THE JOURNAL OF PHYSICAL CHEMISTRY Letters

pubs.acs.org/JPCL Letter

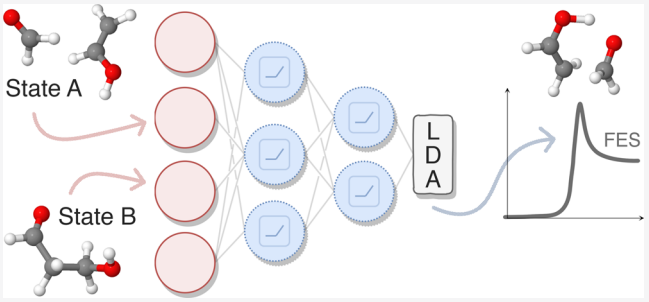
Data-Driven Collective Variables for Enhanced Sampling

Luigi Bonati, Valerio Rizzi, and Michele Parrinello*

Cite This: J. Phys. Chem. Lett. 2020, 11, 2998−3004 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Designing an appropriate set of collective variables is crucial to the success of several enhanced sampling methods. Here we focus on how to obtain such variables from information limited to the metastable states. We characterize these states by a large set of descriptors and employ neural networks to compress this information in a lower-dimensional space, using Fisher's linear discriminant as an objective function to maximize the discriminative power of the network. We test this method on alanine dipeptide, using the nonlinearly separable data set composed by atomic distances. We then study an intermolecular aldol reaction characterized by a concerted mechanism. The resulting variables are able to promote sampling by drawing nonlinear paths in the physical space connecting the fluctuations between metastable basins. Lastly, we interpret the behavior of the neural network by studying its relation to the physical variables. Through the identification of its most relevant features, we are able to gain chemical insight into the process.



In recent decades, enhanced sampling methods have become one of the main tools of computational science. They have extended the scope of atomistic simulations, allowing for long time scale phenomena to be simulated. Starting from the pioneering work of Torrie and Valleau,[1] a large class of such methods relies on the introduction of collective variables (CVs) defined as functions of the atomic coordinates. Once these variables are identified, a bias potential that is a function of the selected CVs is added to the interaction potential to accelerate sampling.[2] The CVs are chosen so as to describe the hard-to-sample modes of the system, but this is nontrivial, especially when the system under investigation is complex and its transition pathways are unknown. Not surprisingly, much effort has been devoted to the identification and improvement of useful CVs. Very recently, machine learning methods of varying complexity have been also used to this effect.[3−11] Building on our previous studies, we introduce a new method that uses deep neural networks (NNs) to perform a nonlinear featurization of a large set of input descriptors in order to build effective CVs. However, before describing our approach we recall some recent findings that give a clue to what we plan to do.

The free-energy landscape of physical systems can be described as being made up of islands of metastability in a sea of improbable configurations. The metastable states are connected by narrow passageways that allow rare but crucial transitions from one state to another to take place. Very recently our group and others[12,13] have developed a family of efficient CVs based only on the fluctuations in the metastable basins. This defies the conventional wisdom that for a variable to be effective it must contain explicit information about the whole reaction path or at least the transition state.[3,14]

As an example of this approach, we briefly recall how the method called harmonic linear discriminant analysis (HLDA)[12] works. One first identifies a small set of descriptors capable of discriminating between the states. The expectation value of the descriptors and their covariance matrices are computed from short unbiased runs in the two basins. With the knowledge of only these quantities, and by using a variant of the classification method that goes under the name of linear discriminant analysis (LDA),[15] one obtains CVs that are linear combinations of the input descriptors.

This simple approach has proven to be highly effective in a variety of cases, such as chemical reactions,[16,17] crystallization,[18] ligand unbinding,[19] and small peptides folding,[20] but it has limitations. At first, the states must be linearly separated in the descriptors space, as the HLDA CV is built upon their linear combination. Thus, HLDA crucially relies on the identification of a small set of uncorrelated descriptors. While this is a far less demanding task than finding CVs, it requires knowledge of the system and physical intuition. This intuition can be of great help in leading to an accelerated sampling, but sometimes it might reflect more our prejudice than the actual system behavior, possibly preventing the

Received: February 18, 2020
Accepted: March 28, 2020
Published: April 2, 2020

https://dx.doi.org/10.1021/acs.jpclett.0c00535
J. Phys. Chem. Lett. 2020, 11, 2998−3004

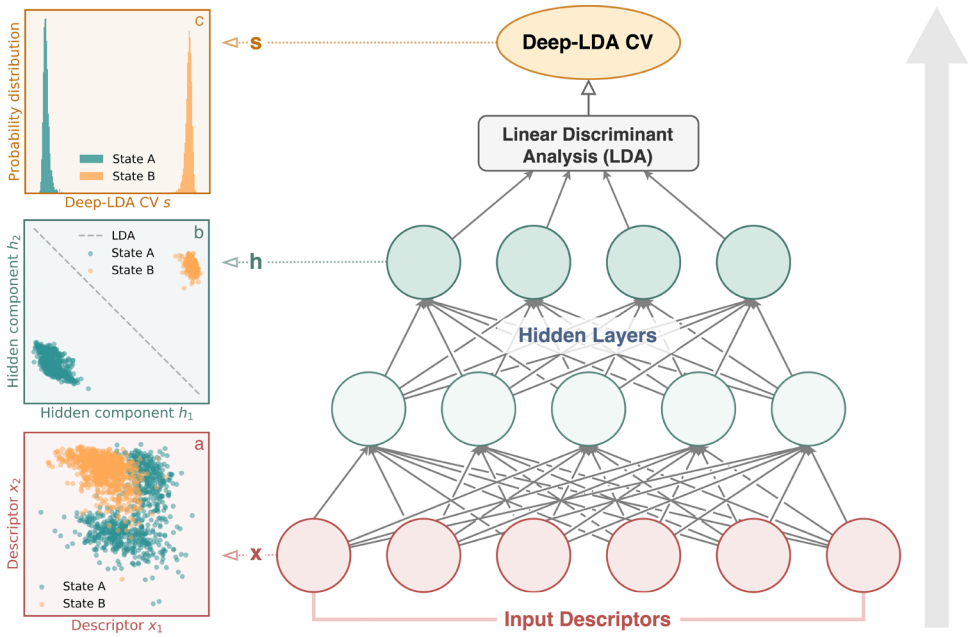


Figure 1. Scheme of the construction of the Deep-LDA CV. A set of physical descriptors are used as inputs of a feed-forward neural network. A nonlinear transformation is made by the NN via the composition of several hidden layers. In the last layer, a linear discriminant analysis is performed, the direction of maximal separation between classes is determined, and the CV is obtained. To better illustrate the workings of Deep-LDA, we report three panels on the left-hand side of the figure. In panel a, we show the distribution of a pair of typical input descriptors and note how they do not distinguish state A from state B. In panel b, we plot the distribution of a pair of variables from the topmost hidden layer, along with the LDA boundary projected in this space. At this stage, after the nonlinear transformation, states A and B are linearly separable. In panel c, we report the probability distribution of the Deep-LDA CV for the two states.

exploration of some of the relevant transition pathways. With the use of an appropriately designed neural network, we want to lift these limitations.

To achieve this result, we employ a NN to perform a nonlinear transformation of the inputs, optimized by Fisher's linear discriminant as objective function. The idea of improving linear techniques by combining them with NNs has been applied to LDA for classification purposes[21] and to other problems in different contexts.[9,11,22,23] Here we want to apply it to the design of CVs.

Method. LDA was first introduced as a classification method. In this context, one searches for the linear combination of the input features that best separates the data in given classes. Let us consider a set of N data points $\mathbf{x}_1, ..., \mathbf{x}_N$ (observations) of dimension d (local descriptors) belonging to C classes (metastable states). The covariance matrix over all these samples can be decomposed in two terms, the so-called within class $\mathbf{S}_w$ and between class $\mathbf{S}_b$ scatter matrices. The former takes into account the fluctuations inside the basins and corresponds to the average of the class covariances $\mathbf{S}_i$:

$$\mathbf{S}_w = \frac{1}{C}\sum_i^C \mathbf{S}_i \tag{1}$$

On the other hand, $\mathbf{S}_b$ is defined as the covariance of the class means $\boldsymbol{\mu}_i$ and thus measures the fluctuations between classes:

$$\mathbf{S}_b = \frac{1}{C}\sum_i^C (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T \tag{2}$$

where $\boldsymbol{\mu}$ is the average of the C class means. Fisher's criterion seeks a linear projection $\mathbf{W}$ into a $C-1$-dimensional space, such that the samples show a high variance between classes and a low variance within. This is achieved by maximizing Fisher's ratio

$$\underset{\mathbf{w}}{\text{argmax}} \frac{\mathbf{W}\mathbf{S}_b\mathbf{W}^T}{\mathbf{W}\mathbf{S}_w\mathbf{W}^T} \tag{3}$$

that measures the degree of separation between the classes.

In order to find the combination $\mathbf{W}$ which maximizes eq 3, one has to solve the generalized eigenvalue problem:

$$\mathbf{S}_b\mathbf{w}_i = v_i\mathbf{S}_w\mathbf{w}_i \quad \forall\, i = 1, ..., d \tag{4}$$

where the eigenvectors $\mathbf{w}_i$ form the projection matrix $\mathbf{W}$ and the eigenvalues v_i quantify the separation in the corresponding directions. We then obtain the compressed representations s_i with $i = 1, ..., C - 1$ by projecting the data points $\mathbf{x}$ along the corresponding eigenvectors as $s_i = \mathbf{w}_i^T\mathbf{x}$.

By choosing s_i as a set of CVs, LDA has been used in enhanced sampling applications. In previous work, a harmonic average of the covariance matrices was performed in eq 1, hence the name HLDA.[12] However, in this Letter we use the standard LDA version.

We now illustrate how LDA can be combined with a neural network (see Figure 1). We feed a number of descriptors to the NN that reduces the dimensionality of the data through a succession of continuous nonlinear transformations represented by $f_\theta(\mathbf{x})$ with parameters θ. We then perform LDA on the topmost hidden space representation $\mathbf{h} = f_\theta(\mathbf{x})$ and use the eigenvalues of eq 4 to optimize the NN.

From now on, we restrict the discussion to the typical case of two-classes, where the loss function reads:

$$\mathcal{L} = -v_1 \tag{5}$$

https://dx.doi.org/10.1021/acs.jpclett.0c00535
J. Phys. Chem. Lett. 2020, 11, 2998−3004

and v_1 is the only nonzero eigenvalue of eq 4. By maximizing the LDA eigenvalue, one increases the discriminative power of the NN and learns at the same time linearly separable latent features (see Figure 1b). The resulting Deep-LDA CV, reported in Figure 1c, is obtained by projecting the output of the NN $\mathbf{h}$ along the LDA eigenvector $\mathbf{w}_1$ as follows: $s = \mathbf{w}_1^{\mathrm{T}}\mathbf{h}$. Generalizing the loss function to multistate problems is simple and can be done either by maximizing the smallest of the $C - 1$ eigenvalues or their sum.[21]

To make the optimization more stable we regularize the within scatter matrix by adding a multiple of the identity matrix

$$\mathbf{S}'_w = \mathbf{S}_w + \lambda \mathbf{I} \tag{6}$$

as done in ref 21. Furthermore, we transform the generalized eigenvalue problem into a standard one by performing a Cholesky decomposition of $\mathbf{S}'_w = \mathbf{L}\,\mathbf{L}^{\mathrm{T}}$.[15] By simple manipulation, we rewrite eq 4 as

$$\tilde{\mathbf{S}}\tilde{\mathbf{w}}_i = v_i \tilde{\mathbf{w}}_i \quad i = 1, 2 \tag{7}$$

where $\tilde{\mathbf{S}} = \mathbf{L}^{-1}\mathbf{S}_{\mathrm{b}}(\mathbf{L}^{\mathrm{T}})^{-1}$ and $\tilde{\mathbf{w}}_i = (\mathbf{L}^{\mathrm{T}})^{-1}\,\mathbf{w}_i$. In this way, we have a symmetric eigenvalue problem which can be solved using the Pytorch library.[24] This allows training Deep-LDA networks with backpropagation and mini-batch gradient descent.

We note that, in the context of classification, eq 5 is used as loss function and training is stopped when the configurations are correctly labeled.[21] However, during the optimization process, the separation between the projected classes of eq 3 tends to increase without bounds either by enlarging the distance between classes or by collapsing them into delta-like distributions. Because we intend to use the compressed representation in enhanced sampling methods, neither very large distances between states nor too narrow basins are suitable for our purpose. We regulate the width of the projected classes with λ (eq 6), which provides a lower bound of the within class covariances, and we control the distance by modifying the loss function as follows:

$$\mathcal{L} = -v_1 - \frac{\alpha}{1 + (\overline{s^2} - 1)^2} \tag{8}$$

where α represents the magnitude of the regularization (see the Supporting Information) and $\overline{s^2} = \frac{1}{n_b}\sum_{j=1}^{n_b} s_j^2$ is the average of the squared CV over a batch of n_b configurations. The second term in eq 8 is maximum for $\overline{s^2} = 1$; thus, it acts as an effective constraint on the distance between the basins. More details about the loss function and the optimization procedure are reported in the Supporting Information.

After the training reaches convergence, we use the Deep-LDA CV in combination with enhanced sampling methods. In principle, this could be done with any method, such as metadynamics[25] or variational enhanced sampling.[26,27] In this work, we choose to employ a recently developed evolution of metadynamics, called on-the-fly probability enhanced sampling (OPES),[28] which builds the bias from an on-the-fly estimation of the probability distribution.

In OPES, the probability distribution at iteration n is given by

$$P_n(\mathbf{s}) = \frac{\sum_k^n w_k G(\mathbf{s}, \mathbf{s}_k)}{\sum_k^n w_k} \tag{9}$$

where $G(\mathbf{s}, \mathbf{s}_k)$ is a multivariate Gaussian and the weights are computed from the previously deposited bias potential $w_k = e^{\beta V_{k-1}(\mathbf{s}_k)}$, with β the inverse temperature. In turn, the bias potential is defined as

$$V_n(\mathbf{s}) = \left(1 - \frac{1}{\gamma}\right)\frac{1}{\beta}\log\left(\frac{P_n(\mathbf{s})}{Z_n} + \epsilon\right) \tag{10}$$

where Z_n is a normalization factor and the bias factor γ is a parameter that, as in well-tempered metadynamics,[29] determines the broadening of the biased distribution. Finally, ϵ is a regularization term that sets a limit to the maximum value of the bias, thus limiting the exploration of higher free-energy regions.

Once the iterative process converged, the free-energy surface (FES) can be computed as

$$F_n(\mathbf{s}) = -\frac{1}{\beta}\log P_n(\mathbf{s}) \tag{11}$$

and simple umbrella-sampling reweighting gives access to all the static properties of interest. We refer the reader to ref 28 for further details.

The combination of Deep-LDA and OPES presents a few noteworthy features. Enhancing the dynamics along the Deep-LDA CV typically spreads the action of the bias over a fairly large number of degrees of freedom. This helps greatly in promoting transitions but also increases the risk of exploring unwanted regions of the phase space. Therefore, the capability of OPES to restrict the exploration by choosing appropriately the ϵ parameter in eq 10 helps in focusing the sampling. Other benefits of OPES include its robustness due to a small number of free parameters and a convergence faster than metadynamics.

Before proceeding to the applications, we summarize the method for clarity:

1. Perform short unbiased MD simulations in the metastable states and compute the descriptors
2. Construct a CV by training a NN with LDA as the objective function
3. Use the Deep-LDA CV to enhance the sampling and obtain the FES

Alanine Dipeptide. The first example that we choose is alanine dipeptide, a small molecule often used to benchmark sampling methods. Alanine has two metastable states that are well-described using the pair of Ramachandran angles ϕ and ψ, which forms a nearly ideal set of CVs.

However, because here we wish to show the strength of the method, we deliberately put ourselves in a situation more complex than necessary and choose a general set of descriptors. This is more similar to the situation one encounters in practice, where optimal CVs are difficult to find. The key here is Deep-LDA's ability to handle a large number of descriptors. Unfortunately, we cannot directly use the atomic coordinates, because the resulting CV would not be rotationally and translationally invariant. A natural choice is to use coordinate combinations that are invariant under such symmetries, for instance distances, angles, and dihedrals. To make things even more challenging, we choose only distance-based descriptors. In such a base the states cannot be linearly separated. Instead, Deep-LDA, which is intrinsically nonlinear, overcomes this difficulty and leads to satisfactory results.

https://dx.doi.org/10.1021/acs.jpclett.0c00535
J. Phys. Chem. Lett. 2020, 11, 2998−3004

We run short unbiased trajectories in the two basins, using as descriptors the 45 distances between heavy atoms, and with the data thus accumulated we train a Deep-LDA CV (see SI-3). We developed an interface to load the Pytorch model in the open-source plug-in PLUMED2[30] that allows using the Deep-LDA CV with enhanced sampling methods. The code and the input files needed to reproduce the simulations are openly available in the PLUMED-NEST repository with plumID:20.004.

Enhancing the fluctuations of the Deep-LDA CV leads to a highly diffusive behavior comparable to the simulations where the pair of Ramachandran angles is biased (see Figure SI-3.1). In Figure 2a we report the FES along the Deep-LDA CV.

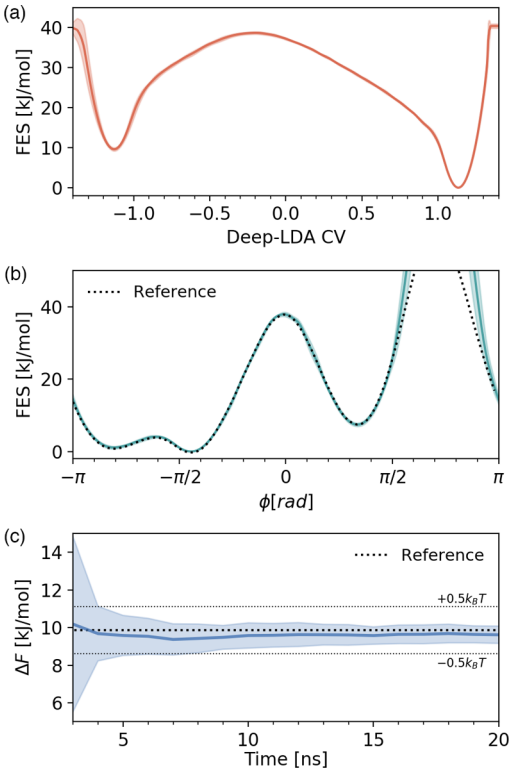


Figure 2. We show here alanine dipeptide convergence tests in which we run 10 independent simulations with OPES and the Deep-LDA CV, from which we extract the mean (solid line) and the standard deviation (shadow region). In panel a, we report the FES along the Deep-LDA CV, calculated from eq 11. In panel b, we show the FES along the ϕ angle, computed with a reweighting technique (see eq SI-3.1). The dotted line represents the reference obtained when biasing directly ϕ and ψ. In panel c, we present the free-energy difference between the two metastable states as a function of time (see eq SI-3.2).

Furthermore, we compare the projection of the FES along ϕ with more standard calculations that use ϕ and ψ as CVs (see Figure 2b). The full landscape in terms of ϕ and ψ is reported in Figure SI-3.3. We then monitor the free-energy difference between the two basins as a function of time (Figure 2c) and observe a rapid convergence to the reference value. These results show that Deep-LDA is able to reproduce the two dihedrals which represent the slowest degrees of freedom.

To illustrate this point further, we present in Figure 3 the isolines of the Deep-LDA CV on top of the FES, both projected onto ϕ and ψ (see also Figure SI-3.2). As there is no direct correspondence between the dihedral angles and the CV

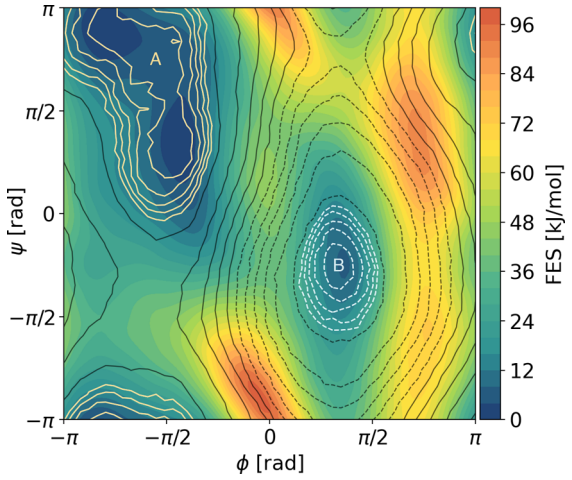


Figure 3. We show the isolines of the conditional probability of the Deep-LDA CV with respect to the Ramanchadran angles, on top of a reference free-energy surface. The highlighted lines correspond to the regions of the two metastable states used for training, and they are spaced by 0.02. The black lines are extrapolated by the network and have a much larger spacing of 0.2. The isolines are solid or dashed depending on whether they are above or below zero.

s, the isolines have been computed from the conditional probability $p(s \mid \phi, \psi)$ in the statistical ensemble generated by an OPES calculation with a uniform target distribution in the (ϕ, ψ) space. Note that the training is performed using data from the two basins alone, and the isolines corresponding to the intermediate regions are extrapolated by the network. When the Deep-LDA fluctuations are enhanced by OPES, the system is driven along the direction perpendicular to the isolines of the CV. It is remarkable how well this direction is correlated to the free-energy path that connects the two basins.

Similar results were obtained with different NN architectures and regularization parameters, as well as by using well-tempered metadynamics (see Figure SI-3.5), demonstrating the robustness of the method on the choice of parameters and the enhanced sampling algorithm used.

Aldol Reaction. The second example is the aldol reaction between vinyl alcohol and formaldehyde. This reaction presents several pathways and products.[31] Here, for simplicity, we focus on the most probable one, which has a barrier of ∼150 kJ/mol, as estimated in static calculations. This represents a challenging test because it is not obvious whether the reaction's concerted mechanism can be captured with information coming exclusively from unbiased simulations of reactants and products. Once again, in the spirit of performing the computation in a blind manner, we build the descriptor set out of all the interatomic distances. We rely on the ability of Deep-LDA to combine them in a meaningful and optimal way and to point out the most relevant ones. Previous experience[17] has shown that a good input set for chemical reactions can be obtained from the interatomic distances r, using the following contact function:

$$c_{ij}(r) = \frac{1 - \left(\frac{r}{\sigma_{ij}}\right)^n}{1 - \left(\frac{r}{\sigma_{ij}}\right)^m} \tag{12}$$

where σ_{ij} are typical bonding lengths between atoms of species i and j.

https://dx.doi.org/10.1021/acs.jpclett.0c00535
J. Phys. Chem. Lett. 2020, 11, 2998−3004

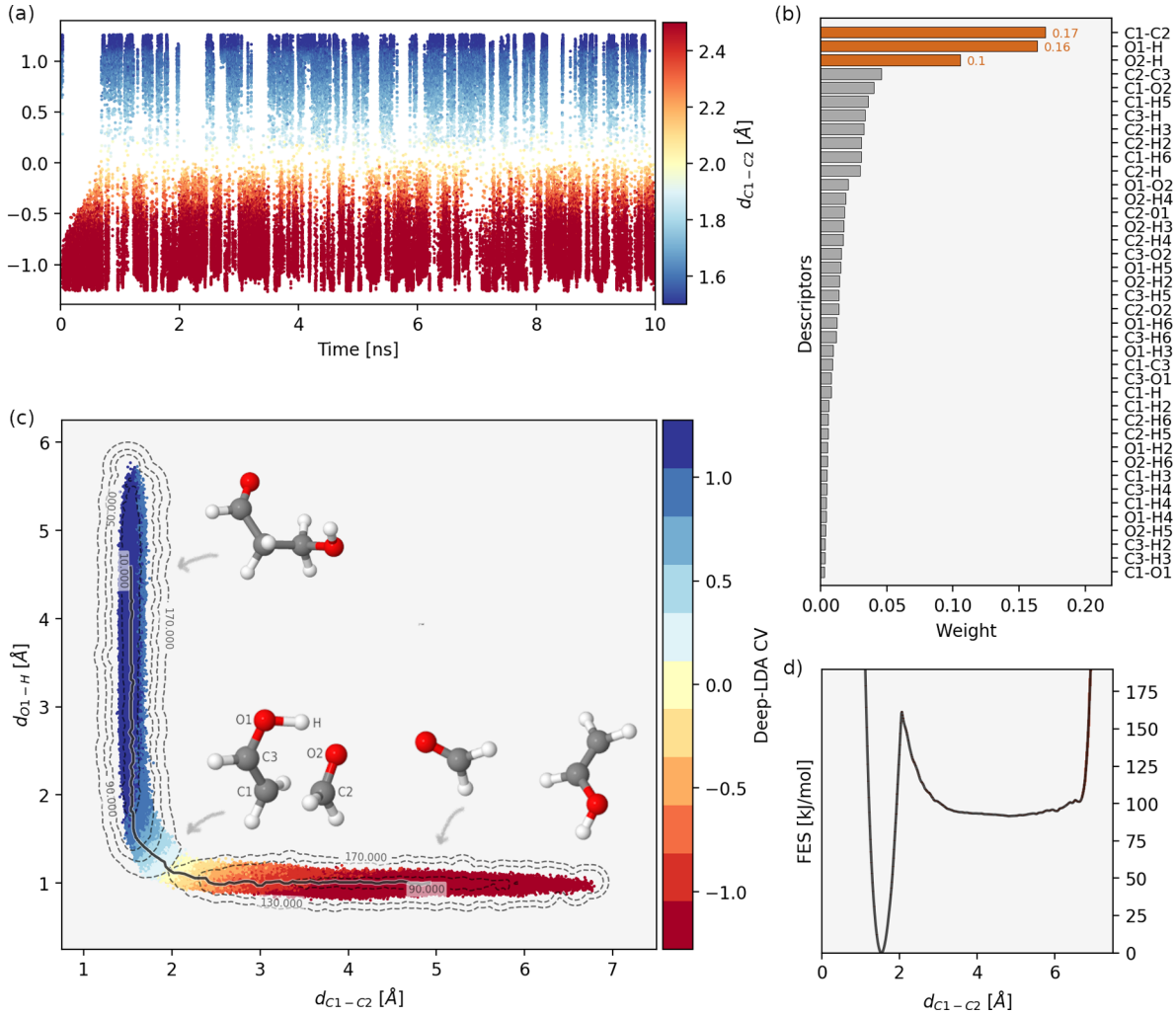


Figure 4. Results of the OPES simulation when enhancing the Deep-LDA CV. (a) Time evolution of the Deep-LDA CV. The points are colored according to the C1−C2 distance. (b) Features importance analysis, according to the magnitude of the weights of the first layer. The rankings are normalized so that their sum is equal to 1. The first three inputs, separated by a gap from the following ones, are colored in orange. (c) Distribution of the visited configurations in the plane of the C1−C2 and O1−H distances, colored by the Deep-LDA CV. The gray dashed lines correspond to the isolines of the free-energy surface projected onto this space, while the black solid one corresponds to the minimum free-energy path, computed using the nudged elastic band method.[33] (d) Free-energy surface projected along the C1−C2 distance, computed with a block average over the second part of the simulation, every 1 ns. The standard error is below 1 kJ/mol and is thus not visible.

This nonlinear preprocessing of the distances is crucial in the case of linear methods. In the present context, this step is not strictly necessary because of the nonlinearity provided by the NN. Nevertheless, it can be of help in focusing on the relevant degrees of freedom. As in the case of NN-based interatomic potentials,[32] the use of chemically informed descriptors facilitates learning from the data. In Figure SI-4.4 we show a simulation employing the distances as inputs of the Deep-LDA CV.

Having defined the input descriptors, we train the Deep-LDA network on unbiased simulations of the two states and use the resulting CV in combination with OPES to enhance the process (see SI-4 for the computational details). The time evolution of the Deep-LDA CV is reported in Figure 4a, which illustrates how the system is reversibly driven from reactants to products.

To assess the convergence of the simulation, we compute the FES along the C1−C2 distance (see inset of Figure 4c) with a block average and report the results in Figure 4d. The uncertainty is 0.25 kJ/mol in the regions of the metastable minima and about 1 kJ/mol close to the free-energy barrier.

In order to gain a physical understanding of the process, we perform a feature importance analysis on the NN. More precisely, we rank the features by summing the modulus of the weights between the input and the first layer, multiplied by the standard deviation of the inputs in the training set. Another option is to calculate the derivatives of the Deep-LDA CV with respect to each input, averaged over all inputs in the training set. We found that these two ways of estimating feature relevance produce similar results (see Figure SI-4.3). As shown in Figure 4b, there are three relevant descriptors, separated by a large gap from the following ones. These are the C1−C2 contact associated with the carbon−carbon bond together with the O1−H and O2−H contacts that characterize the proton transfer. This result has been obtained without any *a priori* information on the reaction pathway and it is in agreement with chemical intuition.

The FES projected on two of these variables has an easily understandable structure (Figure 4c and also Figure SI-4.2).

https://dx.doi.org/10.1021/acs.jpclett.0c00535
J. Phys. Chem. Lett. 2020, 11, 2998−3004

The two basins are clearly separated and the Deep-LDA CV shows different values in the two basins. As in the alanine dipeptide example, the isolines of Deep-LDA resemble quite closely the ones of the FES. This implies that the direction along which the system is driven is correlated with the minimum free-energy path, thus acting as a committor-like collective variable.

We introduce in this Letter a method that compresses the information from the metastable states into CVs. The method relies on a nonlinear dimensionality reduction, performed by a NN, followed by a linear transformation. This is achieved by maximizing the LDA objective function, which searches for the representation that best separates the states. We show that the method effectively draws a path in the descriptors' space. The Deep-LDA CV can be used in combination with any CV-based enhanced sampling method. We wish to remind the reader that our purpose here is not to find the ideal CV, but rather to build a variable that is good enough to promote transitions between the metastable states, given only a limited amount of information. We expect this approach to work best when the starting basins are adjacent in the phase space. If the system presents intermediate metastable states, they can be added iteratively as new classes for Deep-LDA, either by building a CV for each pair of states[17] or by using a multiclass approach.[16] This CV could be further refined to follow the reaction pathway closely, either by including weighted data from biased simulations or by combining it with methods designed to extract the slowest relaxation modes.[34,35] Another result is that the Deep-LDA features ranking can be used to identify the relevant descriptors in a data-driven way, as well as filtering them. For all these reasons we believe that our method could be of help in studying a large variety of rare events, including but not limited to chemical reactions, nucleation events, and ligand-binding processes.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.0c00535.

Details about the neural network optimization, the PLUMED interface with LibTorch, and the computational details about the simulations of alanine dipeptide and the aldol reaction; supplementary results, including a test of the robustness of the method with respect to the training parameters and the choice of the enhanced sampling method (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Michele Parrinello − *Department of Chemistry and Applied Biosciences, ETH Zurich, 8092 Zurich, Switzerland; Institute of Computational Sciences, Università della Svizzera italiana, 6900 Lugano, Switzerland; Italian Institute of Technology, 16163 Genova, Italy*; Email: parrinello@phys.chem.ethz.ch

Authors

Luigi Bonati − *Department of Physics, ETH Zurich, 8092 Zurich, Switzerland; Institute of Computational Sciences, Università della Svizzera italiana, 6900 Lugano, Switzerland;* orcid.org/0000-0002-9118-6239

Valerio Rizzi − *Department of Chemistry and Applied Biosciences, ETH Zurich, 8092 Zurich, Switzerland; Institute of Computational Sciences, Università della Svizzera italiana, 6900 Lugano, Switzerland;* orcid.org/0000-0001-5126-8996

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.0c00535

Notes

The authors declare no competing financial interest.

The results reported in this Letter are openly available in the Materials Cloud Archive (www.materialscloud.org).

■ ACKNOWLEDGMENTS

The authors thank Sandro Bottaro, Michele Invernizzi, and GiovanniMaria Piccini for carefully reading the paper; Manyi Yang for providing the inputs of the aldol reaction; and Riccardo Capelli and Emanuele Grifoni for useful discussions. This research was supported by the NCCR MARVEL, funded by the Swiss National Science Foundation, the Swiss National Science Foundation Grant No. 200021_169429/1 and the European Union Grant No. ERC-2014-AdG-670227/VARMET. The calculations were carried out on the Euler cluster of ETH Zurich.

■ REFERENCES

(1) Torrie, G.; Valleau, J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.

(2) Valsson, O.; Tiwary, P.; Parrinello, M. Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. *Annu. Rev. Phys. Chem.* **2016**, *67*, 159−184.

(3) Ma, A.; Dinner, A. R. Automatic Method for Identifying Reaction Coordinates in Complex Systems. *J. Phys. Chem. B* **2005**, *109*, 6769−6779.

(4) Ribeiro, J. M. L.; Bravo, P.; Wang, Y.; Tiwary, P. Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). *J. Chem. Phys.* **2018**, *149*, 072301.

(5) Chen, W.; Ferguson, A. L. Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration. *J. Comput. Chem.* **2018**, *39*, 2079−2102.

(6) Schöberl, M.; Zabaras, N.; Koutsourelakis, P.-S. Predictive collective variable discovery with deep Bayesian models. *J. Chem. Phys.* **2019**, *150*, 024109.

(7) Rogal, J.; Schneider, E.; Tuckerman, M. E. Neural-Network-Based Path Collective Variables for Enhanced Sampling of Phase Transformations. *Phys. Rev. Lett.* **2019**, *123*, 245701.

(8) Ravindra, P.; Smith, Z.; Tiwary, P. Automatic mutual information noise omission (AMINO): generating order parameters for molecular systems. *Mol. Syst. Des. Eng.* **2020**, *5*, 339−348.

(9) Chen, W.; Sidky, H.; Ferguson, A. L. Nonlinear discovery of slow molecular modes using state-free reversible VAMPnets. *J. Chem. Phys.* **2019**, *150*, 214114.

(10) Wehmeyer, C.; Noé, F. Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *J. Chem. Phys.* **2018**, *148*, 241703.

(11) Hernández, C. X.; Wayment-Steele, H. K.; Sultan, M. M.; Husic, B. E.; Pande, V. S. Variational encoding of complex dynamics. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2018**, *97*, 062412.

(12) Mendels, D.; Piccini, G.; Parrinello, M. Collective Variables from Local Fluctuations. *J. Phys. Chem. Lett.* **2018**, *9*, 2776−2781.

(13) Sultan, M. M.; Pande, V. S. Automated design of collective variables using supervised machine learning. *J. Chem. Phys.* **2018**, *149*, 094106.

(14) Peters, B. *React. Rate Theory Rare Events*; Elsevier, 2017; pp 1−619.

(15) Welling, M. *Fisher linear discriminant analysis*. Dep. Comput. Sci. Univ. Toronto, 2005.
(16) Piccini, G.; Mendels, D.; Parrinello, M. Metadynamics with Discriminants: A Tool for Understanding Chemistry. *J. Chem. Theory Comput.* **2018**, *14*, 5040−5044.
(17) Rizzi, V.; Mendels, D.; Sicilia, E.; Parrinello, M. Blind Search for Complex Chemical Pathways Using Harmonic Linear Discriminant Analysis. *J. Chem. Theory Comput.* **2019**, *15*, 4507−4515.
(18) Zhang, Y. Y.; Niu, H.; Piccini, G.; Mendels, D.; Parrinello, M. Improving collective variables: The case of crystallization. *J. Chem. Phys.* **2019**, *150*, 094509.
(19) Capelli, R.; Bochicchio, A.; Piccini, G.; Casasnovas, R.; Carloni, P.; Parrinello, M. Chasing the Full Free Energy Landscape of Neuroreceptor/Ligand Unbinding by Metadynamics Simulations. *J. Chem. Theory Comput.* **2019**, *15*, 3354−3361.
(20) Mendels, D.; Piccini, G.; Brotzakis, Z. F.; Yang, Y. I.; Parrinello, M. Folding a small protein using harmonic linear discriminant analysis. *J. Chem. Phys.* **2018**, *149*, 194113.
(21) Dorfer, M.; Kelz, R.; Widmer, G. Deep linear discriminant analysis. *4th Int. Conf. Learn. Represent. ICLR*. 2016.
(22) Andrew, G.; Arora, R.; Bilmes, J.; Livescu, K. *Deep canonical correlation analysis. 30th International Conference on Machine Learning*; ICML 2013. 2013; pp 2284−2292.
(23) Mardt, A.; Pasquali, L.; Wu, H.; Noé, F. VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* **2018**, *9*, 5.
(24) Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; Facebook, Z. D.; Research, A. I.; Lin, Z.; Desmaison, A.; Antiga, L.; et al. Automatic differentiation in PyTorch. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 8024−8035.
(25) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(26) Valsson, O.; Parrinello, M. Variational Approach to Enhanced Sampling and Free Energy Calculations. *Phys. Rev. Lett.* **2014**, *113*, 090601.
(27) Bonati, L.; Zhang, Y.-Y.; Parrinello, M. Neural networks-based variationally enhanced sampling. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 17641−17647.
(28) Invernizzi, M.; Parrinello, M. Rethinking Metadynamics: from Bias Potentials to Probability Distributions. *J. Phys. Chem. Lett.* **2020**, 2731.
(29) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(30) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(31) Yang, M.; Zou, J.; Wang, G.; Li, S. Automatic Reaction Pathway Search via Combined Molecular Dynamics and Coordinate Driving Method. *J. Phys. Chem. A* **2017**, *121*, 1351−1361.
(32) Carleo, G.; Cirac, I.; Cranmer, K.; Daudet, L.; Schuld, M.; Tishby, N.; Vogt-Maranto, L.; Zdeborová, L. Machine learning and the physical sciences. *Rev. Mod. Phys.* **2019**, *91*, 045002.
(33) Jonsson, H.; Mills, G.; Jacobsen, K. W. Nudged elastic band method for finding minimum energy paths of transitions. *Class. Quantum Dyn. Condens. Phase Simulations.* **1997**, 385−404.
(34) McCarty, J.; Parrinello, M. A variational conformational dynamics approach to the selection of collective variables in metadynamics. *J. Chem. Phys.* **2017**, *147*, 204109.
(35) Tiwary, P.; Berne, B. J. Spectral gap optimization of order parameters for sampling complex molecular systems. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 2839−2844.

https://dx.doi.org/10.1021/acs.jpclett.0c00535
J. Phys. Chem. Lett. 2020, 11, 2998−3004