Different Files in MD Simulation

Ahmad Alqaisi, Graduate Student

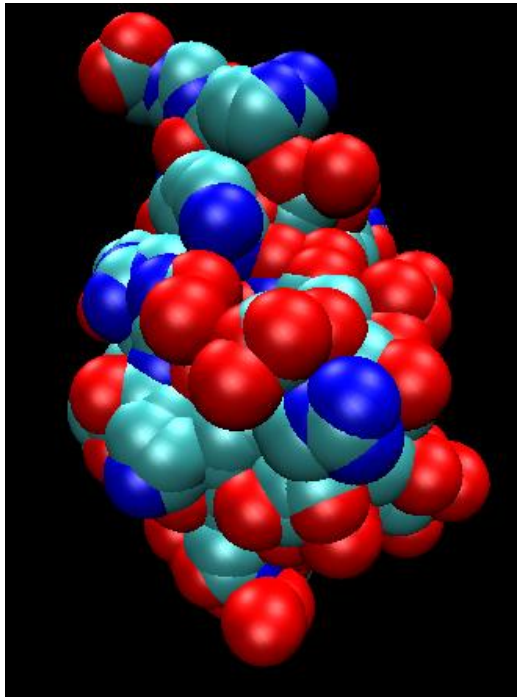
Quantum Mechanical Engineering Lab., Arizona State University

Theoretical and Computational Chemistry Lab., Jordan University

Protein Data Bank (PDB) File

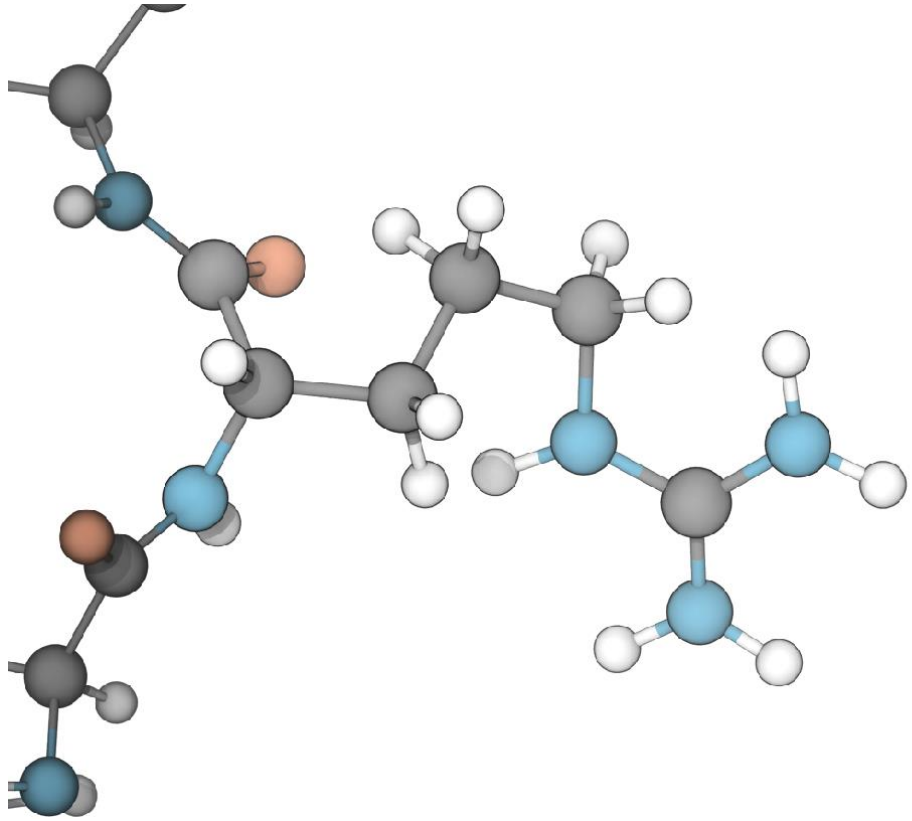
- The Protein Data Bank (pdb) file format is a textual file format describing the three-dimensional structures of molecules held in the Protein Data Bank.

- The pdb format accordingly provides for description and annotation of protein and nucleic acid structures including atomic coordinates, secondary structure assignments, as well as atomic connectivity. In addition experimental metadata are stored.



Protein Data Bank (PDB) File

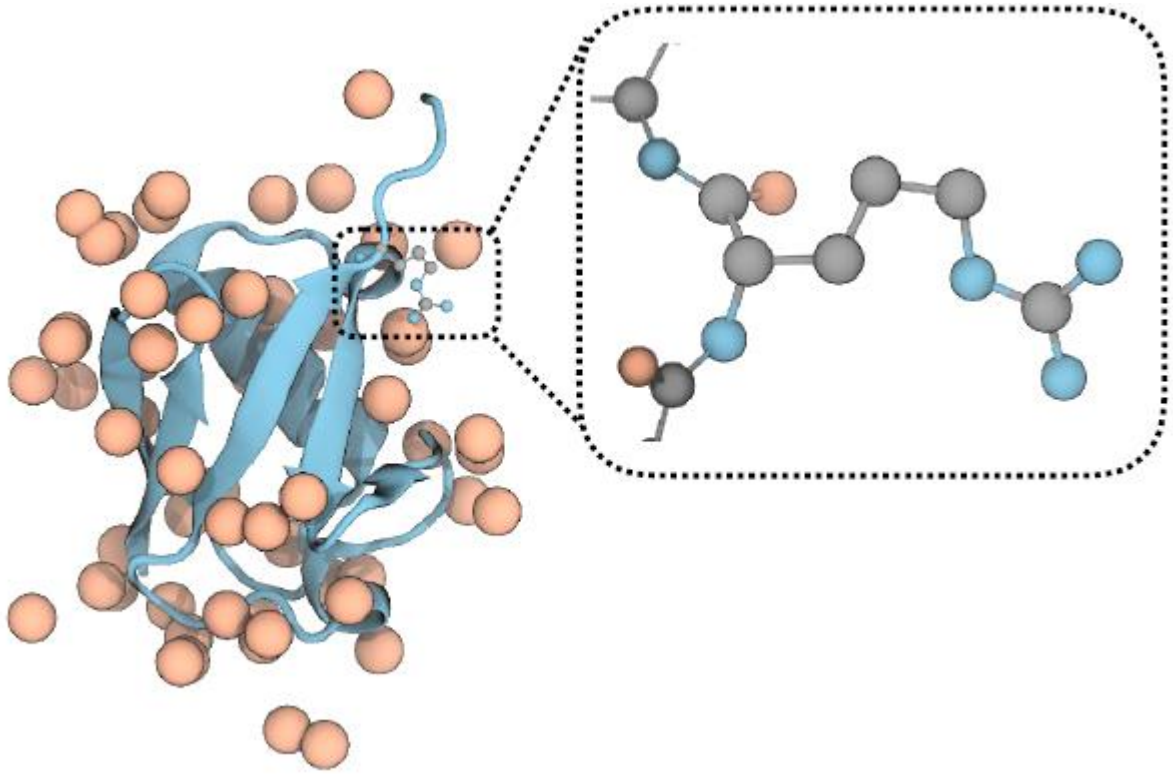
```
ATOM      1  N   MET A   1      27.340  24.430   2.614  1.00  9.67           N
ATOM      2  CA  MET A   1      26.266  25.413   2.842  1.00 10.38           C
ATOM      3  C   MET A   1      26.913  26.639   3.531  1.00  9.62           C
ATOM      4  O   MET A   1      27.886  26.463   4.263  1.00  9.62           O
ATOM      5  CB  MET A   1      25.112  24.880   3.649  1.00 13.77           C
ATOM      6  CG  MET A   1      25.353  24.860   5.134  1.00 16.29           C
ATOM      7  SD  MET A   1      23.930  23.959   5.904  1.00 17.17           S
ATOM      8  CE  MET A   1      24.447  23.984   7.620  1.00 16.11           C
ATOM      9  N   GLN A   2      26.335  27.770   3.258  1.00  9.27           N
ATOM     10  CA  GLN A   2      26.850  29.021   3.898  1.00  9.07           C
ATOM     11  C   GLN A   2      26.100  29.253   5.202  1.00  8.72           C
ATOM     12  O   GLN A   2      24.865  29.024   5.330  1.00  8.22           O
ATOM     13  CB  GLN A   2      26.733  30.148   2.905  1.00 14.46           C
ATOM     14  CG  GLN A   2      26.882  31.546   3.409  1.00 17.01           C
ATOM     15  CD  GLN A   2      26.786  32.562   2.270  1.00 20.10           C
ATOM     16  OE1 GLN A   2      27.783  33.160   1.870  1.00 21.89           O
ATOM     17  NE2 GLN A   2      25.562  32.733   1.806  1.00 19.49           N
ATOM     18  N   ILE A   3      26.849  29.656   6.217  1.00  5.87           N
ATOM     19  CA  ILE A   3      26.235  30.058   7.497  1.00  5.07           C
ATOM     20  C   ILE A   3      26.882  31.428   7.862  1.00  4.01           C
ATOM     21  O   ILE A   3      27.906  31.711   7.264  1.00  4.61           O
ATOM     22  CB  ILE A   3      26.344  29.050   8.645  1.00  6.55           C
ATOM     23  CG1 ILE A   3      27.810  28.748   8.999  1.00  4.72           C
ATOM     24  CG2 ILE A   3      25.491  27.771   8.287  1.00  5.58           C
ATOM     25  CD1 ILE A   3      27.967  28.087  10.417  1.00 10.83           C
ATOM     26  N   PHE A   4      26.214  32.097   8.771  1.00  4.55           N
ATOM     27  CA  PHE A   4      26.772  33.436   9.197  1.00  4.68           C
ATOM     28  C   PHE A   4      27.151  33.362  10.650  1.00  5.30           C
ATOM     29  O   PHE A   4      26.350  32.778  11.395  1.00  5.58           O
ATOM     30  CB  PHE A   4      25.695  34.498   8.946  1.00  4.83           C
ATOM     31  CG  PHE A   4      25.288  34.609   7.499  1.00  7.97           C
ATOM     32  CD1 PHE A   4      24.147  33.966   7.038  1.00  6.69           C
ATOM     33  CD2 PHE A   4      26.136  35.346   6.640  1.00  8.34           C
ATOM     34  CE1 PHE A   4      23.812  34.031   5.677  1.00  9.10           C
ATOM     35  CE2 PHE A   4      25.810  35.392   5.267  1.00 10.61           C
ATOM     36  CZ  PHE A   4      24.620  34.778   4.853  1.00  8.90           C
ATOM     37  N   VAL A   5      28.260  33.943  11.096  1.00  4.44           N
```



Protein Data Bank (PDB) File

- Things to know about the PDB files:

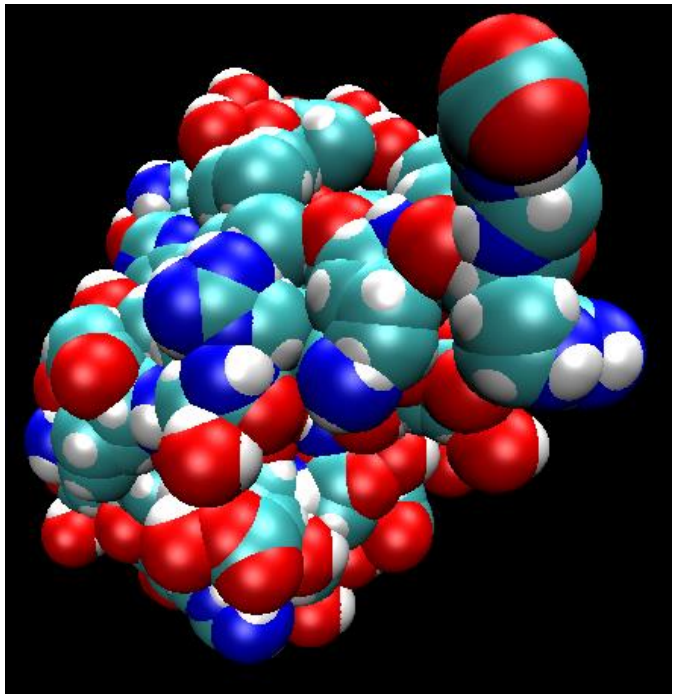
- No Hydrogen Atoms
- Crystallographic Water Molecules and Ions
- Other Molecules
- Missing residues (Possible Sequence Gaps)
- Poorly Defined Atom’s Position



Protein Structure File (PSF)

- The PSF file contains all of the basic molecular details needed to apply a specific force field to a molecular system. And stores structural information of the protein, such as various types of bonding interactions.

- In addition, a PSF often contains hydrogens and other atoms that may be missing from a crystal pdb file to automatically assign coordinates



PSF file has the information of connectivity

Protein Structure File (PSF)

charge *mass*

1	AP1	1	MET	N	NH3	-0.300000	14.0070	0
2	AP1	1	MET	HT1	HC	0.330000	1.0080	0
3	AP1	1	MET	HT2	HC	0.330000	1.0080	0
4	AP1	1	MET	HT3	HC	0.330000	1.0080	0
5	AP1	1	MET	CA	CT1	0.210000	12.0110	0
6	AP1	1	MET	HA	HB1	0.100000	1.0080	0
7	AP1	1	MET	CB	CT2	-0.180000	12.0110	0
8	AP1	1	MET	HB1	HA2	0.090000	1.0080	0
9	AP1	1	MET	HB2	HA2	0.090000	1.0080	0
10	AP1	1	MET	CG	CT2	-0.140000	12.0110	0
11	AP1	1	MET	HG1	HA2	0.090000	1.0080	0
12	AP1	1	MET	HG2	HA2	0.090000	1.0080	0
13	AP1	1	MET	SD	S	-0.090000	32.0600	0
14	AP1	1	MET	CE	CT3	-0.220000	12.0110	0
15	AP1	1	MET	HE1	HA3	0.090000	1.0080	0
16	AP1	1	MET	HE2	HA3	0.090000	1.0080	0
17	AP1	1	MET	HE3	HA3	0.090000	1.0080	0
18	AP1	1	MET	C	C	0.510000	12.0110	0
19	AP1	1	MET	O	O	-0.510000	15.9990	0
20	AP1	2	GLN	N	NH1	-0.470000	14.0070	0
21	AP1	2	GLN	HN	H	0.310000	1.0080	0
22	AP1	2	GLN	CA	CT1	0.070000	12.0110	0
23	AP1	2	GLN	HA	HB1	0.090000	1.0080	0
24	AP1	2	GLN	CB	CT2	-0.180000	12.0110	0
25	AP1	2	GLN	HB1	HA2	0.090000	1.0080	0
26	AP1	2	GLN	HB2	HA2	0.090000	1.0080	0
27	AP1	2	GLN	CG	CT2	-0.180000	12.0110	0
28	AP1	2	GLN	HG1	HA2	0.090000	1.0080	0
29	AP1	2	GLN	HG2	HA2	0.090000	1.0080	0
30	AP1	2	GLN	CD	CC	0.550000	12.0110	0
31	AP1	2	GLN	OE1	O	-0.550000	15.9990	0
32	AP1	2	GLN	NE2	NH2	-0.620000	14.0070	0
33	AP1	2	GLN	HE21	H	0.320000	1.0080	0
34	AP1	2	GLN	HE22	H	0.300000	1.0080	0
35	AP1	2	GLN	C	C	0.510000	12.0110	0

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
14	13	14	15	14	16	14	17
18	5	18	20	19	18	20	21
20	22	22	23	24	22	24	25
24	26	27	24	27	28	27	29
30	27	30	31	32	30	32	33
32	34	35	22	35	37	36	35
37	38	37	39	39	40	41	39
41	42	43	41	43	44	43	45
43	46	47	41	47	48	47	49
50	47	50	51	50	52	50	53
54	39	54	56	55	54	56	57
56	58	58	59	60	58	60	61
60	62	63	60	64	65	64	63

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
5	18	20	5	7	9	5	7	8
5	7	10	7	10	12	7	10	11
7	10	13	7	5	6	7	5	18
7	5	1	8	7	9	10	13	14
10	7	9	10	7	8	11	10	12
13	14	17	13	14	16	13	14	15
13	10	12	13	10	11	15	14	17
15	14	16	16	14	17	18	5	6
20	22	23	20	22	35	20	18	19
21	20	18	21	20	22	22	35	36
22	35	37	22	24	26	22	24	25
22	24	27	22	20	18	24	27	29
24	27	28	24	27	30	24	22	23
24	22	35	24	22	20	25	24	26
27	30	31	27	30	32	27	24	26
27	24	25	28	27	29	30	32	34
30	32	33	30	27	29	30	27	28

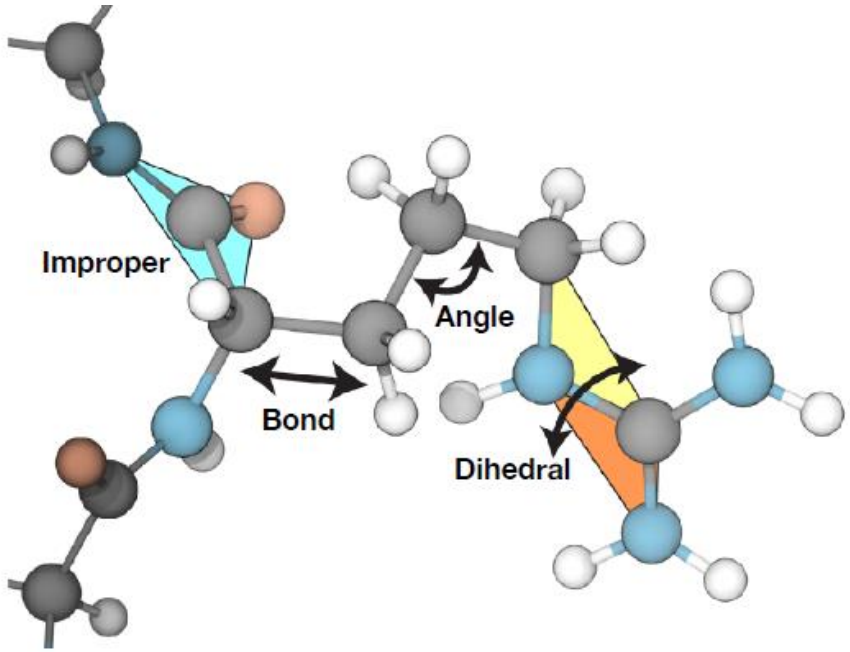
Topology Files (TOP)

We read the connectivity from here to convert .pdb to .psf file

- A CHARMM forcefield topology file contains all of the information needed to convert a list of residue names into a complete PSF structure file.

- In addition, a PSF often contains internal coordinates that allow hydrogen and other atoms that may be missing from a crystal pdb file to automatically assign coordinates

```
MASS    31 H      1.00800 ! polar H
MASS    32 HC     1.00800 ! N-ter H
MASS    33 HA     1.00800 ! nonpolar H
MASS    34 HP     1.00800 ! aromatic H
MASS    35 HB1    1.00800 ! backbone H
MASS    36 HB2    1.00800 ! aliphatic backbone H, to CT2
MASS    37 HR1    1.00800 ! his hel, (+) his HG,HD2
MASS    38 HR2    1.00800 ! (+) his HE1
MASS    39 HR3    1.00800 ! neutral his HG, HD2
MASS    40 HS     1.00800 ! thiol hydrogen
MASS    41 HE1    1.00800 ! for alkene; RHC=CR
MASS    42 HE2    1.00800 ! for alkene; H2C=CR
MASS    43 HA1    1.00800 ! alkane, CH, new LJ params (see toppar_all22_prot_aliphatic_c27.str)
MASS    44 HA2    1.00800 ! alkane, CH2, new LJ params (see toppar_all22_prot_aliphatic_c27.str)
MASS    45 HA3    1.00800 ! alkane, CH3, new LJ params (see toppar_all22_prot_aliphatic_c27.str)
MASS    46 C     12.01100 ! carbonyl C, peptide backbone
MASS    47 CA    12.01100 ! aromatic C
MASS    48 CT    12.01100 ! aliphatic sp3 C, new LJ params, no hydrogens
MASS    49 CT1   12.01100 ! aliphatic sp3 C for CH
MASS    50 CT2   12.01100 ! aliphatic sp3 C for CH2
MASS    51 CT2A  12.01100 ! from CT2 (asp, glu, hsp chi1/chi2 fitting)
MASS    52 CT3   12.01100 ! aliphatic sp3 C for CH3
MASS    53 CPH1  12.01100 ! his CG and CD2 carbons
MASS    54 CPH2  12.01100 ! his CE1 carbon
```



Topology Files (TOP)

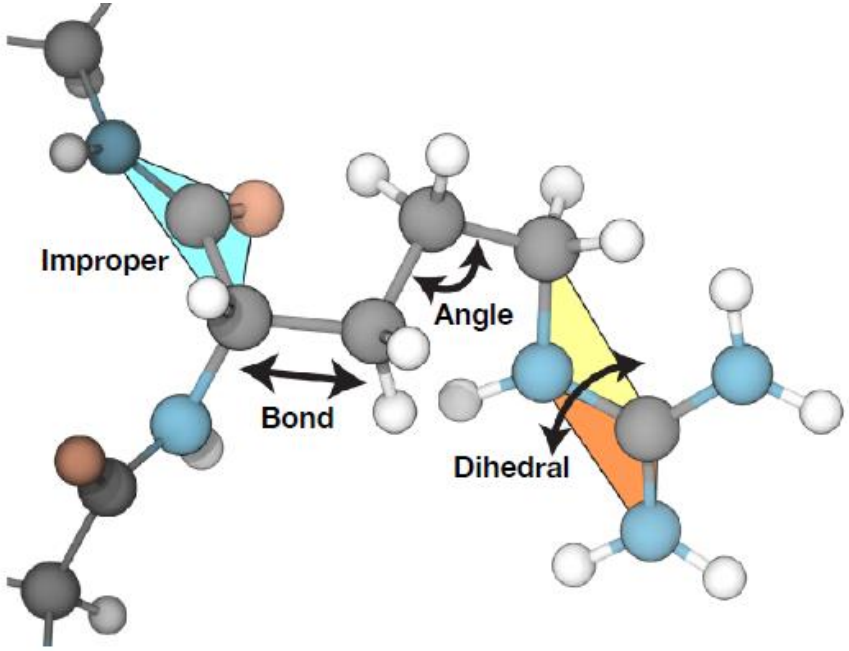
```
MASS    31 H      1.00800 ! polar H
MASS    32 HC     1.00800 ! N-ter H
MASS    33 HA     1.00800 ! nonpolar H
MASS    34 HP     1.00800 ! aromatic H
MASS    35 HB1    1.00800 ! backbone H
MASS    36 HB2    1.00800 ! aliphatic backbone H, to CT2
MASS    37 HR1    1.00800 ! his he1, (+) his HG,HD2
MASS    38 HR2    1.00800 ! (+) his HE1
MASS    39 HR3    1.00800 ! neutral his HG, HD2
MASS    40 HS     1.00800 ! thiol hydrogen
MASS    41 HE1    1.00800 ! for alkene; RHC=CR
MASS    42 HE2    1.00800 ! for alkene; H2C=CR
MASS    43 HA1    1.00800 ! alkane, CH, new LJ params (see toppar_all22_prot_aliphatic_c27.str)
MASS    44 HA2    1.00800 ! alkane, CH2, new LJ params (see toppar_all22_prot_aliphatic_c27.str)
MASS    45 HA3    1.00800 ! alkane, CH3, new LJ params (see toppar_all22_prot_aliphatic_c27.str)
MASS    46 C     12.01100 ! carbonyl C, peptide backbone
MASS    47 CA    12.01100 ! aromatic C
MASS    48 CT    12.01100 ! aliphatic sp3 C, new LJ params, no hydrogens
MASS    49 CT1   12.01100 ! aliphatic sp3 C for CH
MASS    50 CT2   12.01100 ! aliphatic sp3 C for CH2
MASS    51 CT2A  12.01100 ! from CT2 (asp, glu, hsp chi1/chi2 fitting)
MASS    52 CT3   12.01100 ! aliphatic sp3 C for CH3
MASS    53 CPH1  12.01100 ! his CG and CD2 carbons
MASS    54 CPH2  12.01100 ! his CE1 carbon
```

```
RESI ASN          0.00
GROUP
ATOM N    NH1    -0.47  !     |
ATOM HN   H       0.31  !  HN-N
ATOM CA   CT1     0.07  !     |   HB1 OD1    HD21 (cis to OD1)
ATOM HA   HB1     0.09  !     |   |   ||    /
GROUP                   !  HA-CA--CB--CG--ND2
ATOM CB   CT2    -0.18  !     |   |         \
ATOM HB1  HA2     0.09  !     |   HB2        HD22 (trans to OD1)
ATOM HB2  HA2     0.09  !   O=C
GROUP                   !     |
ATOM CG   CC      0.55
ATOM OD1  O      -0.55
GROUP
ATOM ND2  NH2    -0.62
ATOM HD21 H       0.32
ATOM HD22 H       0.30
GROUP
ATOM C    C       0.51
ATOM O    O      -0.51
BOND CB CA  CG CB   ND2 CG
BOND N  HN  N  CA   C   CA    C +N
BOND CA HA  CB HB1  CB  HB2  ND2 HD21  ND2 HD22
DOUBLE C  O   CG  OD1
IMPR N   -C  CA   HN    C   CA +N   O
IMPR CG  ND2 CB   OD1   CG  CB ND2  OD1
IMPR ND2 CG  HD21 HD22  ND2 CG HD22 HD21
CMAP -C  N  CA  C   N  CA  C  +N
DONOR HN N
DONOR HD21 ND2
DONOR HD22 ND2
ACCEPTOR OD1 CG
```

Parameter Files – Force Fields (PRM)

- Force field parameter file is a mathematical expression of the potential which atoms in the system experience

- The parameter file defines bond strengths, equilibrium lengths and contains all of the numerical constants required to determine forces and energies, given that a PSF structure file and atomic coordinates have been provided .pdb file



Parameter Files – Force Fields (PRM)

```
BONDS
!
!V(bond) = Kb(b - b0)**2
!
!Kb: kcal/mole/A**2
!b0: A
!
!atom type Kb          b0
!
NH2  CT1   240.000     1.4550 ! From LSN NH2-CT2
!
!Indole/Tryptophan
CA   CAI   305.000     1.3750 ! from CA CA
CAI  CAI   305.000     1.3750 ! atm, methylindole, fit CCDSS
CPT  CA    300.000     1.3600 ! atm, methylindole, fit CCDSS
CPT  CAI   300.000     1.3600 ! atm, methylindole, fit CCDSS
CPT  CPT   360.000     1.3850 ! atm, methylindole, fit CCDSS
CY   CA    350.000     1.3650 ! trj, adm jr., 5/08/91, indole CCDB structure search
CY   CAI   350.000     1.3650 ! from CY CA
CY   CPT   350.000     1.4300 ! atm, methylindole, fit CDS data
CY   CT3   375.000     1.4920 ! atm, methylindole, fit CDS data
CY   CT2   375.000     1.4920 ! atm, methylindole, fit CDS data
HP   CAI   340.000     1.0800 ! from HP CA
HP   CY    350.000     1.0800 ! trp, adm jr., 12/30/91
```

```
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
!
H    NH2  CT1   50.000    111.00              ! From LSN HC-NH2-CT2
NH2  CT1  CT1   67.700    110.00              ! From LSN NH2-CT2-CT2
NH2  CT1  CT2   67.700    110.00              ! From LSN NH2-CT2-CT2
NH2  CT1  CT3   67.700    110.00              ! From LSN NH2-CT2-CT2
CT1  CD   OH1   55.000    110.50              ! From ASPP CT2-CD-OH1
CT3  CT1  CD    52.000    108.00              ! Ala cter
NH2  CT1  HB1   38.000    109.50   50.00   2.1400 ! From LSN NH2-CT2-HA
NH2  CT1  C     50.000    107.00              ! From ALA Dipep. NH1-CT2-C
!
!Indole/Tryptophan
CAI  CAI  CA    40.000    120.00   35.00   2.41620 ! from CA CA CA
CAI  CA   CA    40.000    120.00   35.00   2.41620 ! from CA CA CA
CPT  CA   CA    50.000    113.20 ! atm, methylindole, 1/17/04
CPT  CPT  CA    50.000    110.00 ! atm, methylindole, 1/17/04
CPT  CAI  CA    50.000    113.20 ! atm, methylindole, 1/17/04
CPT  CPT  CAI   50.000    110.00 ! atm, methylindole, 1/17/04
```

Configuration File (CONF)

- The user defines all the options that NAMD should consider while running the simulation
- The configuration file tells NAMD how the simulation should be run
- NAMD Configuration File is a TCL Script File
- NAMD Configuration File Manual: https://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-unix-html/node26.html

Configuration File (CONF)

Define Initial Structure

Define Variable “temperature”

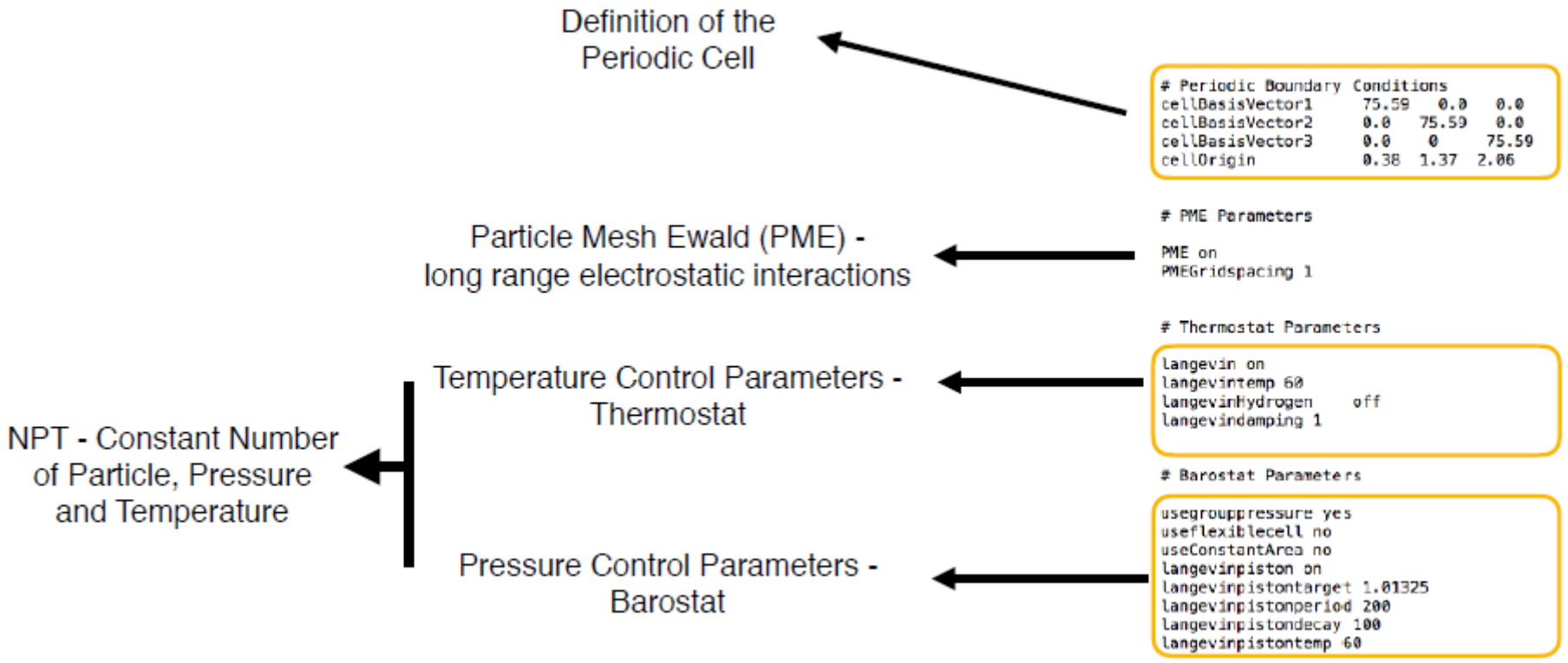
Define Initial Temperature
Temperature Distribution

Define Restraints (e.g. backbone)
Uses an extra PDB file with beta column 1 to identify the atoms subject to the restraint

Output parameters
Printing into files or to the console is very expensive in terms of computation, harming the performance of NAMD if done to often

```
# Initial pdb and psf files

coordinates Ubiquitin_QwikMD.pdb
structure Ubiquitin_QwikMD.psf

# Simulation conditions
set temperature 300.00; # Conversion of 27 degrees Celsius + 273
temperature 0

# Harmonic constraints

constraints on
consref qwikmd_equilibration_0_constraints.pdb
conskfile qwikmd_equilibration_0_constraints.pdb
constraintScaling 2
consexp 2
conskcol B

# Output Parameters

binaryoutput no
outputname qwikmd_equilibration_0
outputenergies 400
outputtiming 400
outputpressure 400
binaryrestart yes
dcdfile qwikmd_equilibration_0.dcd
dcdfreq 10000
XSTFreq 10000
restartfreq 10000
restartname qwikmd_equilibration_0.restart
```

Configuration File (CONF)



Configuration File (CONF)

Keep Molecules Inside the Periodic Cell. No Physical Value Changes, Only Its Representation

Multi Time Stepping
Integration time 2fs
Short Non-Bonded Interaction (2fs)
Full Electrostatics (PME - 4 fs)

Input Parameters Files

What Allows the 2fs Integration Step
Water Oxygen-Hydrogen bonds (opt = water)
Heavy Atom-Hydrogen bonds (opt = all)

Non-Bonded Parameters
Cutoff - VDW interaction
Switching - Smoothly Tend to 0
Pairlistdist - Dist. To Pair of Atoms in VDW Interactions
Stepspercyle - Cycles between evaluating PairList

```
wrapAll on
wrapWater on


# Integrator Parameters

timestep 2
fullElectFrequency 2
nonbondedfreq 1


# Force Field Parameters

paratypecharmm on
parameters par_all36_carb.prm
parameters par_all36_cgenff.prm
parameters par_all36_lipid.prm
parameters par_all36_na.prm
parameters par_all36_prot.prm
parameters toppar_all36_carb_glycopeptide.str
parameters toppar_water_ions_namd.str
exclude scaled1-4
1-4scaling 1.0
rigidbonds all
cutoff 12.0
pairlistdist 14.0
stepspercycle 10
switching on
switchdist 10.0
```

PBC: https://www.ks.uiuc.edu/Research/namd/2.13/ug/node33.html
Temperature Control: https://www.ks.uiuc.edu/Research/namd/2.13/ug/node36.html
Pressure Control: https://www.ks.uiuc.edu/Research/namd/2.13/ug/node37.html
Non-Bonded Int. : https://www.ks.uiuc.edu/Research/namd/2.13/ug/node23.html
Pair List: https://www.ks.uiuc.edu/Research/namd/2.13/ug/node92.html

Configuration File (CONF)

Minimize Potential Energy

Re-Set Temperature

Temperature Ramp (example of scripting)

Run Simulation for 1 ns (2 * 500.000 fs)

Example of Checking for Normal Termination of the Simulation

```
# Script

minimize 1000
reinitvels 60
for {set t 60} {$t <= $temperature} {incr t} {
     langevinpistontemp $t
     run 500
     langevintemp $t
}
run 500000
set file [open qwikmd_equilibration_0.check w+]
set done 1
if {[file exists qwikmd_equilibration_0.restart.coor] != 1 ||
[file exists qwikmd_equilibration_0.restart.vel] != 1 ||
[file exists qwikmd_equilibration_0.restart.xsc] != 1 } {
      set done 0
}
if {$done == 1} {
     puts $file "DONE"
    flush $file
  close $file
} else {
     puts $file "One or more files filed to be written"
   flush $file
  close $file
}
```

Modifications for NAMD Configuration File (CONF)

.log => Energies, timing .dcd => trajectory file

- Add: *stepspercycle 10 or stepspercycle 20* to Minimization.conf, Annealing.conf, Equalibration.conf, and MD.conf
- Modify Output parameters to an appropriate frequencies:
- Recommended is *1000* all for Minimization
- And *1000 for energy related, and 50000 for coordinates related* output

For other procedures: Annealing, Equilibration, MD

```
#Implicit Solvent Parameters

gbis off
alphaCutoff          14.0
ionConcentration     0.15

stepspercycle 10

# Script

minimize 2000

set file [open Minimization.check w+]
set done 1
if {[file exists Minimization.restart.coor] !=
     set done 0
}
if {$done == 1} {
    puts $file "DONE"
    flush $file
  close $file
} else {
    puts $file "One or more files failed to be
   flush $file
  close $file
```

```
# Output Parameters

binaryoutput no
outputname Minimization
outputenergies 1000
outputtiming 1000
outputpressure 1000
binaryrestart yes
XSTFreq 1000
restartfreq 1000
restartname Minimization.restart
```

```
# Output Parameters

binaryoutput no
outputname MD
outputenergies 1000
outputtiming 1000
outputpressure 1000
binaryrestart yes
dcdfile MD.dcd
dcdfreq 50000
XSTFreq 50000
restartfreq 50000
restartname MD.restart
```

Restart/Resume or Continue a NAMD Simulation

- If you want Restart/Resume a NAMD simulation that stopped due to any reason, you do the following:
- Go to the .conf file of the last step
- Make a Copy of that with a new name (e.g.: MD2.conf)
- In the bin---- part, You need to use the .restart files from the last step in this step

  Xxxx.restart.coor

  Xxxx.restart.vel

  Xxxx.restart.xsc

- To know the remaining time, go for the .log file of the last step, in the last line that contains the word "TIMING". Remaining time = Original time – $TIMING_{last}$

Restart/Resume or Continue a NAMD Simulation

- If you want increase the time for a NAMD simulation that has already finished, you do the following:
- Go to the .conf file of the last step
- Copy that
- You need to use the files from the last step in your new .conf file, without .restart

  Xxxx.coor

  Xxxx.vel

  Xxxx.xsc

- You also decide the run time