

**EMBO Practical Course:**  
**Short MD Simulation of the Villan Headpiece Protein**

**Pre-processing of the PDB File**

**First protonate the protein using the ‘protonate’ subroutine to get the proton names into the correct AMBER nomenclature:**

```
[phineus@nidhogg EMBOCOURSE]$ protonate -i 1QQV.pdb > HPH.pdb
```

**Output:**

Here are the mystery protons from input file:

```
H PRO 10 0.71 -6.53 20.78
H3 PRO 10 2.23 -6.03 20.18
HD1 HIE 41 -0.37 -3.70 8.01
```

**Now edit the protonated PDB file: Remove the first 40 residues, to leave just the headpiece:**

```
[phineus@nidhogg EMBOCOURSE]$ nedit HPH.pdb
```

**Now protonate again to add the protons to the N-terminal residue (NH3+ moiety):**

```
[phineus@nidhogg EMBOCOURSE]$ protonate -i HPH.pdb > HPH2.pdb
```

**Use TLEAP to Generate the Coordinate and Topology Files from the PDB File**

```
[phineus@nidhogg EMBOCOURSE]$ tleap
-I: Adding /Bis/shared/centos-3_x86_64/amber9_intel8.1/dat/leap/prep to search path.
-I: Adding /Bis/shared/centos-3_x86_64/amber9_intel8.1/dat/leap/lib to search path.
-I: Adding /Bis/shared/centos-3_x86_64/amber9_intel8.1/dat/leap/parm to search path.
-I: Adding /Bis/shared/centos-3_x86_64/amber9_intel8.1/dat/leap/cmd to search path.
```

```
Welcome to LEaP!
(no leaprc in search path)
```

**Define the desired solute force-field (we are using the ff99 force-field with the Stoney-Brook (SB) correction):**

```
> source leaprc.ff99SB
----- Source: /Bis/shared/centos-3_x86_64/amber9_intel8.1/dat/leap/cmd/leaprc.ff99SB
----- Source of /Bis/shared/centos-3_x86_64/amber9_intel8.1/dat/leap/cmd/leaprc.ff99SB done
Log file: ./leap.log
Loading parameters: /Bis/shared/centos-3_x86_64/amber9_intel8.1/dat/leap/parm/parm99.dat
Reading title:
PARM99 for DNA, RNA, AA, organic molecules, TIP3P wat. Polariz. & LP incl. 02/04/99
Loading parameters: /Bis/shared/centos-3_x86_64/amber9_intel8.1/dat/leap/parm/frcmod.ff99SB
Reading force field modification type file (frcmod)
Reading title:
Modification/update of parm99.dat (Hornak & Simmerling)
Loading library: /Bis/shared/centos-3_x86_64/amber9_intel8.1/dat/leap/lib/all_nucleic94.lib
Loading library: /Bis/shared/centos-3_x86_64/amber9_intel8.1/dat/leap/lib/all_amino94.lib
Loading library: /Bis/shared/centos-3_x86_64/amber9_intel8.1/dat/leap/lib/all_aminoc94.lib
```

Loading library: /Bis/shared/centos-3\_x86\_64/amber9\_intel8.1/dat/leap/lib/all\_aminont94.lib  
Loading library: /Bis/shared/centos-3\_x86\_64/amber9\_intel8.1/dat/leap/lib/ions94.lib  
Loading library: /Bis/shared/centos-3\_x86\_64/amber9\_intel8.1/dat/leap/lib/solvents.lib

### **Now we load in the PDB File:**

```
> HP = loadpdb HPH2.pdb
Loading PDB file: ./ECtest2.pdb
total atoms in file: 579
```

### **We now solvate the system with TIP3P water molecules. We define a solvent box size with 10 angstroms between the protein and the edge of the box:**

```
> solvatebox HP TIP3PBOX 10.0
Solute vdw bounding box:      27.421 34.085 19.578
Total bounding box for atom centers: 47.421 54.085 39.578
Solvent unit box:           18.774 18.774 18.774
Total vdw box size:         50.755 57.504 42.851 angstroms.
Volume: 125065.858 A^3
Total mass 54955.946 amu, Density 0.730 g/cc
Added 2825 residues.
```

### **Next we ask tleap what the total charge of the system is:**

```
> charge HP
Total unperturbed charge: 2.000000
Total perturbed charge: 2.000000
```

### **The total charge is +2, so we have to add two counter-ions (Cl-). We ask tleap to place these in a reasonable position using the subroutine 'addions2':**

```
> addions2 HP Cl- 0
2 Cl- ions required to neutralize.
Adding 2 counter ions to "HP" using 1A grid
Grid extends from solute vdw + 2.47 to 8.47
Resolution: 1.00 Angstrom.
grid build: 0 sec
Calculating grid charges
charges: 47 sec
Placed Cl- in HP at (24.86, -21.45, -8.13).
Placed Cl- in HP at (0.86, 11.55, -21.13).
```

Done adding ions.

### **Next we have to check that the system (the unit called HP) is OK using the command 'check':**

```
> check HP
Checking 'HP'...
Checking parameters for unit 'HP'.
Checking for bond parameters.
Checking for angle parameters.
Unit is OK.
```

**The unit is OK, so now we can create the topology and coordinate file: the topology file contains all the information concerned with the force-field. The topology file is called HP.parm, the coordinate file is called HP.crd:**

```
> saveamberparm HP HP.parm HP.crd
Checking Unit.
Building topology.
Building atom parameters.
Building bond parameters.
Building angle parameters.
Building proper torsion parameters.
Building improper torsion parameters.
total 116 improper torsions applied
Building H-Bond parameters.
Not Marking per-residue atom chain types.
Marking per-residue atom chain types.
(Residues lacking connect0/connect1 -
these don't have chain types marked:

    res    total affected

    CPHE    1
    NLEU    1
    WAT    2825
)
(no restraints)
```

**The program builds the topology file. It tells us that the terminal residues and the water molecules (2825 of them) are not connected on both sides. Finally we quit tleap using the command 'quit':**

```
> quit
Quit
```

**We can generate a pdb file from the coordinate and topology files using 'ambpdb':**

```
ambpdb -p HP.parm < HP.crd > HP_leap.pdb
```

**HP\_leap.pdb contains all the water molecules. After removing the water molecules, we can compare the pdb file HP\_leap.pdb with the starting structure (HPH2.pdb) using MolMol. They are (of course) exactly the same.**

### **Energy Minimization of Structure.**

**Before doing any MD simulation, we need to minimize the structure using the defined force-field. This is a multi-step process. First we hold the solute (protein) fixed using a Cartesian restraint, and allow the water molecules to relax. The minimization input file to do this is called 'min1.inp':**

```
min1.inp

Min (1) of HP in solution      !!!! Title
&cntrl                        !!!! Start reading in the input file information
imin=1, maxcyc=200,           !!!! Do minimization (imin=1) for 200 steps
ntpr=5,                       !!!! Print out information every 5 steps
cut=8.0,                      !!!! Direct sum cutoff of 8 angstrom for PME electrostatic interaction
```

```

ntr=1,          !!!! We are using Cartesian restraints
&end           !!!! End of input
Group in put for restraint
10.0           !!!! Force-constant for Cartesian restraint
RES 1 35       !!!! Apply restraint across all atoms in residues 1 to 35
END
END

```

**We can now run the first minimization routine.**

```
sander -i min1.inp -c HP.crd -p HP.parm -o HPmin1.out -ref HP.crd -r HPmin1.rst
```

**Having minimized the solvent, we can now minimize the entire system using the input file min2.inp:**

```

min2.inp

Min (2) of HP in solution
&cntrl
imin=1, maxcyc=200,
ntr=5,cut=8.0,
&end
END
END

```

```
sander -i min2.inp -c HPmin1.rst -p HP.parm -o HPmin2.out -r HPmin2.rst
```

### Equilibration of the System

**We now have to bring the system to equilibrium. This is a two step process: First, we rapidly bring the system to the correct temperature (300K), using a Berendsen thermostat, and then we introduce a pressure-stat in order to bring the density of the the system to 1 g/cc (notice that on adding the solvent the density was only 0.73 g/cc).**

```

equil.inp

Heating up HP equilibration stage 1
&cntrl
nstlim=10000, dt=0.002, ntx=1, irst=0,
tempi =100.0, temp0=300.0, ntt=1, tautp=1.0,
ntb=1, ntp=0, cut=8.0, ig=71277,
ntc=2, ntf=2, ntwprt=579,
ntr=500, ntwr=500, ntwx=500,
&end

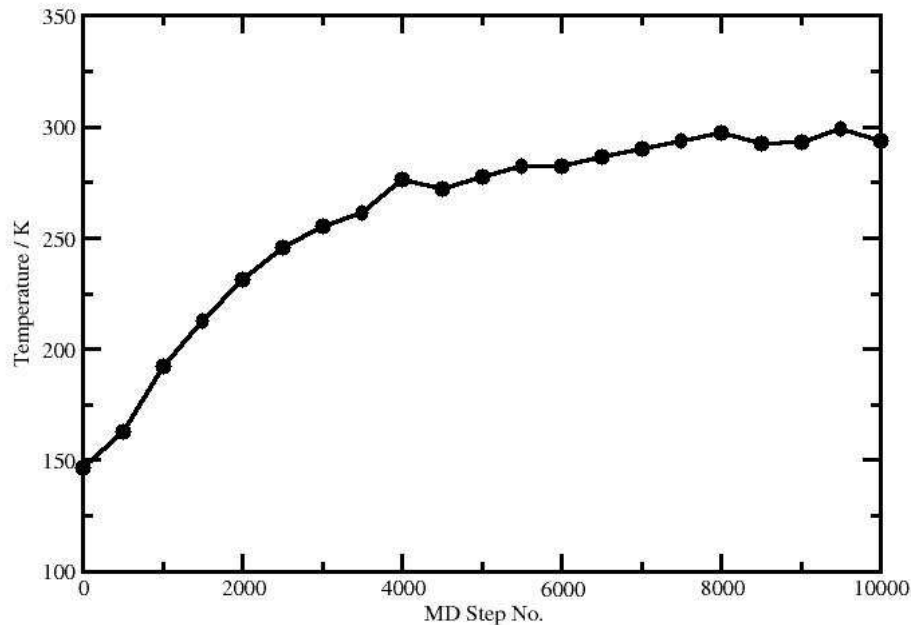
```

```
sander -i equil.inp -c HPmin2.rst -p HP.parm -o HPequil.out -x HPequil.crd -r HPequil.rst
```

**We run a short MD of 10,000 steps with a step-size of 0.002 ps. The command (ntx=1) tells the program that it should only read in the coordinates (as no velocities are as yet available). The command irst=0 tells the program that we are not restarting a simulation. The initial temperature is to be 100K, the desired temperature 300K. The atoms are randomly assigned an initial velocity equivalent to a temperature of 100K. The Maxwellian distribution of velocities is controlled by the random seed generator (ig= 71277). The Berendsen thermostat is defined using ntt=1, and the coupling constant for the thermostat is 1.0 ps (tautp=1.0). We are working under constant volume**

(periodic boundary) conditions ( $ntb=1$ ), and there is no pressure-state ( $ntp=0$ ). The direct space sum cut-off for the PME electrostatic interactions is 8.0 angstrom. We are running the SHAKE algorithm to constrain the length of all bonds involving H atoms ( $ntc=2$ ), and so we don't need to calculate the bond terms for these bonds ( $ntf=2$ ). If we were not running SHAKE, we would have to define a smaller time-step ( $dt$ ) of 0.001 ps. We print out information every 500 steps ( $ntpr$ ,  $ntwr$ ,  $ntwx$ ), but we only require coordinates for the solute (atoms 1-579) ( $ntwprt=579$ ).

Here is the change in temperature variation during the short equilibration:



Having brought the system to 300K under constant volume conditions, we now have to get the system to the correct density using a pressure-stat.

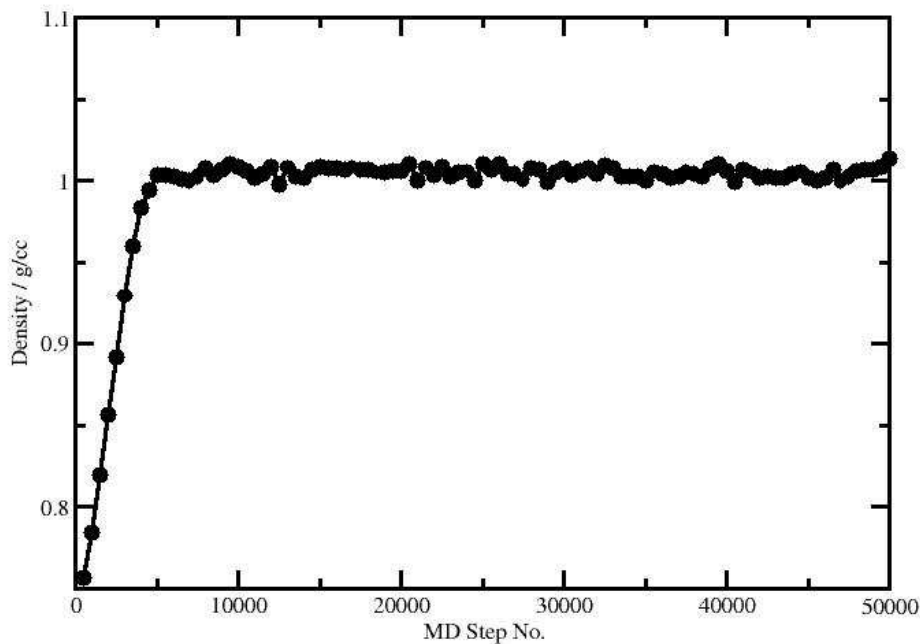
equi2.inp

```
Equilibration HP stage 2
&cntrl
  nstlim=50000, dt=0.002, ntx=5, irest=1,
  temp0=300.0, ntt=1, tautp=2.0, taup=0.2,
  ntb=2, ntp=1, cut=8.0,
  ntc=2, ntf=2, ntwprt=579,
  ntp=500, ntwr=500, ntwx=500,
&end
```

```
sander -i equi2.inp -c HPequi1.rst -p HP.parm -o HPequi2.out -x HPequi2.crd -r HPequi2.rst
```

This input file is similar to `equi1.inp`, but notice that we are now restarting a simulation ( $irest=1$ ), and reading in both coordinate and velocity data ( $ntx=5$ ). The simulation is now to be performed under constant pressure (periodic boundary) conditions ( $ntb=2$ ), and the pressure-stat is on ( $ntp=1$ ) with a tight pressure coupling constant of 0.2 ps. The thermostat coupling constant has been reduced to 2.0 ps.

The density of the system varies over the MD simulation like this:



### Production Run Molecular Dynamics Simulation

The system is now at the correct temperature and density. The average pressure is approximately 1 bar. Before performing a production run MD, one would like to let the system properly equilibrate with weak thermo- and pressure-stats for at least 1-2 nanoseconds. As we don't have time to do this now, an equilibrated system restart file has been provided. The MD input file looks like this:

mdl.inp

```
MD production run for HP
&cntrl
  nstlim=500000, dt=0.002, ntx=5, irest=1,
  temp0=300.0, ntt=1, tautp=2.0, taup=2.0,
  ntb=2, ntp=1, cut=8.0,
  ntc=2, ntf=2, ntwprt=579,
  ntp=500, ntwr=500, ntwx=500,
&end
```

```
sander -i mdl.inp -c HPequi2.rst -p HP.parm -o HPmdl.out -x HPmdl.crd -r HPmdl.rst
```

We want to run a 1 ns MD simulation with a time-step of 0.002 ps using weak thermo- and pressure-stats (notice taup and tautp are now set to 2 ps). This simulation will take some time. Later, we can analyze the trajectory and extract some useful information about the protein dynamics which we can compare to experimental data.