Part I. Strategy for running molecular dynamics simulations

This part of the lecture describes the general steps in setting and running molecular dynamics (MD) simulations using NAMD for a short polyvaline peptide (VVVV). This peptide is one of the two peptides to be examined in the MD project. More information on running MD simulations with NAMD can be found at www.ks.uiuc.edu/Research/namd/current/ug/ .

Stages in a typical MD trajectory

The stages in MD simulations are shown in Fig. 1.

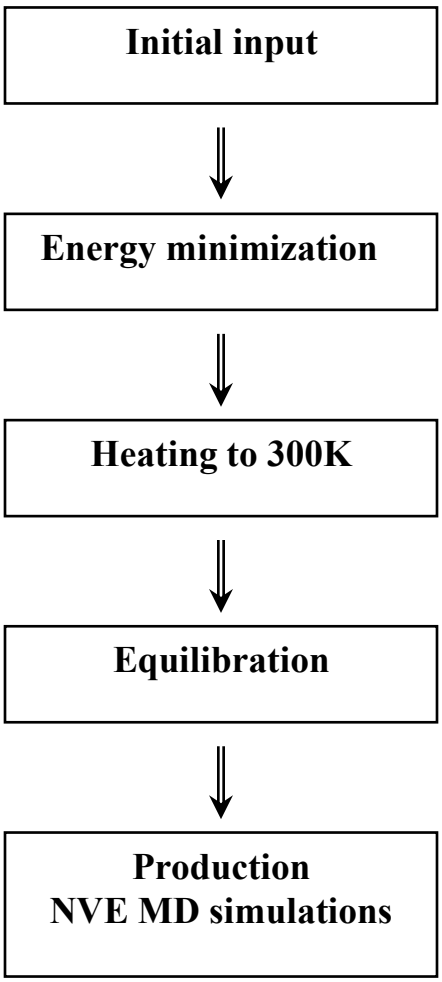

Figure 1. Strategy for performing generic MD simulations.

In what follows it is assumed that the initial solvated structure of polyvaline (the file val_solv.pdb) and the topology file val_solv.psf are already prepared. The goal of MD simulations is to obtain a single 1 ns production trajectory, from which conformational properties of polyvaline can be computed.

1. Initialization: structure/topology/force field files

Three input files are needed to start the simulations. These are the topology file val_solv.psf, the coordinate file val_solv.pdb, and the force field file par_all27_prot.inp (part of CHARMM27 force field). The topology file contains all the information about the structure and connectivity of atoms in the system as well as few parameters of the force field (i.e., those which are incorporated in the top_all27_prot.inp file). In all, the solvated polyvaline system contains 2419 atoms and 788 residues, which include peptide amino acids, capping terminals, and water molecules. The peptide itself contains only 73 atoms in four residues and two terminal groups.

Structure of the topology file is as follows:

2419 !NATOM

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
1	P1	1	VAL	N	NH3	-0.300000	14.0070	0
2	P1	1	VAL	HT1	HC	0.330000	1.0080	0
3	P1	1	VAL	HT2	HC	0.330000	1.0080	0
4	P1	1	VAL	HT3	HC	0.330000	1.0080	0
5	P1	1	VAL	CA	CT1	0.210000	12.0110	0
6	P1	1	VAL	HA	HB	0.100000	1.0080	0
7	P1	1	VAL	CB	CT1	-0.090000	12.0110	0
8	P1	1	VAL	HB	HA	0.090000	1.0080	0
9	P1	1	VAL	CG1	CT3	-0.270000	12.0110	0
10	P1	1	VAL	HG11	HA	0.090000	1.0080	0
11	P1	1	VAL	HG12	HA	0.090000	1.0080	0
12	P1	1	VAL	HG13	HA	0.090000	1.0080	0
13	P1	1	VAL	CG2	CT3	-0.270000	12.0110	0
14	P1	1	VAL	HG21	HA	0.090000	1.0080	0
15	P1	1	VAL	HG22	HA	0.090000	1.0080	0
16	P1	1	VAL	HG23	HA	0.090000	1.0080	0
17	P1	1	VAL	C	C	0.510000	12.0110	0
18	P1	1	VAL	O	O	-0.510000	15.9990	0
19	P1	2	VAL	N	NH1	-0.470000	14.0070	0
20	P1	2	VAL	HN	H	0.310000	1.0080	0
21	P1	2	VAL	CA	CT1	0.070000	12.0110	0
22	P1	2	VAL	HA	HB	0.090000	1.0080	0
.....								
51	P1	4	VAL	C	CC	0.340000	12.0110	0
52	P1	4	VAL	OT1	OC	-0.670000	15.9990	0
53	P1	4	VAL	OT2	OC	-0.670000	15.9990	0
54	P1	4	VAL	N	NH1	-0.470000	14.0070	0
55	P1	4	VAL	HN	H	0.310000	1.0080	0

```
      56 P1   4    VAL  CA   CT1    0.070000       12.0110           0
      57 P1   4    VAL  HA   HB     0.090000        1.0080           0
      58 P1   4    VAL  CB   CT1   -0.090000       12.0110           0
      59 P1   4    VAL  HB   HA     0.090000        1.0080           0
      60 P1   4    VAL  CG1  CT3   -0.270000       12.0110           0
      61 P1   4    VAL  HG11 HA     0.090000        1.0080           0
      62 P1   4    VAL  HG12 HA     0.090000        1.0080           0
      63 P1   4    VAL  HG13 HA     0.090000        1.0080           0
      64 P1   4    VAL  CG2  CT3   -0.270000       12.0110           0
      65 P1   4    VAL  HG21 HA     0.090000        1.0080           0
      66 P1   4    VAL  HG22 HA     0.090000        1.0080           0
      67 P1   4    VAL  HG23 HA     0.090000        1.0080           0
.....................................................................
     302 WT1  576  TIP3 OH2  OT    -0.834000       15.9994           0
     303 WT1  576  TIP3 H1   HT     0.417000        1.0080           0
     304 WT1  576  TIP3 H2   HT     0.417000        1.0080           0
.....................................................................
```

(a) atom number;
(b) segment name, e.g., P1 is the name of peptide segment in the solvated system. The other segment is solvent (WT1);
(c) residue number (may not be sequential);
(d) residue name;
(e) atom name;
(f) atom type;
(g) partial charge;
(h) atomic mass;
(i) flag used to indicate constrained atom

Note that the first four atoms correspond to N-terminal protonated $NH3^+$, which is incorporated by CHARMM convention as a general NTER patch. The C-term CTER patch is ther carboxylate (atom 51, 52, 53). The N- and C-terminal are distinguished by the letters "HT" and "OT" in the atom names for the hydrogens at N-term and the oxygens at the C-term, respectively. Topology file also contains information (atom numbers), which defines covalent bonds, bond angles, dihedral angles, and improper torsion angles.

Bond lengths (defined by pairs of atom numbers)……..

```
1634 !NBOND: bonds
       1       5       2       1       3       1       4       1
       5       6       7       5       7       8       9       7
       9      10       9      11       9      12      13       7
```

Bond angles (defined by triplets of atom numbers)……..

```
     907 !NTHETA: angles
       1       5       6       1       5      17       2       1       5
       2       1       4       2       1       3       3       1       5
       3       1       4       4       1       5       5      17      18
```

Dihedral angles (defined by sets of four atom numbers)……..

```
   171 !NPHI: dihedrals
      1       5       7       9       1       5       7      13
      1       5       7       8       1       5      17      19
```

Improper torsion angles (*sets of four atom numbers)*……..

```
      7 !NIMPHI: impropers
     17       5      19      18      19      17      21      20
     33      21      35      34      35      33      37      36
```

……..

The coordinates of the initial structure are taken from val_solv.pdb file (written according to PDB format):

```
          (a) (b)  (c)  (d) (e)____ (f)                          (g)   (h)        (i)  (j)
ATOM      1   N    VAL  P   1       -0.756   6.281   0.591  1.00  0.00       P1   N
ATOM      2   HT1  VAL  P   1       -1.263   7.048   0.197  1.00  0.00       P1   H
ATOM      3   HT2  VAL  P   1        0.204   6.329   0.315  1.00  0.00       P1   H
ATOM      4   HT3  VAL  P   1       -0.825   6.303   1.588  1.00  0.00       P1   H
ATOM      5   CA   VAL  P   1       -1.325   5.039   0.105  1.00  0.00       P1   C
ATOM      6   HA   VAL  P   1       -2.348   4.996   0.459  1.00  0.00       P1   H
ATOM      7   CB   VAL  P   1       -1.234   5.041  -1.392  1.00  0.00       P1   C
ATOM      8   HB   VAL  P   1       -1.412   6.088  -1.739  1.00  0.00       P1   H
……..
ATOM    264   H1   TIP3W 534        0.640  -6.942   7.449  1.00  0.00       WT1  H
ATOM    265   H2   TIP3W 534        0.973  -8.440   7.379  1.00  0.00       WT1  H
ATOM    266   OH2  TIP3W 535        7.872   3.441   1.788  1.00  0.00       WT1  O
……
```

Notes:

(a) atom number;
(b) atom name;
(c) residue name;
(d) chain identifier;
(e) residue number in the chain (may not be consequent);
(f) xyz coordinates;
(g) occupancy (confidence in determination the atom position in X-ray diffraction). In NAMD package, the value of 0.0 is assigned to all positions "guessed" during generation of psf file;
(h) temperature factor (uncertainty due to thermal disorder);
(i) segment name
(j) element symbol

The third input file is par_all27_prot.inp, which provides virtually all the parameters for CHARMM force field (not included in val_solv.psf file). These include the parameters for bond angle, length, dihedral angle, improper and Lennard-Jones potentials. This file is obtained from http://mackerell.umaryland.edu/charmm_ff.shtml. Download the file toppar_c31b1.tar.gz and extract files par_all27_prot.inp and top_all27_prot.inp.

2. Energy minimization

The first "real" step in preparing the system for production MD simulations involves energy minimization. The purpose of this stage is not to find a true global energy minimum, but to adjust the structure to the force field, particular distribution of solvent molecules, and to relax possible steric clashes created by guessing coordinates of atoms during generation of psf file. The following NAMD configuration file is used to minimize the potential energy of val_solv.pdb structure. Fig. 2 shows the decrease in the potential energy E_p of the peptide + solvent system.

```
# input topology and initial structure……..
structure       val_solv.psf            # Reading topology file
coordinates     val_solv.pdb            # Reading initial structure coordinates


#..force field block .....
paratypecharmm  on                      # Selecting the type of force field (CHARMM)
parameters      par_all27_prot.inp      #Getting the force field parameters for proteins (a)
exclude         scaled1-4               # Exclude/scale local (along the sequence)
                                        # non-bonded interactions (b)
1-4scaling      1.0                     # Scale factor for (i,i+3) EL interactions
dielectric      1.0                     # Value of the dielectric constant


# dealing with long-range interactions……..
switching       on                      # Switch VdW interactions and partition EL into
                                        # local and nonlocal terms
switchdist      8.0                     # Distance (=a), at which the switching function is
                                        # first applied (c)
cutoff          12.0                    #Distance (=b), at which VdW interactions become
                                        # zero and electrostatics becomes purely nonlocal
pairlistdist    13.5                    # Maximum distance used for generating Verlet
                                        # lists (aka in NAMD as pair list) of atoms
margin          0.0                     # Extra distance used in selecting the patches (d)
stepspercycle   20                      # Frequency of updating Verlet list (in integration
                                        #  steps)
rigidBonds      all                     # Apply SHAKE algorithm to all covalent bonds
                                        #  involving hydrogens
rigidTolerance  0.00001                 # Desired accuracy in maintaining SHAKEed bond
                                        #  lengths
rigidIterations 500                     # Maximum number of SHAKE iterations


# this block specifies the Ewald electrostatics.....
PME             on                      # Use Particle-Mesh Ewald summation for long-
                                        # range electrostatics
PMETolerance    0.000001                # The accuracy of computing the Ewald real space
```

```
                                # (direct) term
PMEGridSizeX    32              # Setting the grids of points for
PMEGridSizeY    32              # fast calculation of reciprocal term
PMEGridSizeZ    32              # along x,y and z directions

minimization    on              #  Do conjugate gradient minimization of potential
                                #  energy (instead of MD run)
```

 # this block specifies the output….

```
outputenergies  1000            # Interval in integration steps  of writing energies
                                # to stdout
outputtiming    1000            # Interval of writing timing (basically, speed,
                                # memory allocation, etc) to stdout
binaryoutput    no              # Are binary files used for saving last structure?
outputname      output/val_min  # The file name (without extension!), to which final
                                # coordinates and velocities are to be saved
                                # (appended extensions are *.coor or *.vel)
restartname     output/vali     # The file name (without extension), which holds
                                # the restart structure and velocities
                                # (appended extensions are *.coor or *.vel)
restartfreq     10000           # Interval between writing out the restart
                                # coordinates and velocities
binaryrestart   no              # Are restart files binary?
DCDfile         output/val_min.dcd  # Trajectory filename (binary file)
dcdfreq         1000            # Frequency of writing structural snapshots to
                                # trajectory file
numsteps        2000            # Number of minimization steps
```

 # this block defines periodic boundary conditions.....

```
cellBasisVector1   29.4   0.0   0.0    # Direction of the x basis vector for a unit cell
cellBasisVector2    0.0  29.4   0.0    # Direction of the y basis vector for a unit cell
cellBasisVector3    0.0   0.0  29.4    # Direction of the z basis vector for a unit cell
cellOrigin          0.0   0.0   0.0    # Position of the unit cell center
wrapWater           on                 # Are water molecules translated back to the unit
                                       # cell (purely cosmetic option, has no effect on
                                       #  simulations)
```

Notes:

(a) CHARMM27 force field is used;

(b) This setting implies exclusion of all non-bonded interactions for the atoms (*i,i+1,i+2)* and scaling the EL interactions between (*i,i+3*) by the factor 1-4scaling. The scaling of van-der-Waals (VdW) interactions between the atoms *i* and *i+3* is set in par_all27_prot.inp file;

(c) All distances are in angstroms. All "times" are expressed in the number of integration steps;

(d) This distance is used for generating patches or groups of atoms, which should not be separated by Verlet lists.

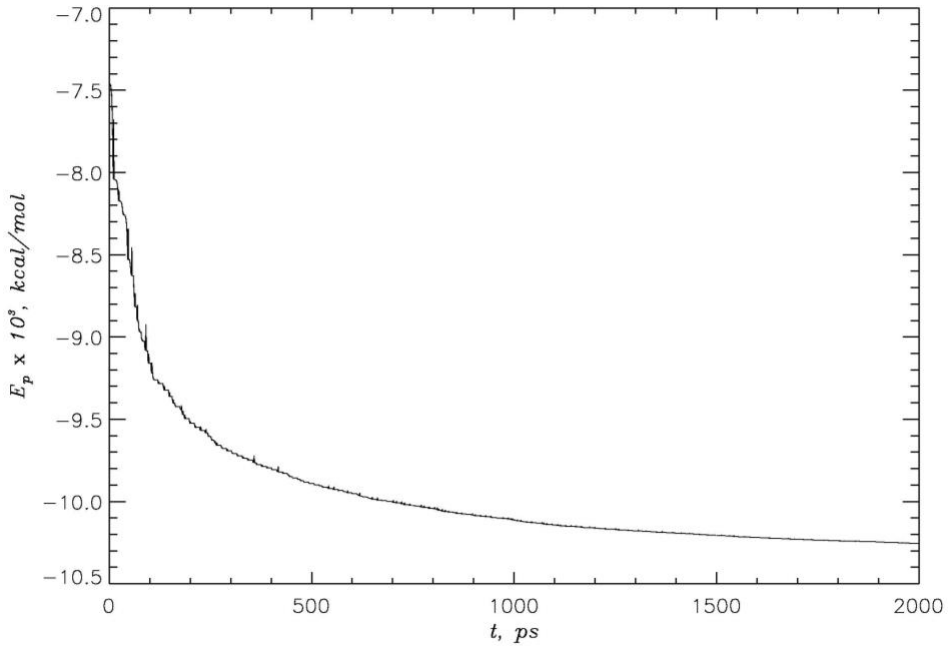

Figure 2. The decrease in potential energy of polyvaline peptide during minimization.

3. Heating the simulation system

During this stage the temperature of the system is linearly increased from 0K to 300K within 300 ps. At each integration step velocities are reassigned (i.e., drawn) from a new Maxwell distribution and the temperature is incremented by 0.001K. Fig. 3 shows the increase in the temperature *T(t)* (upper panel) and potential energy E_p (lower panel) as a function of time. Note that instantaneous temperature *T(t)* and energy E_p fluctuate due to finite system size. The keywords, which have not been changed compared to previous simulations stage, are marked in grey.

```
# input topology and initial structure……
structure       val_solv.psf
coordinates     ./output/val_min.coor     #  Start heating simulations from the minimized
                                          #  structure (a)

#..force field block.....................
paratypecharmm      on
parameters          par_all27_prot.inp
```

```
exclude                 scaled1-4
1-4scaling              1.0
dielectric              1.0


# dealing with long-range interactions……..
switching               on
switchdist              8.0
cutoff                  12.0
pairlistdist            13.5
margin                  0.0
stepspercycle           20
rigidBonds              all
rigidTolerance          0.00001
rigidIterations         100


# this block specifies the Ewald electrostatics.....
PME             on
PMETolerance    0.000001
PMEGridSizeX    32
PMEGridSizeY    32
PMEGridSizeZ    32


#block specifying the parameters for integrator and MTS
timestep                1.0             # Integration time step in fs
fullElectFrequency      4               # The interval between calculation of long-
                                        # range electrostatics using PME method.
                                        # Short-range nonbonded interactions are
                                        # computed at every integration step by default (b)


# this block specifies the output….
outputenergies          1000
outputtiming            1000
binaryoutput             no
outputname      output/val_heat010      #The file name (without extension) to which final
                                         # coordinates and velocities are to be saved
                                         # (appended extensions are *.coor or *.vel)
restartname      output/vali_heat010    # The file name (without extension), which holds
                                          # the restart structure and velocities
                                          # (appended extensions are *.coor or *.vel)
restartfreq     10000
binaryrestart   yes                      # Use binary restart files (c)
DCDfile         output/val_heat010.dcd   # Trajectory filename (binary file)
dcdfreq         1000


#MD protocol block …….
seed            1010            # Random number seed used to generate initial
```

```
                                    # Maxwell distribution of velocities (d)
numsteps        300000              # Number of integration steps
temperature     0                   # Initial temperature (in K), at which initial velocity
                                    # distribution is generated
reassignFreq    1                   # Number of steps between reassignment of
                                    # velocities
reassignIncr    0.001               # Increment used to adjust temperature
                                     # during temperature reassignment
reassignHold    300                 # The value of temperature to be kept after heating is
                                    # completed
```

```
# this block defines periodic boundary conditions.....
cellBasisVector1   29.4    0.0   0.0
cellBasisVector2    0.0   29.4   0.0
cellBasisVector3    0.0    0.0  29.4
cellOrigin          0.0    0.0   0.0
wrapWater             on
```

Notes:

(a) binary format for input files is not used here;
(b) multisteping (r-RESPA) integration of equations of motion is used by default;
(c) restrart files (*.coor and *.vel) are saved in binary format. This option provides better numerical accuracy during writing out/reading in of restart files;
(d) Seed should "identify" (i.e., be unique) for a given trajectory. Only applicable if temperature keyword is present

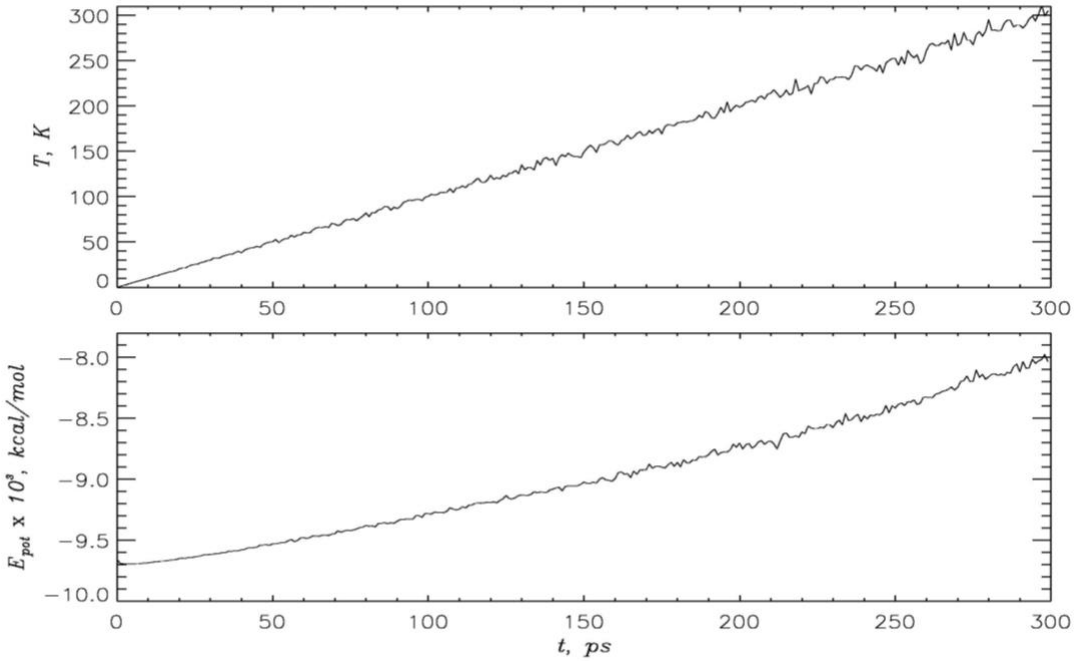

Figure 3. Increase in instantaneous temperature (upper panel) and potential energy (lower panel) during heating

4. Equilibration at constant temperature

Equilibration stage is used to equilibrate kinetic and potential energies, i.e., to distribute the kinetic energy "pumped" into the system during heating among all degrees of freedom. This usually implies that the potential energy "lags on" and must be equilibrated with the kinetic energy. In other words, the kinetic energy must be transferred to potential energy. As soon as potential energy levels off the equilibration stage is completed. Fig. 4 shows the temperature $T(t)$ (upper panel) and the potential energy E_p (lower panel) as a function of time. After initial rapid increase from about -8000 *kcal/mol* the potential energy E_p fluctuates near the constant level. This behavior suggests that the system is equilibrated on a timescale much shorter than 300 *ps*.

```
# input topology and initial structure……..
structure       val_solv.psf
coordinates     ./output/val_heat010.coor     # Start equilibration with the final structure
                                                # from the heating stage of MD trajectory

#..force field block.....................
paratypecharmm              on
parameters                  par_all27_prot.inp
exclude                     scaled1-4
1-4scaling                  1.0
dielectric                  1.0


# dealing with long-range interactions……..
switching                   on
switchdist                  8.0
cutoff                      12.0
pairlistdist                13.5
margin                      0.0
stepspercycle               20
rigidBonds                  all
rigidTolerance              0.00001
rigidIterations             100


# this block specifies the Ewald electrostatics.....................
PME                         on
PMETolerance                0.000001
PMEGridSizeX                32
PMEGridSizeY                32
PMEGridSizeZ                32
```

```
# this block specifies the parameters for integrator and MTS
timestep 1.0
fullElectFrequency  4
# this block deternines the output….
outputenergies                  1000
outputtiming                    1000
binaryoutput                    no
outputname    output/val_equil010      #The file name (without extension) to which final
                                        # coordinates and velocities are to be saved
                                        # (appended extensions are *.coor or *.vel)
restartname    output/vali_equil010      #The file name (without extension), which holds
                                        # the restart structure and velocities
                                        # (appended extensions are *.coor or *.vel)
restartfreq                     10000
binaryrestart                   yes
DCDfile       output/val_equil010.dcd    # Trajectory filename (binary file)
dcdfreq                         1000


#MD protocol block …..…….
seed          2010                      # Random number seed used to generate initial
                                        # Maxwell distribution of velocities
numsteps      300000
temperature   300                       # Temperature of initial velocity generation
rescaleFreq   1                         # Number of integration steps between rescaling
                                        # velocities to a given temperature
rescaleTemp    300                      # Temperature, to which velocities are rescaled


# this block defines periodic boundary conditions.....
cellBasisVector1    29.4    0.0    0.0
cellBasisVector2     0.0  29.4    0.0
cellBasisVector3     0.0    0.0  29.4
cellOrigin           0.0   0.0    0.0
wrapWater             on
```

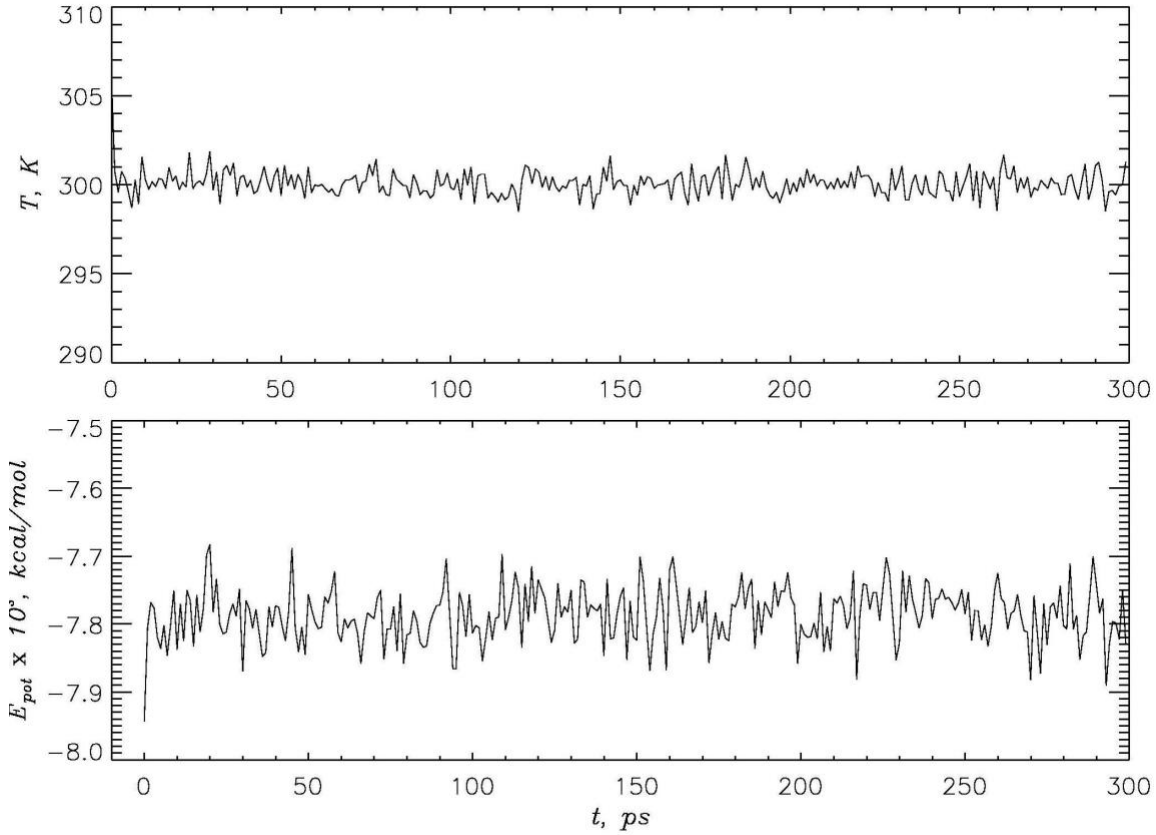

Figure 4. Time dependence of instantaneous temperature (upper panel) and potential energy (lower panel) during equilibration stage.

Note that the potential energy in Fig. 4 is leveled off at about -7800 *kcal/mol,* whereas the final energy during heating stage is somewhat smaller (≈-8000 *kcal/mol*). Although all velocities are rescaled to 300K at every integration step, fluctuations in temperature are still visible in Fig. 4, because the temperatures are plotted after advancing the velocities using Verlet algorithm, but ***before*** actual rescaling.

5. Production stage of MD trajectory (NVE ensemble)

This stage of MD trajectory is used to sample the structural characteristics and dynamics of polyvaline peptide at 300K. Velocity reassigning or rescaling must now be turned off. The fluctuations in instantaneous temperature and nearly constant total energy (***kinetic plus potential*** energies) are shown in Fig. 5.

```
# input topology and initial structure…..
structure       val_solv.psf
coordinates     ./output/val_equil010.coor   # Reading the final structure from
bincoordinates./output/vali_equil010.coor  # equilibration stage in a binary format (a)
#binvelocities  ./output/vali_equil010.vel    # Initial velocities from restart file (b)


#..force field block.......
paratypecharmm      on
parameters          par_all27_prot.inp
exclude             scaled1-4
1-4scaling          1.0
dielectric          1.0


# dealing with long-range interactions……..
switching           on
switchdist          8.0
cutoff              12.0
pairlistdist        13.5
margin              0.0
stepspercycle       20
rigidBonds          all
rigidTolerance      0.00001
rigidIterations     100


# this block specifies the Ewald electrostatics.....
PME                 on
PMETolerance        0.000001
PMEGridSizeX        32
PMEGridSizeY        32
PMEGridSizeZ        32


# this block specifies the parameters for integrator and MTS
timestep            1.0
fullElectFrequency  4


# this block determines the output….
outputenergies      1000
outputtiming        1000
binaryoutput        no
outputname    output/val_quench010    # The file name (without extension) to which final
                                       # coordinates and velocities are to be saved
                                       # (appended extensions are .coor or vel)
restartname   output/vali_quench010   # The file name (without extension), which holds
                                      # the restart structure and velocities
                                      # (appended extensions are *.coor or *.vel)
restartfreq         10000
```

```
binaryrestart           yes
DCDfile         output/val_quench010.dcd  # Trajectory filename (binary file)
dcdfreq                 1000


#MD protocol block.....
seed                    3010            # Random number seed used to generate initial
                                        # Maxwell distribution of velocities
numsteps                1000000         # Number of integration steps during
                                        # production simulations
temperature     300                     #  see also (b)


# this block defines periodic boundary conditions.....
cellBasisVector1    29.4    0.0    0.0
cellBasisVector2     0.0   29.4    0.0
cellBasisVector3     0.0    0.0   29.4
cellOrigin           0.0    0.0    0.0
wrapWater            on
```

Notes:

(a) Initial coordinates are read from equilibration restart file in a binary format. The keyword coordinates has no meaning, although must be used as per NAMD specifications

(b) This particular configuration file forces NAMD to generate a new Maxwell distribution of velocities during initialization of production run (the keyword temperature is present). An alternative option is to read the velocities saved in a binary form from previous simulation stage. In this case the temperature is "imported" with the velocities and the keyword temperature must be omitted, whereas the keyword binvelocities must be uncommented.

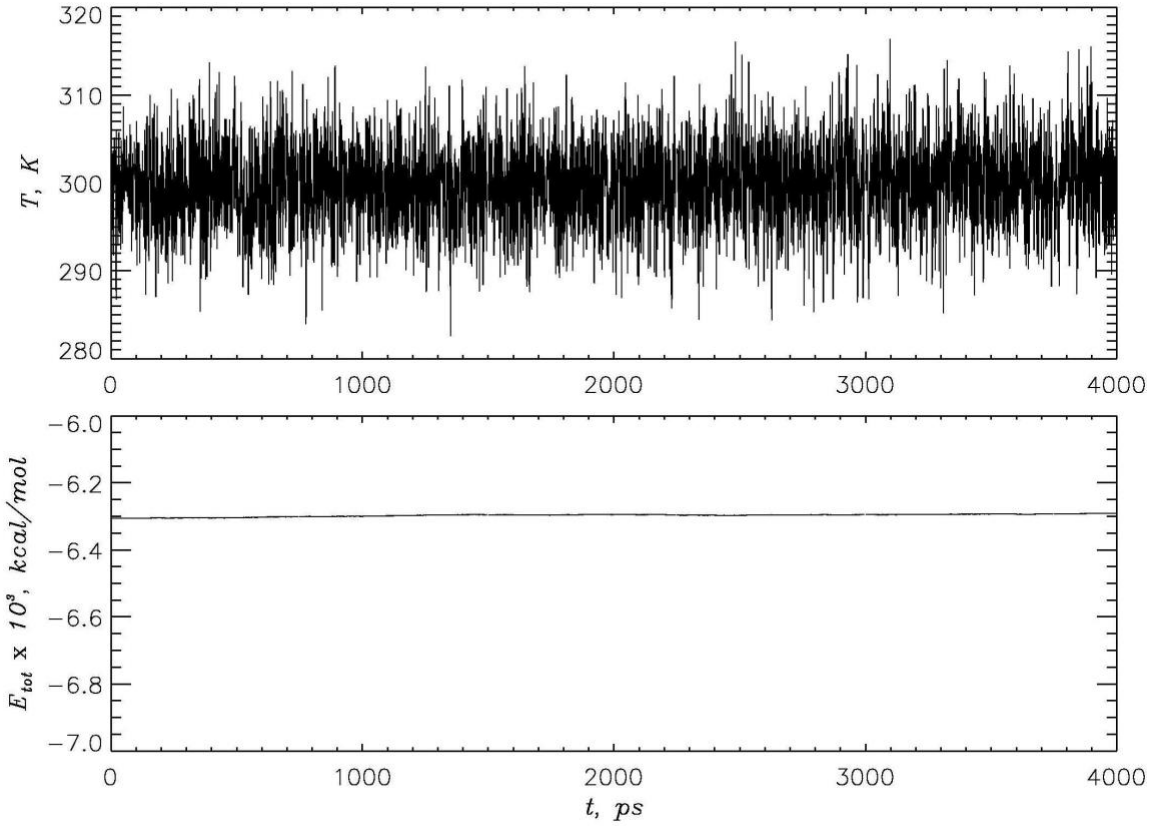

Figure 5. Fluctuations in instantaneous temperature (upper panel) and approximately constant total energy (lower panel) are shown for production NVE simulations of polyvaline.