Enhanced Sampling Techniques

Contents

Chapter 1

Replica Exchange Solute Tempering Methods

Jaya Krishna Koneru and Korey Reid, and Paul Robustelli

1.1 Abstract

Accurate simulations of intrinsically disordered proteins and peptides predictive power and insight into experimental findings extending our predictive capabilities in molecular interactions and understanding biological mechanisms. Unlike equilibrium simulations of ordered proteins in their ground state, simulating disordered proteins, as well as rare allosteric effects in structured proteins, require long continuous simulations which may not be well sampled even out to 1 ms. Replica exchange molecular dynamics simulations with solvent scaling proves to be a powerful group of methods to sample dynamics with limited resources. REST2 and ssREST3 are promising methods in accelerating the sampling of the conformational ensemble, not ommiting the caveat: the simulation is only as good as the forcefield. In addition to discussing the background, we provide an example of REST2 and ssREST3 as it pertains to simulating the C-terminal domain of α-synuclein with and without a small molecular ligand and with an emphasis on assessing convergence of desired collective variables or observables.

Keywords Molecular dynamics, Replica Exchange, Collective variables, Solute-Scaling, α-synuclein

1.2 Introduction

Studying and critically understanding the underlying dynamics of biological systems at the atomistic scale can provide valuable insight into molecular mechanisms. Over many decades experimental techniques have been and continue to be developed in a collective effort to investigate processes at the molecular scale. However, the limitations of experimental techniques often are restricted from producing clear and relevant explaintions at the atomistic level. This is in large part to the ensemble characteristic of experimental measurements and lends to the difficulty of extracting conformational dynamics at various timescales, as well as extracting localized dynamics of say a biological system. Upon entry of molecular dynamics the usefullness was apparent, albiet limited by computational speed. As molecular dynamics engines[2, 3, 5, 17, 20, 24] and their accompanied forcefields[4, 6, 7, 9, 11, 12, 16] have matured, our ability to reach microsecond timescales has become routine. While the product of many decades of effort have resulted in our ability to simulate microseconds for a single biomolecule, vastly extending our comprehension of atomistic dynamics when paired with experimental results, these timescales are in fact beneath the threshold required to study rare events, e.g. allosteric transitions, or sample large degrees of freedom often accompanying IDPs.

When the desire is to study conformational dynamics it becomes a neccessity for long-timescale simulations on supercomputing systems[18, 19], for all other cases where statistical measures of an ensemble enhanced sampling techniques are available[8, 13–15, 22, 23, 25]. The phenomenological

observation desired dictates which simulation methodology will provide usefull information. One such method, Replica Exchange Solute Tempering or REST[10, 23, 25], has arrisen from Hamiltonian Replica Exchange Molecular Dynamics (HREMD) a derivative of Replica Exchange Molecular Dynamics (REMD). For each of these methods, multiple parallel simulations are conducted in parallel. These parallel simulations undergo exchanges at a designated interval. The relevance differences in each exchange method is discussed in the Introduction and Theory section, for a detailed account please refer to the original publications[10, 21, 23].

From Replica Exchange Molecular Dynamics[21], the hamiltonian representing the potential energy of the system can be written as sum of respective contributions, separated into protein-protein, protein-water and water-water:

$$E_n^{REMD}(X_n) = \lambda_n^{pp} E_{pp}(X_n) + \lambda_n^{pw} E_{pw}(X_n) + \lambda_n^{ww} E_{ww}(X_n) \tag{1.1}$$

λ_n^M is the scaling factor, where $M = \{pp, pw, ww\}$ which scales the corresponding energy term. For REST2[23], $\lambda_n^{ww} = 1$, $\lambda_n^{pp} = (\lambda_n^{pw})^2 = \lambda_n$, for simplicity the REST2 hamiltonian simplifies to:

$$E_n^{REST2}(X_n) = \lambda_n E_{pp}(X_n) + \sqrt{\lambda_n} E_{pw}(X_n) + E_{ww}(X_n) \tag{1.2}$$

where,

$$\lambda_n = \frac{\beta_n}{\beta_0} \tag{1.3}$$

and $\beta_n = \frac{1}{k_B T_n}$ for $n = \{0, 1, 2, \ldots, n_{replica}\}$.

Upon implementation of metropolis acceptance criteria with detail balance being satisfied, the acceptance probability between state n and $n+1$ is given by:

$$\begin{aligned}\Delta_{n,n+1} = (\beta_n - \beta_{n+1})\Big[&(E_{pp}(X_{n+1}) - E_{pp}(X_n)) \\ &+ \frac{\sqrt{\beta_0}}{\sqrt{\beta_n} + \sqrt{\beta_{n+1}}}(E_{pw}(X_{n+1}) - E_{pw}(X_n))\Big].\end{aligned} \tag{1.4}$$

By excluding the solvent-solvent interactions w.r.t. scaling, the contribution to the acceptance criteria contains less degrees of freedom relative to REMD resulting in better acceptance between replicas.

Upon investigation, disordered proteins containing hydrophobic residues undergo conformational collapse with respect to scaling E^{pw} to higher effective temperatures. This outcome is unfavorable when attempting to capture a representative ensemble as hydrophobic collapse reduces the overall sampling of the proteins conformational ensemble. Zhang, Liu, and Chen 2023 provided a basis for biasing the scaling such that protein collapse is minimized or negated, with the hamiltonian:

$$E_n^{REST3}(X_n) = \lambda_n E_{pp}(X_n) + \sqrt{\lambda_n}\Big[E_{pw}^{elec} + \kappa_n E_{pw}^{lj}\Big](X_n) + E_{ww}(X_n). \tag{1.5}$$

By incorporating the scaling factor κ_n, where $n = \{1, 2, 3, \ldots, n_{replica}\}$, the dampening effect of λ is deminished. To accomplish this κ_n exist in

the range [1.0, 1.1]. Due to the method of implementation the base replica (unscaled topology), following conversion of the topologies to REST3 [25] formalism using combination rule 1 instead of 2, does not match the potential energy of the original topology. The implementation of REST3 by Zhang, Liu, and Chen targeted scaling of protein-water lj parameters by means of computing C6 and C12 parameters. The descrepancy between the unscaled, converted topology and the original topology led to inception of ssREST3, solvent-scaled REST3, where λ_n is still applied to the protein in the same fashion as REST2 [23] and solvent scaling is applied to the water oxygen via the well-depth, ϵ_{OW}, borrowing from the methods described in Best and Mittal 2010. To overcome the inherent mechanics within GROMACS[20], nonbonded overrides are implemented for water-water and water-ion. Additionally, one can override water-cosolute if scaling is not desired between the two. For ssREST3, κ_n is defined by the expression,

$$\kappa_n = \kappa_{low} * \exp\left(n * \frac{\log(\kappa_{high}/\kappa_{low})}{N_r - 1}\right) ; \quad 1.00 \leq \kappa_n \leq 1.10. \tag{1.6}$$

To understand where κ_n is applied, we start with an expression for the Lennard-Jones potential between the i^{th} protein atom and the water oxygen,

$$\sqrt{\lambda_n} \cdot \kappa_n \cdot E^{lj}_{i,OW} = \sqrt{\lambda_n} \cdot \kappa_n \cdot 4 \cdot \epsilon_{i,OW}\left[\left(\frac{\sigma_{i,OW}}{r}\right)^{12} - \left(\frac{\sigma_{i,OW}}{r}\right)^{6}\right] \tag{1.7}$$

The CHARMM and Amber forcefields both conform to the Lorentz-Berthelot rules, i.e. $\epsilon_{i,j} = \sqrt{\epsilon_i \cdot \epsilon_j}$. From Equation 1.7 we can refactor $\sqrt{\lambda_n} \cdot \kappa_n \cdot \epsilon_{i,OW}$ to clarify which forcefield parameters are scaled.

$$\epsilon^{scaled}_{i,OW} = \sqrt{\lambda_n} \cdot \kappa_n \cdot \epsilon_{i,OW} = \sqrt{(\lambda_n \cdot \epsilon_i) \cdot (\kappa_n^2 \cdot \epsilon_{OW})} \tag{1.8}$$

The sections that follow are the Materials, Methods and Notes sections. Within the materials sections we detail the required software and minimum hardware requirements to perform REST based simulations. The Methods sections contains a complete description with examples on how to create and produce REST2 and ssREST3 simulations following an example simulation involving α-synuclein and a small molecule. Additionally, we include a few examples of analysis to test for convergence. Lastly, we provide various Notes in the last section.

1.3 Materials

For any simulations initial positions of the particles are very important. PDB files provides us with the required information about the initial coordinates where each atom is represented as a particle in space. It might get one wondering on how can a bunch of particles arranged in a 3-dimentional space can behave as a protein or a small molecule. This is be achieved with the help of force-field parameters which are a set of basic parameters that governs the interaction potentials between the different atoms of the molecule or protein. These force-field parameters are derived and optimized from quantum mechanical calculations and experimental data which helps us in replicating the macroscopic observables.

1.4 Methods

Implementing the ssREST3 in molecular dynamics simulations requires the use of the PLUMED2.8 plugin for GROMACS version 2023.5. The first step in the ssREST3 implementation is to generate the a processed topology file your solvated protein of interest using *-pp* option in *gmx grompp*.

```
gmx grompp -f *.mdp -c *.gro -r *.gro -p topol.top -o
    *.tpr -pp
```

This topology files will be used as to implement REST2 scaling procedure using PLUMED2.8 where one will supply the processed topology file and the corresponding λ value as such :

```
plumed partial\_tempering  $\lambda_{n}$ < processed.
   top > scaled.top
```

After the topology files are generated for the respective replicates the heavy atoms of the solvent, for example OW_{tip4pd} which is the water oxygen atom of amber99sb-*disp* force field whose ϵ will be scaled by a factor of κ_n^2 to satisfy the scaling condition as mentioned in Eq. 1.6, 1.8. If κ_n is set to 1.0 across all replicas one can observe the it behaves like a conventional REST2. Along with the scaling of the water oxygen we also made sure that the self interaction of the solvent molecules and there interactions with the ions will not be effected by adding three additions lines in the *[nonbonded]* section of the GROMACS topology file as shown in 1.2. Once this changes are made to the topology files we are ready to run the simulations. Using the below command all the replicates are run in parallel and the the conformational exchange between the replicas is set for every 800 steps which equates to 1.6 ps.

```
gmx  grompp -f *.mdp -c *.gro -r *.gro -p scaled.top
   -o scaled.tpr
gmx mdrun -s scaled.tpr -multi <replica folders> -
   replex 800 -deffnm replica -plumed plumed.dat
```

At the start of the simulations it is always a best practice to check the acceptance ratio between the replicas to ensure that the scaling is not too aggressive or too weak. A minimum of 20% acceptance ratio is recommended for the simulations to be considered valid.

To observe if the solvent scaling is preventing the collapse in the conformations of an IDP we use α-synuclein in absence and presence of a ligand "Fasudil".

Table 1.1: Table showing the differences and similarities of ϵ scaling between the REST2 and ssREST3 methods. In case of ssREST3 the water ϵ gets scaled along with the solute ϵ by a factor if κ^2 where as solvent parameters are not scaled during REST2.

Method	$T_{max}(K)$	λ	ϵ_{CA}	κ	ϵ_{OW}
–	300	1.0	0.359824	1.0	0.998989
REST2	450	0.666667	0.239883	1.0	0.998989
ssREST3	450	0.666667	0.239883	1.1	1.20878

Table 1.2: Table with $(\sigma_{ij}, \epsilon_{ij})$ showing the solvent interactions with itself and ions remain unaffected by the scaling factor κ_n.

Atom types	OW_{tip4pd}	$NA^{+}_{C22^*}$	$CL^{-}_{C22^*}$
OW_{tip4pd}	(0.3165, 9.98989)	(0.279746, 0.442754)	(0.360484, 0.791812)