Stochastic Neural Network Approach for Learning High-Dimensional Free Energy Surfaces

Elia Schneider,[1] Luke Dai,[1] Robert Q. Topper,[2] Christof Drechsel-Grau,[1] and Mark E. Tuckerman[1,3,4,*]
[1]Department of Chemistry, New York University, New York, New York 10003, USA
[2]Department of Chemistry, The Cooper Union for the Advancement of Science and Art, 41 Cooper Square, New York, New York 10003, USA
[3]Courant Institute of Mathematical Science, New York University, New York, New York 10003, USA
[4]NYU-ECNU Center for Computational Chemistry at NYU Shanghai, 3663 Zhongshan Road North, Shanghai 200062, China
(Received 28 February 2017; published 11 October 2017)

The generation of free energy landscapes corresponding to conformational equilibria in complex molecular systems remains a significant computational challenge. Adding to this challenge is the need to represent, store, and manipulate the often high-dimensional surfaces that result from rare-event sampling approaches employed to compute them. In this Letter, we propose the use of artificial neural networks as a solution to these issues. Using specific examples, we discuss network training using enhanced-sampling methods and the use of the networks in the calculation of ensemble averages.

DOI: 10.1103/PhysRevLett.119.150601

One of the outstanding challenges in the computational molecular sciences is the accurate generation of free energy surfaces (FESs) associated with conformational equilibria of complex systems. FESs are ubiquitous, whether the problem is to determine the conformational preferences of peptides and proteins, to predict and rank polymorphs of a molecular crystal, or to identify the binding sites of molecules on a surface. FESs are typically represented in terms of a set of order parameters or coarse-grained variables (CGVs) whose values correspond to a set of collective functions of the primitive atomic coordinates; these functions are known as *collective variables* (CVs). Since the CVs should ideally be chosen to capture the slow modes of the system, as their number grows, so does the number of minima and saddle points that characterize the FES. If the saddles are of sufficiently high energy that the probability to cross them is low, then rare-event or enhanced-sampling methods such as metadynamics [1–3], adiabatic free energy dynamics [4] and its variants [5,6], and temperature-accelerated molecular dynamics [7] are needed in order to generate a FES.

If a system is sufficiently complex, determining an optimal set of CVs is far from trivial and remains a considerable challenge. In some instances, it might be possible to discover optimal CVs in the course of a calculation using automated manifold learning techniques if sufficient sampling can be achieved [8–11]. Often, however, these are chosen *a priori* based on an intuitive guess guided by an understanding of the physics of a system. In such cases, in order to capture the essential conformational equilibria in a system, it can be useful to employ a set of CVs that contains some redundancy. However they are chosen, the number of CVs or CGVs needed to characterize a FES can be rather large, giving rise to a high-dimensional FES (HDFES). Although the aforementioned enhanced-sampling methods are generally capable of generating HDFESs, the volume of data associated with such a surface is big, which gives rise to a problem of representing and storing a HDFES. Moreover, as free energy is the generator of numerous equilibrium properties, it is often necessary to be able to perform calculations with a HDFES, which, in principle, requires having an analytical form for it.

In this Letter, we propose the use of machine learning (ML) [12] techniques, specifically artificial neural networks (ANNs), as a solution to the aforementioned issues. ANNs, which have recently been employed in the generation of model potential energy surfaces [13] and characterization of local structure in polymorphic systems [14], provide a compact representation of a HDFES, are flexible in their mathematical structure, can be trained using stochastic optimization techniques on free energy or gradient data generated via enhanced-sampling methods, and are sufficiently smooth that they can be subsequently employed in Monte Carlo or molecular dynamics calculations in order to compute equilibrium observables. Although ML methods such as Gaussian process regression have been used to explore and represent FESs [15–17], to our knowledge, this is the first attempt to employ ANNs in this context. Here, using training data generated from enhanced-sampling calculations [5], we construct ANN representations of the FESs of small peptides, specifically the alanine di- and tripeptides, as test two- and four-dimensional surfaces, respectively, and evaluate their performance against accurate benchmark FESs [18]. We then employ ANNs to represent two HDFES: the ten-dimensional FES corresponding to the five-residue oligopeptide met-enkephalin and the five-dimensional FES describing different crystal phases of xenon. For both examples, we use the trained networks to compute a relevant equilibrium observable: the NMR J-coupling parameters for met-enkephalin and the isothermal compressibility for crystal xenon.

Consider a system of N atoms having positions $\mathbf{r}_1, \ldots, \mathbf{r}_N \equiv \mathbf{r}$ interacting via a potential $U(\mathbf{r})$ at temperature T. The conformational space of interest is assumed to be characterized by a set of n CVs denoted $q_1(\mathbf{r}), \ldots, q_n(\mathbf{r})$, and the FES is then given by $A(s_1,\ldots,s_n) = -\beta^{-1} \ln P(s_1,\ldots,s_n)$, where $P(s_1, \ldots, s_n)$ is the marginal probability distribution

$$P(s) = \mathcal{N}^{-1} \int d^N\mathbf{r} e^{-\beta U(\mathbf{r})} \prod_{\alpha=1}^{n} \delta(q_\alpha(\mathbf{r}) - s_\alpha). \quad (1)$$

Here, $\mathcal{N} = \int d^N\mathbf{r} \exp[-\beta U(\mathbf{r})]$, $\beta^{-1} = k_B T$, and $s_1, \ldots, s_n \equiv s$ denotes the set of CGVs. Direct sampling of $P(s)$ on an n-dimensional grid rapidly becomes infeasible as n grows beyond three or four dimensions, and, indeed, enhanced-sampling methods like adiabatic free energy dynamics and temperature-accelerated molecular dynamics are designed to sample $P(s)$ "on the fly." Given this, the data generated by such methods can serve as more than mere samples of the marginal distribution: They can be used as the training data for ML models of the FES from which free energy values at points not sampled in the calculations can be predicted. If the chosen ML model is an artificial neural network with K hidden layers and M nodes in each layer, then the FES is represented in the form

$$A_{\mathrm{ANN}}(s;w) = H\Bigg[\sum_{j_K=1}^{m_K} h\Bigg(\ldots h\Bigg\{\sum_{j_2=1}^{m_2} h\Bigg[\sum_{j_1=1}^{m_1} h\Bigg(\sum_{\alpha=1}^{n} s_\alpha w^0_{\alpha,j_1} + w^0_{0,j_1}\Bigg) w^1_{j_1,j_2} + w^1_{0,j_2}\Bigg] w^2_{j_2,j_3} + w^2_{0,j_3}\Bigg\}\ldots\Bigg) w^K_{j_K} + w^K_0\Bigg]. \quad (2)$$

The parameter set w denotes a complete set of fitting parameters, with w^ν_{ik} connecting node i of layer ν with node k of layer $\nu+1$, and $h(x)$ and $H(x)$ are activation functions. In general, we allow the number of nodes in each hidden layer to be different, as denoted by the values of $m_1, m_2, \ldots, m_K$. A schematic of a neural network is shown in Fig. 1. In essence, the ANN is being employed to "learn" the integrals in Eq. (1).

Suppose an enhanced-sampling calculation generates M values of the free energy $A^{(\lambda)}$ at values $s^{(\lambda)} \equiv s_1^{(\lambda)}, \ldots, s_n^{(\lambda)}$ of the CVs, where $\lambda = 1, \ldots, M$. Then, the fitting is accomplished by setting up the cost function

$$E(w) = \frac{1}{2M} \sum_{\lambda=1}^{M} |A_{\mathrm{ANN}}(s^{(\lambda)};w) - A^{(\lambda)}|^2, \quad (3)$$

which is then minimized with respect to w, i.e., find the solution of $\nabla_w E(w) = 0$ to yield an optimal set of parameters w for the given training set.

In some cases, it might be more advantageous to generate free energy gradient data from enhanced-sampling calculations [19,20], in which case, training of the network can also be achieved using a cost function:

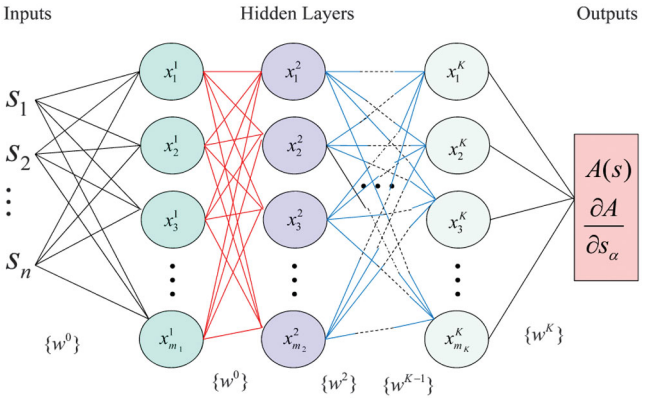

FIG. 1. Schematic representation of an artificial neural network that takes in n CGV values and outputs the free energy $A(s)$ and possibly its gradient $\partial A/\partial s_\alpha$. The network contains $m_1, m_2, \ldots, m_K$ nodes each of K hidden layers. In each node, x^i_j are the arguments of the activation functions in Eq. (2).

$$E_G(w) = \frac{1}{2M} \sum_{\lambda=1}^{M} \sum_{\alpha=1}^{n} \left| \frac{\partial}{\partial s_\alpha^{(\lambda)}} A_{\mathrm{ANN}}(s^{(\lambda)};w) + F_\alpha^{(\lambda)} \right|^2, \quad (4)$$

where $F_\alpha^{(\lambda)} = -\partial A/\partial s_\alpha^{(\lambda)}$ denotes the gradient data generated from the enhanced-sampling calculation. As before, fitting is obtained by minimizing E_G with respect to w. Derivatives of the cost function with respect to the fitting parameters are computed using the *backpropagation* method [21,22], which is described in more detail in Supplemental Material [23].

As a test case, we train ANNs to represent the FESs of the alanine di- and tripeptides in the gas phase, systems for which we have high-quality benchmark FESs [19]. In these examples, interatomic interactions are described by the CHARMM22 [24,25] force field. The results are compared to benchmarks as calculated by Chen, Cuendet, and Tuckerman [19]. As CVs, we use the backbone dihedral angles (ϕ, ψ), which give two- and four-dimensional FESs for the di- and tripeptides, respectively. For each system, free energy data are generated using the driven adiabatic free energy dynamics method [5]. The CV temperature is set to 1500 K, the CV mass is 168.0 amu $\text{Å}^2/\text{rad}^2$, and the harmonic coupling constant is 2.78×10^3 kcal/mol/rad^2. For the alanine dipeptide, data were generated on a two-dimensional grid of 300×300 points for the two CVs. Of these, $M = 40\,000$ randomly chosen grid points were used to train the network, and the remaining 50 000 were used for validation. For the case of the alanine tripeptide, 2×10^5 random CV values were generated, and their free energy values obtained using the Gaussian fit of Chen, Cuendet, and Tuckerman [19]. Of these, $M = 10^5$ were used for training, and the remaining 10^5 were used for validation. In both cases, the architecture of the ANN contained two hidden layers with 20 nodes in each layer for the alanine dipeptide and 40 nodes in each layer for the alanine tripeptide (we discuss these architecture choices in Ref. [23]).

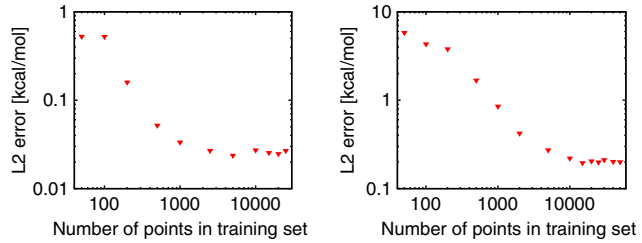

FIG. 2. The L^2 global error versus the number of training points in the training set for the alanine dipeptide (left) and the alanine tripeptide (right).

The activation functions were chosen to be $h(x) = 1/(1+x^2)$, $H(x) = x$. In Fig. 2, we show the L^2 error $= \sqrt{(1/N_v)\sum_{\lambda=1}^{N_v}[A_{\mathrm{ANN}}(s^{(\lambda)};w_{\mathrm{opt}}) - A_{\mathrm{bench}}(s^{(\lambda)})]^2}$, of the FESs represented by the ANNs with respect to the benchmark from Ref. [19] as a function of the number of free energy points used to train the networks for both the alanine dipeptide (left panel) and the alanine tripeptide (right panel). In this expression, N_v is the number of validation points. Here w_{opt} denotes the optimal parameter set. The figure shows a clear convergence of the ANN representation with 10^3 training points for alanine dipeptide and 10^4 points for the alanine tripeptide, leading to an overall accuracy of 0.03 kcal/mol for the dipeptide and 0.2 kcal/mol for the tripeptide.

For both systems, we also trained gradient-based ANNs using gradient data generated from the same driven adiabatic free energy dynamics calculations [5,19]. In the case of the alanine dipeptide, we used 1600 gradients on a 40×40 grid, and for the alanine tripeptide, we used 10^5 random gradient vectors. In both cases, we obtain L^2 errors similar to those obtained with the ANNs trained from the free energy. Figure 3 shows the two-dimensional FES produced by the gradient-trained ANN (left panel) and a comparison between the FES of the alanine tripeptide obtained by the gradient-trained ANN and the free-energy-trained ANN (right panel). Both results show the

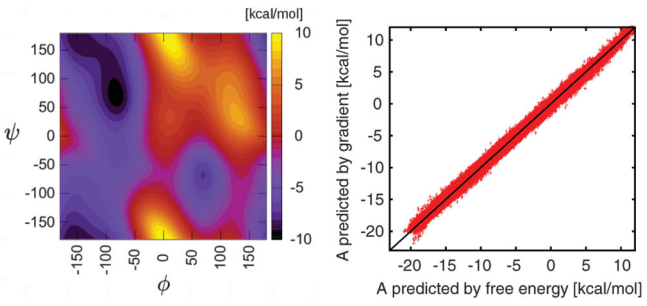

FIG. 3. FES of the alanine dipeptide in the gas phase generated by a trained ANN based on the free energy gradient using training data of Chen, Cuendet, and Tuckerman [19] (left). Comparison of free energies generated using the gradient-based ANN for the gas-phase alanine tripeptide and free energies obtained directly from the free-energy-based ANN used to generate Fig. 2. The black line represent the identity function $f(x) = x$.

ability of the gradient-based approach to obtain an accurate FES.

Once trained, the ANNs for the alanine di- and tripeptides were employed to compute the ensemble average of a physical observable following the procedure of Ref. [20]. Given an observable $O(\mathbf{r})$, the FES $A_{\mathrm{ANN}}(s;w_{\mathrm{opt}})$ evaluated at the optimal parameter values w_{opt} can be used to compute the ensemble average $\langle O\rangle$ via

$$\langle O\rangle = \frac{\int d^n s \langle O\rangle_{\mathbf{r}}(s) e^{-\beta A_{\mathrm{ANN}}(s;w_{\mathrm{opt}})}}{\int d^n s e^{-\beta A_{\mathrm{ANN}}(s;w_{\mathrm{opt}})}}, \tag{5}$$

where $\langle O\rangle_{\mathbf{r}}(s)$ is defined by

$$\langle O\rangle_{\mathbf{r}}(s) = \frac{\mathcal{N}^{-1}}{P(s)} \int d^N \mathbf{r} O(\mathbf{r}) e^{-\beta U(\mathbf{r})} \prod_{\alpha=1}^{n} \delta(q_\alpha(\mathbf{r}) - s_\alpha). \tag{6}$$

If $O(\mathbf{r})$ can be expressed entirely in terms of the CVs, then Eq. (6) becomes unnecessary. For the alanine di- and tripeptides, we computed the ensemble average of the root-mean-square deviation (RMSD) of the dihedral angles defined by

$$O(\phi,\psi) \equiv \mathrm{RMSD}(\phi,\psi) = \sqrt{\frac{1}{2n}\sum_{i=1}^{n}[(\phi_i - \phi^{\min})^2 + (\psi_i - \psi^{\min})^2]}, \tag{7}$$

where $n = 1, 2$ for the alanine di- and tripeptide, respectively, and the set of dihedral angles $(\phi^{\min}, \psi^{\min})$ represents the global minimum of the FES. The motivation for choosing this "observable" is that it is a function of all of the chosen CVs and, hence, tests the reliability of the full ANN-based FES. The integrals in Eq. (5) were computed using a Metropolis Monte Carlo algorithm at temperature $T = 300$ K using 5×10^7 steps. The use of Monte Carlo requires that the ANN be able to predict free energy values for any proposed trial move, which here is made by sampling a uniform distribution in the angle space. In Fig. 4, we show the convergence of the RMSD in Eq. (7) as a function of the number of points in the training set. For the dipeptide, we observe convergence with just a few hundred training points, obtaining a RMSD average of $56.0° \pm 0.2°$. For the tripeptide, convergence is obtained

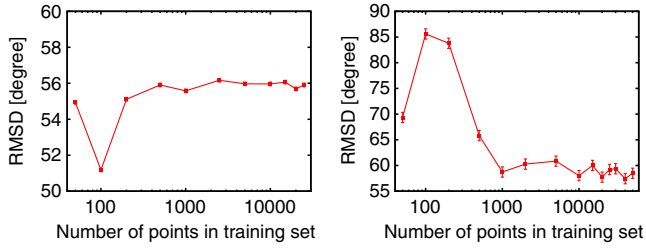

FIG. 4. The ensemble average of the RMSD observable defined in Eq. (7) for the alanine dipeptide (left) and alanine tripeptide (right) as a function of the number of data in the training set.

after a few thousand training points around a value of $59.5° \pm 0.5°$. Comparing the value obtained using an ANN against the ensemble average of the RMSD based on the FES obtained with a Gaussian fit ($56.3° \pm 0.1°$ and $58.3° \pm 0.3°$ for the alanine di- and tripeptide, respectively), we find that ANNs can faithfully reproduce the RMSD values.

As an example of a HDFES for a peptide, we consider the pentapeptide met-enkephalin, which has the sequence tyr-gly-gly-phe-met. This small peptide is well known as an endogenous ligand of opioid receptors and is distributed throughout the central nervous system. In Ref. [18], the complete set of free energy minima and saddle points as a function of the ten backbone dihedral angles and converged free energy differences between these points were computed using the driven adiabatic free energy dynamics algorithm [5] where the CV temperature is set to 400 K, the CV mass is 2.8 amu $Å^2/rad^2$, and the harmonic coupling constant is 2.78×10^3 kcal/mol/rad^2. In particular, 1081 minima and 1431 saddles were uncovered in these calculations for a total of 2512 "landmark" points on the FES. The data from a set of 7.5×10^5 random CVs from a 500 ns driven adiabatic free energy dynamics simulation were used to train a neural network based on their free energy values. The architecture of the ANN contained three hidden layers with 100 nodes in the first two layers and 50 in a third, the addition of which gives significantly higher accuracy. The trained ANN was subsequently used to compute the ensemble average of the NMR spin-spin J-coupling constants for each ϕ, ψ Ramachandran dihedral angle pair in the five-residue sequence using the Karplus equation [26] $J(\phi) = A\cos^2(\phi - \phi_0) + B\cos(\phi - \phi_1) + C$, where $\phi_0 = \phi_1 = 60°$, $A = 7.09$ Hz, $B = -1.42$ Hz, and $C = 1.55$ Hz, which correspond to the parameters reported in Ref. [27] for the NMR spin-spin J coupling between hydrogen atoms H_N and H_α in each peptide backbone. The average in Eq. (5) is computed for each residue using Metropolis Monte Carlo simulations of 5×10^7 steps at a temperature of 300 K with the trial values sampled from a uniform distribution in the ten-dimensional dihedral-angle space. In order to benchmark the J couplings from the ANN, we additionally performed a long NVT simulation, collecting 17 μs and averaging $J(\phi)$ using snapshots written out every 100 fs. The results are shown in Fig. 5. The figure shows that this observable is accurately predicted by the ANN for each residue after roughly 3×10^4 training points. The NVT results and the converged integrals computed with a trained ANN match within an error of 5%, confirming the accuracy of our calculation. Considering the high dimensionality of the FES and the large number of landmark points on it, the roughly 30 000 training points needed to converge the J couplings constitutes a remarkably sparse distribution of points on this HDFES required to reach convergence. In general, the ten-dimensional HDFES associated with met-enkephalin is far too large to represent explicitly, which renders the direct computation of observables such as the J couplings from fully atomistic molecular dynamics calculations significantly more laborious. Clearly, then, ANNs, once trained, provide a sufficiently compact representation that the calculation can be performed straightforwardly and efficiently.

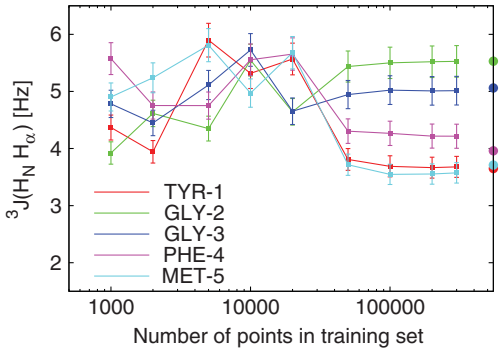

FIG. 5. NMR spin-spin J couplings between H_N and H_α for each of the five residues in met-enkephalin using the trained ANN together with Monte Carlo simulations in order to evaluate Eq. (5). Dots on the right are converged NVT results.

As an additional application, we employ an ANN to learn the free energy landscape of a crystal system. The possibility of a crystallographic fcc-bcc phase transition in solid xenon at a high pressure (25–30 GPa) near the melting point (2700–2900 K) has been debated in the literature [28–32]. In order to sample the HDFES, we performed periodic driven adiabatic free energy dynamics calculations on a system of 4000 xenon atoms interacting via a Buckingham potential [33] at 2700 K and 25 GPa pressure. The CVs consisted of the Steinhardt order parameters Q_4 and Q_6 [34] and the cell lengths $|\mathbf{a}|$, $|\mathbf{b}|$, and $|\mathbf{c}|$, yielding a five-dimensional FES. The data from a set of 5×10^5 random CV configurations chosen from the 200 ns driven adiabatic free energy dynamics run were used to train a neural network based on free energy values. The architecture of the ANN consisted of two hidden layers with 80 nodes in each layer. The left panel in Fig. 6 shows the two-dimensional projection onto the Q_4-Q_6 plane of the FES generated by the trained ANN. The two deepest minima represent bcc and fcc stable structures, while the local minima between them represent an hcp metastable state.

The irregular integration domain suggested by the projection in Fig. 6 represents the thermodynamically accessible region accessed in the driven adiabatic free

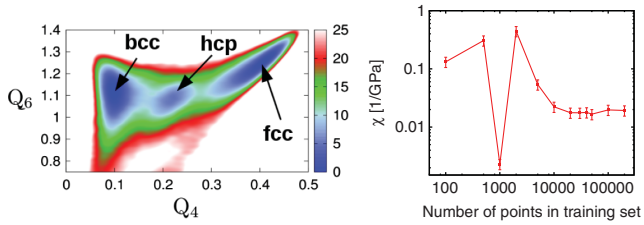

FIG. 6. Left: Projection onto the $Q_4 - Q_6$ plane of the FES (in eV) generated by a trained ANN. Thermodynamically inaccessible regions are shown as white. Right: Dependence of the isothermal compressibility on the number of training points.

energy dynamics simulation. This unusual domain renders the application of Eq. (5) less straightforward. Consequently, we employed a second ANN to classify all points on the five-dimensional landscape in terms of their thermodynamic accessibility, thereby defining an integration domain (see [23] for details). We collected a training set of 4×10^5 points using a free energy threshold of 20 eV to label each point. The trained ANNs were subsequently used to compute the ensemble average of the isothermal compressibility χ of the system defined as $\chi = -(1/\langle V\rangle)(\partial\langle V\rangle/\partial P) = \beta\langle(V - \langle V\rangle)^2\rangle/\langle V\rangle$, where $\langle V\rangle$ is the average volume of the system. The required ensemble averages were computed using Monte Carlo simulations of 10^8 random points. The right panel in Fig. 6 shows the dependence of the compressibility on the number of point in the training set. We observe an accurate result after roughly 2×10^4 points with a final value of $\chi = 0.020 \pm 0.005$ GPa^{-1}. The computed value is slightly higher than the measured compressibility $\chi \sim 0.008$ GPa^{-1} at room temperature and 25 GPa of Ref. [35].

We have shown that machine learning techniques such as ANNs can provide a smooth and compact way to represent HDFESs in complex systems, not only allowing the free energy differences to be obtained easily but also permitting ensemble averages to be computed directly from the machine learning model. In future work, we will explore other machine learning models [15–17] and investigate the application of machine learning techniques to the prediction of structure and polymorphism in molecular crystals [36,37] and to the study of conformational transitions in macromolecules, where high dimensionality limits the capabilities of enhanced-sampling approaches.

The authors acknowledge L. Vogt and J. Rogal for useful discussions. This work was supported by the National Science Foundation partially through the Materials Research Science and Engineering Center (MRSEC) program under Grant No. DMR-1420073 and partially through Grant No. CHE-1565980.

*mark.tuckerman@nyu.edu

[1] A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. **99**, 12562 (2002).
[2] A. Barducci, G. Bussi, and M. Parrinello, Phys. Rev. Lett. **100**, 020603 (2008).
[3] M. Bonomi and M. Parrinello, Phys. Rev. Lett. **104**, 190601 (2010).
[4] L. Rosso, P. Mináry, Z. Zhu, and M. E. Tuckerman, J. Chem. Phys. **116**, 4389 (2002).
[5] J. B. Abrams and M. E. Tuckerman, J. Phys. Chem. B **112**, 15742 (2008).
[6] A. T. Tzanov, M. A. Cuendet, and M. E. Tuckerman, J. Phys. Chem. B **118**, 6539 (2014).
[7] L. Maragliano and E. Vanden-Eijnden, Chem. Phys. Lett. **426**, 168 (2006).
[8] G. Tribello, M. Ceriotti, and M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. **107**, 17509 (2010).
[9] M. Rohrdanz, W. Zheng, and C. Clementi, Annu. Rev. Phys. Chem. **64**, 295 (2013).
[10] L. Nedialkova, M. Armat, I. Kevrekidis, and G. Hummer, J. Chem. Phys. **141**, 114102 (2014).
[11] C. Dsilva, R. Talmon, C. Gear, R. Coifman, and I. Kevrekidis, SIAM J. Appl. Dyn. Syst. **15**, 1327 (2016).
[12] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2006).
[13] J. Behler, Int. J. Quantum Chem. **115**, 1032 (2015).
[14] P. Geiger and C. Dellago, J. Chem. Phys. **139**, 164105 (2013).
[15] E. Chiavazzo, R. Covino, R. R. Coifman, C. W. Geard, A. G. Georgiou, G. Hummer, and I. G. Kevrekidis, Proc. Natl. Acad. Sci. U.S.A. **114**, E5494 (2017).
[16] T. Stecher, N. Bernstein, and G. Csányi, J. Chem. Theory Comput. **10**, 4079 (2014).
[17] L. Mones, N. Bernstein, and G. Csányi, J. Chem. Theory Comput. **12**, 5100 (2016).
[18] M. Chen, T.-Q. Yu, and M. E. Tuckerman, Proc. Natl. Acad. Sci. U.S.A. **112**, 3235 (2015).
[19] M. Chen, M. A. Cuendet, and M. E. Tuckerman, J. Chem. Phys. **137**, 024102 (2012).
[20] M. A. Cuendet and M. E. Tuckerman, J. Chem. Theory Comput. **10**, 2975 (2014).
[21] K. Hornik, M. Stinchcombe, and H. White, Neural Netw. **2**, 359 (1989).
[22] S. Ferrari and R. F. Stengel, IEEE Trans. Neural Networks **16**, 24 (2005).
[23] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.119.150601 for description of the backpropagation method, a discussion about the dependence of the efficiency of the neural networks used in this study on their architecture, and a discussion of the use of classification networks to define thermodynamically accessible regions on the free energy surfaces generated in this study.
[24] A. D. MacKerell *et al.*, J. Phys. Chem. B **102**, 3586 (1998).
[25] A. D. Mackerell, M. Feig, and C. L. Brooks, J. Comput. Chem. **25**, 1400 (2004).
[26] M. Karplus, J. Chem. Phys. **30**, 11 (1959).
[27] K. A. Beauchamp, Y.-S. Lin, R. Das, and V. S. Pande, J. Chem. Theory Comput. **8**, 1409 (2012).
[28] A. B. Belonoshko, R. Ahuja, and B. Johansson, Phys. Rev. Lett. **87**, 165505 (2001).
[29] A. B. Belonoshko, O. LeBacq, R. Ahuja, and B. Johansson, J. Chem. Phys. **117**, 7233 (2002).
[30] F. Saija and S. Prestipino, Phys. Rev. B **72**, 024113 (2005).
[31] A. B. Belonoshko, Phys. Rev. B **78**, 174109 (2008).
[32] T. Q. Yu, P. Y. Chen, M. Chen, A. Samanta, E. Vanden-Eijnden, and M. E. Tuckerman, J. Chem. Phys. **140**, 214109 (2014).
[33] M. Ross and A. K. McMahan, Phys. Rev. B **21**, 1658 (1980).
[34] D. Quigley and P. M. Rodger, J. Chem. Phys. **128**, 154518 (2008).
[35] A. N. Zisman, I. V. Aleksandrov, and S. M. Stishov, Phys. Rev. B **32**, 484 (1985).
[36] T. Q. Yu and M. E. Tuckerman, Phys. Rev. Lett. **107**, 015701 (2011).
[37] A. M. Reilly *et al.*, Acta Crystallogr. Sect. B **72**, 439 (2016).