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 show promise in accelerating the sampling of

1.2 Introduction

Studying and understanding the underlying molecular dynamics of biological systems is very helpful. Very many experimental techniques are developed over the course of scientific history. However, the there are limitations to experimental techniques for example we can't explain the atomistic details of molecular interactions, it will hard to extract the conformational dynamics with fast transitions between them, hard to understand the localized dynamics of a biological system etc. From the dawn of molecular computational techniques it opened a new frontier to study and explore the atomistic regiem of molecular interactions with the use of computational techniques. From simulating a small molecule containing a couple of atoms to large biological systems was made possible with integration of experimental data and optimizing the base parameters used in the simulations because of which we are able to reproduce physical relevant ensembles similar to experimental observables. Development of computational techniques and advancement in computational resources made it possible to study the bio-molecular systems of various length scales and there dynamics at

atomistic level. Still computational techniques are limited by various factors such as athenticity of the force field used, the time scale of the simulation, the size of the system, the sampling of the conformational space etc. To overcome some of these limitations computational techniques widely know as enhanced sampling methods are developed which are helpful in reducing the simulation time scales required to generate the conformational ensembles. There are many techniques in the literature in which we will be discussing about an method called solvent-scaled Replica Exchange with Solute Scaling which is an optimized version of Replica Exchange with Solute Scaling(REST2) method.

1.3 Theory

From Replica Exchange Molecular Dynamics[2], 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[3], $\lambda_n^{ww} = 1$, $\lambda_n^{pp} = (\lambda_n^{pw})^2 = \lambda_n$, 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,

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

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

Upon 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 degrees of freedom. Zhang, Liu, and Chen 2023 provided a basis for biasing the scaling such that protein collapse is minimized or negated. The formalism they proposed,

$$E_n^{REST3}(X_n) = \lambda_n E_{pp}(X_n) + \sqrt{\lambda_n}\Big[E_{pw}^{elec} + \kappa_n E_{pw}^{lj}\Big](X_n) + E_{ww}(X_n). \tag{1.4}$$

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 [4] 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 atoms, this led to inception of ssREST3 where λ_n is still applied to the protein in the same fashion as REST2 [3] and apply solvent scaling on the water oxygen borrowing from the methods described in Best and Mittal 2010. κ_n is defined by the expression,

$$\kappa_n = \kappa_{low} * \exp\left(n * \frac{\log(\kappa_{high}/\kappa_{low})}{N_r - 1}\right) \; ; \quad 1.00 \leq \kappa_n \leq 1.10. \tag{1.5}$$

To correctly identify 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,

$$\sqrt{\lambda_n} \cdot \kappa_n \cdot E^{lj}_{i,OW} = \sqrt{\lambda_n} \cdot \kappa_n \cdot 4 \cdot \epsilon_{i,OW} \left[\left(\frac{\sigma_{i,OW}}{r}\right)^{12} - \left(\frac{\sigma_{i,OW}}{r}\right)^{6} \right] \tag{1.6}$$

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 1.6 we can refactor $\sqrt{\lambda_n} \cdot \kappa_n \cdot \epsilon_{i,OW}$ to clarify which forcefield parameters are scaled.

$$\epsilon^{scaled}_{i,OW} = \sqrt{\lambda_n} \cdot \kappa_n \cdot \epsilon_{i,OW} = \sqrt{(\lambda_n \cdot \epsilon_i) \cdot (\kappa_n^2 \cdot \epsilon_{OW})} \tag{1.7}$$

1.4 Methods

Implementing the ssREST3 in molecular dynamics simulations requires the use of the PLUMED2.8 plugin for GROMACS. The first step in the ssREST3 implementation is to generate the topology files your protein of interest which is solvated. This topology files will be used as to implement REST2 scaling procedure using PLUMED2.8. After the topology files are generated for the respective replicates the heavy atoms of the solvent, for example OW_{tip4p} 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. **??**, 1.7. If κ_n is set to 1.0 across all replicas on 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 no tbe effected by adding 3 additions lines in the *[nonbonded]* section of the GROMACS topology file as shown in 1.2.

1.5 Simulations

We are going to discuss a case study of asyn with and without a ligand.

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_{tip4p}	$\mathrm{NA}^{+}_{C22^*}$	$\mathrm{CL}^{-}_{C22^*}$
OW_{tip4p}	(0.3165, 9.98989)	(0.279746, 0.442754)	(0.360484, 0.791812)