

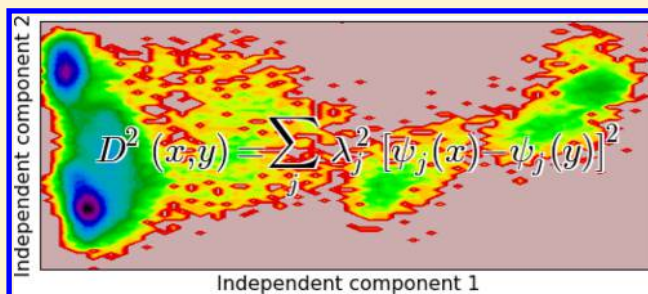
Kinetic Distance and Kinetic Maps from Molecular Dynamics Simulation

Frank Noé^{*,†} and Cecilia Clementi^{*,‡}

[†]FU Berlin, Department of Mathematics, Computer Science and Bioinformatics, Arnimallee 6, 14195 Berlin, Germany

[‡]Center for Theoretical Biological Physics, and Department of Chemistry, Rice University, Houston, Texas 77005, United States

ABSTRACT: Characterizing macromolecular kinetics from molecular dynamics (MD) simulations requires a distance metric that can distinguish slowly interconverting states. Here, we build upon diffusion map theory and define a kinetic distance metric for irreducible Markov processes that quantifies how slowly molecular conformations interconvert. The kinetic distance can be computed given a model that approximates the eigenvalues and eigenvectors (reaction coordinates) of the MD Markov operator. Here, we employ the time-lagged independent component analysis (TICA). The TICA components can be scaled to provide a kinetic map in which the Euclidean distance corresponds to the kinetic distance. As a result, the question of how many TICA dimensions should be kept in a dimensionality reduction approach becomes obsolete, and one parameter less needs to be specified in the kinetic model construction. We demonstrate the approach using TICA and Markov state model (MSM) analyses for illustrative models, protein conformation dynamics in bovine pancreatic trypsin inhibitor and protein-inhibitor association in trypsin and benzamidine. We find that the total kinetic variance (TKV) is an excellent indicator of model quality and can be used to rank different input feature sets.



INTRODUCTION

Characterizing the metastable (long-lived) states, their equilibrium probabilities and the transition rates between them are essential aims of molecular dynamics (MD) simulation. Because of the increasing availability of high-performance computing resources for mass production of MD data, there is an increasing interest in systematic and automatic construction of models of the metastable dynamics, such as Markov state models (MSMs)^{1–6} and diffusion maps.^{7,8} Alternative methods to analyze the space explored by MD simulations include sketch-map,⁹ PCA,¹⁰ kernel-PCA,^{11,12} and maximum-likelihood choice of reaction coordinates.¹³ (For a more extensive review, see ref 14.)

Critical components in any analysis of metastable dynamics from MD data are (1) the choice of suitable coordinates and (2) a distance metric on these coordinates, such that slowly interconverting states are distinguished. Conformation dynamics theory^{5,15–17} suggests a natural choice for the unsupervised reaction coordinates of the system—these are reaction coordinates without defining specific reactant and product end-states, which should be distinguished from supervised reaction coordinates between given reactants and products.¹³ The optimal unsupervised reaction coordinates are given by the eigenfunctions of the backward Markov propagator underlying the MD. Projecting the dynamics upon these eigenfunctions will give rise to a maximum estimate of the time scales^{17–19} and an optimal separation of metastable states. Several approaches are available to approximate these reaction coordinates from

MD data, including diffusion maps,⁸ Markov state models⁵ and Markov transition models,²⁰ time-lagged independent component analysis (TICA),^{21,22} and kernel TICA.²³ The variational approach for conformation dynamics (VAC)^{17,19} is a generalization to all aforementioned models, except for diffusion maps, and describes a general approach to combine and parametrize basis functions, to optimally define the true eigenfunctions of the backward propagator.

Different strategies have been proposed to address the second question, i.e., the choice of a distance metric.^{24–26} Here, we build on the results obtained in the context of diffusion maps. Diffusion maps model the observed data as emerging from a diffusion process.^{7,27} This approach has been further developed and used to model molecular conformation dynamics⁸ and to guide further sampling.²⁸ The diffusion distance, introduced in refs 7, 27, defines a distance metric that measures how slowly conformations interconvert in diffusion processes.

Here, we generalize the idea of the diffusion distance to Markov processes that have a unique equilibrium distribution, such as thermostated MD simulations. We define the *kinetic distance* as a measure of how slowly configurations interconvert in this more general setup. For reversible Markov processes, the kinetic distance takes the same simple form as the diffusion distance and can be computed from any model that provides an

Received: June 13, 2015

Published: September 2, 2015

approximation to the eigenvalues and eigenfunctions of the MD backward propagator. Here, we employ the TICA methodology,^{21,22} which approximates these eigenvalues and eigenfunctions in terms of a linear combinations of molecular coordinates such as atomic positions, distances, or angles. While TICA does not provide the best possible approximation to the true eigenfunctions, it can be directly applied to molecular dynamics data, requires few modeling decisions to be made, and is readily available in the Markov modeling packages PyEMMA (www.pyemma.org)²⁹ and MSMbuilder.³⁰

Using the TICA eigenvalues and eigenvectors, we employ the kinetic distance as a distance metric. As in diffusion maps, the approximate eigenvectors can be rescaled to define a kinetic map in which the Euclidean distance is equivalent to the kinetic distance. This provides an optimal space to perform clustering, Markov modeling, and diffusion map, or to use the VAC processes and perform other analyses of the metastable dynamics. The kinetic map gives a clear answer to the previously arbitrary decision on how many dimensions should be kept in the TICA transformation: In principle, all dimensions are maintained; however, as a result of the scaling, the dimensions with small eigenvalues contribute very little to the kinetic distance. In order to reduce the computational cost, we can choose to keep only the eigenvectors that account for a certain fraction of the total variation in kinetic distance, as it is commonly practiced in principal component analysis (PCA).^{31,32}

We demonstrate the approach using TICA and MSM analyses for illustrative models, protein conformation dynamics in bovine pancreatic trypsin inhibitor and protein–inhibitor association in trypsin and benzamidine. By invoking the Variational Principle of Conformation Dynamics¹⁷ we show that the total kinetic variance (TKV) that arises naturally from kinetic maps is an adequate way of comparing kinetic models. It is shown that the kinetic map generally provides better MSMs than by using unscaled TICA with manually selected dimension.

THEORY

Diffusion Distance and Kinetic Distance. Suppose that we have a dynamical system (here molecular dynamics) with a state space Ω . Ω formally includes positions, momenta, and, if needed, other state variables such as the simulation box size, although we will later make approximations, such as working with only a subset of the configuration coordinates. Our dynamics generates a time sequence of states $\mathbf{x} \in \Omega$ by means of a Markovian algorithm, i.e., some implementation of time-step integrator, thermostat, etc. that allows us to compute the system state in the subsequent time step as a function of the current state. In this framework, there is a transition density $p_\tau(\mathbf{y}|\mathbf{x})$, the probability density of finding the system at state \mathbf{y} at time τ given that we have started it at state \mathbf{x} at time 0. Given these preliminaries, we can write the propagation of a probability density of states $\rho_t(\mathbf{x})$ in time as

$$\rho_{t+\tau}(\mathbf{y}) = \int_{\mathbf{x} \in \Omega} \rho_t(\mathbf{x}) p_\tau(\mathbf{y}|\mathbf{x}) d\mathbf{x} \quad (1)$$

$$= \mathcal{P} \circ \rho_t(\mathbf{x}) \quad (2)$$

where \mathcal{P} is the dynamical propagator (i.e., the Markov operator that describes the action of the integral in eq 1). Finally, we require that there is a unique equilibrium distribution (usually the Boltzmann distribution) defined by

$$\pi(\mathbf{x}) = \mathcal{P} \circ \pi(\mathbf{x}) \quad (3)$$

The above requirements are minimal, since they are fulfilled by almost every implementation of molecular dynamics. There is an alternative description to eq 1 by using the backward propagator, \mathcal{T} , (also known as transfer operator¹⁵), which propagates the weighted densities $v_t(\mathbf{x}) = \rho_t(\mathbf{x})/\pi(\mathbf{x})$. Thus,

$$v_{t+\tau}(\mathbf{y}) = \frac{1}{\pi(\mathbf{y})} \int_{\mathbf{x}} p_\tau(\mathbf{y}|\mathbf{x}) \pi(\mathbf{x}) v_t(\mathbf{x}) d\mathbf{x} \quad (4)$$

$$= \mathcal{T} \circ v_t(\mathbf{x}) \quad (5)$$

We define the kinetic distance $D_\tau(\mathbf{x}_1, \mathbf{x}_2)$ as the diffusion distance introduced in ref 27 in the more general context of the above irreducible Markov processes. $D_\tau(\mathbf{x}_1, \mathbf{x}_2)$ is a distance measure between any two states $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$ and is parametrically dependent on a lag time τ . $D_\tau(\mathbf{x}_1, \mathbf{x}_2)$ is defined as the distance between the system's probability densities at time τ , given that we have initialized the system either at state \mathbf{x}_1 or at state \mathbf{x}_2 at time 0:

$$D_\tau^2(\mathbf{x}_1, \mathbf{x}_2) = \left\| p_\tau(\mathbf{y}|\mathbf{x}_1) - p_\tau(\mathbf{y}|\mathbf{x}_2) \right\|_{\pi}^2 \quad (6)$$

$$= \int_{\mathbf{y} \in \Omega} \frac{|p_\tau(\mathbf{y}|\mathbf{x}_1) - p_\tau(\mathbf{y}|\mathbf{x}_2)|^2}{\pi(\mathbf{y})} d\mathbf{y} \quad (7)$$

The above definition is identical to the concept of the diffusion distance except for the fact that, in principle, we allow $\mathbf{x}_1, \mathbf{x}_2$, and \mathbf{y} to be not only positions but also momenta and other state variables. This modification is needed in order to derive expressions of the kinetic distance that also apply to nonreversible dynamics. Furthermore, the original name “diffusion distance” comes from the fact that it has been derived in a context where the dynamics $p_\tau(\mathbf{y}|\mathbf{x})$ come from a diffusion process. Here, we will apply 6 to a wider class of Markovian dynamics, and therefore use the term “kinetic distance” rather than “diffusion distance” throughout the article.

Spectral Decomposition. In order to derive practically useful expressions of the kinetic distance, we need to conduct a spectral decomposition of 1. Suppose we are given the propagator \mathcal{P} with eigenfunctions ϕ_i and the backward propagator \mathcal{T} with eigenfunctions ψ_j , and we suppose that our propagator has a number of n discrete eigenpairs. The rest of the spectrum is bounded by the ball with radius $|\lambda_{n+1}|$ and it will be called $\mathcal{P}^{\text{fast}}$. We can write the propagation of densities ρ as follows:

$$\rho_{t+\tau}(\mathbf{y}) = \sum_{j=1}^n \lambda(\tau) \langle \psi_j(\mathbf{x}) | \rho_t(\mathbf{x}) \rangle \phi_j(\mathbf{y}) + \mathcal{P}^{\text{fast}}(\tau) \circ \rho_t(\mathbf{x}) \quad (8)$$

The norm of all eigenvalues decay exponentially with increasing lag time. We suppose that we operate at a lag time τ , such that $|\lambda_{n+1}(\tau)| \approx 0$ and, thus, $\mathcal{P}^{\text{fast}}(\tau) \circ \rho_t(\mathbf{x})$ is ~ 0 everywhere. Then, we can effectively describe the dynamics as

$$\rho_{t+\tau}(\mathbf{y}) = \sum_{j=1}^n \lambda(\tau) \langle \psi_j(\mathbf{x}) | \rho_t(\mathbf{x}) \rangle \phi_j(\mathbf{y}) \quad (9)$$

and for the choice $\rho_t(\mathbf{x}) = \delta(\mathbf{x})$, this becomes

$$\rho_{t+\tau}(\mathbf{y}) = \sum_{j=1}^n \lambda(\tau) \psi_j(\mathbf{x}) \phi_j(\mathbf{y}) \quad (10)$$

We sort the eigenvalues by nonincreasing norm:

$$\lambda_1 = 1 > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n| \quad (11)$$

The eigenfunction $\varphi_1(\mathbf{x}) = \pi(\mathbf{x})$ corresponding to the eigenvalue $\lambda_1 = 1$ is given by the Boltzmann equilibrium distribution.

Note that we have not made any restriction with respect to the reversibility of the dynamics as it is often done in Markov modeling.⁵ Thus, the eigenvalues $\lambda(\tau)$ can either be all real-valued, or consist of a mix of real eigenvalues and complex conjugate pairs. For the calculation of the kinetic distance, these cases must be treated slightly differently.

Computing Kinetic Distances and Kinetic Maps. Nonreversible Dynamics. Inserting eq 10 into eq 6 yields

$$D_\tau^2(\mathbf{x}_1, \mathbf{x}_2) = \left\| \sum_{j=1}^n \lambda_j(\tau) (\psi_j(\mathbf{x}_1) - \psi_j(\mathbf{x}_2)) \phi_j(\mathbf{y}) \right\|_{\pi^{-1}}^2 \quad (12)$$

As a result of $\psi_1(\mathbf{x}) \equiv 1$, the first term with $j = 1$ disappears. Evaluating the square norm leads to the following double sum:

$$D_\tau^2(\mathbf{x}_1, \mathbf{x}_2) = \sum_{j=2}^n \sum_{k=2}^n (\psi_j(\mathbf{x}_1) - \psi_j(\mathbf{x}_2)) \lambda_j \langle \phi_j | \phi_k \rangle_{\pi^{-1}}^{-1} \lambda_k (\psi_k(\mathbf{x}_1) - \psi_k(\mathbf{x}_2)) \quad (13)$$

The above expression is applicable to any Markov process with a unique stationary distribution. Unfortunately, eq 13 requires both the propagator and the backward propagator eigenvectors, and there are only few models for nonreversible dynamics that provide both sets of eigenvectors. Markov state models provide both eigenvectors (as left and right eigenvectors of the transition matrix). Moreover, it is possible to design Markov transition models,²⁰ such that both sets of eigenvectors can be computed.

For nonreversible dynamics, there can be complex eigenvalues and eigenvectors that come in complex conjugate pairs $(\lambda_j, \phi_j, \psi_j)$ and $(\bar{\lambda}_j, \bar{\phi}_j, \bar{\psi}_j)$. These are handled as follows. A complex conjugate pair is rewritten into a real pair of eigenvalues/eigenvectors by separating their real (Re) and imaginary (Im) parts:

$$\lambda_j, \bar{\lambda}_j \rightarrow \text{Re}(\lambda_j), \text{Im}(\lambda_j) \quad (14)$$

$$\phi_j, \bar{\phi}_j \rightarrow \text{Re}(\phi_j), \text{Im}(\phi_j) \quad (15)$$

$$\psi_j, \bar{\psi}_j \rightarrow \text{Re}(\psi_j), \text{Im}(\psi_j) \quad (16)$$

Note that the sign of the eigenvector has no effect on the kinetic distance, so it does not matter if we use $(\lambda_j, \phi_j, \psi_j)$ or $(\bar{\lambda}_j, \bar{\phi}_j, \bar{\psi}_j)$ for the imaginary part. The transformed eigenvalue/vector set, which is still the same rank but now real-valued, is inserted into eq 13 to compute the diffusion distance.

Reversible Dynamics. For reversible dynamics, the detailed balance equations hold. In that case, every pair of points \mathbf{x}, \mathbf{y} satisfies the relation

$$\pi(\mathbf{x}) p_\tau(\mathbf{y}|\mathbf{x}) = \pi(\mathbf{y}) p_\tau(\mathbf{x}|\mathbf{y}) \quad (17)$$

In this case, all eigenvalues $\lambda_i(\tau)$ are real and can be associated with relaxation rates κ_i and time scales t_i as

$$\lambda_i(\tau) = e^{-\tau\kappa_i} = e^{-\tau/t_i} \quad (18)$$

Moreover, all associated eigenvectors ψ_i, ϕ_i are real and related by the expression

$$\phi_i(\mathbf{x}) = \pi(\mathbf{x}) \psi_i(\mathbf{x}) \quad (19)$$

Given eq 19, we have $\langle \phi_j | \phi_k \rangle_{\pi^{-1}} = \delta_{jk}$, where δ_{jk} is the Dirac delta. Thus, only the $j = k$ terms survive in eq 10, which simplifies to

$$D_\tau^2(\mathbf{x}_1, \mathbf{x}_2) = \sum_{j=2}^n (\lambda_j \psi_j(\mathbf{x}_1) - \lambda_j \psi_j(\mathbf{x}_2))^2 \quad (20)$$

which is identical to the expression for the diffusion distance.²⁷

Note, however, that we can apply it to any dynamics that is reversible in the state space used in our model. Diffusion maps employ Smoluchowski or Brownian dynamics, which are reversible. Langevin dynamics, which is more commonly used for thermostating MD simulations, fulfills a generalized detailed balance in phase space, with respect to momentum inversion. Moreover, it has been recently shown that Langevin dynamics fulfills eq 17 in position space when integrating over momenta.³³ Since kinetic models for estimating eigenvalues and eigenvectors are usually estimated in position space (or a subset thereof), Langevin dynamics is consistent with the reversible diffusion distance (eq 20).

Based on eq 20, we can define the weighted coordinates

$$\tilde{\psi}_i(\mathbf{x}) = \lambda_i(\tau) \psi_i(\mathbf{x}) \quad (21)$$

which define the kinetic map:

$$\tilde{\Psi} = (\tilde{\psi}_1, \dots, \tilde{\psi}_n)^T \quad (22)$$

The kinetic map is a new set of coordinates in which data points \mathbf{x} have been transformed such that their Euclidean distance corresponds to their diffusion distance:

$$D_\tau^2(\mathbf{x}_1, \mathbf{x}_2) = \|\tilde{\Psi}(\mathbf{x}_1) - \tilde{\Psi}(\mathbf{x}_2)\|^2 \quad (23)$$

When the transition density $p(\mathbf{y}|\mathbf{x})$ originates from a diffusion process, then $\tilde{\Psi}(\mathbf{x})$ is called a diffusion map.²⁷

If the number of available eigenvectors n is large, then evaluating distances in eq 23 is computationally costly. However, this is often unnecessary, because many eigenvalues may be small and the corresponding dimensions in the kinetic map contribute little to the overall distance. We can employ an approach commonly used in PCA with geometric distances: We compute the variance of the kinetic distance along each coordinate:

$$\langle \tilde{\psi}_i, \tilde{\psi}_i \rangle_\pi = \lambda_i^2(\tau) = e^{-2\tau/\kappa_i} \quad (24)$$

The total kinetic variance (TKV) explained by all coordinates is given by

$$\text{TKV} = \sum_{i=2}^n \lambda_i^2(\tau)$$

while the cumulative variance fraction (remember that the eigenvalues are sorted by decreasing norm) is given by

$$c_k = \frac{\sum_{i=2}^k \lambda_i^2(\tau)}{\text{TKV}} \quad (25)$$

We can then decide to truncate the distance after $k = K$ terms where a certain cumulative variance threshold (e.g., 95%) is achieved. The approximate diffusion distance is then given by

$$D_{\tau}(\mathbf{x}_1, \mathbf{x}_2) \approx \sqrt{\sum_{i=2}^K (\tilde{\psi}_i(\mathbf{x}_1) - \tilde{\psi}_i(\mathbf{x}_2))^2} \quad (26)$$

Algorithmic Approach Using TICA. In order to apply the kinetic distance and kinetic maps, we need an algorithm that will approximate the eigenvalues λ_i and eigenfunctions ψ . As described in the [Introduction](#), several methods are available for this. Here, we choose to use the TICA method as implemented in PyEMMA. TICA is very easy to use, very robust, and has few parameters to choose. Although TICA provides only a rather rough approximation to individual eigenvalues and eigenfunctions, the dominant space of eigenfunctions is generally well-represented. We will subsequently build MSMs on TICA subspaces and TICA-approximated kinetic maps in order to evaluate the effect of the kinetic map scaling. Please note that we explicitly recommend against using TICA as an approximation method to the kinetic map without the subsequent MSM step (see the concluding discussion).

Given MD data, we chose a (usually large) set of input coordinates $\{r_i(t)\}$, such as Cartesian coordinates (if there is a reference to orient the solute molecule(s) to) or internal coordinates (inter-residue distances, rotamer dihedral angles, etc.). We define the mean-free coordinates as

$$y_i(t) = r_i(t) - \langle r_i(t) \rangle_t \quad (27)$$

and compute the covariance matrix and time-lagged covariance matrix for a given lag time τ :

$$c_{ij}(0) = \langle y_i(t)y_j(t) \rangle_t \quad (28)$$

$$c_{ij}(\tau) = \langle y_i(t)y_j(t + \tau) \rangle_t \quad (29)$$

In practice, means and covariances are computed by their empirical estimators, and the time-lagged covariance matrix is symmetrized. We then solve the generalized eigenvalue problem:

$$\mathbf{C}(\tau)\mathbf{r}_i = \mathbf{C}(0)\mathbf{r}_i\hat{\lambda}_i(\tau) \quad (30)$$

We have the following approximations for the eigenvalues and eigenfunctions of the backward propagator:

$$\hat{\lambda}_i \lesssim \lambda_i \quad (31)$$

$$\hat{\psi}_i = \sum_j r_{ij}y_j \approx \psi_i \quad (32)$$

The approach above is a special case of the VAC methodology,^{17,19} using the choice defined by [eq 27](#) as a basis set. According to the variational principle,¹⁷ the eigenvalues $\hat{\lambda}_i$ will only be exactly λ_i if $\hat{\psi}_i = \psi_i$; otherwise, it will be underestimated. For finite data, there are statistical errors on top of these approximation errors, and therefore the underestimation is denoted as an approximation in [eq 31](#).

We then apply the following approach:

(1) Perform TICA and compute the eigenvectors $\hat{\psi}_i$ and eigenvalues $\hat{\lambda}_i$.

(2) Define the kinetic map by scaling all coordinates as

$$\tilde{\psi}_i = \lambda_i(\tau)\psi_i \quad (33)$$

(3) The kinetic distance is defined by

$$D_{\tau}(\mathbf{x}_1, \mathbf{x}_2) = \|\tilde{\psi}_i(\mathbf{x}) - \tilde{\psi}_i(\mathbf{x})\|_2 \quad (34)$$

Note that TICA already projects the stationary eigenvector ψ_1 , as a consequence of its construction with mean-free coordinates ([eq 27](#)). As a result, all approximated eigenvectors and eigenvalues are maintained and the diffusion distance ([eq 20](#)) runs through all terms 1, ..., n .

APPLICATIONS

We demonstrate the behavior of the proposed kinetic distance using time-series of two pedagogical dynamical systems and MD simulations of two protein systems. All analyses are run with PyEMMA (www.pyemma.org).

In all examples, we conduct the following data analysis:

(1) Transformation of the input coordinates to an approximation of eigenvalues and eigenfunctions ($\lambda_2, \psi_2, \dots, \lambda_n, \psi_n$) using TICA.

(2) Cluster discretization of (i) the transformed space Ψ , (ii) dimension-reduced versions of it (TICA projections) as previously practiced,^{21,22} and (iii) the full-dimensional kinetic map (scaled TICA transformation) are proposed here. For a given system, we use the same clustering method and an equal number of clusters for the same system, to get comparable results.

(3) Compute Markov models using these different discretization and compare them using the variational principle.¹⁷

In order to compare different Markov models, we employ the VAC approach,¹⁷ which states that the approximated eigenvalues computed via [eq 30](#) underestimate the true eigenvalues:

$$\hat{\lambda}_i(\tau) \leq \lambda_i(\tau) \quad (35)$$

and that equality is only obtained when the corresponding eigenfunction is correct ($\hat{\psi}_i \equiv \psi_i$). One consequence is that (i) the estimated relaxation time scales $t_i(\tau) = -\tau/\ln|\lambda_i|$ are also underestimated,

$$\hat{t}_i(\tau) \leq t_i(\tau)$$

and (ii) a metric that gives us larger estimates for the relaxation time scales (within statistical error) should be preferred. Moreover, we can conclude, from [eq 35](#), the following variational principle for the sum of squares of eigenvalues:

$$\sum_{i=2}^m \hat{\lambda}_i^2(\tau) \leq \sum_{i=2}^m \lambda_i^2(\tau) \quad (36)$$

We leave the term λ_1^2 out of the sum because λ_1 is always equal to 1 for every transition matrix and, therefore, does not help in selecting models. The left-hand side of [eq 36](#) is just the cumulative kinetic variance of the Markov model. If we compare Markov models of equal numbers of states, we can set m to include all eigenvalues of these Markov models and thus use their total kinetic variances (TKVs). The TKVs will compare the quality of the different projections/metrics (full TICA, truncated TICA, or scaled TICA/kinetic map), with respect to their ability to resolve the metastable dynamics using a fixed clustering approach.

In order to compute statistical significances, we employed the recently described Bayesian approach for reversible Markov models described in [ref 34](#) and implemented in PyEMMA.

Two-State Hidden Markov Models. We start by illustrating two pedagogical examples that are realized by Hidden Markov models (HMMs) with Gaussian output. In the first example, our approach works especially well; however, in

the second example, it fails, because of a poor TICA approximation.

In order to have a known reference for the time scales, we fix a transition matrix between two metastable states:

$$P = \begin{bmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{bmatrix}$$

which leads to a single relaxation time scale of $t_2 = 49.5$ steps. In order to generate coordinates, we let each of the two states sample from two-dimensional Gaussian distributions, using means μ_i and covariance matrices Σ_i given by

$$\mu_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad \Sigma_1 = \begin{pmatrix} 0.3^2 & \\ & 2^2 \end{pmatrix}$$

$$\mu_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \Sigma_2 = \begin{pmatrix} 0.3^2 & \\ & 2^2 \end{pmatrix}$$

In order to analyze systematic rather than statistical behavior, we simulated this model for an extensive 250 000 steps, starting from state 1. In every simulation step, we take a step using the transition matrix P (upon average transitioning to the other state every 100 steps), and generate a point from Gaussian distribution 1 or 2, depending which state we are in. Figure 1a shows a scatterplot for the resulting simulation. Additional details on how to construct a HMM can be found in ref 35.

In this example, TICA works especially well. Figure 1a shows that while the main geometric variance is along the x -direction (PCA would find that as the principal component), the slow process is along the y -direction. Indeed, the first independent

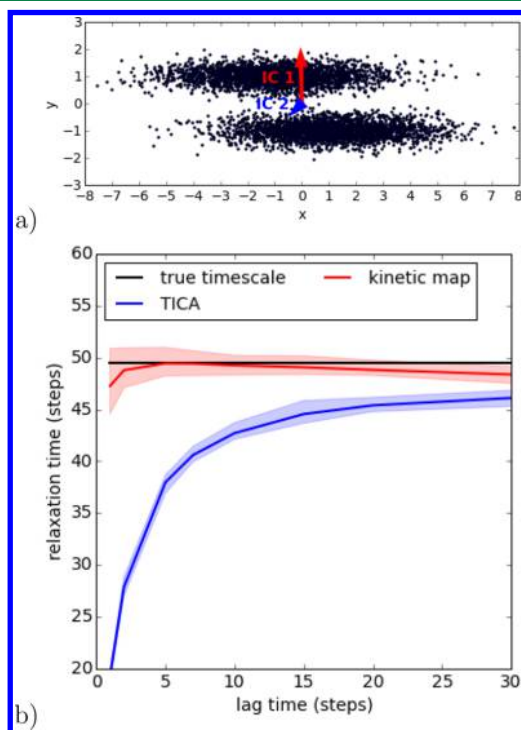


Figure 1. Comparison of Markov models using scaled and unscaled TICA. (a) The data was generated by a two-state Hidden Markov model with Gaussian output distributions. (b) Implied relaxation time scales using unscaled TICA and the kinetic map followed by regular-space clustering with 28 clusters each; 95% error intervals are shown as shaded regions.

component (IC) points exactly along the y -direction, while the second IC points toward a mixture of x - and y -directions (note that the ICs are not orthogonal in Cartesian space but rather in the space weighted by the equilibrium distribution). The arrows are drawn proportional to the TICA eigenvalues, which are $\lambda_1 \approx 0.75$ and $\lambda_2 \approx 0$.

We compare two metrics: Euclidean distance in the two-dimensional TICA space $\Psi = (\hat{\psi}_1, \hat{\psi}_2)^T$ and the kinetic map $\tilde{\Psi} = (\hat{\lambda}_2 \hat{\psi}_1, \hat{\lambda}_2 \hat{\psi}_2)^T$. In this example, the correlation coefficient between true and TICA-approximated kinetic distances is 0.92.

We conduct regular space clustering⁵ using 27 clusters in both cases. Figure 1b shows the mean relaxation time scales as a function of the lag time for the two cases. It is seen that the Euclidean distance in the kinetic map performs much better than the unscaled TICA space. The reason for this behavior is that the second eigenvalue is ~ 0 , and thus the kinetic map effectively ignores $\hat{\psi}_2$. Hence, in the kinetic map version, the 27 clusters perform a fine discretization of the slow reaction coordinate, which has been accurately found by TICA. In contrast, in the unscaled TICA version, the 27 clusters are scattered over a two-dimensional space, leading to a much poorer resolution of the reaction coordinate y . Thus, the kinetic map is better because TICA has already identified the right coordinates—all we needed to do was adjust the metric.

Let us look at a counterexample, where the current approach fails. Although the theory of kinetic distance and kinetic map is correct, the crucial point is that we must be able to generate a sufficiently good approximation of eigenfunctions and eigenvalues (λ_i, ψ_i) in order to apply it. TICA does that, by finding a linear combination of input coordinates, and will fail when the true eigenfunctions ψ_i are still highly nonlinear in these coordinates. So let us design a pathological example. We use the four-state transition matrix

$$P = \begin{bmatrix} 0.9 & 0.1 & & \\ 0.1 & 0.89 & 0.01 & \\ & 0.01 & 0.89 & 0.1 \\ & & 0.1 & 0.9 \end{bmatrix}$$

and define Gaussian distributions that output into two-dimensional Cartesian space:

$$\mu_1 = \begin{pmatrix} -4 \\ 0 \end{pmatrix} \quad \Sigma_1 = \begin{pmatrix} 0.2^2 & \\ & 1 \end{pmatrix}$$

$$\mu_2 = \begin{pmatrix} -1 \\ 0.5 \end{pmatrix} \quad \Sigma_2 = \begin{pmatrix} 1 & \\ & 0.1^2 \end{pmatrix}$$

$$\mu_3 = \begin{pmatrix} 1 \\ 0.5 \end{pmatrix} \quad \Sigma_3 = \begin{pmatrix} 1 & \\ & 0.1^2 \end{pmatrix}$$

$$\mu_4 = \begin{pmatrix} 4 \\ 0 \end{pmatrix} \quad \Sigma_4 = \begin{pmatrix} 0.2^2 & \\ & 1 \end{pmatrix}$$

The simulation is run again for 250 000 steps, and the coordinates from the TICA transformation or the kinetic map are discretized with 25 clusters using k -means.

Figure 2a shows the distribution of simulated points. The slow transition occurs between the two interlaced “T” motives and requires a zigzag path. Thus, the reaction coordinate is highly nonlinear in the given input coordinates and TICA cannot find a linear combination that approximates the true reaction coordinate ψ_2 well. Consequently, both TICA

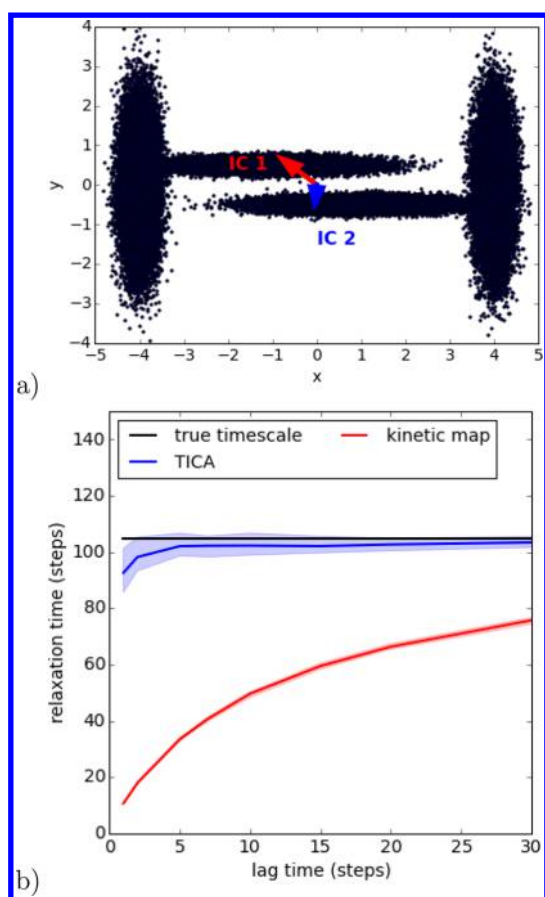


Figure 2. Comparison of Markov models using scaled and unscaled TICA. (a) The data was generated by a two-state Hidden Markov model with Gaussian output distributions. (b) Implied relaxation time scales using unscaled TICA and the kinetic map followed by regular-space clustering with 28 clusters each; 95% error intervals are shown as shaded regions.

coordinates are needed in order to resolve the reaction coordinate. Unfortunately, TICA projects the stationary process ψ_2 out by construction and the corresponding approximated eigenvalue λ_2 becomes approximately zero. Since there are only two eigenvalues in this example, the second coordinate is lost. As a result, the kinetic map is poor and gives rise to a much poorer MSM than the discretization of the unscaled TICA coordinates (Figure 2b). In this example, the correlation coefficient between true and TICA-approximated kinetic distances is 0.69.

Note that this example is extremely pathological and is only supposed to show that there are rare combinations of poor choices in which the current approach can break down. The kinetic map has one dimension less than the original input space, as the first eigenfunction is constant (in TICA, this coordinate is projected out as a result of removing the mean in the basis functions (eq 27)). Since, in our example, TICA required both input coordinates for a good approximation of the reaction coordinate, a poor kinetic map was obtained. When a third input coordinate is added, this problem disappears.

Molecular problems are usually high-dimensional and are, therefore, not expected to cause the observed problem. However, a general lesson from this example is that kinetic map will only be a good approximation of the real kinetic map ($\lambda_i \psi_i$) _{$i=2,\dots,n$} and thus lead to near-optimal distance metric, if we

have a good approximation of the eigenvalues and eigenfunctions (λ_i , ψ_i). Although TICA and, more generally, the method of linear variation^{17,19} only find linear combinations of input coordinates, we can turn these methods into excellent approximators of nonlinear eigenfunctions by providing suitable input coordinates. In MD simulations, coordinates such as interaction distances, contacts, or torsion angles are expected to play a role in the optimal reaction coordinates. Fortunately, we do not need to make a restrictive choice of coordinates, but can simply add all promising coordinates to the input set and then run TICA or the method of linear variation in order to find good combinations.

Bovine Pancreatic Trypsin Inhibitor (BPTI). Let us turn to protein simulations. We analyzed a 1-ms simulation of BPTI produced by D. E. Shaw Research, using the Anton supercomputer³⁶ (see ref 36 for a detailed description of the simulation setup). The trajectory was subsampled every 10 ns, providing 100 000 frames that are sufficient for our analysis. We then used the 174 coordinates of the 58 C_α -atoms after aligning them to their means as an input dataset. We consider four metrics: (i) Euclidean metric in the full TICA space; (ii) projection onto the first two independent components (ICs); (iii) projection onto the first six ICs (in metrics (i) and (ii), gaps are found in the TICA time scales); and (iv) the kinetic map of all scaled TICA coordinates. *k*-means clustering with 100 clusters was used in all cases.

Figure 3a shows that using two and six ICs results in similar estimates for the two slowest time scales, but the subsequent, faster time scales are dramatically underestimated with two ICs (they are below the plotted range). In this case, TICA does an excellent job in approximating the two slowest eigenfunctions with the two first ICs, which consequently means that the faster processes associated with further eigenfunctions are lost by the projection onto the first two ICs. Using all TICA coordinates results in a larger estimate of the slowest time scale for large lag times, but has poorer convergence properties in τ , is worse in all faster time scales, and has much greater statistical error. Presumably, the reason for this poor performance is that the space is so high-dimensional that the 100 cluster centers cannot efficiently discretize it. The kinetic map results appear best for all time scales.

In order to get an idea of the effective dimensionality of the kinetic map, Figure 3b shows the cumulative kinetic variance as a function of the kinetic map dimension. 95% of the variance is obtained after only 13 dimensions, indicating that the data analysis pipeline can work with relatively low-dimensional data after the TICA step.

Figure 3c shows the first two dimensions of the BPTI kinetic map with correctly scaled coordinates. Three metastable states are apparent in this projection whose structures are depicted in Figures 3e.1–e.3. The slowest conformational transition between the pair (1, 2) and state (3) ($\sim 60 \mu\text{s}$) involves an outward motion of the loop around residue 10 (top right in the structure). The second-slowest transition ($\sim 20 \mu\text{s}$) involves minor concerted motions in the loop region and an exchange between an ordered set of structures (1) and a less ordered set of structures (2). Qualitatively, this analysis agrees with previous analyses of that system,^{36,37} but the relaxation time scale found here are larger than previously estimated. Following the variational principle this means a better model was found here.

Figure 3d finally shows the TKV of the four different setups, as a function of lag time τ . The curves exhibit 95% error

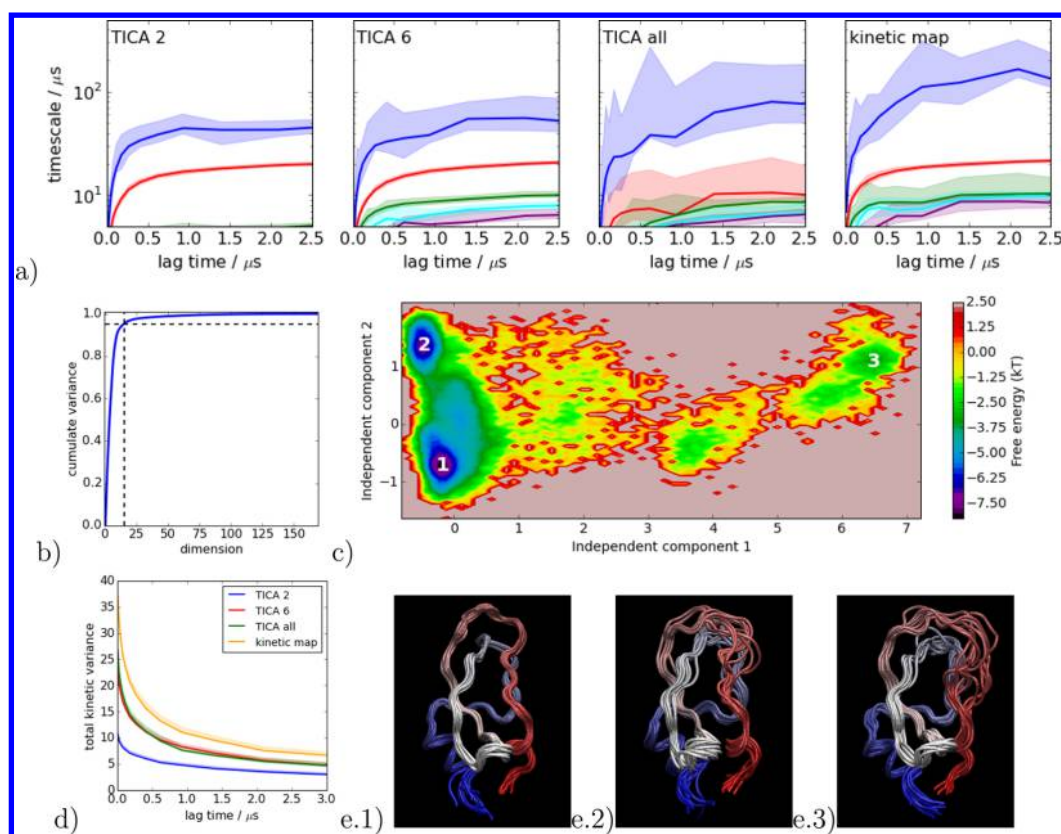


Figure 3. Comparison of Markov models of BPTI (1 ms Anton trajectory³⁶) using different TICA projections and the kinetic map (TICA using the C_{α} coordinates of oriented BPTI configurations; MSMs built based on projections onto two, six, and all dimensions and the kinetic map were compared): (a) the slowest implied relaxation time scales, shown as different colors with 95% error intervals as shaded regions; (b) cumulative variance of the diffusion distance; 95% is reached by using 13 eigenvectors; (c) kinetic map (first two dimensions); and (d) total kinetic variance (TKV), with 95% error intervals shown as shaded regions. (e) Structures of the metastable states as indicated in panel c (panel e.1 corresponds to region “1” in panel c, panel e.2 corresponds to region “2” in panel c, and panel e.3 corresponds to region “3” in panel c).

intervals that are very small, presumably because the significant uncertainties in individual eigenvalues (compare Figure 3a) are strongly coupled with each other. TICA with two coordinates performs worst as it can only resolve two slow processes, TICA with six and all coordinates perform similar. All three setups are clearly outperformed by the kinetic map approach.

Trypsin–Benzamidine. Finally, we analyzed protein–ligand association in benzamidine and trypsin using 491 trajectories of length 100 ns each that have been generated on GPUgrid³⁸ (see ref 38 for simulation details). The trajectories were saved every 100 ps, providing 491 000 frames for the analysis. As coordinates, we chose the distance between the C_{γ} atom of benzamidine with the C_{α} atoms of Trypsin, providing 223 distances. We consider four metrics: (i) Euclidean metric in the full TICA space, (ii) projection onto the first 2 ICs, (iii) projection onto the first 10 ICs, and (iv) the kinetic map of all scaled TICA coordinates. k -means clustering using 100 clusters was used in all cases.

Figure 4a shows that the MSMs using two dimensions converge nicely to time scales of ~ 500 and 50 ns. For larger number of ICs (10 and all), the results become worse. On one hand, 100 clusters are no longer sufficient to discretize these higher-dimensional spaces; on the other hand, it seems that a different second-slowest process is found when looking at higher numbers of dimensions. The kinetic map results show significantly larger time scales than any of the other metrics. These time scales do not clearly converge within the range of lagtimes shown, but it is known that the benzamidine

coordinates relative to trypsin are actually not sufficient to characterize the slowest processes in the system, which are comprised of trypsin conformational switches.^{39,40}

Figure 4b shows the cumulative kinetic variance as a function of the kinetic map dimension; 95% of the variance is obtained after 52 out of 233 dimensions, indicating that the data can be reduced by a factor of ~ 4 –5 with little losses.

Figure 4c shows the TKV of the four different setups, as a function of lag time (τ). Again, the 95% error intervals are very small and all differences are statistically significant. Interestingly, TICA with two coordinates performs worst overall. Although Figure 4a suggests that this setup gives a good approximation of the two slowest processes, it fails overall for the other processes. Again, the kinetic map clearly delivers the best overall MSM.

Figure 4d shows the first two dimensions of the trypsin–benzamidine kinetic map with correctly scaled coordinates. Three metastable states are apparent in this projection whose structures are depicted in Figures 4e.1–e.3. According to the MSM on the input coordinates used here, the slowest conformational transition is the binding unbinding transition, although we know that slower transitions exist in the protein conformation.³⁹ The second-slowest process involves exchange with a binding intermediate, where the ligand interacts with trypsin residues close to the binding site. Qualitatively, this analysis agrees with previous analyses of that system that used the trypsin coordinates,³⁸ but the kinetic map shows larger relaxation time scales. Following the variational principle, this means that a better model was found here.

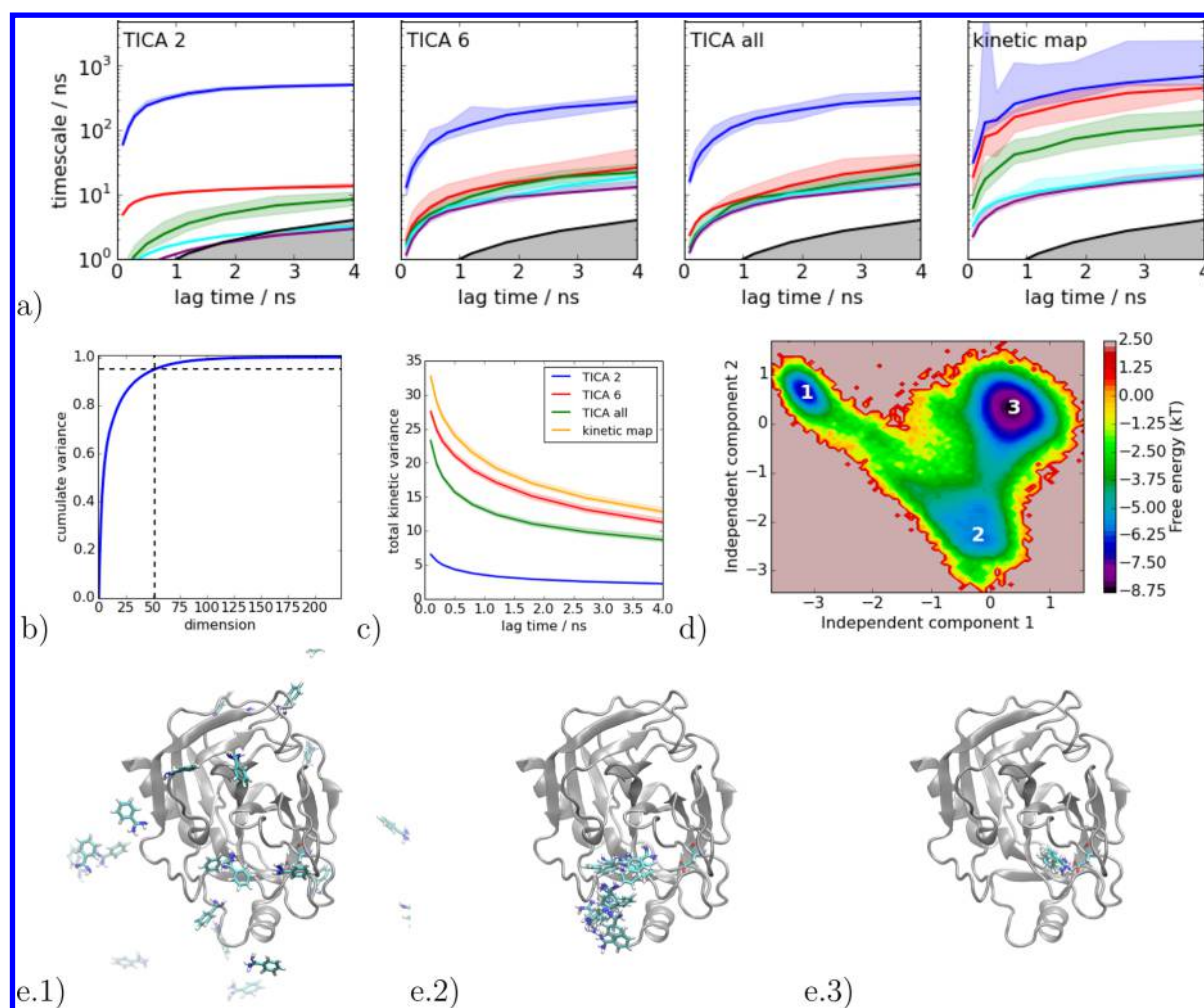


Figure 4. Comparison of Markov models of trypsin–benzamidin dynamics (491 GPUgrid trajectories of 100 ns each³⁸), using different TICA projections and scaled TICA. TICA using the distances from benzamidin to all trypsin- C_{α} coordinates had 135 usable eigenvalues. Projections onto 2, 5, 10 and all 135 dimensions (unscaled) were compared to scaled TICA: (a) the slowest implied relaxation time scales with 95% error intervals as shaded regions; (b) cumulative variance of the diffusion distance (95% is reached by using 50 eigenvectors); (c) total kinetic variance (TKV), with 95% error intervals as shaded regions; and (d) kinetic map (first two dimensions). (e) MD configurations sampled from the three metastable states visible in panel d (region 1, dissociated; region 2, prebound, region 3, bound).

DISCUSSION

The kinetic distance defined here is the optimal distance metric for analyzing metastable molecular dynamics. In practice, the kinetic distance can only be computed approximately and requires a method that approximates the eigenvalues and eigenfunctions of the Markov backward propagator underlying the MD simulation. Here, we suggest using time-lagged independent component analysis (TICA) as a fast and convenient method to transform a large set of Cartesian or internal coordinates of the molecule into such an approximation. Subsequently, the kinetic distance can be computed. The TICA coordinates can be transformed to a kinetic map by weighting them with their eigenvalues. In this kinetic map, kinetic distances and Euclidean distances are approximately equivalent, which makes it an excellent space for visualization and further analyses such as clustering, Markov modeling, or diffusion map.

We have shown that as long as pathological cases are avoided, the kinetic distance provides a distance metric that builds better Markov models, because the TICA components are weighted in such a way that the clustering algorithm can

optimally concentrate on the slow coordinates. Instead of projecting onto an arbitrary number of TICA dimensions as in previous work, or trying to select optimal values using machine learning techniques, the present theoretical insights lead to a unique and indisputable choice of using all coordinates in a weighted form. To reduce computational effort, a controlled truncation of the TICA space can be made by defining the percentage of the cumulative variation in kinetic distance (e.g., 95%).

We generally do not recommend using TICA as a method to approximate kinetic distances without the subsequent step of cluster-based MSMs. TICA can provide arbitrarily good or arbitrarily bad approximations to individual eigenfunctions and eigenvalues. Since TICA is a variational method, its approximation quality will be critically dependent on the choice of the basis set, i.e., the input features fed into TICA. If these input features can be linearly combined to form the true eigenfunctions, the TICA approximation is exact. If the eigenfunctions cannot be expressed as linear combinations of the input features, the kinetic map can be quite misleading (see the second application example). Nonetheless, the subsequent MSM step can recover from a poor kinetic map, because its

eigenfunctions are constant on the clusters, thus breaking the linear dependency between input features and eigenfunction approximations.

If a good direct approximation to the kinetic map is required, we suggest selecting a nonlinear method (for example, the Variational Approach for conformation dynamics (VAC) approach, in conjunction with product bases, diffusion maps, Markov transition models, or kernel-based TICA.

TICA-based kinetic maps are implemented in PyEMMA versions 1.2.2 or later. See www.pyemma.org for download instructions, documentation and examples.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: frank.noe@fu-berlin.de (F. Noé).

*E-mail: cecilia@rice.edu (C. Clementi).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Kyle Beauchamp (MSKCC New York), John Chodera (MSKCC New York), Benedict Leimkuhler (University of Edinburgh), and Alessandro Laio (SISSA Trieste) for inspiring discussions. Special thanks to Ben for organizing a great workshop, which was the starting point for this paper. We are grateful to Gianni De Fabritiis, for sharing the trypsin–benzamidine data, and D. E. Shaw Research, for sharing the BPTI data. Author C.C. is supported by the National Science Foundation (Grant Nos. CHE-1152344 and CHE-1265929) and the Welch Foundation (Grant No. C-1570). Author F.N. is supported by European Commission ERC (Starting Grant No. 307494-pcCell) and Deutsche Forschungsgemeinschaft (Grant No. 825/3-1).

REFERENCES

- (1) Swope, W. C.; Pitera, J. W.; Suits, F. Describing protein folding kinetics by molecular dynamics simulations: 1. Theory. *J. Phys. Chem. B* **2004**, *108*, 6571–6581.
- (2) Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C. Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states. *J. Chem. Phys.* **2007**, *126*, 155102.10.1063/1.2714539.
- (3) 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*, 155101.10.1063/1.2714538.
- (4) Buchete, N.-V.; Hummer, G. Coarse master equations for peptide folding dynamics. *J. Phys. Chem. B* **2008**, *112*, 6057–6069.
- (5) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B. G.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov models of molecular kinetics: Generation and Validation. *J. Chem. Phys.* **2011**, *134*, 174105.
- (6) Bowman, G. R.; Pande, V. S.; Noé, F., Eds. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*; Advances in Experimental Medicine and Biology, Vol. 797; Springer: Dordrecht, The Netherlands, 2014.
- (7) Coifman, R. R.; Lafon, S.; Lee, A. B.; Maggioni, M.; Nadler, B.; Warner, F.; Zucker, S. W. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 7426–7431.
- (8) Rohrdanz, M. A.; Zheng, W.; Maggioni, M.; Clementi, C. Determination of reaction coordinates via locally scaled diffusion map. *J. Chem. Phys.* **2011**, *134*, 124116.10.1063/1.3569857.
- (9) 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.

- (10) Altis, A.; Otten, M.; Nguyen, P. H.; Hegger, R.; Stock, G. Construction of the free energy landscape of biomolecules via dihedral angle principal component analysis. *J. Chem. Phys.* **2008**, *128*, 245102.
- (11) Pozun, Z. D.; Hansen, K.; Sheppard, D.; Rupp, M.; Muller, K.-R.; Henkelman, G. Optimizing transition states via kernel-based machine learning. *J. Chem. Phys.* **2012**, *136*, 174101.1–174101.9.
- (12) Antoniou, D.; Schwartz, S. D. Toward identification of the reaction coordinate directly from the transition state ensemble using the kernel PCA method. *J. Phys. Chem. B* **2011**, *115*, 2465–2469.
- (13) Peters, B.; Beckham, G. T.; Trout, B. L. Extensions to the likelihood maximization approach for finding reaction coordinates. *J. Chem. Phys.* **2007**, *127*, 034109.
- (14) Rohrdanz, M. A.; Zheng, W.; Clementi, C. Discovering Mountain Passes via Torchlight: Methods for the Definition of Reaction Coordinates and Pathways in Complex Macromolecular Reactions. *Annu. Rev. Phys. Chem.* **2013**, *64*, 295–316.
- (15) Schütte, C.; Fischer, A.; Huisinga, W.; Deuffhard, P. A Direct Approach to Conformational Dynamics based on Hybrid Monte Carlo. *J. Comput. Phys.* **1999**, *151*, 146–168.
- (16) Sarich, M.; Noé, F.; Schütte, C. On the approximation quality of Markov state models. *Multiscale Model. Simul.* **2010**, *8*, 1154–1177.
- (17) Noé, F.; Nüske, F. A variational approach to modeling slow processes in stochastic dynamical systems. *Multiscale Model. Simul.* **2013**, *11*, 635–655.
- (18) Prinz, J.-H.; Chodera, J. D.; Noé, F. Spectral rate theory for two-state kinetics. *Phys. Rev. X* **2014**, *4*, 011020.
- (19) Nüske, F.; Keller, B. G.; Pérez-Hernández, G.; Mey, A. S. J. S.; Noé, F. Variational Approach to Molecular Kinetics. *J. Chem. Theory Comput.* **2014**, *10*, 1739–1752.
- (20) Wu, H.; Noé, F. Gaussian Markov transition models of molecular kinetics. *J. Chem. Phys.* **2015**, *142*, 084104.
- (21) Perez-Hernandez, 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.
- (22) 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.
- (23) Schwantes, C. R.; Pande, V. S. Modeling Molecular Kinetics with tICA and the Kernel Trick. *J. Chem. Theory Comput.* **2015**, *11*, 600–608.
- (24) Cossio, P.; Laio, A.; Pietrucci, F. Which Similarity Measure Is Better for Analyzing Protein Structures in a Molecular Dynamics Trajectory? *Phys. Chem. Chem. Phys.* **2011**, *13*, 10421–10425.
- (25) McGibbon, R. T.; Pande, V. S. Learning Kinetic Distance Metrics for Markov State Models of Protein Conformational Dynamics. *J. Chem. Theory Comput.* **2013**, *9*, 2900–2906.
- (26) Zhou, T.; Caflisch, A. Distribution of Reciprocal of Interatomic Distances: A Fast Structural Metric. *J. Chem. Theory Comput.* **2012**, *8*, 2930–2937.
- (27) Nadler, B.; Lafon, S.; Kevrekidis, I. G.; Coifman, R. R. Diffusion Maps, Spectral Clustering and Eigenfunctions of Fokker-Planck Operators. *Adv. Neural Inf. Process. Syst.* **2005**, *18*, 955–962.
- (28) Preto, J.; Clementi, C. Fast recovery of free energy landscapes via diffusion-map-directed molecular dynamics. *Phys. Chem. Chem. Phys.* **2014**, *16*, 19181–19191.
- (29) Senne, M.; Trendelkamp-Schroer, B.; Mey, A. S. J. S.; Schütte, C.; Noé, F. EMMA—A software package for Markov model building and analysis. *J. Chem. Theory Comput.* **2012**, *8*, 2223–2238.
- (30) Beauchamp, K. A.; Bowman, G. R.; Lane, T. J.; Maibaum, L.; Haque, I. S.; Pande, V. S. MSMBuilder2: Modeling Conformational Dynamics at the Picosecond to Millisecond Scale. *J. Chem. Theory Comput.* **2011**, *7*, 3412–3419.
- (31) Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Edu. Psych.* **1933**, *24*, 417–441.
- (32) Amadei, A.; Linssen, A. B.; Berendsen, H. J. C. Essential dynamics of proteins. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412–225.

- (33) Bittracher, A.; Koltai, P.; Junge, O. Pseudo generators of spatial transfer operators. **2014**, *arXiv:1412.1733* (<http://arxiv.org/abs/1412.1733>).
- (34) Trendelkamp-Schroer, B.; Wu, H.; Paul, F.; Noé, F. Estimation and uncertainty of reversible Markov models. Submitted to *J. Chem. Phys.*, **2015** (preprint available at *arXiv:1507.05990*, <http://arxiv.org/abs/1507.05990>).
- (35) Rabiner, L. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **1989**, *77*, 257–286.
- (36) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R.; Eastwood, M.; Bank, J.; Jumper, J.; Salmon, J.; Shan, Y.; Wriggers, W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341–346.
- (37) Noé, F.; Wu, H.; Prinz, J.-H.; Plattner, N. Projected and Hidden Markov Models for calculating kinetics and metastable states of complex molecules. *J. Chem. Phys.* **2013**, *139*, 184114.
- (38) Buch, I.; Giorgino, T.; de Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 10184–10189.
- (39) Plattner, N.; Noé, F. Protein conformational plasticity and complex ligand binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun.* **2015**, *6*, 7653.
- (40) Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M. Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E386–E391.