



Machine learning approaches for analyzing and enhancing molecular dynamics simulations

Yihang Wang¹, João Marcelo Lamim Ribeiro² and Pratyush Tiwary³

Molecular dynamics (MD) has become a powerful tool for studying biophysical systems, due to increasing computational power and availability of software. Although MD has made many contributions to better understanding these complex biophysical systems, there remain methodological difficulties to be surmounted. First, how to make the deluge of data generated in running even a microsecond long MD simulation human comprehensible. Second, how to efficiently sample the underlying free energy surface and kinetics. In this short perspective, we summarize machine learning based ideas that are solving both of these limitations, with a focus on their key theoretical underpinnings and remaining challenges.

Addresses

¹ Biophysics Program and Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA

² Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1677, New York, NY 10029, USA

³ Department of Chemistry and Biochemistry and Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA

Corresponding author: Tiwary, Pratyush (ptiwary@umd.edu)

Current Opinion in Structural Biology 2020, 61:139–145

This review comes from a themed issue on **Theory and simulation**

Edited by **Alan E Mark** and **Christine Peter**

For a complete overview see the [Issue](#) and the [Editorial](#)

Available online 20th January 2020

<https://doi.org/10.1016/j.sbi.2019.12.016>

0959-440X/© 2020 Elsevier Ltd. All rights reserved.

Introduction

With the ever-increasing power and availability of computing resources and accurate interaction models, molecular dynamics (MD) simulations have become an indispensable tool for the study of biophysical systems. MD has allowed us to probe, with high spatiotemporal resolution, complex processes such as protein folding, ligand dissociation and countless others. Not dissimilar to many other fields in science and engineering, here as well we have had an explosion of data, easily reaching hundreds of gigabytes for a standard microsecond long MD simulation of a protein. This immediately leads to two pressing questions. First, what do we do with this much data — how do we store it and

how do we make sense of it? And if a microsecond long simulation of a single solvated protein can generate this much data, what would happen if we tried to simulate an entire cell for similar duration? Second, microsecond long simulations can take weeks, what will it take to reach timescales actually relevant to various biological processes such as ligand dissociation and slow conformational exchanges, namely milliseconds and beyond? Thus MD suffers from two, at first glance opposing but deeply connected problems — enormous amounts of data which can be difficult to analyze, and yet, the inability to generate data at timescales that we actually care about when it comes to interpreting or designing laboratory experiments.

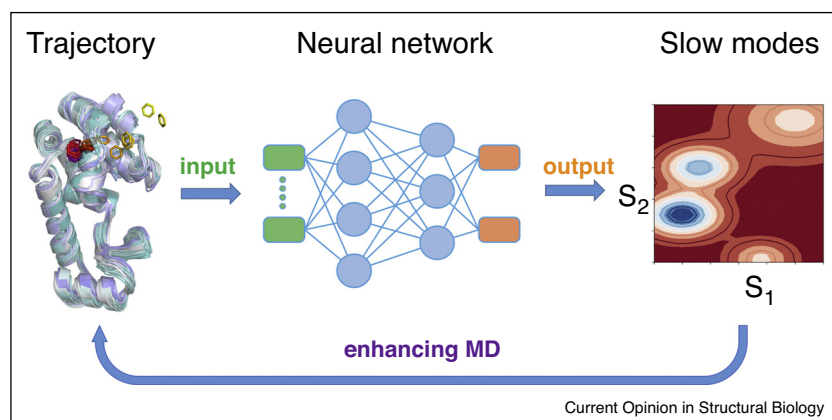
The purpose of this review is to summarize how we are starting to see breakthroughs in the endeavor of surmounting both these problems by using ideas from machine learning (ML), motivating how the above two problems are actually two sides of the same coin. By ML here we mean methods incorporating artificial neural networks (ANN). Most of these methods follow some variant of the scheme shown in [Figure 1](#). That is, ML is used to find a projection from the high dimensional structure space to a low dimensional feature space. We focus on theoretical underpinnings and do not discuss applications. In addition, ML approaches to designing force-field [1–3] itself will also not be covered. We conclude with caution highlighting that albeit the future ahead is exciting, much work remains.

Underlying terms and constructs

We are interested in a generic atomic/molecular system comprising N atoms, where the $3N$ position coordinates are collectively denoted by \mathbf{x} , which in this review we call a microstate. In classical MD, the N atoms execute classical dynamics at a temperature T (or inverse temperature β) under the influence of an interaction potential $U(\mathbf{x})$. Further, we are typically interested in cases relevant to biophysical systems characterized by the existence of metastable states where the system spends extended amounts of time, infrequently moving between any two such states. Typical objective in performing MD is to evaluate the following two broad types of quantities for some generic low-dimensional collective variable (CV) $s(\mathbf{x})$:

- 1 equilibrium properties, such as the free energy $F(s)$, where $P(s) \equiv e^{-\beta F(s)} = \int d\mathbf{x} \delta[s - s(\mathbf{x})] P_0(\mathbf{x})$, where $P_0(\mathbf{x}) \propto e^{-\beta U(\mathbf{x})}$ is the equilibrium probability of a microstate \mathbf{x} , and $P(s)$ is the probability of the CV.

Figure 1



A schematic illustrating the typical workflow of some of the methods that use machine learning to analyze and enhance MD simulation. High dimensional data which describes the time evolution of the system in configuration space is used as the input of an ANN. The ANN is trained to project the input to a low dimensional space. Depending on the structure of NN and objective function, the low dimensional representation captures different features that are considered to be important, such as slow modes. In some methods, the feature learnt are used to further enhanced the sampling of MD simulation as shown by the arrow in the bottom.

2 dynamic properties, such as the mean first passage time for escaping any metastable state in \mathbf{x} -space.

The dynamics in the full high-dimensional \mathbf{x} -space can be safely assumed to be Markovian, and thus the time-propagation of the system in \mathbf{x} -space can be described for a given lag time τ using a transfer operator [4] K with eigenvalues $1 = \lambda_0 > \lambda_1 \geq \dots$ and corresponding eigenvectors $\psi_0(\mathbf{x}) = 1, \psi_1(\mathbf{x}), \psi_2(\mathbf{x}), \dots$. The eigenvectors with indices $i \geq 1$ correspond to the slow modes of the system with corresponding timescales $t_i = -\frac{\tau}{\log \lambda_i}$. A central objective of the methods in this review is to learn the so-called low-dimensional *reaction coordinate* (RC) for the high-dimensional dynamics. Loosely speaking, this RC can be defined as a special CV which encapsulates slow modes of the high-dimensional transfer operator such that the modes not captured are either fast or irrelevant. Although in principle one could construct the RC as a function of Cartesian coordinates, in practice it is more efficient to construct it as a function of low-dimensional molecular descriptors often called *order parameters* (OP). These OPs are system observables that can be defined based on some pre-knowledge of the system or using methods like AMINO [5]. In this review, OPs will be denoted as \mathbf{v} .

Using ML for analyzing MD trajectories

ML has been used in various forms to analyze long MD trajectories and learn relevant slow modes. Ma and Dinner in 2005 used ANNs for constructing a RC [6], the idea being to sample configurations along transition pathways and tabulate an ensemble of committor values, with ANNs trained to fit CVs to these values. Recent

work has revived and somewhat generalized this framework [7].

While a committor might be the most rigorous description of a slow mode, there are other reasonably accurate and computationally cheaper principles defining what constitutes slow modes. One such formalism is the variational approach to conformation dynamics (VAC) [8^{••}]. VAC calculates the eigenvalues and eigenvectors of the transfer operator starting with ψ_1 , by solving a variational principle which states that any trial eigenvector ψ'_1 will have a time-lagged autocorrelation $\langle \psi'_1 | K | \psi'_1 \rangle \leq 1$, as long as it is orthonormal to ψ_0 , with equality holding iff $\psi'_1 = \psi_1$. Thus we can approximate the first nontrivial eigenfunction ψ_1 by searching for a ψ'_1 that maximizes its autocorrelation function subject to the orthonormal conditions. Refs. [8^{••},9] show how to calculate such matrix elements given unbiased MD data. Further modes can be learned in a similar manner but with more orthonormality conditions. Methods such as TICA [10] implement VAC directly by learning slow eigenvectors as linear combinations of pre-selected basis functions, and are at the heart of building MSMs. The Variational approach for Markov processes (VAMP) principle [11^{••}] generalizes the mathematical framework in the VAC principle to nonstationary and nonreversible processes, while keeping the same key underlying intent of learning a transformation of \mathbf{x} in which the dynamics is as Markovian as possible. VAMPnets use ANNs to implement the VAMP principle and automate the various steps involved in construction of a MSM. Even more recent methods such as state-free reversible VAMPnets (SRV) combine strengths of VAC and VAMPnets for equilibrium systems, and use ANNs to construct a full hierarchy

of slow modes expressible as non-linear functions of input coordinates.

Slow modes s as per VAC may also be equivalently defined as a mapping $s(\mathbf{x})$ that maximizes autocorrelation $A(s)$ for a lag time τ :

$$A(s) = \frac{\mathbb{E}[\tilde{s}(\mathbf{x}_t)\tilde{s}(\mathbf{x}_{t+\tau})]}{\sigma(s(\mathbf{x}_t))\sigma(s(\mathbf{x}_{t+\tau}))} \quad (1)$$

where $\tilde{s} = s - \mathbb{E}(s)$ is the mean-free latent variable and $\sigma(s)$ is the standard deviation of s . Time-lagged autoencoders (TAE) use an encoder-decoder framework to find the slow component so defined. However, instead of calculating the precise expectation value in Equation 1, TAE approximates the slow mode by minimizing the reconstruction loss L_R :

$$L_R \equiv L_{TAE} = \sum_t \|\mathbf{x}_{t+\tau} - D(E(\mathbf{x}_t))\|^2 \quad (2)$$

Here E is the encoder which maps the input configuration \mathbf{x} to low-dimensional s and D is the decoder mapping it back to coordinate space. Maximizing Equation 1 and minimizing Equation 2 are identical if the encoder and decoder are linear [12*].

Variational dynamic encoder (VDE) [13**] also uses an encoder-decoder pair with time-lagged data, but instead of having only a term related to reconstruction error L_R , VDE includes two additional terms. Firstly, the latent variable s_t is not directly used as an input of the decoder to predict $\mathbf{x}_{t+\tau}$. Instead, a sample s'_t is drawn from a Gaussian distribution $N(\mu(\mathbf{x}_t), \sigma^2(\mathbf{x}_t))$, with mean $\mu(\mathbf{x}_t) = E(\mathbf{x}_t)$ as in TAE and the variance $\sigma^2(\mathbf{x}_t)$ learnt through another ANN. This forces the decoder to be tolerant to small variances of signals from latent space, thus increasing model generalizability. The use of a Gaussian prior also helps the distribution of s'_t to be smooth, allowing meaningful interpolation between states in latent space. This is done by introducing the Kullback-Leibler divergence loss:

$$L_{KL} = \mathbb{E} \left[\frac{1 + \log \sigma(\mathbf{x}_t)^2 - \mu(\mathbf{x}_t)^2 - \sigma(\mathbf{x}_t)^2}{2} \right] \quad (3)$$

Secondly, an autocorrelation loss $L_{AC} = -A(s)$ is introduced. It encourages the learning of modes with high autocorrelation and makes the training process easier to converge. The VDE objective function is then a sum of these three loss terms. Recent work has included a probabilistic framework similar in spirit to VDE but within a full Bayesian approach [14].

The EncoderMap approach of Lemke and Peter is another method that makes use of a NN encoder-decoder architecture [15]. In addition to a reconstruction loss analogous to Equation 2 measuring the distance between input configurations \mathbf{x}_t and their reconstructions $D(E(\mathbf{x}_t))$, the NN loss function used for training includes an additional term meant to force s to be interpretable. This additional term is the sketch-map cost function [16], which aims to ‘focus’ the network’s low-dimensional central bottleneck on learning a CV capable of reproducing distances between adjacent metastable basins as opposed to intra-basin and basins separated from each other. In addition, a recent ML approach [17] has been proposed that does not leverage ANNs like most of the above approaches. Instead, it combines the XGBoost algorithm together with an exclusion loop to determine a set of essential internal coordinates.

In recent work [18], Olsson and Noé have extended the notion of encoding a global molecular configuration, \mathbf{x} , into encoding several local configurations $\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^j$, each representing a partition of the original global molecular structure into a local substructure. To model the time evolution of \mathbf{x} the propagator or conditional distribution $p(\mathbf{x}_t|\mathbf{x}_{t-\tau})$ is written in terms of the substructures:

$$p(\mathbf{x}_t|\mathbf{x}_{t-\tau}) \propto e^{\sum_i \mathbf{x}_i^j (\sum_j J_{ij}(\tau) \mathbf{x}_{t-\tau}^j + h_i(\tau))} \quad (4)$$

where J_{ij} is the coupling parameter between the i th and j th subsystem and h_i describes the coupling between the i th subsystem with an external field. With the choice of this model, the problem of determining the coupling parameters can then be reduced to N logistic regression problems. A notable feature of this approach is that it seems capable of predicting molecular configurations that have not been incorporated into the training data. Prediction of unsampled but likely configurations along with a quantitative measure of their relative likelihoods is exactly the theme of our next section.

Using ML and related data-driven approaches to enhance sampling

In this section, we review approaches that use ML to not just analyze existing MD generated structures and trajectories, but also actively enhance the sampling capacity of MD. In other words, these approaches use ML to not just learn from given data, but actually generate statistically accurate information when the underlying processes are so slow that they simply cannot be sampled in unbiased MD even with the best available computing resources. In order to generate novel low-probability structures one can simply try heating the system. A much more formidable problem is generating them so that their statistics, typically through some corrective reweighting scheme, is in accordance with the underlying Boltzmann probability, and even more ambitiously, with the underlying kinetics

at least in terms of inter-conversion timescales between different metastable states. This is a complex problem for which many exciting solutions have recently been proposed, and here we outline some of them. Essentially, the success of ML highly relies on the abundance of data. However if the events of interest are rare, one faces paucity of relevant data to train the ML upon. One could train ML models on data from enhanced sampling methods, but these methods themselves need an estimate of the RC, and a wrong choice of RC used to generate the data could mislead the ML being trained upon it. One solution to this dilemma is to iterate between rounds of sampling and ML, where every ML round generates a progressively improved RC, which is used to perform better enhanced sampling and so on. Many (though not all) of the schemes described in this section are built around this central idea, differing in the precise forms of (a) the ML procedure, (b) enhanced sampling scheme and (c) how exactly information from ML is used to perform sampling and vice-versa.

The molecular enhanced sampling with autoencoders (MESA) approach uses an ANN with non-linear encoder and decoder to learn the RC from input data, which itself is generated through umbrella sampling along trial RCs [19]. Every round of ML leads to an improved RC along which new umbrella sampling is performed. The iterations are continued until the free energy from umbrella sampling no longer varies with further iterations. Similar to MESA, nonlinear RCs learned by methods like VDE [13^{••}] can also be used to perform enhanced sampling, typically using TICA modes as input variables.

Reweighted autoencoded variational Bayes for enhanced sampling (RAVE) [20,21[•]] is an iterative ML-MD method motivated by the observation that many feature learning methods, in addition to classifying features, also provide the probability distribution in feature space [22]. The learnt features and their probability distribution can then respectively be used as RC and its fixed or static bias can then be applied to $U(\mathbf{x})$ leading to more ergodic exploration. One crucial distinguishing feature of RAVE is that it avoids additional biasing along the RC as in umbrella sampling [23,24] or metadynamics [25], as these could forcefully lead to enhanced sampling through non-equilibrium ways. In other words, even if the RC from ML was very far from the truth, when used in umbrella sampling especially with the typical post-processing WHAM protocol, one could still obtain some sort of free energy profile and have no way to tell how erroneous the RC from ML was. On the other hand, a static bias would lead to enhanced exploration only if the orthogonal hidden modes are not relevant. This serves as a test in RAVE that helps weed out spurious local minima solutions that often plague deep learning. Keeping the transition states between different features devoid of bias also allows obtaining pathways and rate constants. To learn these

features and their probabilities, RAVE uses a past-future information bottleneck optimization scheme that outputs a minimally complex yet maximally predictive model. An ANN decoder is trained to predict the future state of the system instead of only trying to recover the input data, and a linear encoder is used to get an interpretable projection from the space of order parameters to the RC. The active enhanced sampling (AES) approach [26] is another approach similar to RAVE, that uses well-tempered metadynamics [25,27] and a stochastic kinetic embedding formalism.

Instead of using enhanced sampling methods to explore new possible configurations, deep generative Markov state models (DeepGenMSM) [28] use a generative NN to predict the future evolution of the system and thus propose new configurations. The encoder has a SoftMax layer which gives the probability of mapping an input configuration to different discretized states in latent space. A generative model is fit to predict the time-delayed evolution of the system by minimizing a suitably defined energy distance, which measures the difference between the transition density of the system and that of the generative model. The generative model is then essentially extrapolated to produce high dimensional structures that were not in the training database.

Boltzmann Generators are a very recent deep learning based approach that learns the equilibrium probability $P_0(\mathbf{x})$ without resorting to running long trajectories [29]. It leverages recent advances in probabilistic generative modeling in which invertible coordinate transformations mapping \mathbf{x} onto a random variable \mathbf{x}' whose distribution is straightforward to sample are learnt [30,31]. Using \mathbf{x}' together with the fact that \mathbf{x} follows the Boltzmann distribution, $P_0(\mathbf{x}) \propto e^{-\beta U(\mathbf{x})}$, the ANN can be trained (in principle) without using maximum likelihood on a pre-existing dataset. Starting with the KL divergence of the probability predicted by the ANN relative to the exact (and simple) distribution in \mathbf{x}' -space, the loss function becomes

$$\mathbb{E}[\beta U(\mathbf{x}) - \log |J(\mathbf{x}')|] \quad (5)$$

where the expectation is calculated with respect to samples drawn from the exact distribution in \mathbf{x}' -space, \mathbf{x} is the output of the generative network that describes the inverse mapping from \mathbf{x}' back to \mathbf{x} , and the Jacobian J describes the effect of the coordinate transformation on the distribution. Notice that lowering the loss function given in Equation 5 tends to lead the network towards learning a transformation that results in sampling low-energy configurations (i.e. approach the Boltzmann distribution). In practice, however, Boltzmann generators do also use maximum likelihood on a pre-existing dataset in order force the network to give non-negligible

probabilities to other metastable states in addition to the global minima.

The NN based variationally enhanced sampling (VES) method also stands out with respect to many of the other ML methods mentioned here as it does not try to learn slow modes, but instead tries to express the bias as a smoothly differentiable ANN potential as a function of pre-selected small number of CVs [32^{••}]. To learn such a bias, it optimizes the objective function introduced in [33].

We now mention some data driven approaches which while not using ANNs, are connected with ML ideas. We begin with the diffusion map directed MD (DM-d-MD) [34] and variants thereof [35], which are a series of pioneering methods that use non-linear manifold learning techniques such as diffusion maps [36] to gradually build up the slow modes of a system. The central idea in these methods is to start with a short unbiased MD run, perform diffusion map (or a similar method) on it, and then use this diffusion map to select coordinates in configuration space for launching new rounds of unbiased MD (unbiased apart from the implicit bias in selection of initiation points). The key differences in various flavors of such an approach arise in how the new launch-off points are selected. Although these approaches lead to enhanced exploration, it is non-trivial to obtain a Boltzmann-weighted ensemble directly from such approaches. The extended DM-d-MD method in principle alleviates this problem, but can lead to a drop in the computational speed-up relative to unbiased MD.

SGOOP [37,38] is an iterative method similar to RAVE and MESA that uses rounds of enhanced sampling to learn a progressively improved RC. However to learn the RC, instead of ML, a maximum path entropy or Caliber model [39] is learnt that identifies the RC as the low-dimensional projection with maximum separation of timescales between visible and hidden modes.

Finally, the reinforcement learning based adaptive sampling (REAP) method of Ref. [40] learns relevant CVs on-the-fly as exploration of the landscape is carried out. REAP starts with a dictionary of OPs \mathbf{v} and associated trial weights. A round of unbiased MD is carried out, clustered into states, after which the weights are adjusted in order to maximize a reward function. In a nutshell, the reward function is designed to favor the least populated clusters and is iteratively adjusted as new clusters are visited.

Software

While the development of new algorithms is important and thus was the focus of this review, it is equally important to have efficient software implementing these algorithms in an accurate manner. Thankfully there is no

dearth of such software. PYEMMA, PLUMED and ANNCOLVAR, as well as associated modules and scripts provided in GitHub repositories of various publications [41–43], make it possible to implement many of the algorithms listed in this review.

Conclusions

In this overview we have summarized some recent ML-based methods for analyzing and enhancing MD simulations. This is a very lively field with multiple approaches published even during the course of this review being written. These approaches are making it possible to compress high dimensional data generated during MD into low dimensional models, arguably in a more robust and automatic manner than achievable with previous non-ML methods, and revealing hidden patterns that might not have been discernible otherwise. Thus while there is clear progress, we would argue that the field is still full of several difficult, exciting open questions. First, what did the ML model learn, or can we peer into the black box that is ML and interpret what it has learnt. This is the interpretability challenge. Second, would the ML model still work for small (or large) perturbations in the system being studied. In other words, would ML models trained on given systems be transferable to new systems of interest. This is the transferability challenge. Thirdly, can a fitted ML model be used to generate Boltzmann-weighted samples that were previously unexplored, or the sampling challenge. These challenges are not very different from those faced by ML in application domains outside biology, and thus a lot is to be gained from cross-pollination between ideas from active ML experts across different domains. We thus conclude with cautious optimism for the future.

Conflict of interest

None declared.

Acknowledgements

The authors thank Deepthought2, MARCC, and XSEDE (Projects CHE180007P and CHE180027P) for helping our group with computational resources used to develop some of the methods reported in this work. YW would like to thank NCI-UMD Partnership for Integrative Cancer Research for financial support.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
 - of special interest
1. Behler J, Parrinello M: **Generalized neural-network representation of high-dimensional potential-energy surfaces.** *Phys Rev Lett* 2007, **98**:146401.
 2. Bartók AP, Payne MC, Kondor R, Csányi G: **Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons.** *Phys Rev Lett* 2010, **104**:136403.

3. Smith JS, Isayev O, Roitberg AE: **Ani-1: an extensible neural network potential with DFT accuracy at force field computational cost.** *Chem Sci* 2017, **8**:3192-3203.
4. Klus S, Nüske F, Koltai P, Wu H, Kevrekidis I, Schütte C, Noé F: **Data-driven model reduction and transfer operator approximation.** *J Nonlinear Sci* 2018, **28**:985-1010.
5. Ravindra P, Smith Z, Tiwary P: **Automatic mutual information noise omission (AMINO): generating order parameters for molecular systems.** *Mol Syst Des Eng* 2020.
6. Ma A, Dinner AR: **Automatic method for identifying reaction coordinates in complex systems.** *J Phys Chem B* 2005, **109**:6769-6779 PMID:16851762.
7. Jung H, Covino R, Hummer G: **Artificial Intelligence Assists Discovery of Reaction Coordinates and Mechanisms from Molecular Dynamics Simulations.** 2019arXiv:1901.04595.
8. Noé F, Nuske F: **A variational approach to modeling slow processes in stochastic dynamical systems.** *Multisc Model Simul* 2013, **11**:635-655

Methodological framework introduces a variational principle for estimating slow Markov processes in complex dynamical systems from simulation data

9. Nüske F, Keller BG, Pérez-Hernández G, Mey ASJS, Noé F: **Variational approach to molecular kinetics.** *J Chem Theory Comput* 2014, **10**:1739-1752.
10. Pérez-Hernández G, Paul F, Giorgino T, De Fabritiis G, Noé F: **Identification of slow molecular order parameters for Markov model construction.** *J Chem Phys* 2013, **139**:07B604_1.
11. Mardt A, Pasquali L, Wu H, Noé F: **Vampnets for deep learning of molecular kinetics.** *Nat Commun* 2018, **9**:5

Artificial neural network framework for automatic MSM construction, developed by combining the VAMP principle with ANNs. The framework was used to investigate the folding dynamics of the NTL9 protein.

12. Chen W, Sidky H, Ferguson AL: **Capabilities and limitations of time-lagged autoencoders for slow mode discovery in dynamical systems.** *J Chem Phys* 2019, **151**:064123

A detailed discussion on the relationships between TAE and other methods.

13. Hernández CX, Wayment-Steele HK, Sultan MM, Husic BE, Pande VS: **Variational encoding of complex dynamics.** *Phys Rev E* Jun 2018, **97**:062412

Encoder-decoder AN, framework for learning the RC. A time-lagged variational autoencoder is used to approximate the propagator and extract a low dimensional kinetic model. This framework was applied to the folding process of the villin headpiece subdomain, with the aim of determining its ability to extract a good low-dimensional model for protein folding.

14. Schöberl M, Zabarar N, Koutsourelakis P-S: **Predictive collective variable discovery with deep Bayesian models.** *J Chem Phys* 2019, **150**:024109.
15. Lemke T, Peter C: **Encodermap: dimensionality reduction and generation of molecule conformations.** *J Chem Theory Comput* 2019, **15**:1209-1215 PMID:30632745.
16. Ceriotti M, Tribello GA, Parrinello M: **Simplifying the representation of complex free-energy landscapes using sketch-map.** *Proc Natl Acad Sci U S A* 2011, **108**:13023-13028.
17. Brandt S, Sittel F, Ernst M, Stock G: **Machine learning of biomolecular reaction coordinates.** *J Phys Chem Lett* 2018, **9**:2144-2150 PMID:29630378.
18. Olsson S, Noé F: **Dynamic graphical models of molecular kinetics.** *Proc Natl Acad Sci U S A* 2019, **116**:15001-15006.
19. Chen W, Ferguson AL: **Molecular enhanced sampling with autoencoders: on-the-fly collective variable discovery and accelerated free energy landscape exploration.** *J Comput Chem* 2018, **39**:2079-2102.
20. Ribeiro JML, Bravo P, Wang Y, Tiwary P: **Reweighted autoencoded variational bayes for enhanced sampling (rave).** *J Chem Phys* 2018, **149**:072301.
21. Wang Y, Ribeiro JML, Tiwary P: **Past-future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics.** *Nat Commun* 2019, **10**:3573

Encoder-decoder AN, framework for learning the RC as well as a static bias. First the past-future information bottleneck is used to define a RC, with an encoder-decoder AN, implemented to learn an optimal RC given simulation data. The framework is then used to determine the kinetics of ligand dissociation from the T4 lysozyme protein as well as to predict critical mutations by scanning the mutual information between the RC and protein residues

22. Wetzel SJ: **Unsupervised learning of phase transitions: from principal component analysis to variational autoencoders.** *Phys Rev E* 2017, **96**:022140.
23. Torrie GM, Valleau JP: **Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling.** *J Comput Phys* 1977, **23**:187-199.
24. Mezei M: **Adaptive umbrella sampling: self-consistent determination of the non-Boltzmann bias.** *J Comput Phys* 1987, **68**:237-248.
25. Valsson O, Tiwary P, Parrinello M: **Enhancing important fluctuations: rare events and metadynamics from a conceptual viewpoint.** *Ann Rev Phys Chem* 2016, **67**:159-184 PMID:26980304.
26. Zhang J, Chen M: **Unfolding hidden barriers by active enhanced sampling.** *Phys Rev Lett* 2018, **121**:010601.
27. Tiwary P, Parrinello M: **A time-independent free energy estimator for metadynamics.** *J Phys Chem B* 2015, **119**:736-742 01.
28. Wu H, Mardt A, Pasquali L, Noé F: **Deep Generative Markov State Models.** 2018arXiv:1805.07601.
29. Noé F, Olsson S, Köhler J, Wu H: **Boltzmann generators: sampling equilibrium states of many-body systems with deep learning.** *Science* 2019, **365**.
30. Dinh L, Sohl-Dickstein J, Bengio S: **Density Estimation Using Real NVP.** 2016arXiv:1605.08803.
31. Kingma DP, Dhariwal P: **Glow: generative flow with invertible 1x1 convolutions.** *Advances in Neural Information Processing Systems* 2018:10215-10224.
32. Bonati L, Zhang Y-Y, Parrinello M: **Neural networks-based variationally enhanced sampling.** *Proc Natl Acad Sci U S A* 2019, **116**:17641-17647

The variationally enhanced sampling (VES) method is combined with ANNs. Here, the bias potential is expressed as a powerful ANN, and the neural network is trained in order to produce a rich bias potential that is close minimizing the functional defined in VES

33. Valsson O, Parrinello M: **Variational approach to enhanced sampling and free energy calculations.** *Phys Rev Lett* 2014, **113**:090601.
34. Preto J, Clementi C: **Fast recovery of free energy landscapes via diffusion-map-directed molecular dynamics.** *Phys Chem Chem Phys* 2014, **16**:19181-19191.
35. Chiavazzo E, Covino R, Coifman RR, Gear CW, Georgiou AS, Hummer G, Kevrekidis IG: **Intrinsic map dynamics exploration for uncharted effective free-energy landscapes.** *Proc Natl Acad Sci U S A* 2017, **114**:E5494-E5503.
36. Coifman RR, Lafon S, Lee AB, Maggioni M, Nadler B, Warner F, Zucker SW: **Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps.** *Proc Natl Acad Sci U S A* 2005, **102**:7426-7431.
37. Tiwary P, Berne BJ: **Spectral gap optimization of order parameters for sampling complex molecular systems.** *Proc Natl Acad Sci U S A* 2016, **113**:2839-2844.
38. Smith Z, Pramanik D, Tsai S-T, Tiwary P: **Multi-dimensional spectral gap optimization of order parameters (SGOOP) through conditional probability factorization.** *J Chem Phys* 2018, **149**:234105.
39. Dixit PD, Jain A, Stock G, Dill KA: **Inferring transition rates of networks from populations in continuous-time Markov processes.** *J Chem Theory Comput* 2015, **11**:5464-5472.

40. Shamsi Z, Cheng KJ, Shukla D: **Reinforcement learning based adaptive sampling: reaping rewards by exploring protein conformational landscapes.** *J Phys Chem B* 2018, **122**:8386-8395.
41. Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, Wehmeyer C, Prinz J-H, Noé F: **Pyemma 2: a software package for estimation, validation, and analysis of Markov models.** *J Chem Theory Comput* 2015, **11**:5525-5542.
42. Bonomi M, Bussi G, Camilloni C, Tribello GA, Banáš P, Barducci A, Bernetti M, Bolhuis PG, Bottaro S, Branduardi D, Capelli R, Carloni P, Ceriotti M, Cesari A, Chen H, Chen W, Colizzi F, De S, De La Pierre M, Donadio D, Drobot V, Ensing B, Ferguson AL, Filizola M, Fraser JS, Fu H, Gasparotto P, Gervasio FL, Giberti F, Gil-Ley A, Giorgino T, Heller GT, Hocky GM, Iannuzzi M, Invernizzi M, Jelfs KE, Jussupow A, Kirilin E, Laio A, Limongelli V, Lindorff-Larsen K, Löhr T, Marinelli F, Martin-Samos L, Masetti M, Meyer R, Michaelides A, Molteni C, Morishita T, Nava M, Paissoni C, Papaleo E, Parrinello M, Pfaendtner J, Piaggi P, Piccini G, Pietropaolo A, Pietrucci F, Pipolo S, Provati D, Quigley D, Raiteri P, Raniglo S, Rydzewski J, Salvalaglio M, Sosso GC, Spiwok V, Sponer J, Swenson DWH, Tiwary P, Valsson O, Vendruscolo M, Voth GA, White A, The PLUMED Consortium: **Promoting transparency and reproducibility in enhanced molecular simulations.** *Nat Methods* 2019, **16**:670-673.
43. Trapl D, Horvancanin I, Mareska V, Ozcelik F, Unal G, Spiwok V: **Anncolvar: approximation of complex collective variables by artificial neural networks for analysis and biasing of molecular simulations.** *Front Mol Biosci* 2019, **6**:25.