

# Enhanced Sampling Techniques

## Contents

<b>Contents</b>	<b>1</b>
<b>1 Replica Exchange Solute Tempering Methods</b>	<b>3</b>
<i>Jaya Krishna Koneru and Korey Reid, and Paul Robustelli</i>	
1.1 Abstract . . . . .	3
1.2 Introduction . . . . .	3
1.3 Theory . . . . .	4
1.4 Methods . . . . .	5
1.5 Simulations . . . . .	6



## 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  $\alpha$ -synuclein with and without a small molecular ligand and with an emphasis on assessing convergence of desired collective variables or observables.

### 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 explanations 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 usefulness was apparent, albeit 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 necessity 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 useful information. One such method, Replica Exchange Solute Tempering or REST[10, 23, 25], has arisen from Hamiltonian Replica Exchange Molecular Dynamics derived from Replica Exchange Molecular Dynamics.

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[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:

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

$\lambda_n^M$  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:

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

where,

$$\lambda_n = \sqrt{\frac{\beta_n}{\beta_0}} \quad (1.3)$$

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

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,

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

By incorporating the scaling factor  $\kappa_n$ , where  $n = \{1, 2, 3, \dots, n_{\text{replica}}\}$ , the dampening effect of  $\lambda$  is diminished. To accomplish this  $\kappa_n$  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 atoms, this led to inception of ssREST3 where  $\lambda_n$  is still applied to the protein in the same fashion as REST2 [23] and apply solvent scaling on the water oxygen borrowing from the methods described in Best and Mittal 2010.  $\kappa_n$  is defined by the expression,

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

To correctly identify where  $\kappa_n$  is applied, we start with an expression for the Lennard-Jones potential between the  $i^{th}$  protein atom and the water oxygen,

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

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.

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

## 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

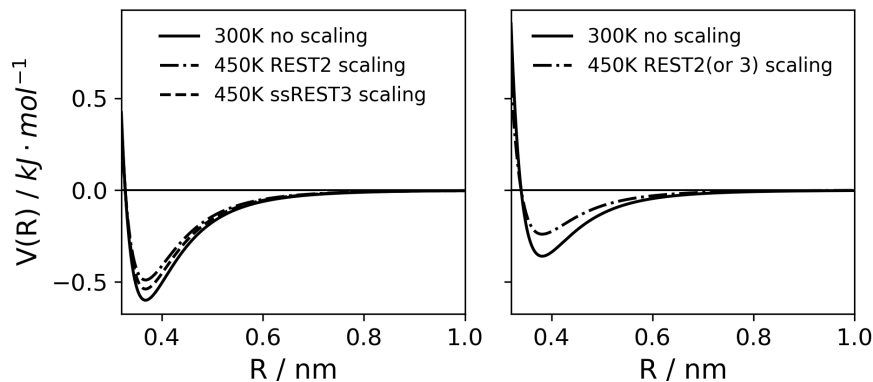


Figure 1.1: LJ potential for differing values of epsilon for the same atom-atom pair.

in the ssREST3 implementation is to generate the topology files your protein of interest which is solvated in a box. 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  $\epsilon$  will be scaled by a factor of  $\kappa_n^2$  to satisfy the scaling condition as mentioned in Eq. ??, 1.7. If  $\kappa_n$  is set to 1.0 across all replicas one can observe that 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 their interactions with the ions will not be affected by adding 3 additional lines in the *[nonbonded]* section of the GROMACS topology file as shown in 1.2. Once these changes are made to the topology files we are ready to run the simulations. The conformational exchange between the replicas is set for every 800 steps which equates to 1.6 ps.

## 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  $\epsilon$  scaling between the REST2 and ssREST3 methods. In case of ssREST3 the water  $\epsilon$  gets scaled along with the solute  $\epsilon$  by a factor of  $\kappa^2$  where as solvent parameters are not scaled during REST2.

Method	$T_{max}(K)$	$\lambda$	$\epsilon_{CA}$	$\kappa$	$\epsilon_{OW}$
–	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  $\kappa_n$ .

Atom types	$OW_{tip4p}$	$NA_{C22^*}^+$	$CL_{C22^*}^-$
$OW_{tip4p}$	(0.3165, 9.98989)	(0.279746, 0.442754)	(0.360484, 0.791812)