Replica exchange MD of Alanine Dipeptide with Implicit Solvent

In this exercise you will perform a **REMD simulation with 8 replicas between 300 and 500 K, 10 ns MD per replica of alanine dipeptide in implicit solvent** using Gromacs with the Amber99sb-ILDN force field and the GBSA solvent model.

General Notes:
Information about REMD with GROMACS can be found at
http://www.gromacs.org/Documentation/How-tos/REMD

Preparation of the starting structure
The necessary steps for preparing the structure for a REMD simulation depend on whether or not an explicit solvent will be used.

- In the present case an implicit solvent (GBSA) is intended to be used and so the preparatory steps are limited to generating the Gromacs topology (*gmx pdb2gmx*) and energy minimization (*gmx grompp + gmx mdrun*).
- Using explicit solvent would require the setting up of a box (*gmx editconf*), adjustment of salt concentration and charge neutralization (*gmx grompp + gmx genion*, depending on the particular system), energy minimization (*gmx grompp + gmx mdrun*), and an equilibration step (*gmx grompp + gmx mdrun*) at the desired temperatures.

a) Topology generation:
gmx pdb2gmx -f ala2.pdb -ter -ignh
The default output files *conf.gro*, *topol.top* and *posre.itp* will be generated.

b) Energy minimization:
gmx grompp -f min-implicit.mdp -c conf.gro -p topol.top -o min.tpr
gmx mdrun -deffnm min -v >& min.out &
Hint: Use the same mdp file as in the implicit solvent exercise!

c) REMD:
i. Depending on the number of processors available and the range of temperature to sample (they are actually extremely dependent on each other), choose a distribution of the temperatures. Use an exponential distribution:

$$T_i = T_0 * e^{ki}$$

where k and T_0 can be tuned to obtain reasonable temperature intervals to allow exchanges. The exponential allows an increase of temperature interval as the temperature increases. This is necessary because the distribution of the total energy increases with the temperature and thus the exchange rate increases. Keep the exchange rate constant across the temperature range.
Use http://folding.bmc.uu.se/remd/ to choose the temperature of the replicas depending on T_{min}, T_{max} and the number of replicas, N.

ii. In case of explicit solvent: Once you have the temperature distribution, equilibrate the systems at the N temperatures separately using a separate *.mdp* file for each.

iii. Construct a series of N input files (*.tpr*) from the minimized (in case of explicit solvent: equilibrated) structures, using different *.mdp* files to generate the different *.tpr* files.

Example:
For instance, we want to use **8 replicas** with temperatures at 300.00, 322.60, 347.34, 374.38, 403.97, 436.28, 471.67, 510.41.
8 .mdp files named remd_0.mdp, ... remd_7.mdp were prepared. All the parameters contained in the 8 .mdp files are the same with the exception of:

ref_t = 300.0 ; for replica 1, i.e. remd_0.mdp
gen_temp = 300.0 ; for replica 1, i.e. remd_0.mdp
.
.
.
ref_t = 510 ; for replica 8, i.e. remd_7.mdp
gen_temp = 510 ; for replica 8, i.e. remd_7.mdp

A simple loop to generate 8 .tpr files can be used:
for i in {0..7}; do gmx grompp -f remd_${i}.mdp -c min.gro -p topol.top -o remd_${i}.tpr; done

With all the 8 .tpr files present in the working directory, the REMD job can be submitted:
mpirun -np 8 gmx_mpi mdrun -s remd_.tpr -multi 8 -replex 1000 -deffnm remd_ -v >& md.out &

np = the number of processors used
md_.tpr = Note this is not a mistake, '_' followed by a '.' is meant.
multi = instructs the program to perform 8 multi-runs
replex = instructs the system to attempt an exchange every 1000 steps, i.e., every 2 ps as the time step is 2 fs.
reseed = random seed for REMD, using -1 tells the system to generate a random seed for you.

If you open the .log files, you will see the replica exchange statistics such as the exchange probability, and the replicas involved in the exchange per 100 steps.

However, you will run the simulation as batch job (also available on the website!):
qsub qsub.job

With *qstat* you can the state of job in the queue.
Gromacs writes trajectories per temperature (***temperature replicas***), which are not continuous with respect to coordinates (as the coordinates get exchanged).

Exercise:
All files, which you need, are provided on our website: http://www.strodel.info/teaching.php
For the following exercise, we will use our **computer cluster JUBIO**. To copy something to **JUBIO** (for instance ala2.pdb), use the command on your computer (**not** in the **JUBIO** shell):
scp -r ala2.pdb biosimXY@134.94.118.40*:/home/biosimXY*

To copy the files from **JUBIO** to your computer, type the following command on your computer (**not** in the **JUBIO** shell; It will copy the files in your current directory!):
scp -r biosimXY@134.94.118.40*:/home/biosimXY/ala2.pdb .*

Running the REMD simulation:

1) Perform a REMD simulation of alanine dipeptide using **Amber99SB-ILDN** as force field, **GBSA** as implicit solvent model, **eight 10 ns replicas between 300 and 500 K**. The *.mdp file* for the lowest temperature (*remd_0.mdp*) is provided via our website: http://www.strodel.info/teaching.php
Report the temperatures of the 8 replicas. Produce the 8 other input files *remd_1.mdp,..., remd_7.mdp*.

Analysis of the REMD simulation:

2) Report the average acceptance for exchanges between replicas. Are these values acceptable?
3) Produce and analyze a plot showing the temperature for each replica.
Here, you can also use a for loop in bash to generate the plots:
for i in {0..7}; do echo 14 | gmx energy -f remd_${i}.edr -s remd_${i}.tpr -o temp_${i}.xvg ; done &> temp.out &
echo 14 |: A pipeline will be used and directly deliver your choices (14 for Temperature) to Gromacs.
&> temp.out &: All output, which will print in the shell, will be written in the temp.out

For plotting the data using xmgrace: *xmgrace Epot_*.xvg*
To get a better curve, use the transformation tool in xmgrace: *Data > Transformations > running averages*
Select all sets and 100 for length of average to get a proper plot.

4) Produce and analyze a plot showing the potential energy for each replica.
You can also use a for loop in bash to generate the plots (Use 10 for potential energy instead of 14; Change the output name temp to Epot!).

5) Extract the Ramachandran angles Φ and Ψ from each replica.
for i in {0..7}; do echo 1 | gmx rama -f remd_${i}.trr -s remd_${i}.tpr -o rama_${i}.xvg ; done &> rama.out &

6) Produce the files *energies.data*, *phi.dat* and *psi.dat*, containing the corresponding values for each replica:
for i in {0..7}; do awk '{if ($1 !~ /[#@]/) print $2 }' Epot_${i}.xvg > Epot_${i}.dat ; done
paste Epot_0.dat Epot_1.dat Epot_2.dat Epot_3.dat Epot_4.dat Epot_5.dat Epot_6.dat Epot_7.dat > energies.data

for i in {0..7}; do awk '{if ($1 !~ /[#@]/) print $1 }' rama_${i}.xvg > phi_${i}.dat ; done
paste phi_0.dat phi_1.dat phi_2.dat phi_3.dat phi_4.dat phi_5.dat phi_6.dat phi_7.dat > phi.dat

for i in {0..7}; do awk '{if ($1 !~ /[#@]/) print $2 }' rama_${i}.xvg > psi_${i}.dat ; done
paste psi_0.dat psi_1.dat psi_2.dat psi_3.dat psi_4.dat psi_5.dat psi_6.dat psi_7.dat > psi.dat

7) Produce the free energy surface (FES) along Φ and Ψ for T = 300, 400 und 500 K.

- Can you see differences between them? Describe in **2-3 sentences**.
- Compare and describe these FESs with those obtained from the single-temperature MD runs at 300 K using explicit solvent (TIP3P) and implicit solvent (GBSA) in **4-8 sentences**.

For the production of the FES, you need the *WHAM program* and *histo.para* file, which can both be obtained from http://www.strodel.info/teaching.php.
Make sure that the *WHAM program* is executable: *chmod 744 wham*

Adjust the settings in *histo.para* (Using temperatures for each replica and number of data points).
And then execute the *WHAM program*: *./wham phi.dat psi.dat <temperature>*
where <temperature> has to be replaced by the temperature value at which the FES should be generated, e.g., 300 K. The *WHAM program* produces several files of which *F_R1_R2.dat* contains the free energies along Φ and Ψ, which can be plotted using *gnuplot*.

Hint: Before you generate the next FES.dat with wham, rename the *F_R1_R2.dat to FES_temperature.dat*, otherwise your previous file will be overwritten! Please, copy all data to your computer and then plot the FES via *gnuplot*!