Tutorial 2
Replica Exchange Solute Scaling
(REST2)

Anji Babu Kapakayala
Department of Chemistry
Indian Institute of Technology Kanpur, India.

March 2019

This tutorial is designed to help the reader to understand the underlying theory of Hamiltonian Replica Exchange Methods and to learn how to perform REST2 simulation for various λ values, where λ is a scaling factor between 1 and 0. Firstly, the theory behind the Replica Exchange Solute Scaling(REST2) simulations will be discussed briefly. Then we will learn how to setup and perform REST2 simulation using gromacs patched with plumed for simple system Alanine di-peptide in explicit solvent (Water) with λ values 1.0, 0.7, 0.5, 0.4, 0.3. (i.e 5 Replica). Where $\lambda = 1.0$ is corresponds to the unscaled replica. The later part of this tutorial teaches you basic analysis of the outputs and construction of free energy surface.

Attention: This tutorial assumes that the reader have a clear understanding on how to build the initial structure, topology and perform the basic runs of energy minimization and equilibration using gromacs.

1 Theory

Conformational sampling to simulate protein folding, drug binding processes using molecular dynamics is hampered by the slow barrier crossing conformational transitions. To overcome this, several enhanced sampling methods have been devised. Among these, global tempering approaches enhances the sampling of all the degrees of freedom of the system. Parallel tempering replica exchange molecular dynamics is a widely used global tempering method, where several copies of the system are simulated at different temperatures simultaneously and independently, while exchange of coordinates between two adjacent replicas is attempted after certain time intervals based on metropolis criteria. In this

manner, high temperature replicas deliver new conformations to the lower temperature replicas which are otherwise very rarely sampled at low-temperatures. Unfortunately, the efficiency of this approach decreases with system size, as the energy fluctuations are proportional to the square root of the number of degrees of freedom of the system. Which should require to use a large set of temperatures (Replicas) to simulate larger systems like protein in explicit solvent. Therefore the number of replicas will exponentially grows with system size in REMD.

To alleviate this limitation, **Prof. B. J. Borne** has proposed a method named "**Replica Exchange Solute Scaling: REST2**" in 2011. Unlike t-REMD, in REST2 the temperature of the system remains same for all the replicas instead the potential energies of selected atoms in each replica will be scaled by some parameter λ. As we know that the canonical ensemble probability is proportional to the boltzmann factor which is dependent on the potential energy and temperature of the system.

$$P(R) = \exp \frac{U(R)}{k_B T} \tag{1}$$

The probability at half the potential energy is equal to the probability at double of the temperature. Which clearly indicates that the scaling of potential energy terms is indeed enhancing the sampling by increasing the effective temperature of the system. The advantage of scaling potential energy over the temperature is related to the fact that the energy is extensive quantity where as the temperature is intensive. Therefore, one can selectively choose the portion of a system to be scaled. The REST2 method devides the system into two regions Hot(H) and Cold(C), where the scaled part comes under HOT region and the unscaled part comes under Cold region.

However, in the classical simulations the potential energy of the system is constructed from defined force field parameters. Therefore, the force field parameters which contributes to the potential energy barriers (i.e charges, proper dihedral angles, electrostatics, Lennard-Jones) of hot region are scaled by some factor λ.

To be specific, the modified potential energy of the replica m is

$$E_m^{REST2}(R) = \frac{\beta_m}{\beta_0} E_{pp}(R) + \sqrt{\frac{\beta_m}{\beta_0}} E_{pw}(R) + E_{ww}(R) \tag{2}$$

For $\frac{\beta_m}{\beta_0} = \lambda$, We can rewrite the above equation as,

$$E_m^{REST2}(R) = \lambda E_{pp}(R) + \sqrt{\lambda} E_{pw}(R) + E_{ww}(R) \tag{3}$$

Where λ is any real number between 1 and 0.

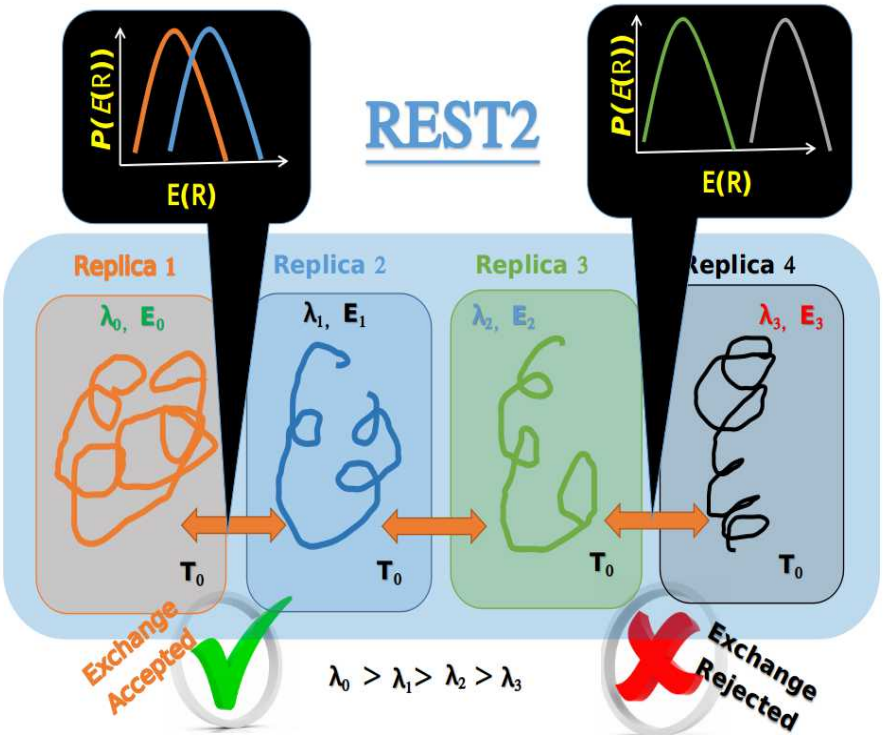

Fig:1. Schematic of Replica Exchange Solute Scaling(**REST2**) Simulation.

In the REST2 simulation(Fig.1), several replicas will run simultaneously and independently with the modified potential corresponds to the λ values. After certain period of time interval the adjacent replicas will attempted to exchange the coordinates based on the metropolis exchange criteria with following acceptance probability:

$$p(i \rightarrow j) = min(1, e^{-\Delta}) \tag{4}$$

Where,

$$\Delta = (\beta_i - \beta_j) \left[(E_{pp}(R_j) - E_{pp}(R_i) + \frac{\sqrt{\beta_0}}{\sqrt{\beta_i} + \sqrt{\beta_j}} (E_{pw}(R_j) - E_{pw}(R_i) \right] \tag{5}$$

We can clearly observe that the solvent-solvent interaction terms were cancelled out in the exchange probability expression. It boosts the acceptance ratio between the neighbour replica during exchange of coordinates.

For more details of the method, it is highly recommended to go through REST2 journal.[1]

(Lingle Wang, Richard. A Friesner, and B. J. Borne, *J. Phys. Chem. B*, **2011**, 115, 30, 9431-9438.)

2 Setting up REST2 Simulation

Throughout the tutorial, We will be using the gromacs-5.1.2 and plumed-2.2.4 versions with the recent implementation of REST2 by **Prof. G. Bussi** in 2014.[2] Make sure that plumed package consists -hrex.

2.1 Initial Structure

We will perform REST2 simulation on Alanine di-peptide in explicit water(Fig.2) with 5 replicas (λ values 1.0, 0.7, 0.5, 0.4, 0.3) using AMBER14SB force fields. All the peptide atoms were chosen as HOT region throughout the tutorial. Exchanges were attempted at every 100 steps. ϕ and ψ angles were chosen as collective variables(**CV**) for constructing free energy.

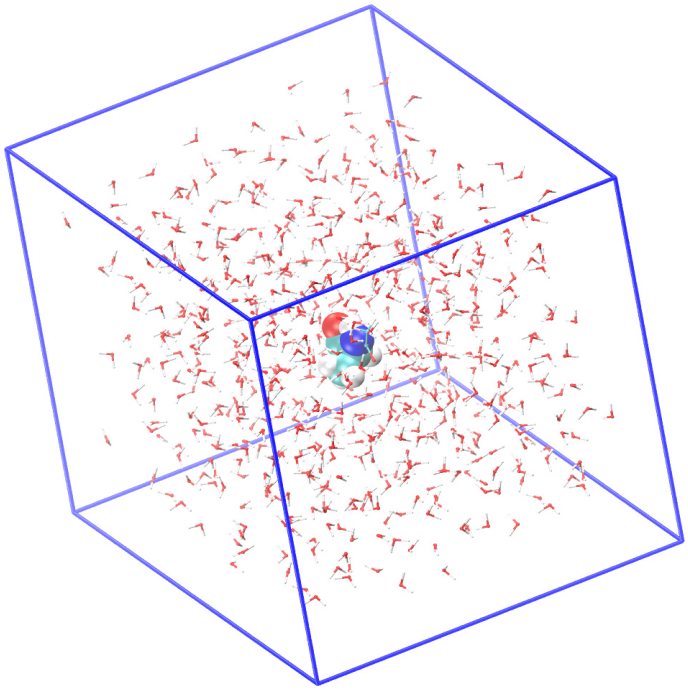

Fig:2. Alanine Di-peptide in explicit solvent (Water)

2.2 Energy minimization & Equilibration

It is assumed that the reader have a clear understanding on how to build the initial structures (alanine.gro, alanine.top) and perform basic runs of energy minimization and equilibration using gromacs.

If not, no worries. I would recommend you to go through the famous lysozome

gromacs tutorial[3](Ref 3) to know about how to build and perform equilibration using gromacs.

2.3 Required Inputs

The following files are required to run the simulation

- Well equilibrated Ala_di-pep.gro
- Ala_di-pep.top
- Gromacs input file(md.mdp)
- Plumed input file(plumed.dat)

Plumed Input: plumed.dat

```
# set up two variables for Phi and Psi dihedral angles
phi: TORSION ATOMS=5,7,9,15
psi: TORSION ATOMS=7,9,15,17

# monitor the two variables
PRINT STRIDE=10 ARG=phi,psi FILE=COLVAR
```

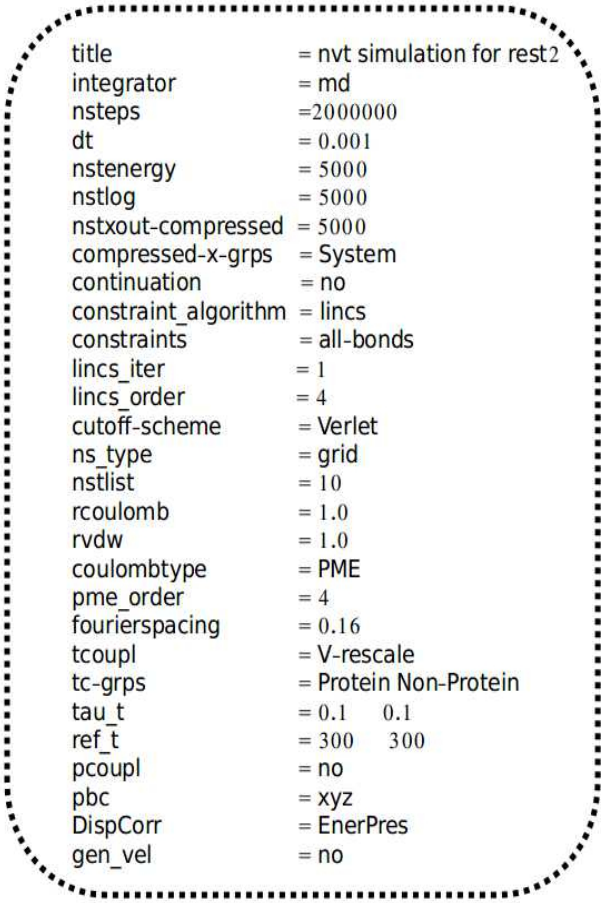

Fig:5. **md.mdp**: Gromacs input file

2.4 Selecting HOT atoms

As we have discussed, REST2 runs on modified potential terms of selectively chosen HOT atoms, for which we need to edit the topology file in order to select the molecules or atoms that you wish to effect (typically only solute) by appending underscore (_) to the each atom of atom type in the [atoms] section. Before attempting to edit the topology file make sure that you are editing the processed gromacs topology file produced by using grompp -pp. Whereas, editing normal topology will not work because it does not contains all the parameters. The processed topology file should not contain any *#include* statements.

For instance, change the [atoms] section as following to select HOT atoms:

Before Editing: Processed.top

```
[ atoms ]
;nr   type   resnr   residue   atom   cgnr   charge   mass   typeB   chargeB
massB
;residue   1   ACE   rtp   ACE   q 0.0
1   HC   1   ACE   HH31   1   0.112300   1.0080   ;qtot   0.1123
2   CT   1   ACE   CH3   2   -0.366200   12.0100   ;qtot   -0.2539
```

After Editing: Processed.top

```
[ atoms ]
;nr   type   resnr   residue   atom   cgnr   charge   mass   typeB   chargeB
massB
;residue   1   ACE   rtp   ACE   q 0.0
1   HC_   1   ACE   HH31   1   0.112300   1.0080   ;qtot   0.1123
2   CT_   1   ACE   CH3   2   -0.366200   12.0100   ;qtot   -0.2539
```

2.5 Scaling Topologies

The modified typology needs to be scaled at each λ value. Plumed comes with a partial_tempering command line tool that can be used to generate scaled topology, which will scale the force fields parameters of the section [atom type] of a given topology file.

The partial_tempering tool can be used as follows:

```
plumed   partial_tempering   λ < processed.top > scaled.top
```

Where, λ is the scaling factor to scale the force field parameters from processed.top file. The scaled parameters will be written to scaled.top.

For example, the scaled topology should something look as below. Here the scaling factor λ has choosen randomly to illestrate.

Before Scaling: Processed.top

```
[ atomtypes ]
; name   at.num   mass   charge   ptype   sigma   epsilon
HC   1   1.00800   0.000000   A   0.264953   0.065688
CT   6   12.01000   0.000000   A   0.339967   0.45773
```

After Scaling: Processed.top

```
[ atomtypes ]
;name   at.num   mass   charge   ptype   sigma   epsilon
HC   1   1.00800   0.000000   A   0.264953   0.0656888
HC_   HC   1.00800   0.000000   A   0.264953   0.0486152   ;   scaled
CT   6   12.01000   0.000000   A   0.339967   0.45773
CT_   CT   12.01000   0.000000   A   0.339967   0.338758   ;   scaled
```

In genearal, you can generate the λ values for chosen effective temperature range obtained from geometric progression. Which means keep the exchange rate constant across the temperature range. The effective temperatures can be obtained using

$$T_i = T_0 * e^{ci} \tag{6}$$

Where, c is the desire acceptance ratio $c = \log\left(\frac{\frac{T_0}{T_i}}{n-1}\right)$,T_0 is the starting temperature, T_i is the effective temperature of the replica i and n is the number of replicas.

We can also calculate the λ corresponding to the particular effective temperature as

$$\lambda = \frac{T_0}{T_i}$$

2.6 Run REST2 Simulation

Well, now you are ready to submit REST2 simulation of alanine di-peptide in explicit solvent using 5 λ values (5 replicas) ranging from 1.0 to 0.3. Make sure that, you have processed 5 topology files corresponding to 5 λ values before starting simulation.

2.6.1 Generate tpr files

Use the following gromacs commands to generate gromacs binary input file (tpr).

```
for i in '0 1 2 3 4'; do
gmx_mpi grompp -f nvt.mdp -p scaled_$i.top -c
ala_wat_$i.gro -o res2_$i.tpr -maxwarn 10
;done
```

2.6.2 Submit REST2 using mdrun

Now, you can submit the REST2 simulation using gromacs mdrun and -hrex tags.

```
mpirun -np 5 gmx_mpi mdrun -v -deffnm rest_
-plumed plumed.dat -multi 5 -replex 100 -hrex
```

Where,

-np = No. of processors used
-multi = Instruct the program to preform multi (5) runs
-replex = Instruct the system to attend an exchange at every 100 steps.
-hrex = Instruct the program to perform Hamiltonian Replica Exchange Simulation using plumed.

A complete run can be done by using following shell script (Adopted from ref(3)).

```
# Shell Script to perform REST2 Simulation
#
#!/bin/bash
nrep=5              # No. of Replicas
tmin=300            # Initial Temp. of the system
tmax=1000           # build geometric progression
list=$(
awk -v n=$nrep \
    -v tmin=$tmin \
    -v tmax=$tmax \
  'BEGIN{for(i=0;i<n;i++){
    t=tmin*exp(i*log(tmax/tmin)/(n-1));
    printf(t); if(i<n-1)printf(",");
  }
}'
)
#
for((i=0;i<nrep;i++))
do
lambda=$(echo $list | awk 'BEGIN{FS=",";}{print
$1/$'$((i+1))';}')
#
# Scale the topology
  plumed partial_tempering $lambda < processed.top
> processed_scaled_$i.top
#
# Generate tpr file
gmx_mpi grompp -f nvt.mdp -p
processed_scaled_$i.top -o topol$i.tpr -maxwarn 10
done
#
# Submit Simulation
mpirun -np $nrep gmx_mpi mdrun -v -deffnm rest2_
-plumed plumed.dat -multi $nrep -replex 100 -hrex
```

Fig:6. **Shell Script**

Attention: Option -hrex requires also option -plumed. If you do not care about plumed, just provide an empty plumed.dat file. But, without -plumed flag the simulation will fail.

3 Analyse the Output

3.0.1 Quick Checks

It is very important to check whether everything went well or not, soon after successful completion of the simulation.

1. At first, you should able to find the average exchange probabilities are meaningful. Find the replica exchange statistics such as exchange probabilities and the exchanges of replica involved in every 100 steps at the end of rest2_$i.log files. You can access them by using simple bash command is

   grep -A9 "average probabilities" *.log

   The output of this command for one log file something should look like as bellow.

```
rest2_0.log:Repl   average probabilities:
rest2_0.log-Repl   0   1   2   3   4
rest2_0.log-Repl   1.0   .14   1.0   .40
rest2_0.log-Repl   number of exchanges:
rest2_0.log-Repl   0   1   2   3   4
rest2_0.log-Repl   9975   1343   9981   3963
rest2_0.log-Repl   average number of exchanges:
rest2_0.log-Repl   0   1   2   3   4
rest2_0.log-Repl   1.0   .13   1.0   .40
rest2_0.log-
```

2. Have a look at the system temperature of all the replicas. As we knew that, in REST2 all the replicas runs at same temperature.

   The system temperatures can be obtain from gmx energy tool.

   gmx energy -f rst2_$i.edr -s rest2_$i.tpr -o Temp_$i.xvg

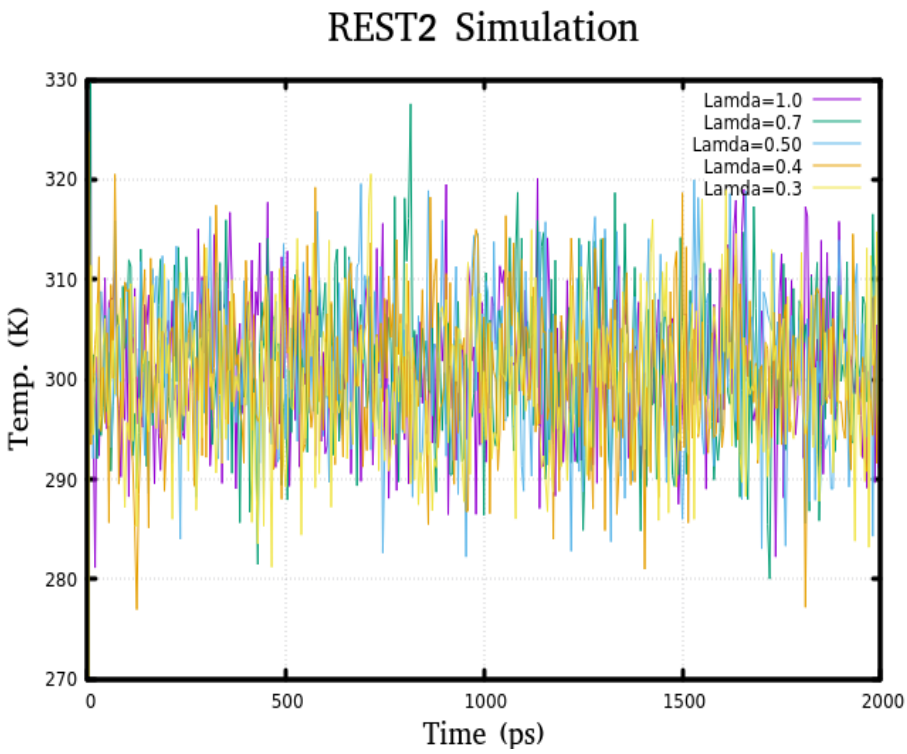

Fig:7. **Temperature of all the replicas**

3. Compare the energy files of replica scaled at $\lambda = 0.5$ with the unscaled replica at $\lambda = 1.0$. In the two resulting energy files you should see: long range electrostatics, LJ, and dihedral energy is half in the scaled case all other terms (bonds/bends) are identical.

   You can use gromacs energy tool to get the energy components of the system.

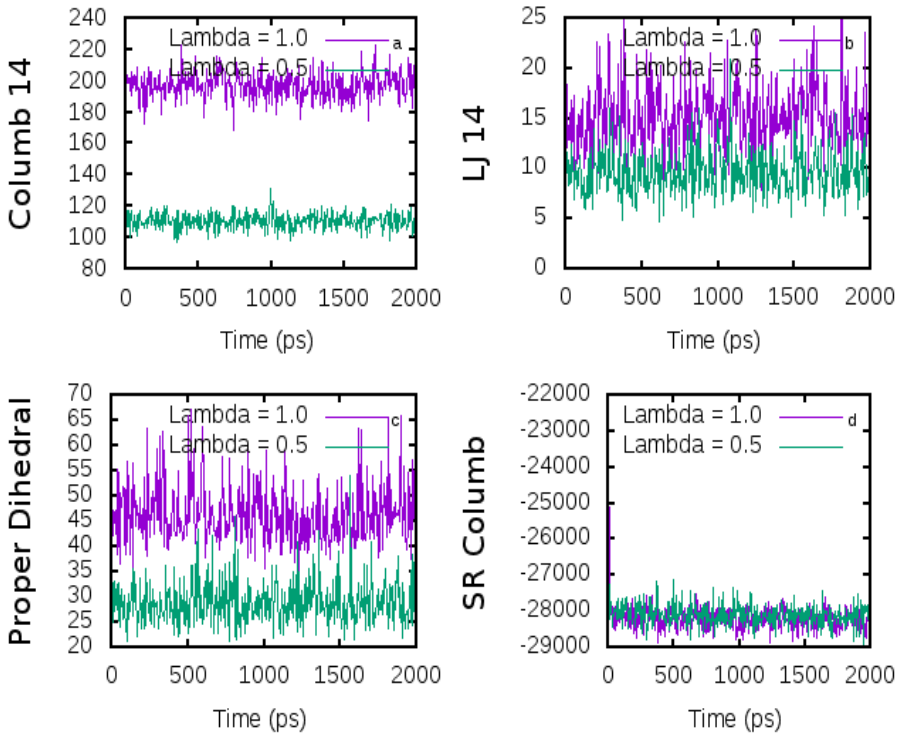

Fig:8. **Energy Components of Replica at $\lambda = 1.0$ and Replica at $\lambda = 0.5$**

3.0.2 Collective Variables Analysis

PLUMED will write the CV values in files names **COLVAR.$i**, check their evaluation by plotting with time using **Gnuplot**. The gnuplot commands as follows:

```
# Open Gnuplot on terminal
Terminal$ gnuplot

plot "COLVAR.$i" u 1:2 w l lw 2 title"PHI"
plot "COLVAR.$i" u 1:3 w l lw 2 title"PSI"
```

The plots should look like this.

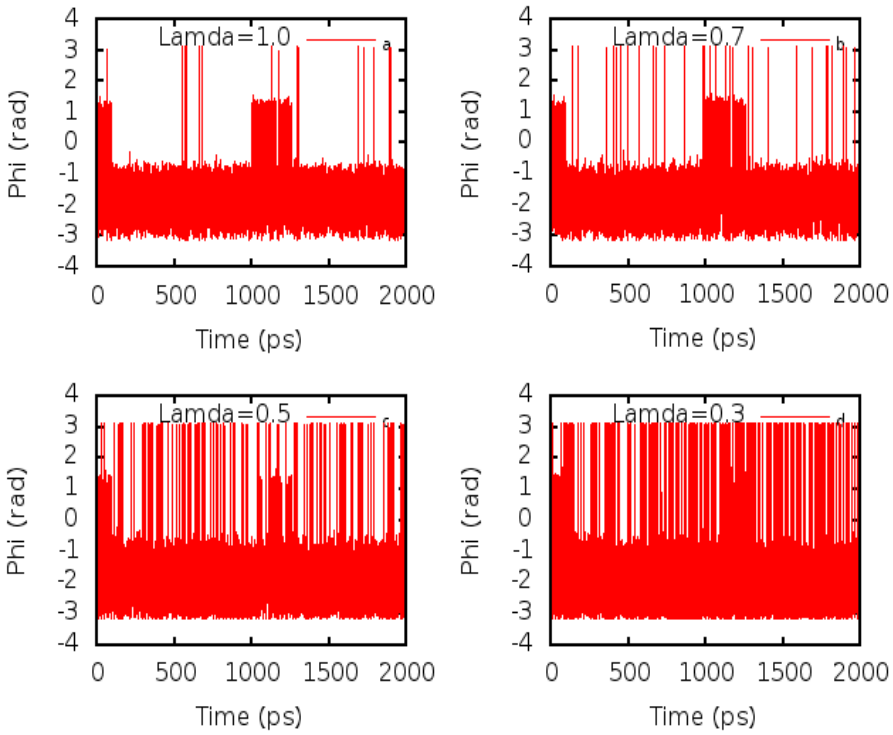

Fig:9. **The Time evolution of ϕ in replicas at λ = 1.0, 0.7, 0.5, 0.3 respectively.**

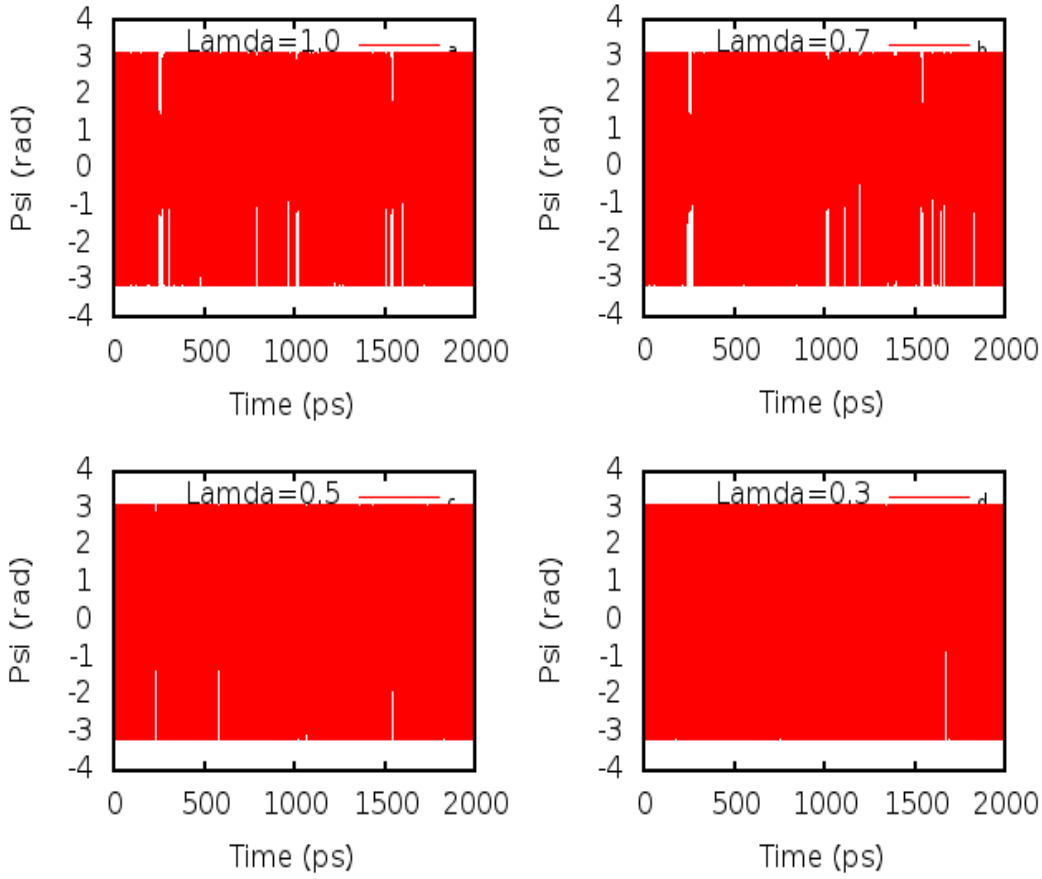

Fig:10. **The Time evolution of ψ in replicas at λ = 1.0, 0.7, 0.5, 0.3 respectively.**

We can clearly observe the sampling of phi & psi angles by plotting ramachandran plot (i.e ϕ v/s ψ)

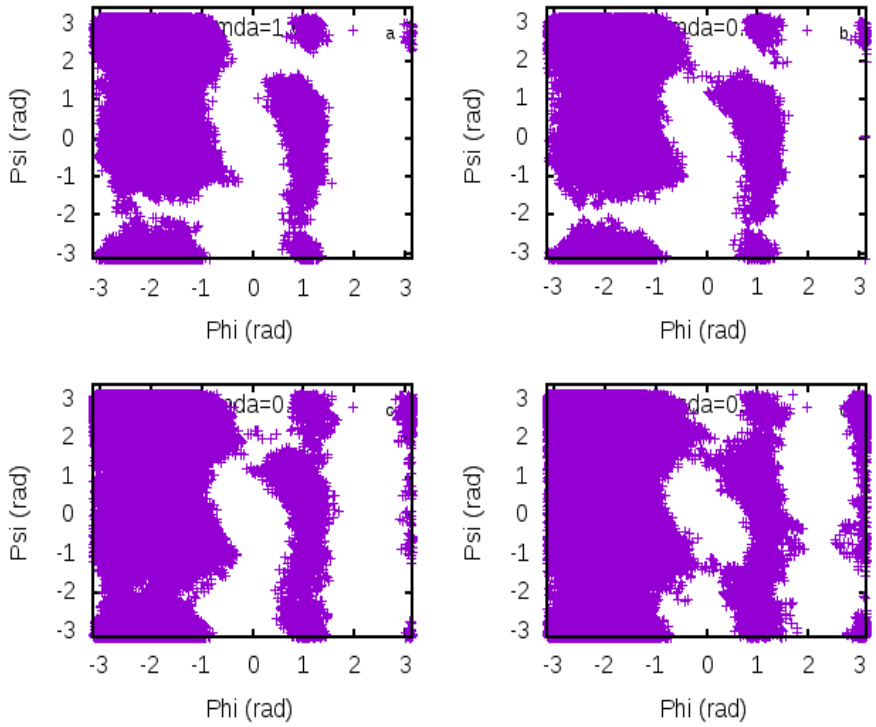

Fig:11. **Ramchandra plot at all replicas at λ = 1.0, 0.7, 0.5, 0.3 respectively.**

3.0.3 Construction of FES

The 1D free energy surface along ϕ and ψ can be constructed using plumed in-built tool **sum_hills** with following command:

 # Construct FES using sum_hills

plumed sum_hills –histo COLVAR.0 –idw phi –sigma 0.2 –kt 2.5 –outhisto fes_phi.dat
plumed sum_hills –histo COLVAR.0 –idw psi –sigma 0.2 –kt 2.5 –outhisto fes_psi.dat

 # Construct 2D FES
mv COLVAR.0 COLVAR
plumed sum_hills –histo COLVAR –sigma 0.2,0.2 –kt 2.5 –outhisto fes_2D.dat

These command will produce the following figures:

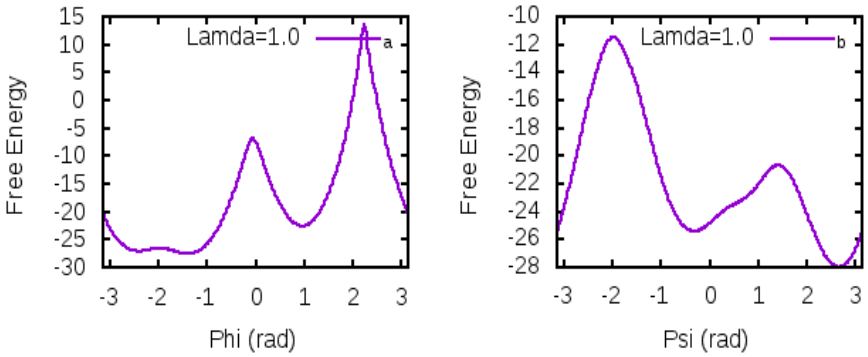

Fig:11. **Ramchandra plot at all replicas at λ = 1.0, 0.7, 0.5, 0.3 respectively.**

2-D Free Energy:
The 2D free enrgy surface(Fig.12) can be constructed using plumed sum_hills tool. Since we had run for 2ns simulation the higher energy states could not be accessed in this tutorial.

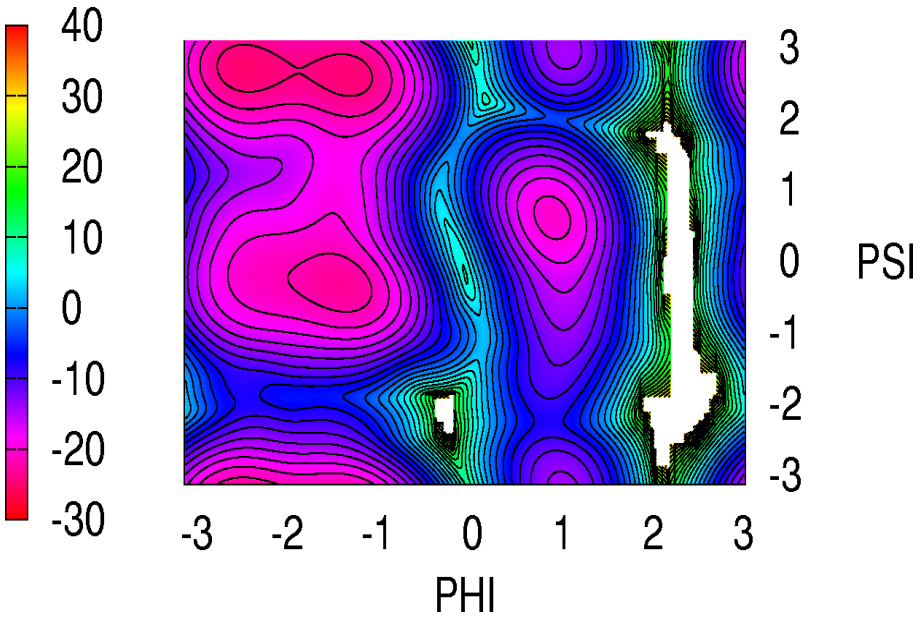

Fig:12. **2D FES of Alanine Di-peptide produced from 2ns REST2 simulation**

4 Summary

To be conclude, through this tutorial we have discussed the underlying theory of REST2 and learned how to setup the REST2 simulation using latest versions of gromacs & plumed then analyzed the results using gmx energy & plumed sum_hills tools by following some of the quick checks to troubleshot the simulation errors which may occur during simulation. A shell script to run this complete tutorial in a single shot by giving the processed topology and ala.gro files as inputs, can be available at Nisanth Nair Research Group Github Page(Ref 4).

5 References:

1. Lingle Wang, Richard. A Friesner, and B. J. Borne, *J. Phys. Chem. B*, **2011**, 115, 30, 9431-9438.

2. Giovanni Bussi, *Molecular Physics*, **2014**, 112:3-4, 379-384.

3. **Gromcas Tutorial**:www.bevanlab.biochem.vt.edu/Pages/Personal/justin/gmx-tutorials/lysozyme/01_pdb2gmx.html.

4. Nisanth Nair Research Group (GitHub Page): Click here.

5. Plumed Tutorial: Hamiltonian Replica Exchange MD.

@———- The End ———@