing the input file names, since we will end up copying and modifying this file several times during the course of a long simulation. The `extendedSystem` file holds the periodic cell dimensions that are needed to continue after a constant pressure simulation. In order to start numbering timesteps with the final step of the previous run, we use `firsttimestep` to specify the timestep on which the input coordinates and velocities were written. The number of steps calculated in this run will be `numsteps - firsttimestep`.

```
# continuing a run
set inputname      myinput          ;# only need to edit this in one place!
binCoordinates     $inputname.coor  ;# coordinates from last run (binary)
binVelocities      $inputname.vel   ;# velocities from last run (binary)
extendedSystem    $inputname.xsc   ;# cell dimensions from last run
firsttimestep      50000            ;# last step of previous run
numsteps           100000           ;# run stops when this step is reached
```

The `outputName` prefix will be used to create all of the trajectory (.dcd, .xst), output (.coor, vel, .xsc), and restart (.restart.coor, .restart.vel, .restart.xsc) files generated by this run. Be careful that this is different from `$myinput`, or the job will overwrite its input files!

```
outputName         myoutput   ;# base name for output from this run
```

We will write restart files, coordinate trajectory .dcd files, and extended system (periodic cell) trajectory .xst files at regular intervals.

```
restartfreq         500     ;# 500 steps = every 1ps
dcdfreq             500
xstFreq             500
```

The default is to print energies every steps, but that makes for a very long file, so we'll cut it down to every 100 steps. We'll also enable the printing of performance information every 1000 steps.

```
outputEnergies      100     ;# 100 steps = every 0.2 ps
outputTiming       1000     ;# shows time per step and time to completion
```

Next are the parameter file itself and the options that control the nonbonded potential functions. These are mostly specified by the CHARMM force field, but the cutoff and other distances may be shortened when full electrostatics are used; Just be sure that you adjust all of the distances together. The pairlistdist can be made longer if you see warnings in the NAMD output, but making it bigger than necessary will reduce performance.

```
# Force-Field Parameters
paraTypeCharmm     on
parameters          par_all27_prot_lipid.inp

# These are specified by CHARMM
exclude             scaled1-4
1-4scaling          1.0
switching           on

# You have some freedom choosing the cutoff
cutoff              12. ;# may use smaller, maybe 10., with PME
switchdist          10. ;# cutoff - 2.

# Promise that atom won't move more than 2A in a cycle
pairlistdist        14. ;# cutoff + 2.
```

The chosen `stepspercycle` is related to `pairlistdist` above, as pairlists are generated at the beginning of each cycle and are assumed to contain every atom that passes within the cutoff distance during entire cycle. Warnings will appear in the output if `pairlistdist` is too short (or `stepspercycle` is too large). The length of the simulation must be an integer number of cycles.

```
stepspercycle       10   ;# redo pairlists every ten steps
```

The integration timestep is normally limited to 1 fs. This can be extended to 2 fs by fixing the length of all bonds in the molecule involving hydrogen atoms via `rigidBonds all`, which also makes water molecules completely rigid. To only make water rigid, use `rigidBonds water` and a 1 fs timestep. For any simulations involving water molecules, one should make water rigid since water molecules have been parametrized as rigid water. Nonbonded forces must be calculated at least every 2 fs, which in this example is every step. Full electrostatics forces (from particle mesh Ewald, discussed below) are evaluated every other step in this example. `nonbondedFreq` and `fullElectFrequency` must evenly divide `stepspercycle`.

```
# Integrator Parameters
timestep            2.0  ;# 2fs/step
rigidBonds          all  ;# needed for 2fs steps
nonbondedFreq       1    ;# nonbonded forces every step
fullElectFrequency  2    ;# PME only every other step
```

Langevin dynamics balances friction with random noise to drive each atom in the system towards a target temperature. The following parameters are good for equilibration, but a `langevinDamping` as low as 1. would be sufficient for simulation. Higher damping may be necessary if, for example, work done on the system by steering forces is driving up the temperature.

```
# Constant Temperature Control
langevin            on            ;# langevin dynamics
```

```
langevinDamping     1.            ;# damping coefficient of 1/ps
langevinTemp        $temperature  ;# random noise at this level
langevinHydrogen    no            ;# don't couple bath to hydrogens
```

If the simulation will use periodic boundary conditions, they are specified as shown below. Like the `temperature` option, these should only be given when starting a simulation from scratch, since the basis vectors will fluctuate during constant pressure simulation and updated values need to be read via `extendedSystem` from a .xsc file.

```
# Periodic Boundary conditions
cellBasisVector1    31.2   0.   0.  ;# vector to the next image
cellBasisVector2     0.   44.8  0.
cellBasisVector3     0.    0   51.3
cellOrigin           0.    0.   0.  ;# the *center* of the cell
```

The cell origin is the one coordinate that does not change during constant pressure simulation. This point plays a role in calculating the pressure contribution due to fixed atoms, harmonic restraints, and other "external" forces. To minimize pressure fluctuations, this point should be close to the center of the macromolecule of interest.

The origin also defines the center of the periodic cell to which coordinates will be wrapped on output if the following options below are used. With wrapping, some molecules will jump between sides of the cell in the trajectory file to yield the periodic image nearest to the origin. Without wrapping, molecules will have smooth trajectories, but water in the trajectory may appear to explode as individual molecules diffuse. Wrapping only affects output, not the correctness of the simulation.

```
wrapWater           on              ;# wrap water to central cell
wrapAll             on              ;# wrap other molecules too
wrapNearest         off             ;# use for non-rectangular cells
```

Particle mesh Ewald (PME) full electrostatics are more accurate and less expensive than larger cutoffs, and are recommended for most work. PME is only applicable to periodic simulations, and the user must specify a grid corresponding to the size of the periodic cell. PME grid dimensions should have small integer factors only and be greater than or equal to length of the basis vector. To manually define the grid size instead of letting NAMD choose the dimensions for you according to these guidelines, replace the GridSpacing command with explicit GridSize commands instead.

```
#PME (for full-system periodic electrostatics)
PME                 yes

# let NAMD determine grid
PMEGridSpacing      1.0
```

```
# manually specify grid
#PMEGridSizeX        32  ;# 2^5, close to 31.2
#PMEGridSizeY        45  ;# 3^2 * 5, close to 44.8
#PMEGridSizeZ        54  ;# 2 * 3^3, close to 51.3
```

Constant pressure is recommended for periodic simulations. Using group-based pressure to control the periodic cell fluctuations is desirable because the atom-based pressure has more high-frequency noise. `useFlexibleCell` is useful for anisotropic systems such as membranes, allowing the height, length, and width of the cell to vary independently, possibly even fixing the lipid cross-sectional (x-y plane) area with `useConstantArea`. For a protein surrounded by water there is nothing to prevent the cell from becoming highly extended in one dimension, so it is better to choose `useFlexibleCell no` in this case.

```
# Constant Pressure Control (variable volume)
useGroupPressure      yes ;# needed for rigid bonds
useFlexibleCell       no  ;# no for water box, yes for membrane
useConstantArea       no  ;# no for water box, maybe for membrane
```

The actual parameters of the Nosé-Hoover Langevin piston method control the target pressure and the dynamical properties of the barostat. A "piston" with a longer period (i.e., larger mass) will better damp out fluctuations in the instantaneous pressure. Langevin dynamics is applied to the piston itself coupling it to a heat bath with a damping constant of 1/`langevinPistonDecay`. We set `langevinPistonDecay` smaller that `langevinPistonPeriod` to ensure that harmonic oscillations in the periodic cell are overdamped.

```
langevinPiston        on
langevinPistonTarget  1.01325      ;# pressure in bar -> 1 atm
langevinPistonPeriod  100.         ;# oscillation period around 100 fs
langevinPistonDecay   50.          ;# oscillation decay time of 50 fs
langevinPistonTemp    $temperature ;# coupled to heat bath
```

When initially assembling a system it is sometimes useful to fix the protein while equilibrating water or lipids around it. These options read a PDB file containing flags for the atoms to fix. The number and order of atoms in the PDB file must exactly match that of the PSF and other input files.

```
fixedAtoms          on
fixedAtomsFile      myfixedatoms.pdb  ;# flags are in this file
fixedAtomsCol       B                 ;# set beta non-zero to fix an atom
```

The interactive MD features of NAMD and VMD allow you to connect to a running simulation to apply steering forces manually. These options affect performance, and should therefore not be used unless you are actually steering the simulation.

```
IMDon           on
IMDport         3000    ;# port number (enter it in VMD)
IMDfreq         1       ;# send every 1 frame
IMDwait         no      ;# wait for VMD to connect before running?
```

Now we minimize the system to eliminate bad initial contacts, reinitialize the velocities to the desired target temperature (since minimization sets velocities to zero), and run for 100 ps. We could accomplish the same thing with two different NAMD runs using the `numsteps` and `minimization` options. Scripting commands such as those below override `numsteps`.

```
minimize            1000          ;# lower potential energy for 1000 steps
reinitvels          $temperature  ;# since minimization zeros velocities
run 50000 ;# 100ps
```

Naturally, these are not all of the configuration options accepted by NAMD, but only a rapid introduction to those you are most likely to encounter. Documentation for these and many other options can be found in the Users Guide.

F NAMD Standard Output

When NAMD runs, important information on the progress of the simulation is written to standard output, appearing on the console unless output from NAMD is redirected to a file. This output is designed to be easy for both humans and programs to interpret.

At the beginning of the file, the version of NAMD being used is presented, along with other useful information, some of which is reported back to the developers over the network to help us track NAMD's popularity.

```
Info: NAMD 2.5b2ss03 for Linux-i686-Clustermatic
Info:
Info: Please visit http://www.ks.uiuc.edu/Research/namd/
Info: and send feedback or bug reports to namd@ks.uiuc.edu
Info:
Info: Please cite James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart,
Info: Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel,
Info: Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with
Info: NAMD. Journal of Computational Chemistry, 26:1781-1802, 2005.
Info: in all publications reporting results obtained with NAMD.
Info:
Info: Built Fri May 30 13:09:06 CDT 2003 by jim on umbriel
Info: Sending usage information to NAMD developers via UDP.
Info: Sent data is: 1 NAMD  2.5b2ss03  Linux-i686-Clustermatic 47 umbriel jim
Info: Running on 47 processors.
```

This is followed by a long list of the various configuration options and summaries of information extracted from the parameter and structure files. This is mostly self-explanatory and just restates the config file options, or the defaults for options not specified.

NAMD goes through several phases to setup data structures for the actual simulation, reporting memory usage (for the first process only!). Patches are the cells that NAMD uses to spread atoms across a parallel simulation, although the size of the patch grid is independent of the number of processors being used.

```
Info: Entering startup phase 0 with 6668 kB of memory in use.
Info: Entering startup phase 1 with 6668 kB of memory in use.
Info: Entering startup phase 2 with 10212 kB of memory in use.
Info: Entering startup phase 3 with 10212 kB of memory in use.
Info: PATCH GRID IS 4 (PERIODIC) BY 4 (PERIODIC) BY 4 (PERIODIC)
Info: REMOVING COM VELOCITY -0.0765505 0.00822415 -0.00180344
Info: LARGEST PATCH (31) HAS 408 ATOMS
Info: Entering startup phase 4 with 13540 kB of memory in use.
Info: Entering startup phase 5 with 13540 kB of memory in use.
Info: Entering startup phase 6 with 13540 kB of memory in use.
Info: Entering startup phase 7 with 13540 kB of memory in use.
Info: COULOMB TABLE R-SQUARED SPACING: 0.0625
Info: COULOMB TABLE SIZE: 705 POINTS
Info: NONZERO IMPRECISION IN COULOMB TABLE: 4.03897e-28 (692) 1.00974e-27 (692)
```

```
Info: NONZERO IMPRECISION IN COULOMB TABLE: 2.5411e-21 (701) 5.92923e-21 (701)
Info: Entering startup phase 8 with 13540 kB of memory in use.
Info: Finished startup with 13540 kB of memory in use.
```

Energies, along with temperatures and pressures, are printed as a single line with a unique prefix for easy processing with utilities such as grep and awk. An ETITLE line is printed once for every ten ENERGY lines. The spacing of the fields, however, is designed to be easy to scan on an 80-column display, as seen below:

```
ETITLE:      TS           BOND          ANGLE          DIHED          IMPRP
          ELECT            VDW       BOUNDARY           MISC       KINETIC
          TOTAL           TEMP         TOTAL2         TOTAL3       TEMPAVG
       PRESSURE     GPRESSURE         VOLUME       PRESSAVG     GPRESSAVG

ENERGY:    1000         0.0000         0.0000         0.0000         0.0000
    -97022.1848      9595.3175         0.0000         0.0000     14319.5268
    -73107.3405       300.2464    -73076.6148    -73084.1411       297.7598
      -626.5205      -636.6638    240716.1374      -616.5673      -616.6619
```

Energy values are in kcal/mol. This example is from a simulation of pure, rigid water, so the BOND, ANGLE, DIHED, and IMPRP terms are all zero. BOUNDARY energy is from spherical boundary conditions and harmonic restraints, while MISC energy is from external electric fields and various steering forces. TOTAL is the sum of the various potential energies, and the KINETIC energy. TOTAL2 uses a slightly different kinetic energy that is better conserved during equilibration in a constant energy ensemble. TOTAL3 is another variation with much smaller short-time fluctuations that is also adjusted to have the same running average as TOTAL2. Defects in constant energy simulations are much easier to spot in TOTAL3 than in TOTAL or TOTAL2.

Pressure values are in bar. Temperature values are in Kelvin. PRESSURE is the pressure calculated based on individual atoms, while GPRESSURE incorporates hydrogen atoms into the heavier atoms to which they are bonded, producing smaller fluctuations. The TEMPAVG, PRESSAVG, and GPRESSAVG are the average of temperature and pressure values since the previous ENERGY output; for the first step in the simulation they will be identical to TEMP, PRESSURE, and GPRESSURE.

To estimate NAMD's performance for a long simulation, look for the "Benchmark time" lines printed in the first several hundred steps. These are a measure of NAMD's performance after startup and initial load balancing have been completed; this number should be very close to the long-time average.

```
Info: Benchmark time: 47 CPUs 0.0475851 s/step 0.275377 days/ns 13540 kB memory
```

Periodic performance output, if enabled, shows the CPU and wallclock time used by the first processor (not the aggregate of all processors in the simulation). The hours remaining are estimated based on performance since the last TIMING output.

```
TIMING: 1000  CPU: 18.35, 0.01831/step  Wall: 50.1581, 0.0499508/step, 6.92374
hours remaining, 14244 kB of memory in use.
```

Lines such as these are printed whenever file output occurs.

```
OPENING COORDINATE DCD FILE
WRITING COORDINATES TO DCD FILE AT STEP 1000
```

It is important to search the NAMD output for Warning and ERROR lines, which report unusual and possibly erroneous conditions. In the example below, the pairlist distance is too short or the cycle length is too long, possibly leading to reduced performance.

```
Warning: Pairlistdist is too small for 1 patches during timestep 17.
Warning: Pairlists partially disabled; reduced performance likely.
Warning: 20 pairlist warnings since previous energy output.
```

It is foolish to ignore warnings you do not understand!