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) capped with acetyl and amide groups. 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.

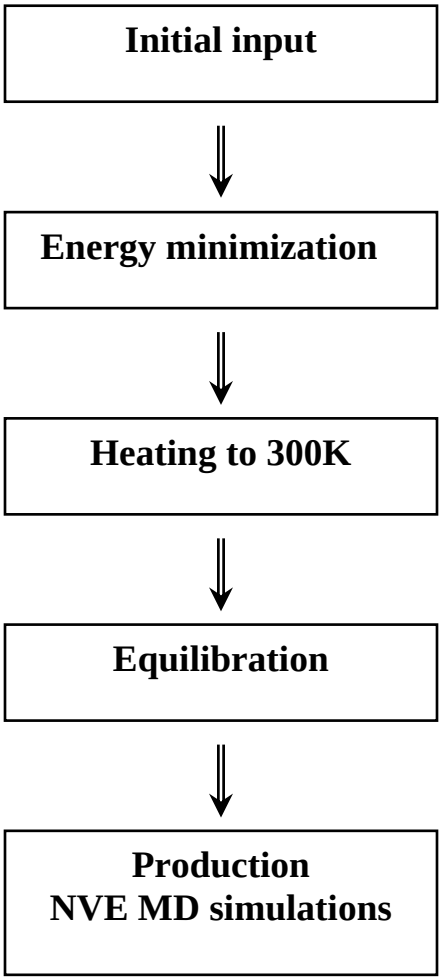

Figure 1. Strategy for performing generic MD simulations.

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_all22_prot.inp (part of CHARMM22 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_all22_prot.inp file). In all, the solvated polyvaline system contains 2476 atoms and 805 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:

2476 !NATOM

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
1	VAL4	1	VAL	CAY	CT3	-0.270000	12.0110	0
2	VAL4	1	VAL	HY1	HA	0.090000	1.0080	0
3	VAL4	1	VAL	HY2	HA	0.090000	1.0080	0
4	VAL4	1	VAL	HY3	HA	0.090000	1.0080	0
5	VAL4	1	VAL	CY	C	0.510000	12.0110	0
6	VAL4	1	VAL	OY	O	-0.510000	15.9990	0
7	VAL4	1	VAL	N	NH1	-0.470000	14.0070	0
8	VAL4	1	VAL	HN	H	0.310000	1.0080	0
9	VAL4	1	VAL	CA	CT1	0.070000	12.0110	0
10	VAL4	1	VAL	HA	HB	0.090000	1.0080	0
11	VAL4	1	VAL	CB	CT1	-0.090000	12.0110	0
12	VAL4	1	VAL	HB	HA	0.090000	1.0080	0
13	VAL4	1	VAL	CG1	CT3	-0.270000	12.0110	0
14	VAL4	1	VAL	HG11	HA	0.090000	1.0080	0
15	VAL4	1	VAL	HG12	HA	0.090000	1.0080	0
16	VAL4	1	VAL	HG13	HA	0.090000	1.0080	0
17	VAL4	1	VAL	CG2	CT3	-0.270000	12.0110	0
18	VAL4	1	VAL	HG21	HA	0.090000	1.0080	0
19	VAL4	1	VAL	HG22	HA	0.090000	1.0080	0
20	VAL4	1	VAL	HG23	HA	0.090000	1.0080	0
21	VAL4	1	VAL	C	C	0.510000	12.0110	0
22	VAL4	1	VAL	O	O	-0.510000	15.9990	0

……

```
  302  SOLV 81   TIP3 OH2  OT    -0.834000       15.9994          0
  303  SOLV 81   TIP3 H1    HT     0.417000       1.0080           0
  304  SOLV 81   TIP3 H2    HT     0.417000       1.0080           0
……
```

(a) atom number;
(b) segment name, e.g., VAL4 is the name of peptide segment in the solvated system. The other segment is solvent (SOLV);
(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 six atoms correspond to acetyl cap (CH3-CO-), which is incorporated by CHARMM convention into the first residue. The amide cap (-NH2) is included in the last residue. The N- and C-terminal caps are distinguished by the letters "Y" and "T" in the atom names, 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)……..

```
2475 !NBOND: bonds  # number of bonds
     1     2     1     3     1     4     5     1
     5     7     6     5     7     8     7     9
     9    10    11     9    11    12    13    11
```

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

```
933 !NTHETA: bond angles # number of bond angles
     1     5     6     1     5     7     2     1     4
     2     1     3     3     1     4     5     7     9
     5     7     8     5     1     4     5     1     3
```

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

```
183 !NPHI: dihedrals # number of dihedral angles
     1     5     7     8     1     5     7     9
     2     1     5     7     2     1     5     6
```

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

```
12 !NIMPHI: impropers # number of improper torsions
     5     1     7     6     7     5     9     8
    21     9    23    22    23    21    25    24
……..
```

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)
ATOM      1   CAY  VAL    1     -3.881  -1.044  -2.178   0.00  0.00     VAL4
ATOM      2   HY1  VAL    1     -3.778  -1.657  -1.394   0.00  0.00     VAL4
ATOM      3   HY2  VAL    1     -3.019  -0.986  -2.681   0.00  0.00     VAL4
ATOM      4   HY3  VAL    1     -4.605  -1.380  -2.780   0.00  0.00     VAL4
ATOM      5   CY   VAL    1     -4.128  -0.133  -1.848   0.00  0.00     VAL4
ATOM      6   OY   VAL    1     -4.231   0.480  -2.632   0.00  0.00     VAL4
ATOM      7   N    VAL    1     -4.185  -0.152  -0.850   1.00  0.00     VAL4
ATOM      8   HN   VAL    1     -4.049  -0.874  -0.158   1.00  0.00     VAL4
……..
ATOM    254   OH2 TIP3   65      0.440  13.787 -14.273   1.00  0.00     SOLV
ATOM    255   H1  TIP3   65      0.953  14.520 -13.933   1.00  0.00     SOLV
ATOM    256   H2  TIP3   65      0.764  13.660 -15.165   1.00  0.00     SOLV
……
```

Notes:

(a) atom number;
(b) atom name;
(c) residue name;
(d) residue number (may not be consequent);
(e) xyz coordinates;
(f) 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;
(g) temperature factor (uncertainty due to thermal disorder);
(h) segment name

The third input file is par_all22_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 www.pharmacy.umaryland.edu/faculty/amackere/force_fields.htm. Download the file toppar_c31b1.tar.gz and extract files par_all22_prot.inp and top_all22_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_all22_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) CHARMM22 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_all22_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.

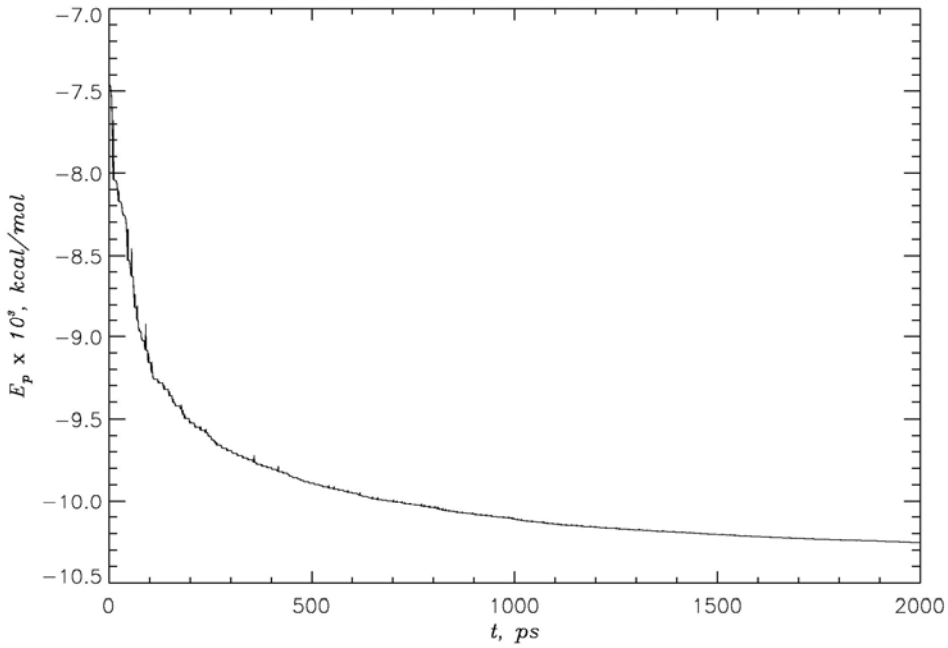

Figure 2. The decrease in potential energy of polyvaline peptide during minimization.

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_all22_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

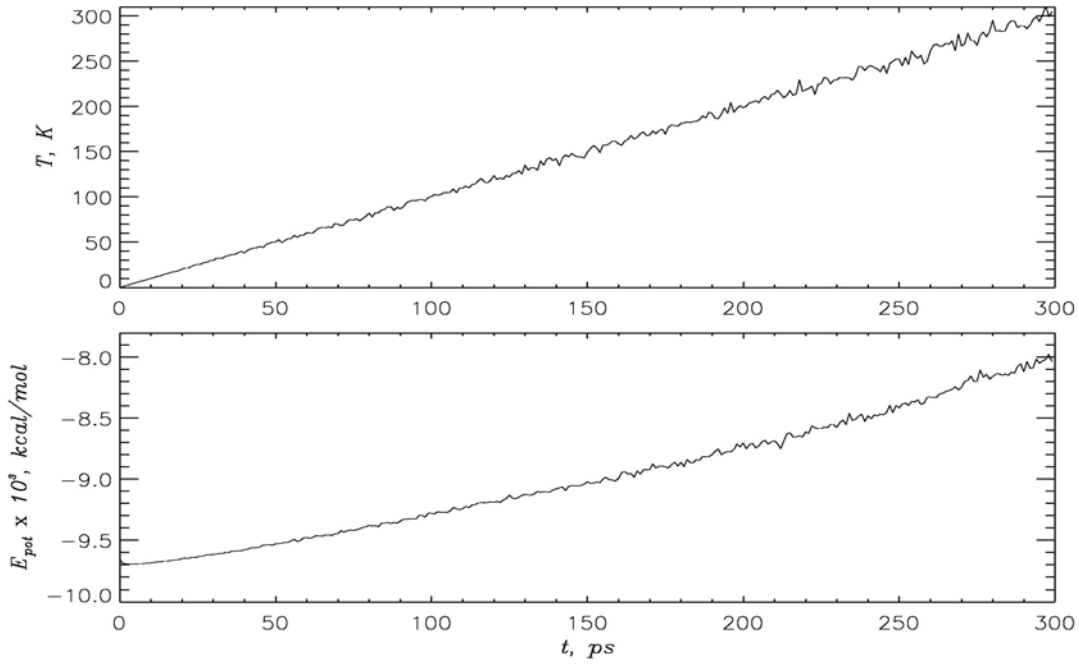

Figure 3. Increase in instantaneous temperature (upper panel) and potential energy (lower panel) during heating

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_all22_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_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
```

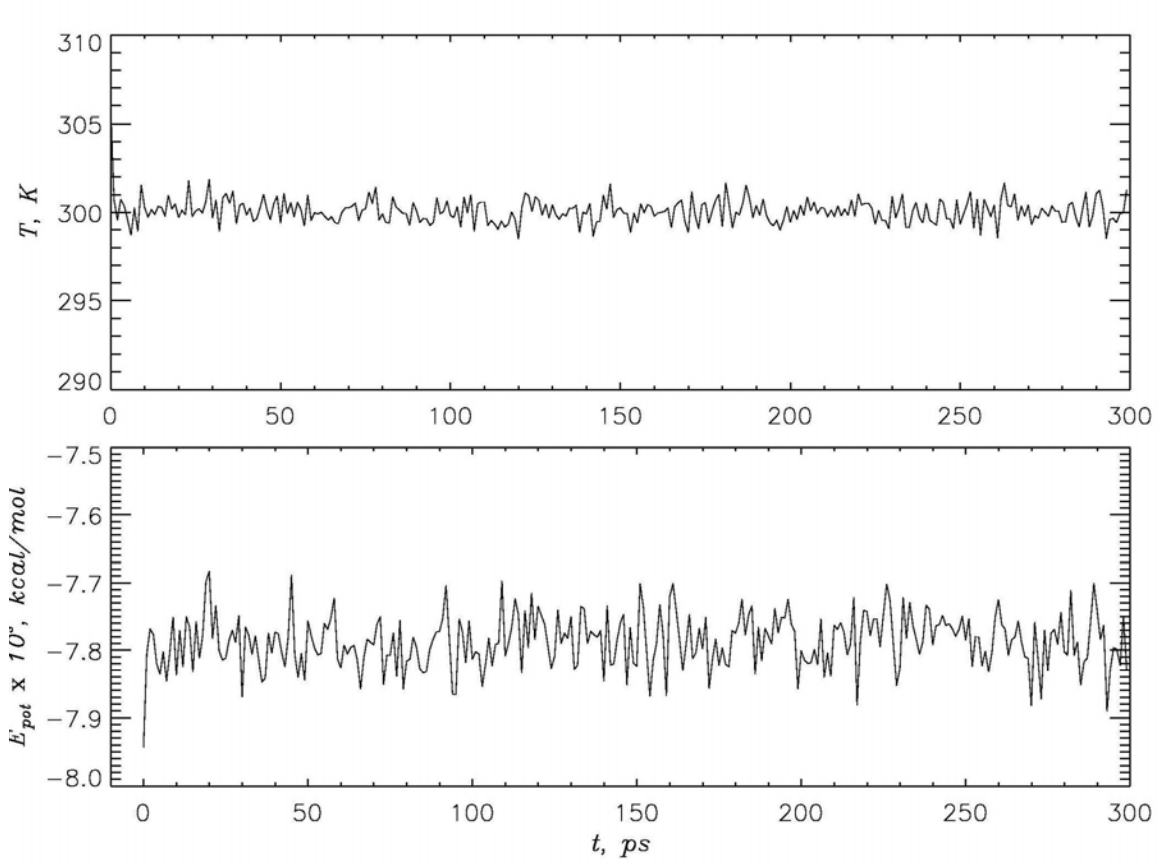

Figure 4. Time dependence of instantaneous temperature (upper panel) and potential energy (lower panel) during equilibration stage.

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_all22_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.

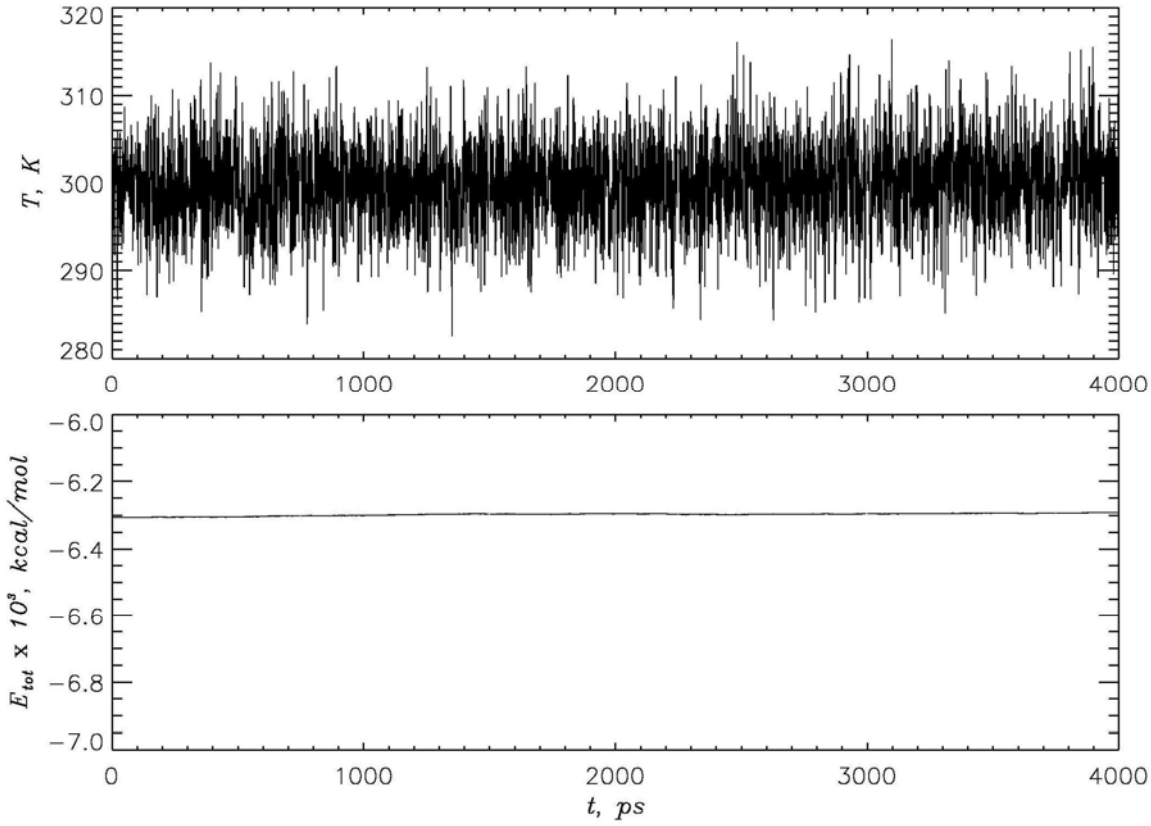

Figure 5. Fluctuations in instantaneous temperature (upper panel) and approximately constant total energy (lower panel) are shown for production NVE simulations of polyvaline.

Part II. Molecular dynamics project: Simulations of peptides in explicit water

The purpose of this project is to perform MD simulations for short peptides and explore their secondary structure propensities. Two peptides, acetyl-VVVV-NH_2 and acetyl-AAAAAAAAA-NH_2, are to be used in the simulations. From the analysis of secondary structure propensities of proteins in structural databases it is known that valine and alanine generally favor β-strand and α-helix conformations, respectively. Therefore, MD simulations at normal physiological conditions (300K, normal pH, no salt) are expected to reflect these propensities.

Note that polyvaline sequence is shorter than polyalanine, because β-strand conformation usually does not rely on interresidue contacts (other than steric clashes). In contrast,

helical structures draw their stability, at least partially, from the (*i*,*i+4*) interactions (usually, hydrogen bonds). Accordingly, polyalanine sequence must be long enough to form at least two full helical turns. Because there are usually 3.6 residues per turn in an α-helix and polyalanine sequence is nine residues long, a short α-helix is expected to form in these simulations.

Secondary structure propensities can be examined by computing the distribution of (ϕ,ψ) backbone dihedral angles collected during production MD simulations. The best way to visualize the distribution $P(\phi,\psi)$ is to plot the Ramachandran plots for both peptides. All data are to be obtained from standard microcanonical (NVE) MD simulations of 1 ns long using NAMD program.

Input files: There are three input files for each peptide. For polyvaline the topology file and the file with the solvated peptide coordinates are val_solv.psf and val_solv.pdb, respectively. For polyalanine the corresponding files are ala2_solv.psf and ala2_solv.pdb. For both peptides the CHARMM22 parameter file par_all22_prot.inp is used. The density of water is approximately 1.0 g/cm^3 for both peptides. The sizes of the unit cells for solvated polyvaline and polyalanine are 29.4 x 29.4 x 29.4 A and 34.7 x 34.7 x 34.7 A, respectively.

Energy minimization: Because the initial structures of the peptides were generated using a template based on the structure of other protein and initial distribution of water molecules may create steric clashes, energy minimization is required to relax "strained" local conformations. The NAMD configuration files for performing energy minimization for polyvaline and polyalanine peptides are val_min.namd and ala2_min.namd, respectively. After completing minimization the potential energy of the system E_p must be plotted as a function of minimization steps to assure that E_p nearly levels off by the end of minimization stage.

The energy can be extracted from the stdout produced by NAMD. For convenience, the output should be redirected to, e.g., val_min.out file, the part of which is given below. In general, various energy terms are output by NAMD. Specifically, look for the lines starting with "ENERGY:' and a single line "ETITLE:", which describes the "identity" of the energy terms in the line "ENERGY:". The total energy (kinetic E_k+potential E_p) corresponds to the eleventh term, while the kinetic energy is given by the tenth term. During minimization total energy coincides with the potential energy.

To start NAMD minimization first copy all input files into a dedicated directory. Then create a subdirectory output, in which trajectory (*.dcd), coordinate (*.coor), and velocity (*.vel) files will be output by NAMD.

To start NAMD minimization type the following command in the terminal window:

```
namd2 val_min.namd > val_min.out &
```

For polyalanine use ala2_min.namd.

Fragment of the file val_min.out (created as a result of redirection of stdout) is shown below:

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

ENERGY:       0       213.4650        82.1573        11.9522         2.8768
      -8786.9485       1018.5613         0.0000         0.0000         0.0000
      -7457.9359          0.0000     -7457.9359     -7457.9359         0.0000
      21899.6229       1661.0030     25412.1840     21899.6229      1661.0030
……..……
```

Heating to 300K: Once energy minimization is completed, the system is to be heated from 0 to 300K as described in the section "Part I. Strategy for running molecular dynamics simulations". The configuration file for polyvaline system is val_heat010.namd. Modify this file to obtain similar configuration file for heating the polyalanine peptide ala2_heat010.namd. Remember to change names of input/output files, random seed value, and the size of the unit cell.

Plot the instantaneous temperature *T(t)* and the potential energy $E_p(t)$ to verify the system behavior. *T(t)* can be extracted from, e.g., val_heat010.out file. Remember to subtract kinetic energy (the tenth term in "ENERGY:" line) from the total energy to get $E_p(t)$. Set the length of heating trajectory to 300 ps.

To start NAMD heating simulations use the command:

```
namd2 val_heat010.namd  > val_heat010.out &
```

For polyalanine use ala2_heat010.namd.

Equilibration at 300K: Perform equilibration of both polypeptides at 300 K by rescaling velocities at every integration step for 300 ps as described in the section "Part I. Strategy for running molecular dynamics simulations". It is likely that the systems will be equilibrated much faster than 300 ps, but this long equilibration stage provides an opportunity to monitor instantaneous temperatures *T(t)* and energies $E_p(t)$.

Plot these quantities to demonstrate that their baselines have no apparent time dependence. The configuration file for equilibrating polyvaline is val_equil010.namd. Modify this file to obtain the configuration file for polyalanine peptide.

To start NAMD equilibration simulations for polyvaline use the command:

```
namd2 val_equil010.namd > val_equil010.out &
```

For polyalanine use ala2_equil010.namd.

Production stage (NVE ensemble): Generate 1 ns trajectories for both peptides using microcanonical (NVE) ensemble. The configuration file for polyvaline simulations is val_quench010.namd (modify this file accordingly to run simulations for polyalanine peptide).

The structural snapshots (trajectory) will be saved in a "DCD" file in the directory ./output. Make the plots and check that the total energy $E_p(t)+ E_k(t)$ and instantaneous temperatures $T(t)$ in the stdout files for both peptides (such as val_quench010.out) remain constant or the baseline remains approximately constant for $T(t)$.

To start NAMD NVE simulations for polyvaline use the command:

```
namd2 val_quench010.namd > val_quench010.out &
```

For polyalanine use ala2_quench010.namd.

If for some unforeseen reason simulations crash (power outage etc), repeat the stage affected by the crash.

Computing the distribution of dihedral angles: The backbone dihedral angles (ϕ,ψ) are to be computed using the positions of heavy backbone atoms saved every 1 ps. To this end the coordinates of peptide backbone in pdb format must be first extracted from binary DCD files. Use for this purpose VMD program, which may be downloaded from http://www.ks.uiuc.edu/Research/vmd/ (use version 1.8.3).

Manual is available at http://www.ks.uiuc.edu/Research/vmd/current/ug/ug.html and an excellent tutorial can be found at http://www.ks.uiuc.edu/Training/Tutorials/vmd/tutorial-html/index.html. VMD will be installed on Linux machines in the computer lab. Type vmd in the terminal window to start the program.

How to extract backbone coordinates:

1. Place tcl script getpdb.tcl in the directory, where dcd files are located.

2. Start vmd program and load in the VMD Main window two files val_solv.psf and val_quench010.dcd.

3. Select File menu and then New Molecule. Browse for the psf (in the directory one level up) and dcd files and load both in this particular order.

4. Open Extensions menu from the VMD Main window and choose Tk console (tcl console), which will open in a new window.

5. Change the directory in Tcl console to that of the dcd file and start tcl script getpdb.tcl using the command source getpdb.tcl. This script reads dcd file and saves the coordinates of N, Ca, and C backbone atoms into pdb file.

For polyvaline peptide this procedure will output pdb file val_quench010.pdb. For polyalanine replace val_quench010.pdb in the tcl script with ala2_quench010.pdb.

Compute the backbone dihedral angles (ϕ,ψ) for polyvaline and polyalanine using the extracted pdb files and the program for calculating dihedral angles. Plot the distributions of dihedral angles $P(\phi,\psi)$ on the Ramachandran plot (Fig. 6). Note that for the first and last residues only one angle (ϕ or ψ) is defined. These residues must be excluded from calculations. For example, for four residue polyvaline only two pairs of (ϕ,ψ) are to be plotted from each conformational snapshot. When the distributions $P(\phi,\psi)$ is computed, compare them with the typical β-strand and α-helix regions. Use the following definitions (see also *Proteins: Struct. Funct. Gen*. **20**, 301 (1994) and *Proc. Natl. Acad. Sci USA* **96**, 9074 (1999)).

β-strand is defined by a polygon with the vertices:
(-180,180),(-180,126),(-162,126),(-162,108),(-144,108),(-144,90),(-50,90),(-50,180)

α-helix is defined by a polygon with the vertices:
(-90,0),(-90,-54),(-72,-54),(-72,-72),(-36,-72),(-36,-18),(-54,-18),(-54,0)

Visualization of simulations: It is instructive to visualize each step of MD simulations using VMD. Once a stage of the MD project is completed, start VMD in the directory output. Load psf and dcd files as described above. The animation will start automatically in the VMD display window. Although the default molecular representation is rather poor, there are many more advanced graphical representations available in VMD. Here a VDW option is illustrated.

1. Select Graphics menu in VMD Main window and choose Representations. There clear the Selected Atoms line, type "protein", and press Enter.

2. Select VDW for the Drawing method. Choose Coloring method and select, e.g., Type. Protein atoms will be colored according to their type.

3. You may want to adjust the size of atom spheres by changing the Sphere Scale. Click Create Representation to store currently created view as a separate representation. Then clear the Selected Atoms line again, type "water", and press Enter.

4. Keep VDW as the Drawing method. Select Coloring Method and choose ResName. Click Create Representation to store the water view as a separate representation.

5. To save the current view (protein+water) in the script file, select File menu in the VMD Main window and choose Save State option. The saved VMD script will have the extension *.vmd. Before loading at a later time a saved state make sure that VMD Main window show no other loaded molecules.

6. If there are loaded molecules, highlight them, select Molecule menu and choose Delete Molecule. As a result for each MD stage there will be a separate VMD script file, containing a graphical representation.

7. Animations of the full system (peptide+water) may be slow depending on CPU power. In this case, either reduce the sphere size or eliminate the water representation.

8. The snapshots may be saved as postscript pictures using the option Render in the File menu.

Project report: The project report must contain a full description and explanation of all the simulation steps taken. The following plots must be included for both polyvaline and polyalanine peptides:

1. Dependence of the potential energy on minimization step

2. Time dependencies of the potential energy and temperature for heating and equilibration stages

3. Time dependence of the total energy and temperature during production run

4. Ramachandran plots showing dihedral angle distributions and boundaries of helical and strand structure (as in Fig. 6)

5. Include pictures of the final production structure (peptide plus water) for each peptide visualized using VMD.

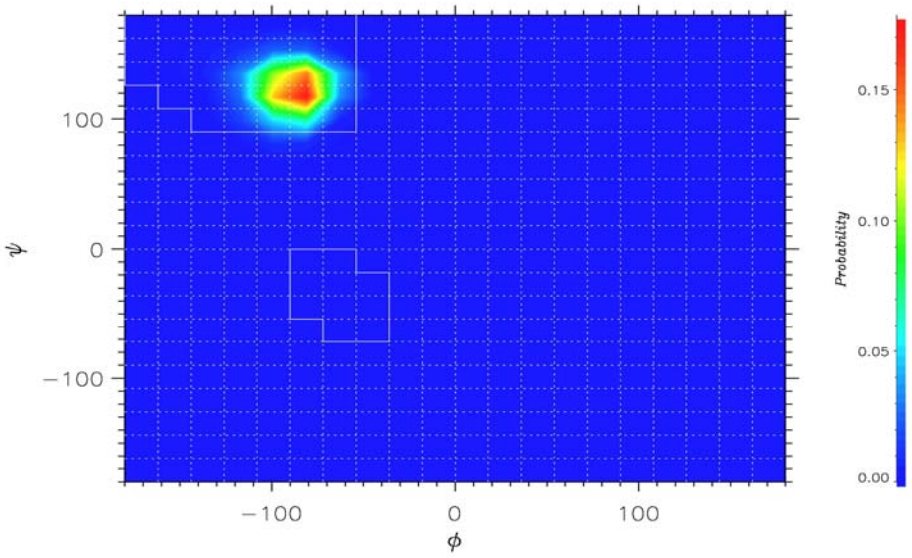

Figure 6. Ramachandran plot for polyvaline peptide. The probability of dihedral angles $P(\phi,\psi)$ is color coded. All dihedral angles are localized in the β-strand region.

Part III. Solvation of peptide structure

In most MD simulations the very first step involves the solvation of a protein structure with solvent (usually water). This means that water molecules must be randomly distributed with some density around a protein to fill the entire unit cell. In what follows the preparation of solvated polyvaline peptide is considered. Initial polyvaline peptide structure (with no water around) is contained in the file val0.pdb. (Note that this file contains only heavy atoms and no hydrogen atoms.) The easy way to solvate this structure is to use VMD program.

There are two main steps in preparing solvated system:

1. In the command line window of VMD (vmd console) type the command play build.vmd. The script build.vmd loads the pdb file and call in the psfgen program, which serves as a plugin for VMD. psfgen is a tcl interpreter, which reads in a CHARMM topology file top_all22_prot.inp and based on the sequence information in pdb file creates a psf peptide topology file. psfgen adds missing atoms (hydrogens) as well as adds terminals (acetyl terminal ACE and standard NH2 terminal CT2). The resulting peptide topology file is saved to val.psf. psfgen then loads again pdb file and guesses the positions of missing atoms. The all-atom structure is finally saved to val.pdb (no water).

2. The next step is to distribute water molecules around the peptide. This is accomplished by the script solv.vmd. This script first reads the topology and the coordinates of polyvaline and then calls the program solvate, which distributes waters around the peptide. (Note solvate is a plugin for VMD and as psfgen comes with VMD package).

The following options determine the details of peptide solvation:

(1) –o val_solv specifies the name for final (pepetide+water) topology and coordinate files;

(2) –s WT indicates the name of the water box to be used for peptide solvation (water box is also included in VMD distribution);

(3) -b 2.4 gives the minimum distance between peptide and waters. All waters positioned closer than 2.4 A to the peptide are deleted;

(4) –minmax { {-15 -15 -15} {15 15 15} } specifies the dimensions of the solvated system using the format {{–xmin –ymin –zmin} {xmax ymax zmax} }.

The final solvated system val_solv.pdb has the dimensions 30 x 30 x 30 Å and contains 2422 atoms (or 787 residues). Because of slightly different water boxes, the dimensions and the number of atoms in this version of solvated polyvaline do not coincide with the

ones used in MD simulations. For polyvaline peptide the scripts build.vmd and solv.vmd can be used without any modifications. For polyalanine remember to change the following:

1. all names val must be replaced with ala2
2. change the name of the segment VAL4 with ALA9
3. to obtain proper size of solvation box set the (x,y,z)min and (x,y,z)max to 18 Å