

CHAPTER II

MATERIALS AND METHODS

2.1 Materials

2.1.1 Hardware

Desktop computers with the 3rd and 4th generation Intel® Core™ i7 processor were mainly used to operate the simulation in this work.

2.1.2 Software

The NAMD 2.9 program

The NAMD 2.9 is a molecular dynamics (MD) program, specially designed for molecular dynamics simulation of large scale biomolecular systems. To run a basic MD simulation with the program NAMD 2.9, four important files are needed including (i) a structure coordinate or PDB file which contains atomic coordinates of the systems. (ii) a protein structure file (PSF) which contains information about the atom type, mass, charge and connectivity (such as bond, angle, dihedral angles improper angle etc.) within the molecular system (iii) Force field parameter file and (iv) a configuration file which contain a set of commands, options and values for controlling the simulation by NAMD.

The Visual molecular dynamics (VMD)

VMD is a molecular graphic program, used for setting up the simulation and analyzing the trajectory file from simulation especially from NAMD.

The SSH client

The ssh client is a remote login program for logging into the remote machine to execute command via network with secure encrypted.



2.2 Methods

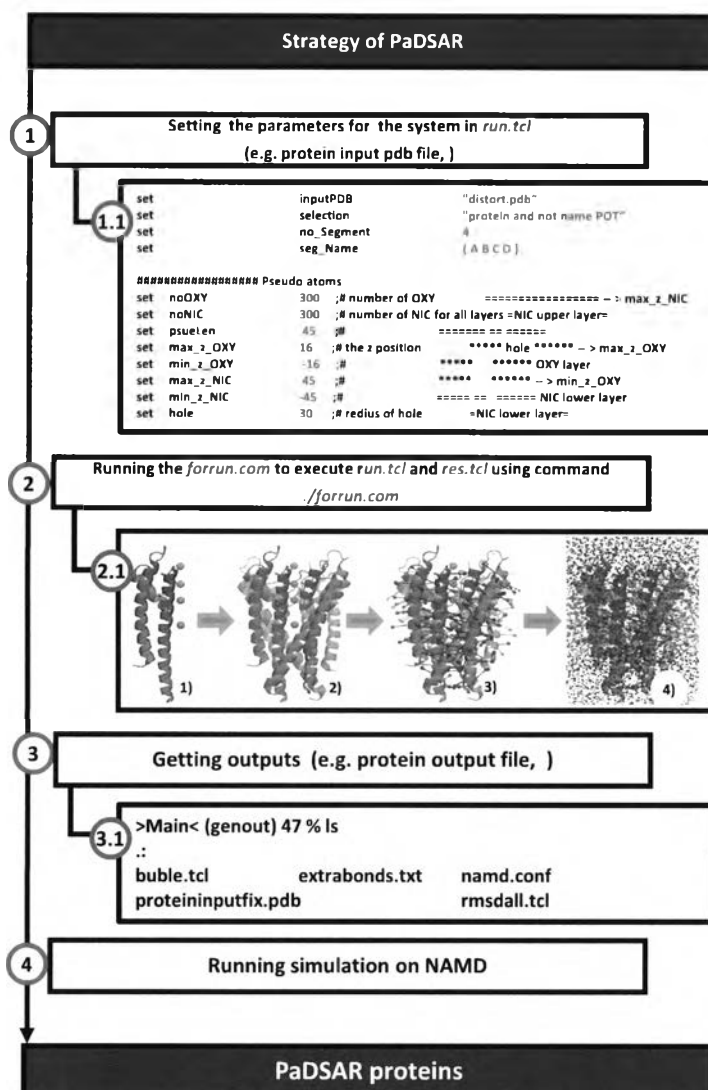


Figure 2.1 Schematic shows the preparation of protein to run PaDSAR on NAMD. 1 Setting initial parameter in the *run.tcl*. 1.1 An example of *run.tcl*. 2 Generating the needed files for running the simulation on NAMD 2.1 Generating the protein coordinate file. 1) Initial coordinates of the protein in monomer. 2) In case of symmetry multi-mers proteins. The monomer of protein will be duplicated and rotated to generate the multi-mers. 3) The EPs types pseudoatoms will be attached to an α -carbon in backbone of protein relate to the SDSL/EPR. 4) Several hundred of OXY and NIC pseudoatoms (PsDAs) are distributed to the system representing the environment background. 3 Getting the output files for running PaDSAR. 3.1 examples of output files. 4 Running the PaDSAR on NAMD and obtaining the PaDSAR protein.



229682773

2.2.1 Writing the Tcl script file for running the simulation in NAMD

To implement PaDSAR on NAMD, it is necessary to build the extra command lines that provide special options of simulation conditions¹⁰ (Figure 2.1). This has been done through the Tcl language¹¹ because the program NAMD and VMD recognize this computer language. A set of Tcl script files for preparing, executing and analyzing the simulation is shown below.

There are three files used for generating needed files for NAMD.

forrun.com* **res.tcl** **run.tcl**

forrun.com is an executable file,

run.tcl is a user defined file and

res.tcl is a Tcl procedure library file

Run forrun.com to excute **res.tcl** and **run.tcl**

./forrun.com

The following files are generated after executing *forrun.com*. These files are subsequently used to run NAMD-PaDSAR.

namd.conf **buble.tcl** **extrabonds.txt** **proteininput.pdb**
proteininput.psf **proteininputfix.pdb** **rmsdall.tcl**

namd.conf is the NAMD configuration file,

buble.tcl is an extra Tcl command file used accompany with the tclBC command in *namd.conf* to restrain the boundary of atoms,

extrabonds.txt is used for keeping the secondary restrain for some residues during the simulation,

proteininput.pdb and *proteininput.psf* are the protein coordinate file and protein structure file respectively,

proteininputfix.pdb was used when need to hold fix atoms and

rmsdall.tcl is the Tcl script file for calculating of RMSD.



229682773

2.2.2 Building the system containing pseudoatoms

For a testing model, the 2Å-resolution X-ray structure of homotetrameric potassium channel from *Streptomyces lividans* (KcsA, pdb code is 1K4C) has been used. Each subunit of KcsA consists of two transmembrane helices TM1 (residue 22 - 51) and TM2 (residue 86 -124) as shown in Figure 2.2.

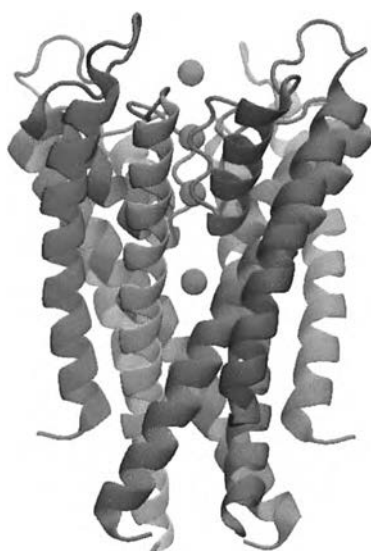


Figure 2.2 Tetramer X-ray structure of a potassium channel from *Streptomyces lividans* (KcsA) TM1 and TM2 helices of a monomer are shown in red and orange helices respectively.

Based on the previous PaDSAR work, the PaDSAR restraints are derived from the solvent accessibility data obtained using electron paramagnetic resonance experiments of a given spin-labeled protein sample. An interpretation of the solvent accessibility values of TM1 and TM2 of the KcsA potassium channel to assign a type of pseudoatoms according to one of the three categories EP1=buried, EP2=water-exposed, EP3=lipid-exposed has been previously achieved in the original work (Figure 2.3 A). The assigned EP-type pseudoatom has been then attached into the PVP-pseudoatom locate at the α -carbon of the residues of KcsA structure (Figure 2.3 B). All assigned EP-type pseudoatoms are shown in the Table 2.1⁹.

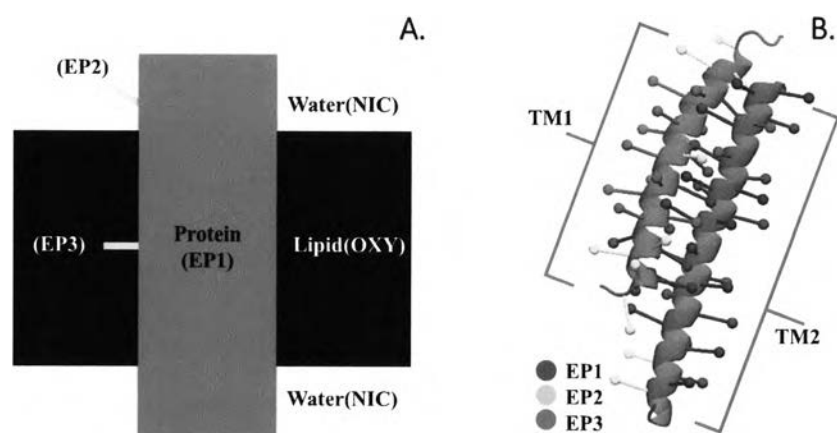


Figure 2.3 A. Scenario of a membrane protein having different surrounding environment B. Attachment of pseudoatom EP-type into transmembrane segments of KcsA channel.

Table 2.1 KcsA residues with the assignments of the pseudoatom types

Pseudoatoms	Property	Residue number
EP1	Buried	25, 26, 29, 30, 33, 36, 40, 44, 47, 48, 86, 87, 89, 91, 92, 95, 96, 97, 98, 99, 100, 101, 102, 103, 105, 106, 107, 108, 109, 110, 111, 113, 114, 115, 118, 119
EP2	Water-	23, 24, 27, 28, 39, 49, 50, 52, 117, 120
EP3	Lipid-	31, 32, 34, 35, 37, 38, 41, 42, 45, 46, 90, 94

300 NIC-pseudoatoms and 300 OXY-pseudoatoms were placed in the system in a way that the OXYs were positioned in a membrane bilayer region whereas the NICs were not in the bilayer (Figure 2.4). The initial configuration of these NIC- and OXY-pseudoatoms has been generated from a molecular dynamics simulation. The interactions between pseudoatoms were specified in the CHARMM parameter¹⁹ in terms of ϵ and R_{\min} which correspond to the minimum energy (kcal/mol) and an equilibrium distance of pair atoms (\AA), respectively. The force field parameter and topology of pseudoatoms; EP1, EP2, EP3, OXY, NIC and PVP were built and defined in accordance with the original PaDSAR version (Table 2.2, Table 2.3 and Figure 2.5). All the interatomic pair potentials were defined according to the 12-6 Lennard-Jones like function (Figure 2.6). For the pair potentials, $\epsilon = -2.00$ kcal/mol and $R_{\min} = 2.00$ \AA are used to compute the interaction energy belonging to EP1-PVP, EP2-NIC and EP3-OXY. This functional form is similar to the Type I function of the original PaDSAR. The

potentials of $\epsilon = -0.05$ kcal/mol and $R_{\min} = 6.00$ Å are used to describe the interaction between mismatch pair of pseudoatoms that includes EP1-NIC, EP1-OXY, EP2-OXY, EP2-PVP, EP3-NIC and EP3-PVP. This function is consistent to the Type III function used in the CHARMM-PaDSAR.

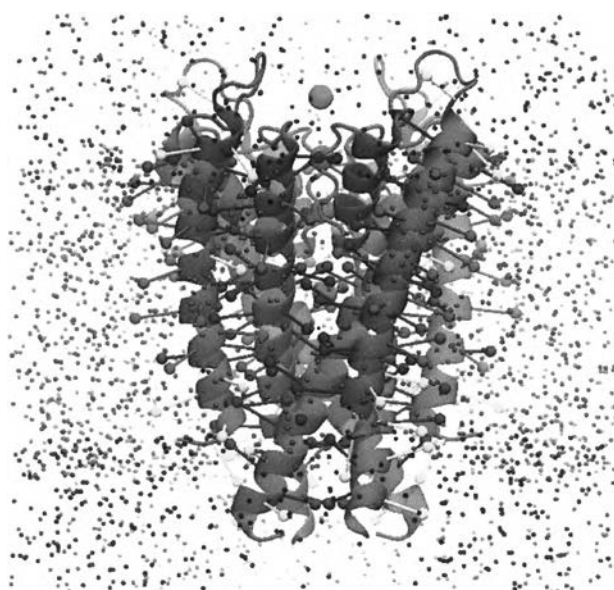


Figure 2.4 NAMD snapshot (a tetramer structure of KcsA channel and the environment pseudoatoms (OXY=red dot, NIC=blue dot)).

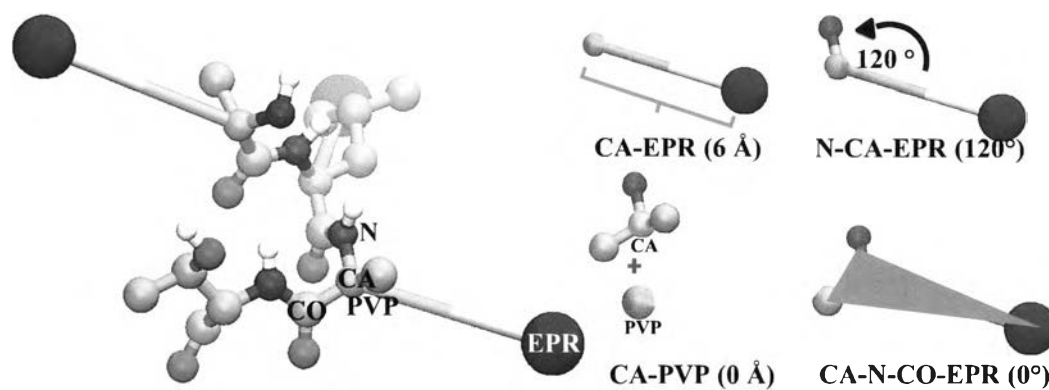


Figure 2.5 Topology of EPR pseudoatoms

Table 2.2 Pseudoatom interactions defined in terms of Lennard-Jones Like potential (adapted from the original CHARMM-PaDSAR⁹.)

EP-Type pseudoatom	Repulsion		
	PVP (protein)	NIC (water)	OXY (lipid)
EP1 (buried)	No	Yes	Yes
EP2 (water)	Yes	No	Yes
EP3 (lipid)	Yes	Yes	No

Table 2.3 Force field parameters for the pseudoatoms

Type		K	$I_0(R_{min})$
Bonds	PVP-EPR*	100.0 (kcal mol ⁻¹ Å ⁻²)	6.0 (Å)
	PVP-CA	100.0 (kcal mol ⁻¹ Å ⁻²)	0.0 (Å)
Angles	N-CA-EPR*	50.0 (kcal mol ⁻¹ rad ⁻²)	120.0 (°)
Improper torsions	CA-N-CO-EPR*	55.0 (kcal mol ⁻¹ rad ⁻²)	0.0 (°)
van der Waals	OXY-OXY, NIC-NIC, OXY-NIC	0.0 (kcal mol ⁻¹)	0.1 (Å)
	EP2-NIC, EP3-OXY	-2.0 (kcal mol ⁻¹)	2.0 (Å)
	EP1-OXY, EP1-NIC, EP2-OXY, EP2-PVP, EP3-NIC, EP3-PVP	-0.05 (kcal mol ⁻¹)	6.0 (Å)

*EPR = EP1, EP2, EP3



2296827773

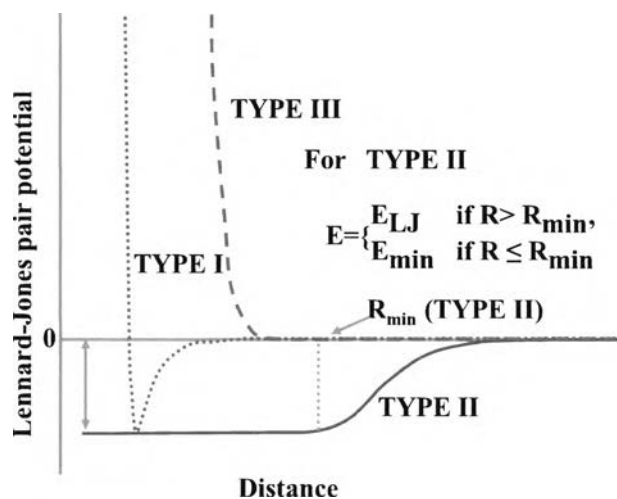


Figure 2.6 Lennard-Jones potential energy represents three types of pseudoatoms interaction.

2.2.3 Molecular dynamics simulation

The molecular (MD) dynamics simulation is a computational method to mimic the dynamics of atoms, molecules systems based on Newton's second law ($\sum F = ma$) that serves the dynamics and thermodynamics properties of system¹⁰.

MD simulations were carried out using the program NAMD. The united-atom CHARMM 19 force field parameter and topology were used for proteins. The tclBC command (Tcl boundary condition) was added in the configuration file to constrain the NIC and OXY pseudoatoms to move within the water and membrane regions, respectively. During simulations the secondary structure of TMs helices were restrained and the pore helix, the selectivity filter, and the potassium ions located in the pore were kept fixed to the native conformation.

MD simulation parameters were shown below in bold. Note that NAMD does not interoperate any command or string after the number sign, #. It was used for input the comment in the Tcl script file.

```

#####          namd configuration file          #####
structure       ./proteininput.psf
coordinates     ./proteininput.pdb
firsttimestep   0

```

To load the protein structure file and a protein coordinate file into NAMD and set the first time step to zero respectively.




```
##### simulation parameters #####
paraTypeCharmm      on
parameters          param19-epr.inp
COMmotion           no
temperature          310
```

paraTypeCharmm on: to use the CHARMM force field parameter, param19-epr.inp, and *COMmotion no*: to keep restrain the center of mass of entire system during the simulation taking place and set the temperature of system to 310 K.

```
##### Force-Field Parameters #####
exclude            scaled1-4
1-4scaling         1.0
cutoff             10
nonbondedScaling  1.0
switching          on
switchdist         9
pairlistdist       12
pairlistsPerCycle  2
margin             0.0
```

This part was used for setting the van de Wals interaction for non-bonded atoms. Generally, van der Waals energy is a function of distance and limit to zero when the distance goes to infinity (Figure 2.7). In NAMD, to reduce the time consuming, the interaction of non-bonded atoms are truncated and specified with cutoff values. If the distance of two atoms is greater than *cutoff* the van der Waals energy will be set to zero. And *switchdist* is the point that NAMD starts using switching function to bring the van der Waals energy to zero at the cutoff point smoothly. The other most important command for non-bonded term is *pairlistdist* which is the longest distant that NAMD still calculate the energy of two non-bonded atoms.

အတိုင်း... ၂၅၅၆
အတိုင်း... ၇၁၂၄
အတိုင်း... ၁၁၆ ချိတ်, ၂၅၅၀

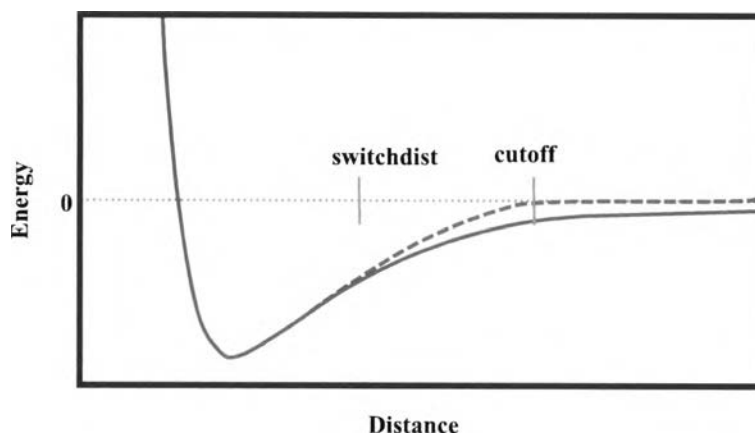


Figure 2.7 The graph represents the van der Waals energy with (red dot) and without (green line) using switching function.

exclude scaled1-4: Non-bonded interactions of types in Figure 2.8 will be excluded. *1-4scaling 1.0*: to set the scaling for *scaled1-4* to 1.0. *cutoff 10*: to set the cutoff distance to 10 Å. *nonbondedScaling 1.0*: to set the van der Waals interaction scaling to 1.0 or no modification. *Switching on*: to turn the switching function on. *Switchdist 9*: to set the switching distance to 9 Å. *pairlistdist 12*: to set the pair list distance to 12 Å. *pairlistsPerCycle 2*: to update the pair list distance every 2 step of circle. *margin 0.0*: The default value of margin of an identical simulation cube space.

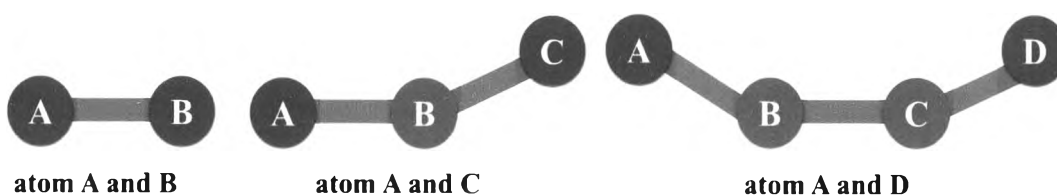


Figure 2.8 The kinds of non-bonded atoms to be excluded when using *exclude scaled1-4* command.

```
#####      tclBC      #####
tclBC on
tclBCScript {
  set bubbleCenter "0.0 0.0 0.0 "
  set tclBCScript   buble.tcl
  source $tclBCScript
}
tclBCArgs {0. 15. .01 5.}
```

To add the tclBC command into NAMD, this command was used to restrain the boundary of group of atoms. *tclBC on*: to turn the tclBC command on. set *bubbleCenter "0.0 0.0 0.0 "*: The center of boundary in Cartesian coordinate (x , y, z) was set to 0.0 0.0 0.0 and set tclBCScript *buble.tcl*: to call the external Tcl script named *buble.tcl* this file contains condition loops to distinguish atoms.

```
#####      fixed atom      #####
fixedAtoms      on
fixedAtomsForces off
fixedAtomsFile   ./proteininputfix.pdb
fixedAtomsCol    B
```

This set of commands was used for holding fix atoms during the simulation. *fixedAtoms on*: to turn the fixatoms command on. *fixedAtomsForces off*: The calculation of total energy will not count the fixed atoms due to the fixedAtomsForces command is off and the proteininputfix.pdb file is almost similar with the protein coordinate file but the beta column in this file was set to 1 for fixed atoms and 0 for free atoms. *fixedAtomsCol B*: Column in proteininputfix.pdb that contain the fix atoms data, in this case B is referred to beta column.

```
#####      2nd structure restrain      #####
extraBonds      on
extraBondsFile   extrabonds.txt
```

To Turn the secondary structure restrain on and call the external text file that contains the secondary structure information for a given protein named *extrabonds.txt*.



#####	Integrator Parameters	#####
timestep	2.0	
rigidBonds	all	
nonbondedFreq	1	
useSettle	on	
fullElectFrequency	2	
stepspercycle	20	

timestep 2.0: Time interval for each step in femtoseconds. *rigidBonds all*: To constrain water as rigid molecules. *nonbondedFreq 1*: to specify how often the nonbonded energy to be calculate. *useSettle on*: to use SETTLE algorithm to rigid water molecules. *fullElectFrequency 2*: to specify the number of each timesteps for evaluation full electrostatics. *stepspercycle 20*: Number of timesteps in each cycle.

#####	Boundary Conditions	#####
#####	Periodic Boundary Conditions	#####
cellBasisVector1	100 0 0	
cellBasisVector2	0 100 0	
cellBasisVector3	0 0 100	
cellOrigin	0.00 0.00 3.87	
XSTfile	protein	
XSTfreq	2000	
wrapAll	on	

To use the periodic boundary conditions, the demonstration of cellBasisVector1, 2 and 3 are the boundary vector in x, y and z direction and cellOrigin is the origin of unit cell. *XSTfile protein*: to set name of (eXtended System Trajectory) XST file this file contain periodic cell boundary and extended variable data during the simulation. *XSTfreq 2000*: to generate the XST data every 2000 steps. *wrapAll on*: If part of atoms or molecules move across the boundary of cells the coordinate will not be translated to the other side of cell.



```
##### EXECUTION SCRIPT Output #####
outputName      protein
binaryrestart   yes
restartfreq     1000
DCDfreq        500
outputEnergies  500
outputPressure  500
outputtiming    500
```

outputName protein: to set name of output files. *binaryrestart yes*: to turn the binary restart on. *restartfreq 1000*: to generate the restart file every 1,000 steps. *DCDfreq 500*: to write position coordinate to the trajectory file every 500 steps. *outputEnergies 500*: to write out put energies every 500 steps. *outputPressure 500*: to write an output pressure ever 500 steps. *outputtiming 500*: to generate CPU and wallclock times and memory usage data every 500 steps.

```
##### Minimization #####
minimize      2000
reinitvels    310
run           500000
```

minimize for 2,000 steps before run the MD simulation 500,000 steps with randomly reinitializes the initial velocities of all atoms at given temperature.

2.2.4 Analysis of MD trajectory

The total energy and root-mean square deviation (RMSD) of α -carbon of TM1 and TM2 were obtained from the MD trajectories using Tcl script files running on the VMD.

