



# DESP: Deep Enhanced Sampling of Proteins' Conformation Spaces Using AI-Inspired Biasing Forces

Emmanuel Oluwatobi Salawu\*

Machine Learning Solutions Lab, Amazon Web Services (AWS), Herndon, VA, United States

The molecular structures (i.e., conformation spaces, CS) of bio-macromolecules and the dynamics that molecules exhibit are crucial to the understanding of the basis of many diseases and in the continuous attempts to retarget known drugs/medications, improve the efficacy of existing drugs, or develop novel drugs. These make a better understanding and the exploration of the CS of molecules a research hotspot. While it is generally easy to computationally explore the CS of small molecules (such as peptides and ligands), the exploration of the CS of a larger biomolecule beyond the local energy well and beyond the initial equilibrium structure of the molecule is generally nontrivial and can often be computationally prohibitive for molecules of considerable size. Therefore, research efforts in this area focus on the development of ways that systematically favor the sampling of new conformations while penalizing the resampling of previously sampled conformations. In this work, we present *Deep Enhanced Sampling of Proteins' Conformation Spaces Using AI-Inspired Biasing Forces* (DESP), a technique for enhanced sampling that combines molecular dynamics (MD) simulations and deep neural networks (DNNs), in which biasing potentials for guiding the MD simulations are derived from the KL divergence between the DNN-learned latent space vectors of [a] the most recently sampled conformation and those of [b] the previously sampled conformations. Overall, DESP efficiently samples wide CS and outperforms conventional MD simulations as well as accelerated MD simulations. We acknowledge that this is an actively evolving research area, and we continue to further develop the techniques presented here and their derivatives tailored at achieving DNN-enhanced steered MD simulations and DNN-enhanced targeted MD simulations.

**Keywords:** conformation space, deep neural network, protein, molecular dynamics simulation, variational autoencoder

## INTRODUCTION

The functions of biomolecules are encoded in their structures and dynamics (Council and others 1989; Karplus and Kuriyan, 2005; Yang et al., 2014). And there are innumerable pieces of evidence linking the basis of many diseases to anomalies in the structures and the dynamics of the molecules that are involved in the biological systems that the diseases affect (McCafferty and Sergeev, 2016; Chiti and Dobson, 2017; Guo et al., 2017; Hartl, 2017; Tramutola et al., 2017; Salawu, 2018a; Ittisoponpisan et al., 2019; Laskowski et al., 2020) because the normal functioning of the biological systems depends on the molecules' proper structures and dynamics. Furthermore, the various structures that a molecule can take (i.e., the molecule's conformation space, CS) and their associated MD are not only of vital

## OPEN ACCESS

### Edited by:

Guang Hu,  
Soochow University, China

### Reviewed by:

Min Zeng,  
Central South University, China  
Jinwei Duan,  
Chang'an University, China

### \*Correspondence:

Emmanuel Oluwatobi Salawu  
esalawu@amazon.com

### Specialty section:

This article was submitted to  
*Biological Modeling and Simulation*,  
a section of the journal  
*Frontiers in Molecular Biosciences*

**Received:** 25 July 2020

**Accepted:** 15 February 2021

**Published:** 04 May 2021

### Citation:

Salawu EO (2021) DESP: Deep  
Enhanced Sampling of Proteins'  
Conformation Spaces Using AI-  
Inspired Biasing Forces.  
*Front. Mol. Biosci.* 8:587151.  
doi: 10.3389/fmolb.2021.587151

importance in deciphering of many diseases (Salawu, 2018a; Salawu, 2018b) but are also crucial in the drug development efforts targeted at curing or managing many diseases (Carlson and McCammon, 2000; Lee et al., 2018; Pawełand and Caffisch, 2018; Wang et al., 2018; Lin et al., 2020). These recognitions have motivated extensive efforts in the field of structural biochemistry and form the rationale for many structural biology studies (such as through X-ray crystallography, NMR, and Cryo-EM) and the creation of the Protein Data Bank (Berman et al., 2000) as well as other databases for molecular structures. Nonetheless, considerable challenges exist because the solely static molecular structures obtained through the wet laboratory approaches alone (such as the ones listed above) often fall short of providing enough insights into the dynamics of the molecules of interest. These challenges have led to the growing roles and the increasing importance of computational approaches, such as molecular dynamics (MD) simulations, that are often used for studying the dynamic behaviors of molecules and their interactions with other molecules as well as for exploring much wider CS of the molecules of interest.

While it is generally easy to computationally explore the CS of small molecules (such as peptides and ligands), the exploration of the CS of larger a biomolecule beyond the local energy well and beyond the initial equilibrium structure of the molecule is generally nontrivial (Shaw et al., 2008; Shaw et al., 2009) and can often be computationally prohibitive for a molecule of considerable size. These difficulties arise from the existence of energy barriers between different states that the molecule could assume, thereby hindering the movement of the molecule from one structural state to another (Hamelberg et al., 2004; Hénin and Chipot, 2004; Salawu, 2020). At this point, it is important to acknowledge existing efforts targeted at removing, avoiding/sidestepping, lowering, or surmounting these energy barriers, thereby achieving enhanced sampling of the CS of molecules. Therefore, we recognize some of the previous publications in this domain and highlight them in the next paragraphs.

Most of the existing popular approaches for achieving enhanced sampling may be broadly viewed in two categories, namely: those that require the user to specify well-defined collective variables (CVs)/reaction coordinates (RCs) (Babin et al., 2008; Laio and Gervasio, 2008; Bussi and Laio, 2020) and those that do not require the user to explicitly specify the CV/RC (Sugita and Okamoto, 1999; Hamelberg et al., 2004; Moritsugu et al., 2012; Harada and Kitao, 2013; Miao et al., 2015; Chen and Ferguson, 2018; Salawu, 2020). Reconnaissance meta-dynamics uses a self-learning algorithm for accelerated dynamics and is capable of handling a large number of collective variables by making use of bias potentials created as a function of individual locally valid CVs that are then patched together to obtain the sampling across a large number of collective variables (Tribello et al., 2010). Some of the challenges of reconnaissance meta-dynamics such as those associated with the creation of bias potentials as a function of individual locally valid CVs could be addressed by any technique that could potentially learn a compressed representation of those CVs and efficiently explore the combined CVs together in the compressed space. This is the subject of an actively growing

research area that leverages the powers of deep neural networks (DNNs)/machine learning (ML). Bridging the fields of enhanced sampling and ML, Bonati et al. (2019) developed DNN-based variationally enhanced sampling that uses neural networks to represent the bias potential in a variational learning scheme that makes it possible for the efficient exploration of even high-dimensional free energy surfaces. In a similar way, reweighted autoencoded variational Bayes (RAVE) models MD simulation trajectories using the VAE whereby the learned distribution of the latent space variable is used to add biasing potentials, thereby penalizing the repeated sampling of the most favorable frequently visited states (Ribeiro et al., 2018). Although other enhanced sampling methods implement the biasing protocol in two steps, RAVE's identification of the RC and its derivation of unbiased probability distribution occur simultaneously. And through the systematic use of the Kullback–Leibler (KL) divergence metric, RAVE can identify physically meaningful RCs from among a group of RCs explored.

In addition to the efforts mentioned above, the combination of well-tempered meta-dynamics and time-lagged independent component analysis to study rare events and explore complex free energy landscapes have also been looked into (McCarty and Parrinello, 2017). Since the initial choice of CVs for meta-dynamics is often suboptimal, the work shows the finding of new and optimal CVs with better convergence properties by the analysis of the initial trajectory using time-lagged independent component analysis (McCarty and Parrinello, 2017). However, a more recent study has shown that rather than using linear dimension reduction methods (such as independent component analysis) a modified autoencoder could more accurately encode the low dynamics of the underlying stochastic processes of MD simulations better than linear dimension reduction methods (Wehmeyer and Noé, 2018). Indeed, there are continuous and growing efforts in the combinations of DNN models and MD simulations in the enhancement of the sampling of molecules' CS and other various aspects of molecular sciences (Allison, 2020; Salawu, 2020; Sidky et al., 2020).

In this work, we present *Deep Enhanced Sampling of Proteins' Conformation Spaces Using AI-Inspired Biasing Forces* (DESP), which also combines DNNs and MD simulations to create a robust technique for enhanced sampling of CS of molecules. Here, a DNN model is trained alongside MD simulations of the molecule of interest such that the models learn a compressed representation of the sampled structures of the molecule. The latent space vectors of the DNN model are then used in ways that provide useful information for inferring appropriate biasing potentials that are then used for guiding the MD simulations, thereby allowing efficient sampling of the molecule's CS. More specifically, the use of the KL divergence between the VAE's latent vectors of the current conformation (obtained from the MD simulations) and the VAE's latent vectors of the known, previously sampled, conformations makes it possible to bias the MD simulation away from visiting previously sampled conformations and rather toward visiting previously unsampled conformations.

The AI-based enhanced sampling approach presented in this work is not dependent on having prior knowledge of the molecule's CS distribution and does not require any careful

selection of collective variables. Therefore, this approach is very promising, given that the selection of appropriate collective variables is often very challenging (Tribello et al., 2010), and there is no well-defined solution that can fit all situations/all molecular systems. Rather than requiring manual specification of the collective variables to use, DESP, by itself, learns the compressed representation of the molecular system of interest and derives biasing potentials based on the distribution of the molecule's conformations in that compressed representation space. The results obtained show that DESP outperforms both conventional and accelerated MD simulations, and efficiently samples wider CS than conventional and accelerated MD simulations. Furthermore, the ideas in DESP are generalizable and may be used for implementing other forms of biased MD simulations including targeted and steered MD simulations. In the next section, we present the methods that make DESP possible and thereafter the overall DESP algorithm.

## MATERIALS AND METHODS

### Protein Molecules Used

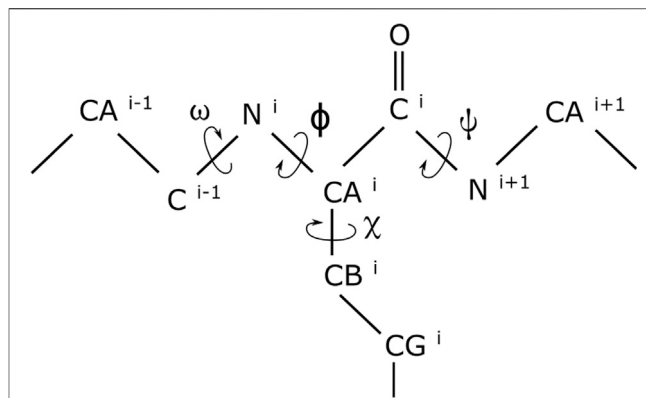
We began with a smaller protein/peptide (alanine dodecapeptide with 12 alanine residues,  $A_{12}$ ) and modeled its 3D structure using RPBS (Alland et al., 2005). The small size of alanine dodecapeptide helped in the initial testing and fine-tuning of DESP. In addition to  $A_{12}$ , we obtained a solution nuclear magnetic resonance (NMR) structure of GB98 that was expressed in *Escherichia coli* BL21 (DE3) from the Protein Data Bank (Berman et al., 2000), PDB ID: 2lhd (He et al., 2012). GB98 was selected because of its relatively small/medium size and because of the presence of the various secondary structure types (namely, alpha-helix, beta-sheet, and coils) in it. On the other hand, any protein could be used for the demonstration of the functionality of DESP, and the ones used here are just examples.

### Creation of the Initial Molecular Systems

Assignment of appropriate residues' charges and protonation states were handled using PDB2PQR (Dolinsky et al., 2007; Jurrus et al., 2018). Using AmberTools18's tLeap (Pearlman et al., 1995; Case et al., 2005; Salomon-Ferrer et al., 2013a; Salomon-Ferrer et al., 2013b), ff14SB (Maier et al., 2015) force-fields for the proteins, and ions234lm\_126\_tip3p for the ions and the water molecules (Li and Merz, 2014), we created explicitly solvated molecular systems for  $A_{12}$ 's and GB98's molecular dynamics (MD) simulations with OpenMM (Eastman et al., 2017) containing 2068 TIP3P water molecules ( $42.38\text{\AA} \times 48.80\text{\AA} \times 47.15\text{\AA}$  box size, for  $A_{12}$ ) or 9,981 ( $101.10\text{\AA} \times 94.35\text{\AA} \times 98.34\text{\AA}$  box size, for GB98) TIP3P water molecules.

### Energy Minimization and Heating

Each of the molecular systems was energy-minimized using OpenMM (Eastman et al., 2017). The energy minimizations were done in two stages—weakly ( $2.5 \text{ kcal/mol/\AA}^2$ ) restraining all the alpha carbon atoms in the first stage, and without any restraints in the second stage. With the weak restraints ( $2.5 \text{ kcal/}$



**FIGURE 1** | Dihedral angles in a short segment of a protein. While the omega ( $\omega$ ), phi ( $\phi$ ), and psi ( $\psi$ ) dihedral angles are in the proteins backbone, the chi<sub>1</sub> ( $\chi_1$ ) dihedral angle is at the beginning of an amino acid's side chain.

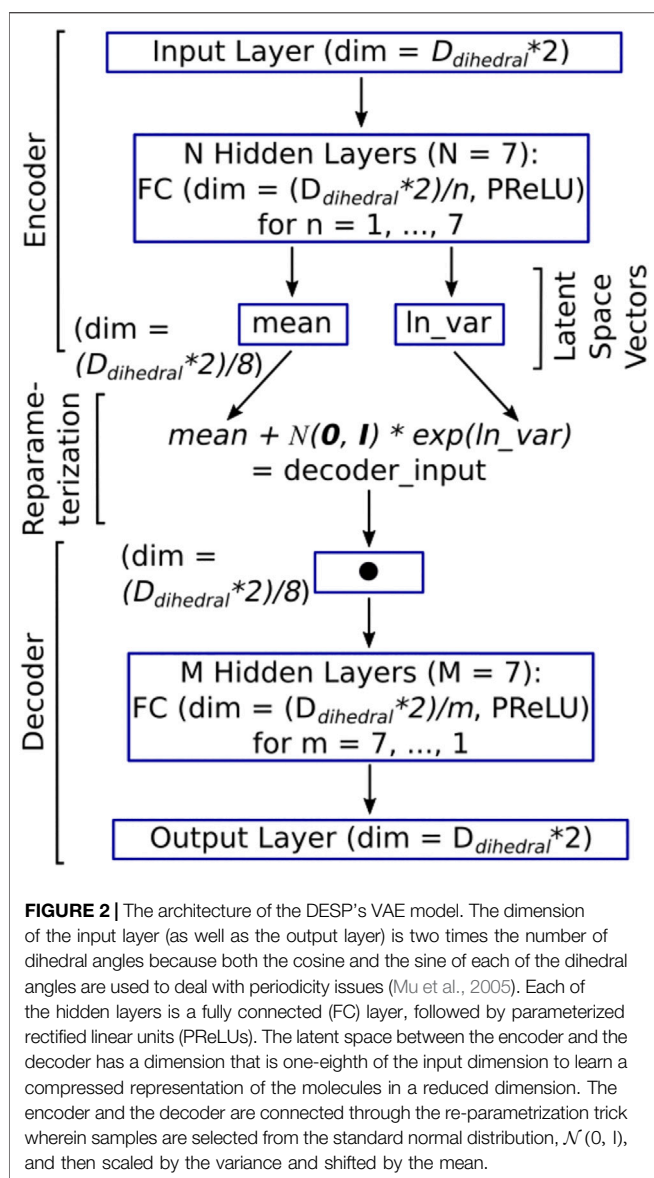
$\text{mol/\AA}^2$ ) reapplied on the alpha carbon atoms, the molecular systems were steadily heated to a temperature of 310 K in a canonical ensemble using the Langevin thermostat (Pastor et al., 1988).

### Conventional Molecular Dynamics Simulations

During both the equilibration and production runs, we controlled the systems' temperatures and pressures using the Langevin thermostat (Pastor et al., 1988) with a collision frequency of  $2\text{ps}^{-1}$  and the Monte Carlo barostat (Chow and Ferguson, 1995; Åqvist et al., 2004), respectively. Full electrostatic interaction energies were calculated using the particle mesh Ewald method (Darden et al., 1993). A cutoff distance of  $10\text{\AA}$  and a cubic spline switch function were used when calculating nonbonded interactions. All bonds in which at least one atom is hydrogen are constrained using the SHAKE algorithm (Ryckaert et al., 1977). All production run MD simulations were performed at 2 femtoseconds time step. Overall, the results from 800 ns of conventional MD simulations, 800 ns of accelerated MD simulations, and 280 ns of DESP MD simulations are presented in this work for each of the  $A_{12}$  and the GB98 molecular systems.

### Representations of the Molecules for Deep Learning Modeling

Since considerable changes in the conformation of biomolecules can be captured by variations in the dihedral angles of the molecules (Salvador, 2014; Cukier, 2015; Ostermeir and Zacharias, 2014; Lemke and Peter, 2019), we represent a molecule's conformation by the cosine and the sine of the dihedral angles (Mu et al., 2005) of that conformation. For these, we make use of the omega ( $\omega$ ), phi ( $\phi$ ), psi ( $\psi$ ), and chi<sub>1</sub> ( $\chi_1$ ) dihedral angles (with examples illustrated in **Figure 1**). Although using both the cosine and the sine of each of the dihedral angles doubles the dimensionality, it helps



in removing the adverse effect that the periodicity of the dihedral angles would have had on the modeling. Extensive details of the benefits of using the dihedral angles (Lemke and Peter, 2019) and of simultaneously using both their cosines and sines have been documented elsewhere (Mu et al., 2005) and, for brevity, are not repeated here.

### DNN Architecture: Variational Autoencoder

Our DNN of the type variational autoencoder (VAE) has a simple architecture, as shown in **Figure 2**. The input layer takes both the cosine and the sine of the dihedral angles' representation of the molecular conformation (giving rise to a vector of dimension  $D_{\text{Dihedrals}} * 2$ ) as input. The input layer is followed by  $N$  hidden layers (where  $N = 7$  in the current case). Each of the hidden layers, numbered  $n = [1, 2 \dots N]$ , has  $D_{\text{Dihedrals}}/n$  nodes. The next layer is made up of two latent space vectors, each of size  $(D_{\text{Dihedrals}} * 2)/(N + 1)$ , which is  $(D_{\text{Dihedrals}} * 2)/8$  in the current case. The first

latent space vector represents the mean for the Gaussian distribution that the latent space encodes (i.e., mean in **Figure 2**), while the second vector represents the natural logarithm of the variance for the Gaussian distribution that the latent space encodes (i.e.,  $\ln\_var$  in **Figure 2**). The DNN architecture up to this point is the encoder (**Figure 2**).

The decoder, which is like a mirror image of the encoder, begins with an input layer with  $(D_{\text{Dihedrals}} * 2)/(N + 1)$  nodes and is followed by  $M$  hidden layers (where  $M = 7$  in the current case). Each of the hidden layers, numbered  $m = [M, M-1 \dots 1]$ , has  $(D_{\text{Dihedrals}} * 2)/m$  nodes. The output (which is the final) layer emits the reconstructed cosine and sine of the dihedral angles of the molecular conformation that was passed in as input. To allow the passage of backpropagation signals through the entire VAE (i.e., from the decoder to the encoder), we connect the encoder and the decoder by a re-parameterization trick that is made up of an equation that takes the output of the encoder (namely, the vector of mean, and the vector of the logarithm of variance) as an input and uses it to sample from the corresponding normal distributions. This is done indirectly by initially drawing samples from the standard normal distribution. The samples drawn are then scaled and shifted accordingly using the variance vector and the mean vector, thereby obtaining the intended distribution (see the re-parameterization expression in **Figure 2**). The output of the re-parameterization is then fed into the decoder's input layer (**Figure 2**). We used PyTorch (Paszke et al., 2019) with CUDA support for building all deep neural network models in this study. Given the architecture of the DNN and its inputs and outputs, we can now examine how the DNN is trained.

### DNN Training

We defined the model's loss function as a weighted combination (**Eq. 1**) of reconstruction loss captured by mean square error (MSE) loss (**Eq. 2**) and the Kullback–Leibler (KL) divergence loss (**Eq. 3**). We set the weighting parameter,  $w$ , to 0.1 so that the MSE loss has a higher weight ( $1 - 0.1 = 0.9$ ) than the KL divergence loss (0.1). We arrived at this weighting scheme from our preliminary experiments through grid search, wherein we observed that setting the KL divergence's weight to 0.1 helped in the faster convergence of the model loss and in achieving a much better reconstruction accuracy for the trained model, on both the training dataset and validation dataset.

$$Loss_{model} = (1 - w) * Loss_{MSE} + (w) * Loss_{KL} \quad (1)$$

$$Loss_{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (2)$$

$$\begin{aligned} Loss_{KL} &= D_{KL} \left\{ N \left[ (\mu_1, \dots, \mu_n)^T, \text{diag}(\sigma_1^2, \sigma_n^2) \right] \parallel \mathcal{N}(0, \mathbf{I}) \right\} \\ &= \frac{1}{2} \sum_{i=1}^n [\sigma_i^2 + \mu_i^2 - \ln(\sigma_i^2) - 1] \end{aligned} \quad (3)$$

The KL Divergence upon which **Eq. 3** is based represents a special case involving the KL-divergence between a multivariate normal distribution,  $N [(\mu_1, \dots, \mu_n)^T, \text{diag}(\sigma_1^2, \dots, \sigma_n^2)]$  with means  $\mu_1, \dots, \mu_n$  and variances  $\sigma_1^2, \dots, \sigma_n^2$ , and a standard normal distribution,  $\mathcal{N}(0, \mathbf{I})$ .

```

Initialize  $N_{Completed} \leftarrow 0$ ;  $N_{Needed} \leftarrow 1e9$ ;  $N_{Short} \leftarrow 1e7$ ;  $N_{Saving} \leftarrow 1e4$ ;
 $N_{Biasing} \leftarrow 50$ ;  $N_{BeforeRetraining} \leftarrow 100$ ;
 $S_{Frames} \leftarrow \{\} = \text{Empty Set}$ ;  $N_{Strata} \leftarrow 10$ 
For  $i$  in range  $\left(1 \text{ to } \frac{N_{Short}}{N_{Saving}}\right)$  do
  Run conventional MD simulations for  $N_{Saving}$  steps
  Save the conventional MD simulations frame to  $S_{Frames}$ 
  Update  $N_{Completed} \leftarrow N_{Completed} + N_{Saving}$ 
Train the VAE using the frames in  $S_{Frames}$  (or a subset of it)
While  $N_{Completed} < N_{Needed}$  do
  Set  $N_{AvailableFrames} \leftarrow \frac{N_{Completed}}{N_{Saving}}$ ;  $N_{EachStartum} \leftarrow \frac{N_{AvailableFrames}}{N_{Strata}}$ 
  For  $i$  in range  $\left(1 \text{ to } \frac{(N_{Needed} - N_{Short})}{(N_{Saving} * N_{BeforeRetraining})}\right)$  do
    For  $j$  in range  $(1 \text{ to } N_{BeforeRetraining})$  do
      For  $k$  in range  $\left(1 \text{ to } \frac{N_{Saving}}{N_{Biasing}}\right)$  do
        Set  $V_{KL} \leftarrow$ 

$$\frac{\sum_{l=1}^{N_{Strata}} KL_{Div}\left(\text{Encoder}(\text{Frame}_{Last}), \text{Encoder}(\text{Frame}_{(l-1)*N_{EachStartum}})\right)}{N_{Strata}}$$

        Set  $V_{Biasing} \leftarrow \left(\frac{V_{KL\_Upper}}{V_{KL}}\right)^2$ 
        Run biased MD simulations for  $N_{Biasing}$  steps while adding
           $V_{Biasing}$  to the protein atoms' potential energy
        Keep track of the  $V_{Biasing}$  used
        Update  $N_{Completed} \leftarrow N_{Completed} + N_{Biasing}$ 
        Save the biased MD simulations frame to  $S_{Frames}$ 
        Train the VAE further using the frames in  $S_{Frames}$  (or a subset of it)
      Reweigh the biased MD simulations trajectory based on  $V_{Biasing}$ 

```

**FIGURE 3** | DESP algorithm. The DESP combines DNNs with MD simulations to achieve enhanced sampling of molecules' conformation spaces.

For the minimization of the loss and, thus, the training of the model, we used the Adam optimizer proposed by (Kingma and Ba, 2014) and with the modifications proposed by (Reddi et al., 2019). We initialized the learning rate to  $1e-4$ , the betas [which are used for computing the running averages of gradient and its square (Paszke et al., 2019)] to 0.9 and 0.999, and the weight decay (which is a form of L2 regularization penalty) to 0.01. We used a multistep learning rate scheduler to gradually reduce the learning rate as the training proceeds through 50 equally distributed epoch milestones. At each of the milestones, the new learning rate is obtained by multiplying the current learning rate by 0.99. We used a batch size of 512 and set out to run 5,000 epochs in the initial training of the model. We adopt early stopping if the model does not improve over 250 consecutive epochs, in which case we would retain the last known best model and stop further training of the model.

## DESP: Deep Enhanced Sampling of Proteins' Conformation Space Algorithm

Having described the individual components of the DESP above, we now present the overall DESP algorithm (Figure 3) that combines DNN with MD simulations to achieve enhanced sampling of the conformation space of macromolecules. It begins with the initialization of the total number of MD simulation steps needed (e.g.,  $N_{Needed} = 1e9$ ), the number of MD simulation steps for the initial short MD run ( $N_{Short} = 1e7$ ) that will be used for the initial DNN model training, the total number of steps completed ( $N_{Completed} = 0$ ), the number of steps to run before saving a frame ( $N_{Saving} = 1e4$ ), and the total number of steps completed before updating the biasing potentials ( $N_{Biasing} = 50$ ). While  $N_{Completed}$  essentially ranges from 0 to  $N_{Needed}$  over time, the other variables are relatively as follows:

$$N_{Biasing} \ll N_{Saving} \ll N_{Short} \ll N_{Needed} \quad (4)$$

We use “<<” to signify a difference of one order of magnitude or more.  $N_{\text{Biasing}}$ ,  $N_{\text{Saving}}$ ,  $N_{\text{Short}}$ ,  $N_{\text{Needed}}$ , and  $N_{\text{Completed}}$  are natural numbers.

We run  $N_{\text{Short}}$  steps of unbiased MD simulation, saving frames for every  $N_{\text{Saving}}$  steps. The saved frames are added to a pool of frames (i.e., set  $S_{\text{Frames}}$ ): increase  $N_{\text{Completed}}$  by  $N_{\text{Short}}$  (i.e.,  $N_{\text{Completed}} \leftarrow N_{\text{Completed}} + N_{\text{Short}}$ ). We use the MD simulation's frames in  $S_{\text{Frames}}$  (or its subset, selected randomly) to train the VAE and save the trained VAE ( $\text{VAE}_{\text{Trained}}$ ). While  $N_{\text{Completed}}$  is less than  $N_{\text{Needed}}$ , we continue the biased MD simulations coupled with the usage of the  $\text{VAE}_{\text{Trained}}$  and its further training as follows. 1) Calculate the KL divergence (using latent vectors of the  $\text{VAE}_{\text{Trained}}$ 's means and variance, based on Eq. 5) between the last frame of the MD simulations and the representative/sampled structures from pool  $S_{\text{Frames}}$ . 2) Run the ongoing MD simulation for  $N_{\text{Biasing}}$  steps, but now by adding a biasing potential ( $V_{\text{Biasing}}$ , as defined in Eq. 6) that is based on the KL divergence. Keep track of the  $V_{\text{Biasing}}$ . Increase  $N_{\text{Completed}}$  by  $N_{\text{Biasing}}$  (i.e.,  $N_{\text{Completed}} \leftarrow N_{\text{Completed}} + N_{\text{Biasing}}$ ). 3) For every  $N_{\text{Saving}}$  steps of the MD simulations, add the new frame to pool  $S_{\text{Frames}}$ . And for every  $N_{\text{Saving}} * 100$  steps of the MD simulations (which means that additional 100 new frames would have been added to  $S_{\text{Frames}}$ ), we use the frames in  $S_{\text{Frames}}$  (or its subset, selected randomly) to further train the  $\text{VAE}_{\text{Trained}}$ . When the  $N_{\text{Completed}}$  is equal to  $N_{\text{Needed}}$ , we stop the MD simulations and use the trajectory of  $V_{\text{Biasing}}$  to reweigh the MD simulation trajectory.

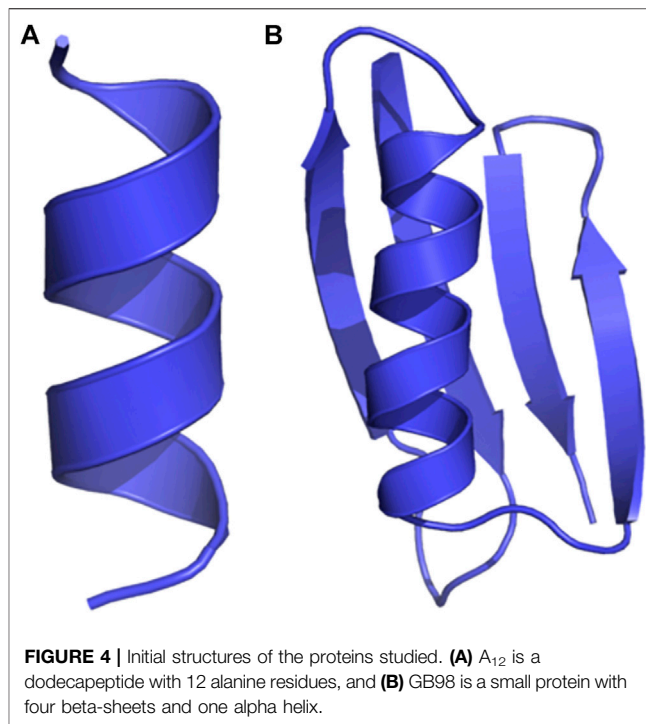
The KL Divergence upon which the biasing potential is based involves pairs of multivariate normal distributions of the same dimension and can be represented by Eq. 5, which denotes the KL divergence of  $\mathcal{N}_1 \sim \mathcal{N}(\mu_1, \Sigma_1)$  from  $\mathcal{N}_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$ .

$$V_{KL} = D_{KL}(\mathcal{N}_0 || \mathcal{N}_1) = \frac{1}{2} \left( \text{tr}(\Sigma_1^{-1}\Sigma_0) + (\mu_1 - \mu_0)^T \Sigma_1^{-1} (\mu_1 - \mu_0) - k + \ln \left( \frac{\det \Sigma_1}{\det \Sigma_0} \right) \right) \quad (5)$$

$$V_{\text{biasing}} = (V_{KL_{\text{upper}}} / V_{KL})^2 \quad (6)$$

where  $V_{KL_{\text{upper}}}$ , which is set to  $1e-5$  in this work, is a weak upper bound of  $V_{KL}$ .  $V_{KL_{\text{upper}}}$  is a settable parameter but can be left at this default value obtained from our preliminary experiments where this provided optimal enhanced sampling without making the system unstable. This value can be tuned up or down to modulate how aggressive (high  $V_{KL_{\text{upper}}}$ ) or conservative (low  $V_{KL_{\text{upper}}}$ ) the enhancement of the sampling should be. The obtained  $V_{\text{Biasing}}$  is added to the potential energy term involving the protein atoms.

At this point, we find it important to further clarify that the use of dihedral angles as input to the VAE in DESP does not mean that dihedral angles are being used directly as the reaction coordinates for biasing the MD simulations. Using all the dihedral angles by themselves would be overwhelming (especially for medium-sized to large-sized molecules) and,



**FIGURE 4 |** Initial structures of the proteins studied. (A) A<sub>12</sub> is a dodecapeptide with 12 alanine residues, and (B) GB98 is a small protein with four beta-sheets and one alpha helix.

more importantly, will not work if used directly even with existing enhanced sampling methods. On the other hand, the VAE learns the compressed representation of the molecular system, and it is the compressed representation (obtainable from the latent space vectors of the VAE, see Figure 2) that is used for achieving the biasing, as presented in the algorithm (see Figure 3). In other words, generally, a bias potential  $V(R)$  used in the MD simulation would depend on  $R$ , the atomistic coordinate, usually through some collective variables. The same is, in principle, true in the current work, except that the bias potential  $V(R)$  used in DESP depends on  $R'$ , where  $R'$  is a compressed representation of  $R$  that is obtained from the DNN.

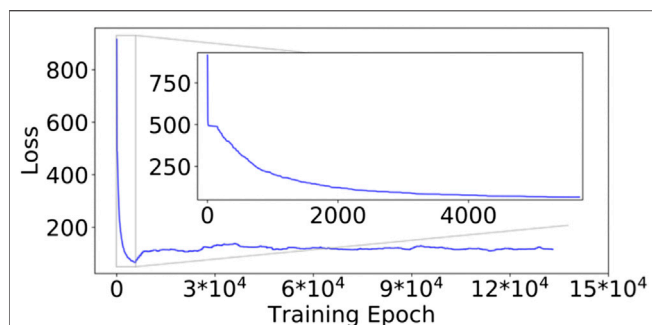
## Reweighting of the Probability Distribution

The probability,  $p'$  (RC), along a reaction coordinate of interest, RC ( $r$ ), where  $r$  represents the atomic coordinates  $r_1^3, \dots, r_n^3$ , based on the biased MD simulations can be reweighed using  $V_{\text{Biasing}}$  to obtain the un-normalized probability distribution,  $p$  (RC), of the canonical ensemble (Sinko et al., 2013; Miao et al., 2015; Salawu, 2018a) as shown in Eq. 7. And the reweighed free energy change can be obtained from Eq. 8.

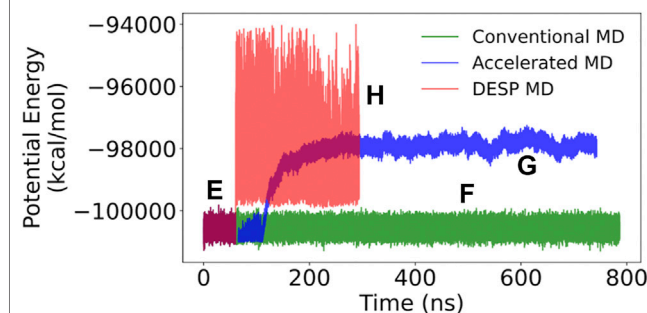
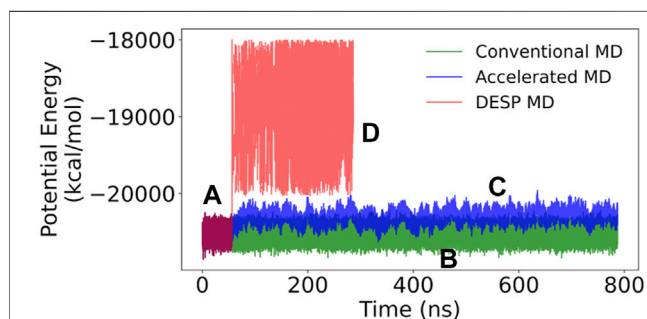
$$p(RC_a) = p'(RC_a) * \frac{e_a^{\beta V_{\text{biasing}}}}{\sum_{a=1}^M e_a^{\beta V_{\text{biasing}}}} \text{ for } a = 1, \dots, M \quad (7)$$

where  $\beta = -\frac{1}{k_B T}$ .

$$F(RC_a) = \beta \ln p(RC_a) \quad (8)$$



**FIGURE 5** | Model loss values during the DESP for the GB98 molecular system. The loss decreased steadily in the first segment of the DESP (see the inset). The model's loss is slightly higher in the subsequent training of the model because the model was exposed to a more diverse molecular structure. The trajectory of the loss for the  $A_{12}$  molecular system is similar in the overall structure/trend to that of the GB98 molecular system and is not shown here for brevity. DESP systematically modifies the molecular system's energy surface.

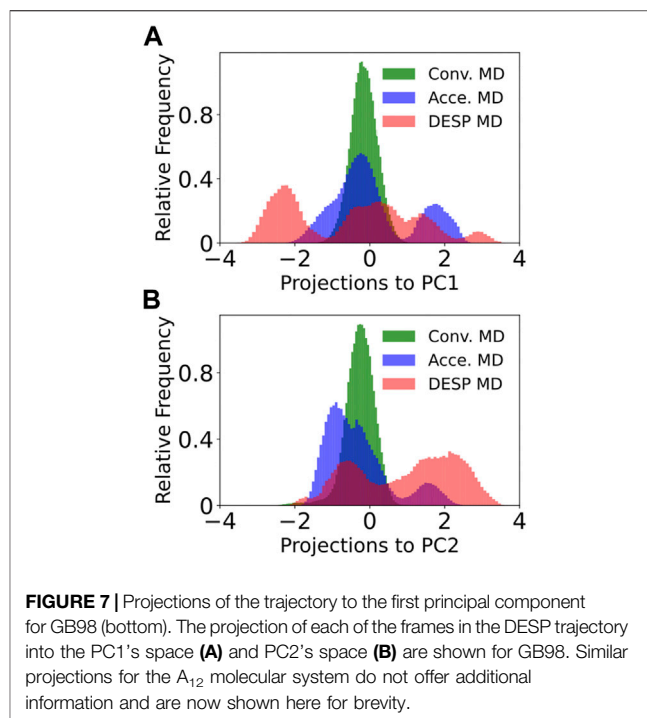


**FIGURE 6** | The potential energy for  $A_{12}$  (top) and GB98 (bottom) molecular systems. **(A and E)** The initial stages of DESP (as well as the initial stages of accelerated) MD simulations are identical to conventional MD simulations and have identical systems' potential energies. **(B and F)** The trajectories of the potential energy for the conventional MD simulations are shown in green; **(C and G)** those for the accelerated MD simulations are shown in blue; while **(D and H)** those for the DESP MD simulations are shown in red.

## RESULTS AND DISCUSSION

### GB98 is a Small Protein With One $\alpha$ - and Four $\beta$ -Folds, While $A_{12}$ is a Dodecalanine

We show the initial 3D structure of the studied molecules,  $A_{12}$  and GB98, in **Figures 4A,B**, respectively.  $A_{12}$  is a peptide with 12



**FIGURE 7** | Projections of the trajectory to the first principal component for GB98 (bottom). The projection of each of the frames in the DESP trajectory into the PC1's space **(A)** and PC2's space **(B)** are shown for GB98. Similar projections for the  $A_{12}$  molecular system do not offer additional information and are now shown here for brevity.

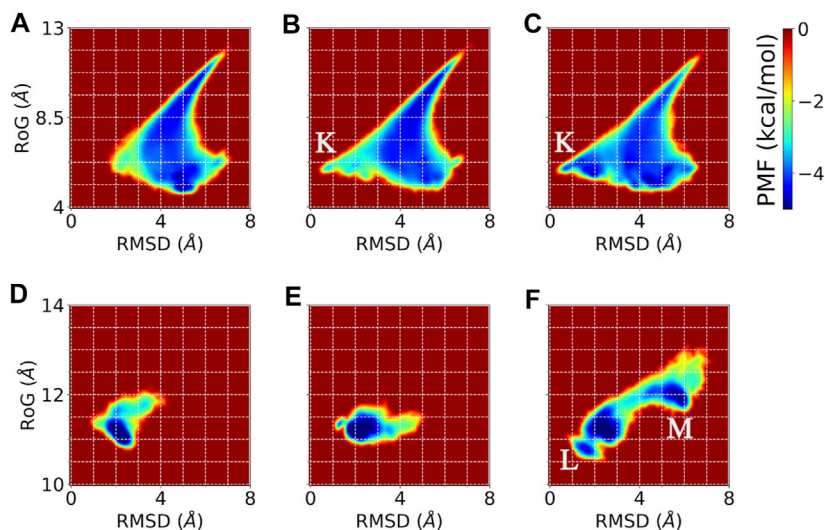
alanine amino acids, while GB98 is a small protein with four beta-sheets, and one alpha-helix (He et al., 2012; Salawu, 2016). These small-sized and medium-sized molecules helped in illustrating the capabilities of DESP.

### DNN Model Loss During the DESP

The initial training of the DESP's VAE started with a high model total loss ( $Loss_{Model}$ , Eq. 1) of approximately 915.71 (**Figure 5**, for the GB98 molecular system), which decreased steadily as the model continued to learn the compressed representation of the molecule under study (inset of **Figure 5**). The initial model training was stopped when the  $Loss_{Model}$  reached 66.65 after 5,664 epochs and would not further decrease for the next 250 epochs. The  $Loss_{Model}$  during the subsequent training of the DNN alongside the DNN-biased MD simulations (using the MD simulation's newly generated molecular structures) is shown in the rest of **Figure 5** from epoch 5,664 to the end.

The reader would notice that the  $Loss_{Model}$  obtained during the subsequent training of the model alongside the DNN-biased MD simulations is slightly higher than the smallest  $Loss_{Model}$  obtained in the initial model training. This is interesting and understandable because the initial training of the DESP's DNNs was done using only the structures/conformations of the molecule obtained from conventional MD simulations in the first segment of the DESP (**Figure 6A**), while the subsequent training of the DNN was done using the more structurally diverse conformations of the molecule obtained during the biasing segment of the DESP (**Figure 6B**).

The initial stages of DESP (as well as the initial stage of the accelerated MD) simulations are identical to those of



**FIGURE 8** | Potentials of mean force (PMF) showing the distribution of the sampled conformations by conventional MD, accelerated MD, and DESP MD simulations of  $A_{12}$  (top) and GB98 (bottom) molecular systems. Reweighting has been done wherever necessary. Comparison using PMF based on physically meaningful collective variables, namely, the root mean square deviation from a known experimental/initial structure (RMSD) and the radius of gyration (RoG) are shown for conventional (**A and D**), accelerated (**B and E**), and DESP (**C and F**) MD simulations for the  $A_{12}$  (top/A, B, C) and the GB98 (bottom/D, E, F) molecular systems. Overall, one would notice that the rightmost panels (**C and D**) show wider and more diverse regions visited by the molecular system, which means that the DESP can explore more conformation spaces than either the conventional (**A and B**) or the accelerated (**B and E**) MD simulations for these collective variables. The regions with stable conformations sampled by both DESP and accelerated MD simulations but not sampled by the conventional MD simulations are marked with “K,” while the regions sampled by DESP alone but not sampled by either the conventional or the accelerated MD simulations are marked with “L” and “M.”

conventional MD simulations, and the molecular systems are observed to have identical systems' energy surfaces/distributions as conventional MD simulations. Indeed, in the current work, for a given molecular system, after equilibration, the ongoing conventional MD simulation is forked/copied into three: one for continuation as a conventional MD simulation, one for continuation as an accelerated MD simulation, and one for continuation as a DESP MD simulation. At the start of the biasing phase of DESP, we observed that the molecular systems' energies are modified and the potential energies increase and change based on the conformation being sampled (Figures 6A–D for the  $A_{12}$  molecular system, and Figures 6E–H for the GB98 molecular system; not drawn to the same scale for all the MD simulations). The modification of the systems' potential energies makes it possible for the system to escape possible energy barriers, thereby encouraging the sampling of wider conformation spaces (Figure 7; Figure 8).

### DESP Efficiently Samples a Wider Range of a Molecule's Conformation Space Than Both Conventional and Accelerated MD Simulations

To compare the conformation spaces sampled by DESP to that sampled by conventional and accelerated MD simulations, we carried out dihedral principal components analysis (dPCA) on the molecule's dihedral angles (namely, phi, psi, omega, and  $\chi_1$ ) by making use of both the cosine and the sine of each of the

dihedral angles (Mu et al., 2005) and projected each of the sampled structures from the DESP and from both the conventional and accelerated MD simulations into the principal components' (PC) space. A visualization of the trajectory in the PC space (see Figure 7 for the first two PCs, PC1 and PC2) shows that DESP samples a wider range of the molecule's conformation spaces than the conventional and the accelerated MD simulations (Figure 7) despite that the DESP is just about one-third as long (i.e., ~280 ns) as the conventional and the accelerated MD simulations (i.e., ~800 ns, Figure 6; Figure 7).

It is worthy of note that the distributions shown in Figure 7 are unweighed and cannot be strictly interpreted in the probability sense most especially for the DESP and for the accelerated MD simulations that involve the use of biasing potentials. It is, therefore, important to reweigh any DESP-obtained (or accelerated-MD-obtained) distribution while considering the biasing potentials (Salawu, 2018a; Sinko et al., 2013; Miao et al., 2015). Such reweighing can be achieved through Eqs 7, 8 or as described in previous publications (Sinko et al., 2013; Miao et al., 2015; Salawu, 2018a).

For the potentials of mean force (PMF) obtainable through the reweighing of the trajectories, we use two physically interpretable/physically meaningful reaction coordinates, namely, the molecule's radius of gyration (RoG) and the molecule's root mean square deviation (RMSD) from the experimentally solved structure (i.e., the NMR structure in the case of GB98) or the initial structure (i.e., the energy minimized modeled structure in the case of  $A_{12}$ ). The PMF obtained from the reweighed trajectory (Figure 8) further establishes that DESP



samples have much wider conformation spaces than the conventional and the accelerated MD simulations. We show the PMF obtained from collective variables (CVs) defined by the combination of the RMSD and the RoG in **Figure 8**.

From the energy landscape one sees regions with stable conformations that are sampled by DESP but are not sampled by the conventional MD simulations (see **Figures 8A–C** for the  $A_{12}$  molecular system, and **Figures 8D–F** for the GB98 molecular system). For the sake of illustration, we mark the regions sampled by both DESP and accelerated MD simulations but not sampled by the conventional MD simulations with “K” (**Figures 8B,C**), and we mark the regions sampled by DESP alone but not sampled by either the conventional MD simulations or the accelerated MD simulations with “L” and “M” (**Figure 8F**). Overall, one would notice that the rightmost panels (C, D) show wider and more diverse regions visited by the molecular system, which means that the DESP can explore more conformation spaces than either the conventional (A, B) or the accelerated (B, E) MD simulations, to the extent of capturing a few global but moderate unfolding and refolding events. The comparison of the energy landscapes shows that while DESP shows a moderately better sampling of a wider range of conformation spaces than both the conventional and the accelerated MD simulations for a small molecular system (namely,  $A_{12}$ , **Figures 8A–C**), the superiority of the sampling efficiency of DESP is more remarkably evident for larger molecules as shown by the medium-sized GB98 molecular system wherein DESP samples much wider regions/conformations spaces than both the conventional and the accelerated MD simulations (**Figures 8D–F**). This is desirable because it is with the larger molecules that highly efficient conformation space sampling methods, such as DESP, are most needed.

## REFERENCES

- Alland, C., Moreews, F., Boens, D., Carpentier, M., Chiusa, S., Lonquety, M., et al. (2005). RPBS: a web resource for structural bioinformatics. *Nucleic Acids Res.* 33, W44. doi:10.1093/nar/gki477
- Allison, J. R. (2020). Computational methods for exploring protein conformations. *Biochem. Soc. Trans.* 48 (4), 1707–1724. doi:10.1042/bst20200193
- Åqvist, J., Wennerström, P., Nervall, M., Bjelic, S., and Brandsdal, B. O. (2004). Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chem. Phys. Lett.* 384 (4–6), 288–294. doi:10.1016/j.cplett.2003.12.039
- Babin, V., Roland, C., and Sagui, C. (2008). Adaptively biased molecular dynamics for free energy calculations. *J. Chem. Phys.* 128 (13), 134101. doi:10.1063/1.2844595
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Res.* 28 (1), 235–242. doi:10.1093/nar/28.1.235
- Bonati, L., Zhang, Y. Y., and Parrinello, M. (2019). Neural networks-based variationally enhanced sampling. *Proc. Natl. Acad. Sci. U.S.A.* 116 (36), 17641–17647. doi:10.1073/pnas.1907975116
- Bussi, G., and Laio, A. (2020). Using metadynamics to explore complex free-energy landscapes. *Nat. Rev. Phys.* 2, 1–13. doi:10.1038/s42254-020-0153-0
- Carlson, H. A., and McCammon, J. A. (2000). Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* 57 (2), 213–218.
- Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., et al. (2005). The amber biomolecular simulation programs. *J. Comput. Chem.* 26 (16), 1668–1688. doi:10.1002/jcc.20290

## CONCLUSION

In this work, 1) It has been shown, with computational experiments and pieces of evidence obtained therefrom, that it is possible to enhance the MD simulation sampling of molecules' conformation spaces using deep learning techniques (VAE in the current case). 2) It has been shown one of the possible ways with which it could be achieved, namely, by biasing the MD simulations based on the VAE's latent space vectors. 3) The use of the KL divergence of the DNN-learned latent space vectors of the most recently sampled conformation from the previously sampled conformations made it possible to bias the MD away from visiting already sampled conformations, and thereby encouraging the sampling of previously unsampled states. 4) It should be noted that the ideas in DESP are generalizable and may be used for implementing other forms of biased MD simulations, including targeted and steered MD simulations, and we explore these in our subsequent articles.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

- Chen, W., and Ferguson, A. L. (2018). Molecular enhanced sampling with autoencoders: on-the-fly collective variable discovery and accelerated free energy landscape exploration. *J. Comput. Chem.* 39 (25), 2079–2102. doi:10.1002/jcc.25520
- Chiti, F., and Dobson, C. M. (2017). Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. *Annu. Rev. Biochem.* 86, 27–68. doi:10.1146/annurev-biochem-061516-045115
- Chow, K. H., and Ferguson, D. M. (1995). Isothermal-isobaric molecular dynamics simulations with Monte Carlo volume sampling. *Comput. Phys. Commun.* 91 (1–3), 283–289. doi:10.1016/0010-4655(95)00059-o
- Council, National Research, and others (1989). Opportunities in biology. National academies. Available at: <https://www.nap.edu/read/742/chapter/3> (Accessed February 22, 2021).
- Cukier, R. I. (2015). Dihedral angle entropy measures for intrinsically disordered proteins. *J. Phys. Chem. B* 119 (9), 3621–3634. doi:10.1021/jp5102412
- Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: AnN-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98 (12), 10089–10092. doi:10.1063/1.464397
- Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., et al. (2007). PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* 35, W522. doi:10.1093/nar/gkm276
- Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Beauchamp, K. A., Wang, L.-P., et al. (2017). OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *Plos Comput. Biol.* 13 (7), e1005659. doi:10.1371/journal.pcbi.1005659
- Guo, T., Noble, W., and Hanger, D. P. (2017). Roles of tau protein in health and disease. *Acta Neuropathol.* 133 (5), 665–704. doi:10.1007/s00401-017-1707-9

- Hamelberg, D., Mongan, J., and McCammon, J. A. (2004). Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* 120 (24), 11919–11929. doi:10.1063/1.1755656
- Harada, R., and Kitao, A. (2013). Parallel cascade selection molecular dynamics to generate conformational transition pathway. *J. Chem. Phys.* 139, 035103. doi:10.1063/1.4813023
- Hartl, F. U. (2017). Protein misfolding diseases. *Annu. Rev. Biochem.* 86, 21–26. doi:10.1146/annurev-biochem-061516-044518
- He, Y., Chen, Y., Alexander, P. A., Bryan, P. N., and Orban, J. (2012). Mutational tipping points for switching protein folds and functions. *Structure* 20 (2), 283–291. doi:10.1016/j.str.2011.11.018
- Hénin, J., and Chipot, C. (2004). Overcoming free energy barriers using unconstrained molecular dynamics simulations. *J. Chem. Phys.* 121 (7), 2904–2914. doi:10.1063/1.1773132
- Ittisoponpisan, S., Khanna, T., Alhuzimi, E., David, A., and Sternberg, M. J. E. (2019). Can predicted protein 3D structures provide reliable insights into whether missense variants are disease associated? *J. Mol. Biol.* 431 (11), 2197–2212. doi:10.1016/j.jmb.2019.04.009
- Jurru, E., Engel, D. F., Star, K. B., Monson, K., Brandi, J., Felberg, L. E., et al. (2018). Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* 27 (1), 112–128. doi:10.1002/pro.3280
- Karplus, M., and Kuriyan, J. (2005). Molecular dynamics and protein function. *Proc. Natl. Acad. Sci.* 102 (19), 6679–6685. doi:10.1073/pnas.0408930102
- Kingma, D. P., and Ba, J. (2014). A method for stochastic optimization. ArXiv [Preprint]. Available at: <http://arXiv:1412.6980> (Accessed December 22, 2014).
- Laio, A., and Gervasio, F. L. (2008). Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* 71 (12), 126601. doi:10.1088/0034-4885/71/12/126601
- Laskowski, R. A., Stephenson, J. D., Sillitoe, I., Orengo, C. A., and Thornton, J. M. (2020). VarSite: disease variants and protein structure. *Protein Sci.* 29 (1), 111–119. doi:10.1002/pro.3746
- Lee, Y., Basith, S., and Choi, S. (2018). Recent advances in structure-based drug design targeting class A G protein-coupled receptors utilizing crystal structures and computational simulations. *J. Med. Chem.* 61 (1), 1–46. doi:10.1021/acs.jmedchem.6b01453
- Lemke, T., and Peter, C. (2019). EncoderMap: dimensionality reduction and generation of molecule conformations. *J. Chem. Theor. Comput.* 15 (2), 1209–1215. doi:10.1021/acs.jctc.8b00975
- Li, P., and Merz, K. M., Jr (2014). Taking into account the ion-induced dipole interaction in the nonbonded model of ions. *J. Chem. Theor. Comput.* 10 (1), 289–297. doi:10.1021/ct400751u
- Lin, S.-M., Lin, S.-C., Hsu, J.-N., Chang, C.-K., Chien, C.-M., Wang, Y.-S., et al. (2020). Structure-based stabilization of non-native protein-protein interactions of coronavirus nucleocapsid proteins in antiviral drug design. *J. Med. Chem.* 63 (6), 3131–3141. doi:10.1021/acs.jmedchem.9b01913
- Maier, J. A., Martinez, C., Kasavajhala, K., WickstromHauser, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theor. Comput.* 11 (8), 3696–3713. doi:10.1021/acs.jctc.5b00255
- McCafferty, C. L., and Sergeev, Y. V. (2016). Silico mapping of protein unfolding mutations for inherited disease. *Sci. Rep.* 6 (1), 1–12. doi:10.1038/srep37298
- McCarty, J., and Parrinello, M. (2017). A variational conformational dynamics approach to the selection of collective variables in metadynamics. *J. Chem. Phys.* 147 (20), 204109. doi:10.1063/1.4998598
- Miao, Y., Feher, V. A., and McCammon, J. A. (2015). Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation. *J. Chem. Theor. Comput.* 11 (8), 3584–3595. doi:10.1021/acs.jctc.5b00436
- Morisugu, K., Terada, T., and Kidera, A. (2012). Disorder-to-order transition of an intrinsically disordered region of sortase revealed by multiscale enhanced sampling. *J. Am. Chem. Soc.* 134 (16), 7094–7101. doi:10.1021/ja3008402
- Mu, Y., Nguyen, P. H., and Stock, G. (2005). Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins* 58 (1), 45–52. doi:10.1002/prot.20310
- Ostermeir, K., and Zacharias, M. (2014). Hamiltonian replica-exchange simulations with adaptive biasing of peptide backbone and side chain dihedral angles. *J. Comput. Chem.* 35 (2), 150–158. doi:10.1002/jcc.23476
- Pastor, R. W., Brooks, B. R., and Szabo, A. (1988). An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.* 65 (6), 1409–1419. doi:10.1080/00268978800101881
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). Pytorch: an imperative style, high-performance deep learning library. ArXiv [Preprint]. Available at: <http://arXiv:1912.01703> (Accessed December 3, 2019).
- Paweland, S., and Caflisch, A. (2018). Protein structure-based drug design: from docking to molecular dynamics. *Curr. Opin. Struct. Biol.* 48, 93–102. doi:10.1016/j.sbi.2017.10.010
- Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., DeBolt, S., et al. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* 91 (1), 1–41. doi:10.1016/0010-4655(95)00041-d
- Reddi, S. J., Kale, S., and Kumar, S. (2019). On the convergence of Adam and beyond. ArXiv [Preprint]. Available at: <http://arXiv:1904.09237> (Accessed April 19, 2019).
- Ribeiro, J. M. L., Bravo, P., Wang, Y., and Tiwary, P. (2018). Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). *J. Chem. Phys.* 149 (7), 72301. doi:10.1063/1.5025487
- Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23 (3), 327–341. doi:10.1016/0021-9991(77)90098-5
- Salawu, E. O. (2020). Enhanced sampling of nucleic acids' structures using deep-learning-derived biasing forces. *IEEE Symp. Ser. Comput. Intel.* 11, 1648–1654. doi:10.1109/SSCI47803.2020.9308559
- Salawu, E. O. (2018b). In silico study reveals how E64 approaches, binds to, and inhibits falcipain-2 of *Plasmodium falciparum* that causes malaria in humans. *Sci. Rep.* 8 (1), 16380. doi:10.1038/s41598-018-34622-1
- Salawu, E. O. (2016). Random forests secondary structure assignment for coarse-grained and all-atom protein systems. *Cogent Biol.* 2 (1), 1214061. doi:10.1080/23312025.2016.1214061
- Salawu, E. O. (2018a). The impairment of TorsinA's binding to and interactions with its activator: an atomistic molecular dynamics study of primary dystonia. *Front. Mol. Biosci.* 5, 64. doi:10.3389/fmolb.2018.00064
- Salomon-Ferrer, R., Case, D. A., and Walker, R. C. (2013a). An overview of the Amber biomolecular simulation package. *Wires Comput. Mol. Sci.* 3 (2), 198–210. doi:10.1002/wcms.1121
- Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013b). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. explicit solvent particle Mesh Ewald. *J. Chem. Theor. Comput.* 9 (9), 3878–3888. doi:10.1021/ct400314y
- Salvador, P. (2014). "Dependencies of J-couplings upon dihedral angles on proteins," in *Annual reports on NMR spectroscopy*. Amsterdam, Netherlands: Elsevier, 185–227.
- Shaw, D. E., Deneroff, M. M., Dror, R. O., Kuskin, J. S., Larson, R. H., Salmon, J. K., et al. (2008). Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM* 51 (7), 91–97. doi:10.1145/1364782.1364802
- Shaw, D. E., Dror, R. O., Salmon, J. K., Grossman, J. P., Mackenzie, K. M., Joseph, A., et al. (2009). Millisecond-scale molecular dynamics simulations on anton. *Proc. Conf. High Perform. Comput. Netw. Storage Anal.* 65, 1–11. doi:10.1145/1654059.1654126
- Sidky, H., Chen, W., and Ferguson, A. L. (2020). Machine learning for collective variable discovery and enhanced sampling in biomolecular simulation. *Mol. Phys.* 118 (5), e1737742. doi:10.1080/00268976.2020.1737742
- Sinko, W., Miao, Y., de Oliveira, C. A. F., and McCammon, J. A. (2013). Population based reweighting of scaled molecular dynamics. *J. Phys. Chem. B* 117 (42), 12759–12768. doi:10.1021/jp401587e
- Sugita, Y., and Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314 (1–2), 141–151. doi:10.1016/s0009-2614(99)01123-9

- Tramutola, A., Lanzillotta, C., Perluigi, M., and Butterfield, D. A. (2017). Oxidative stress, protein modification and Alzheimer disease. *Brain Res. Bull.* 133, 88–96. doi:10.1016/j.brainresbull.2016.06.005
- Tribello, G. A., Ceriotti, M., and Parrinello, M. (2010). A self-learning algorithm for biased molecular dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 107 (41), 17509–17514. doi:10.1073/pnas.1011511107
- Wang, X., Song, K., Li, L., and Chen, L. (2018). Structure-based drug design strategies and challenges. *Curr. Top. Med. Chem.* 18 (12), 998–1006. doi:10.2174/1568026618666180813152921
- Wehmeyer, C., and Noé, F. (2018). Time-lagged autoencoders: deep learning of slow collective variables for molecular kinetics. *J.Chem. Phys.* 148 (24), 241703. doi:10.1063/1.5011399
- Yang, L.-Q., Sang, P., Tao, Y., Fu, Y.-X., Zhang, K.-Q., Xie, Y.-H., et al. (2014). Protein dynamics and motions in relation to their functions: several case studies and the underlying mechanisms. *J. Biomol. Struct. Dyn.* 32 (3), 372–393. doi:10.1080/07391102.2013.770372
- Conflict of Interest:** Author EOS is employed by the company Amazon Web Services (AWS). There are no other authors. The work and substantial parts of the manuscript were completed as an independent research scientist before joining AWS.

Copyright © 2021 Salawu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.