

Nonparametric Determination of the Committor in Multimolecular Systems

Lair F. Trugilho,* Stefan Auer, Leandro G. Rizzi, and Sergei V. Krivov*

Cite This: <https://doi.org/10.1021/acs.jctc.5c01427>

Read Online

ACCESS |



Metrics & More

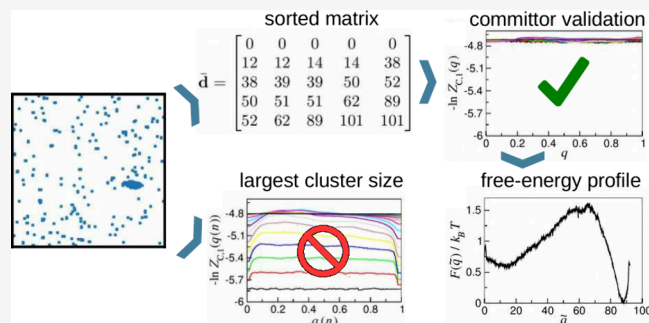


Article Recommendations



Supporting Information

ABSTRACT: A fundamental problem in analyzing large longitudinal data sets modeling dynamics in multimolecular systems is determining the underlying free-energy landscapes as a function of the committor, the optimal reaction coordinate. Here, we demonstrate that by combining a nonparametric approach with a systematic method for generating permutationally invariant collective variables, the committor can be effectively determined to describe multimolecular aggregation in a system with anisotropic interactions. The optimality of the committor is verified by a stringent validation test, and it is shown that the diffusive model along the committor yields kinetic properties identical to those derived from the original dynamics. Our method is general and relevant to the large machine learning community developing methods to determine the committor from longitudinal data sets.



Multimolecular systems undergoing transitions between states with distinct physical properties are ubiquitous in nature, and prominent examples include condensation of small molecules,¹ protein crystallization,² and macromolecular aggregation.^{3–5} The steady-state kinetics of transitions in those systems is usually described by projecting the very-high dimensional configurational space of the molecules onto a few reaction coordinates (RCs). The remaining large number of degrees of freedom are modeled as noise by a stochastic model, such as the diffusive models described by the Smoluchowski and Fokker–Planck-like equations.^{6,7} The RC choice, however, is one of the pivotal difficulties that may lead to discrepancies when comparing simulated transition rates with their experimentally measured counterparts.⁸

In the case of physically motivated RCs,^{9–11} the projection usually introduces memory effects, and because of their non-Markovian character, there is no guarantee that the corresponding diffusive models can be used to describe quantitatively the system's dynamics at arbitrary time scales.^{12,13} Alternatively, the formalism of optimal reaction coordinates¹⁴ selects RCs that minimize non-Markovian effects so that dynamics can be accurately described as simple diffusion on the corresponding free-energy landscape. For equilibrium stochastic dynamics between two boundary states, the committor function is considered an optimal RC.^{15,16} It is defined as the probability for the trajectory to reach one boundary state (B) before reaching the other one (A) starting from any given configuration.¹⁷ It has been shown that, despite the dynamics projected on the committor coordinate q being not strictly Markovian, a diffusive model along it can be used

to exactly compute the equilibrium flux $J = D(q)p_{\text{eq}}(q)$,^{15,18} the mean transition path times (MTPT) $\hat{\tau}$,¹⁴ and the mean first passage times (MFPT)^{15,19,20}

$$\tau_{B \rightarrow A} = \int_0^1 \frac{dq'}{p_{\text{eq}}(q')D(q')} \int_{q'}^1 p_{\text{eq}}(q'')dq'' \quad (1)$$

between any two points on the committor landscape. These results hold for arbitrarily complex equilibrium stochastic dynamics in configuration space and do not require a separation of time scales. Here $D(q)$ and $p_{\text{eq}}(q)$ are the temperature-dependent diffusion coefficient and equilibrium probability density function, respectively. While the description of the system's dynamics using the committor is rather simple since it only depends on the free-energy profile $F(q)/k_B T \propto -\ln p_{\text{eq}}(q)$, as well as the diffusion coefficient $D(q)$, its determination is an unsolved problem for multimolecular systems. In principle, the committor can be computed from the multidimensional Smoluchowski-like equation for diffusing particles,^{14,21} but due to its complexity, it has only been solved for low dimensional systems.²² In this letter, we consider the high-dimensional case of multimolecular aggregation, and

Received: August 26, 2025

Revised: October 9, 2025

Accepted: October 9, 2025

present a systematic way of constructing permutationally invariant collective variables (CVs), which can be used in nonparametric optimization to determine the committor. The optimization scheme has been validated previously in unimolecular systems, such as the analysis of large scale atomistic protein folding simulations.^{23,24}

The presented approach is timely, given the rapid development of computer hardware, which has led to an explosion of large, high-dimensional longitudinal data sets modeling transitions in multimolecular systems. These data sets require sophisticated analysis and interpretation. Recently, numerous machine learning approaches have been proposed to determine the committor, but they remain limited to relatively simple or low-dimensional systems.^{25–27} Other relevant developments include path sampling which uses machine learned committor estimates to improve sampling and to iteratively refine it.²⁸ Additionally, the committor-based bias potential introduced in ref 29 iteratively improves sampling of transition states leading to better committor representations. Independent of these efforts for constructing optimal RCs to describe transitions in multimolecular systems, there is a lack of well-established benchmark systems, which are easy to simulate yet challenging to analyze. Recent studies either considered complex systems with relatively simple RCs (e.g., ion association²⁸), or simple model systems (e.g., alanine dipeptide^{26,27,29} or Mueller potential^{25–27,29}). Here, we considered a system with a low simulation cost, but rather complex RC landscape. In this way, it can serve as one of such benchmark systems, and the described approach together with the stringent optimality criterion, provides state-of-the-art results for aggregating phenomena. As it is feasible to sufficiently sample this system with long equilibrium simulations, scarce sample problems addressed in refs 26–29 are not critical. Nevertheless, our methodology can be combined with nonequilibrium sampling using the nonequilibrium nonparametric approach of ref 30.

Here, we consider stochastic dynamics in a multidimensional configuration space $\vec{X}$, where we are interested in describing the dynamics of a reaction between two specified boundary states, A and B . To achieve this, we introduce a scalar RC, $r(\vec{X})$, which projects $\vec{X}$ onto a one-dimensional coordinate such that $r(\vec{X} \in A) = 0$ and $r(\vec{X} \in B) = 1$. In general, the dynamics projected onto this coordinate is non-Markovian due to the loss of information associated with the dimensionality reduction.^{12,13} This non-Markovian behavior can be accurately described using the generalized Langevin equation with a memory kernel, which can be derived via the Mori-Zwanzig formalism.³¹ However, computing the memory kernel for practical cases is often challenging. Neglecting memory effects, dynamics can be approximated using a simple diffusive model, where the free energy $F(r)$ and diffusion coefficient $D(r)$ are computed as functions of the RC from its time series $r(k\Delta t) = r(\vec{X}(k\Delta t))$ sampled with Δt (e.g., simulation step). Such models generally predict faster kinetics manifested as shorter MFPT, higher flux, and lower apparent free energy barriers.^{23,32–35}

To improve the accuracy of this description, one can optimize the RC to mitigate these non-Markovian effects by employing variational approaches.^{33–38} Consider, specifically, the nonparametric approach of RC optimization,^{23,24,30,39} which improves RC iteratively. It considers iterative variations of the putative RC time series, $r_{m+1}(\vec{X}(k\Delta t)) = r_m(\vec{X}(k\Delta t)) + \delta r(\vec{X}(k\Delta t))$, where

$\delta r(\vec{X}(k\Delta t))$ satisfies $\delta r(\vec{X}(k\Delta t) \in A) = \delta r(\vec{X}(k\Delta t) \in B) = 0$ to preserve boundary conditions. For optimizing the committor, we used the variation

$$\delta r(r_m, x_m) = \sum_{i=0}^l \sum_{j=0}^{l-i} a_{ij}(r_m)^i (x_m)^j \quad (2)$$

where $x_m = x_m(\vec{X}(k\Delta t))$ denotes a randomly chosen CV time series in the m -th iteration which will be described later. This definition is convenient as the optimal variational coefficients a_{ij} can be found analytically by minimizing the total squared displacement (TSD) functional,^{30,39} see also the [Supporting Information \(SI\)](#). We note that as the approach deals directly with the CV time series, there is no need to compute the full RC function $r(\vec{X})$, in this sense it is nonparametric, as opposed to most machine learning approaches, which try to parametrize the function $r(\vec{X})$ with, e.g., an artificial neural network. The convergence of such an optimization scheme can then be verified through the $Z_{C,1}$ validation criterion for the committor, which states that if the putative RC r closely approximates the committor q , function $Z_{C,1}(r, \Delta t)$ is constant, i.e., independent of r and Δt , and equal to the number of transitions N_{AB} between the specified boundary states.¹⁵ $Z_{C,1}(r, \Delta t)$ can be straightforwardly computed from the RC time-series $r(k\Delta t)$ as $\sum_k' |r(k\Delta t + \Delta t) - r(k\Delta t)|$, where the prime indicates that the sum is over all transitions such that r is between $r(k\Delta t + \Delta t)$ and $r(k\Delta t)$ ¹⁵ (further details in the [SI](#)).

It is worth noting that, assuming diffusive dynamics over a putative RC r , the diffusion coefficient can be accurately estimated as¹⁵

$$D(r, \Delta t) = \frac{Z_{C,1}(r, \Delta t)}{\Delta t Z_H(r, \Delta t)} \quad (3)$$

where the histogram $Z_H(r, \Delta t)$ is related to the equilibrium distribution $p_{\text{eq}}(r) \propto Z_H(r, \Delta t)$ and free-energy profile $F(r)/k_B T \propto -\ln Z_H(r, \Delta t)$. In this way, $F(r)/k_B T$ and $Z_{C,1}(r, \Delta t)$ completely determine the diffusive models over the chosen RC, whether it displays Markovian dynamics or not.⁴⁰ In particular, for the optimal coordinate q , the criteria that $Z_{C,1}(q, \Delta t)$ is independent of Δt ensures that the diffusion coefficient $D(q)$ given by eq 3 is also independent of Δt (this is because the non-normalized histogram scales as $Z_H(q, \Delta t) \sim \Delta t^{-1}$), just like it is for normal (i.e., Markovian) diffusion.

What remains is the task of identifying a suitable set of CVs $\{x_m\}$ for a multimolecular system with N molecules. In particular, such CVs need to respect the system's symmetries, which for multimolecular systems are invariance under translations, rotations and permutation of identical molecules.⁴¹ For the first two, one can just consider distances d_{ij} between molecules noting that, if periodic boundary conditions are used in simulations, the minimum image convention should be used. Permutational invariance has proven more difficult to ensure and here we propose a systematic way of constructing such invariant CVs: We take the full matrix of distances $\mathbf{d}$ with $N \times N$ d_{ij} elements and sort it (from small to large) sequentially along its two axis, i.e., first inside each column and then inside each row. In this way, the time series of the elements $\bar{d}_{ij} = \bar{d}_{ij}(k\Delta t)$ of the sorted matrix $\bar{\mathbf{d}}$ compose

the set of invariant CVs that can be used as x_m in eq 2 (see the SI for details).

To demonstrate the applicability of our approach for multimolecular aggregation, we consider a lattice system to compute its committor q along with the diffusive model, i.e., the free-energy profile $F(q)/k_B T$ and diffusion coefficient $D(q)$. As described in refs 40, 42, the system is defined as a regular two-dimensional square lattice of size $L = 200$ containing $N = 400$ molecules. Besides its position, each molecule has an additional degree of freedom that determines its orientation, which can be of two types. The interaction between neighbor molecules is then defined depending on their relative orientation, which can be stronger, ψ_s , for aligned molecules or weaker, ψ_w , for nonaligned ones. In this way, the quantity $\xi = \psi_s/\psi_w$ is a measure of the interaction anisotropy. Similarly to ref 40 we perform Metropolis Monte Carlo (MC) simulations in the canonical (NVT) ensemble at a temperature $T = T^*$ where the system undergoes a first-order phase transition between diluted and aggregated states (in the SI we indicate the value of T^* for different ξ). For the data production runs we performed six long equilibrium MC simulations, where each simulation consists of 10^8 MC steps (MCs), and configurations were saved at every $\Delta t_0 = 400$ MCs.

First, for the sake of comparison, we show results obtained using the number of molecules in the largest aggregate n as RC, which is commonly used in the context of the classical nucleation theory.⁴³ We computed the committor as a function of n , $q(n)$, to evaluate the optimality of n . To do so, we performed the optimization procedure given by eq 2 with $x_m = n(k\Delta t)$ for all iteration steps m until convergence is reached ($m = 20$). As the transformation $n \rightarrow q(n)$ does not change the RC optimality, the profile along $q(n)$ is representative of the original n RC.¹⁵ Moreover, using the $q(n)$ coordinate one can directly apply the committor validation test to see whether $q(n)$ approaches the true multidimensional committor function q . As can be seen in Figure 1 (a), $-\ln Z_{C,1}(q(n), \Delta t)$ increases from -5.8 to -4.8 as Δt increases from Δt_0 to $2^{15}\Delta t_0$, indicating that $q(n)$ failed to pass the committor validation test. For the $q(n)$ coordinate the boundary states A and B are

defined by the corresponding minima on the free-energy profile $F(n)/k_B T$.

Now we turn our attention to the putative committor RC obtained from the optimization procedure (eq 2) using $\{x_m\}$ as the CVs ($\bar{d}_{ij}$) that respects all the symmetries of the system, for details see the SI. It can be seen in Figure 1 (b) that the constructed committor satisfactorily passed the validation test since $Z_{C,1}(q, \Delta t)$ is constant, that is, it only fluctuates near the expected value $-\ln Z_{C,1}(q, \Delta t) \approx -\ln N_{AB} \approx -4.72$ for various Δt up to statistical uncertainty, roughly estimated as $1/\sqrt{2N_{AB}} \approx 0.07$. Here, the boundary states were defined in a more accurate, systematic way, by allowing optimization of the putative RC (soft committor) around the corresponding free-energy minima (see the SI for details). As the constructed optimal coordinate q passed the validation test, eq 3 implies that the diffusion coefficient $D(q, \Delta t)$ is independent of Δt . On the other hand, for the suboptimal coordinate $q(n)$, our estimates for $Z_{C,1}(q(n), \Delta t)$, and hence for the diffusion coefficient $D(q(n), \Delta t)$ depend on sampling interval (observation time scale) Δt , indicating the presence of strong non-Markovian effects. Interestingly, the non-Markovian behavior of $q(n)$ and n is similar to the one observed when considering the energy as a RC.⁴⁰

Next, we include in Figure 1(c) and (d) the free-energy profiles $F(q(n))/k_B T$ and $F(\tilde{q})/k_B T$, where $\tilde{q}(n)$ and $\tilde{q}$ are the rescaled coordinates where the diffusion coefficient is unitary,^{32,44} $D(\tilde{q}, \Delta t_0) = 1$ and $D(\tilde{q}(n), \Delta t_0) = 1$ (see the SI for implementation details), so that the diffusive model and kinetics are specified by the free-energy profile alone. The comparison between the two free-energy profiles indicates that the free-energy barriers are almost the same, so the main difference is due to the ranges observed over $\tilde{q}(n)$ and $\tilde{q}$. While for $\tilde{q}$ the minima are separated by $\Delta\tilde{q} \equiv \tilde{q}_B - \tilde{q}_A \approx 90$, the n -based variable shows a separation around $\Delta\tilde{q}(n) \equiv \tilde{q}(n)_B - \tilde{q}(n)_A \approx 45$. As the diffusion coefficients over those coordinates are unitary, the smaller separation over $\tilde{q}(n)$ implies a higher diffusivity for the n -based coordinate so that the related diffusive model leads to faster kinetics, with e.g., the MFPT $\tau_{B \rightarrow A} = 1.8 \times 10^3 \Delta t_0$ calculated via eq 1 that is smaller than the $\tau_{B \rightarrow A} = 4.7 \times 10^3 \Delta t_0$ determined directly from the time series $n(k\Delta t)$ (See Table 1). On the other hand, for the proposed optimal coordinate q , $\tau_{B \rightarrow A} = 5.4 \times 10^3 \Delta t_0$ calculated via eq 1 is very close to the value $\tau_{B \rightarrow A} = 5.3 \times 10^3 \Delta t_0$ extracted directly from the series $q(k\Delta t)$.

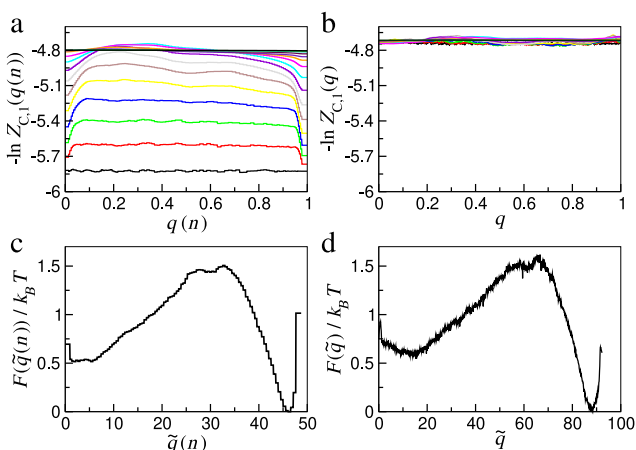


Figure 1. Functions $-\ln Z_{C,1}(r, \Delta t)$ and free-energy profiles $F(r)/k_B T$ for the n -based RC $q(n)$ and the optimized committor q . Panels (a) and (b) show the $Z_{C,1}$ -based committor validation tests for both RCs, where different colors correspond to different $\Delta t = 2^i \Delta t_0$, with $i = 0, 1, 2, \dots, 15$ (from the bottom to the top). In panels (c) and (d), the profiles are presented as a function of the rescaled RCs where $D(r, \Delta t_0) = 1$.

Table 1. Estimates for MFPT $\tau_{B \rightarrow A}$ and Number of Transitions N_{AB} Obtained from Both n -Based RC $q(n)$ and Committor q for Systems with Different Anisotropies ξ^a

	$\xi = 1$	$\xi = 3$	$\xi = 5$	$\xi = 7$
$\tau_{B \rightarrow A}$ from eq 1 with $q(n)$	1.8	3.3	3.6	6.2
$\tau_{B \rightarrow A}$ from time series of $q(n)$	4.7	9.7	13	22
N_{AB} from $Z_{C,1}(q(n), \Delta t_0)$	336	171	161	114
N_{AB} from time series of $q(n)$	122	55	40	28
$\tau_{B \rightarrow A}$ from eq 1 with q	5.4	11	15	26
$\tau_{B \rightarrow A}$ from time series of q	5.3	10	14	25
N_{AB} from $Z_{C,1}(q, \Delta t_0)$	113	52	38	26
N_{AB} from time series of q	112	53	37	27

^aAll times are given in units of $10^3 \times \Delta t_0$.

Next we include results for systems with different anisotropies ξ to show how the estimates for the kinetic properties evaluated through the diffusive model over the committor q compare with those obtained from the physically motivated RC $q(n)$. In addition to the estimates for the MFPTs evaluated through eq 1 and from the time series of the RCs, we include in Table 1 the values of the number of transitions N_{AB} which can be computed not only directly from the times series but also from¹⁵ $N_{AB} = \int Z_{C,1}(r, \Delta t_0) dr$. The values displayed in Table 1 corroborate the fact that the diffusive model along $q(n)$ gives MFPTs and number of transitions that would correspond to aggregation kinetics that is about three to four times faster than the corresponding values determined directly from the time series $n(k\Delta t)$. Conversely, the values of $\tau_{B \rightarrow A}$ and N_{AB} obtained from the diffusive model along our constructed optimal coordinate q give essentially the same results measured directly from the trajectories $q(k\Delta t)$. In the SI we include additional results obtained for larger systems sizes, as well as for the reverse MFPT $\tau_{A \rightarrow B}$ and MTPT $\hat{t}$, which indicates that the agreement between the diffusive model along the proposed committor q and the discrepancies for the $q(n)$ coordinate are observed for all cases.

To demonstrate the generality of our approach, we also present results for a more realistic three-dimensional Lennard-Jones system governed by overdamped Langevin dynamics, as detailed in the Supporting Information. The proposed method successfully computes the committor and associated kinetic properties.

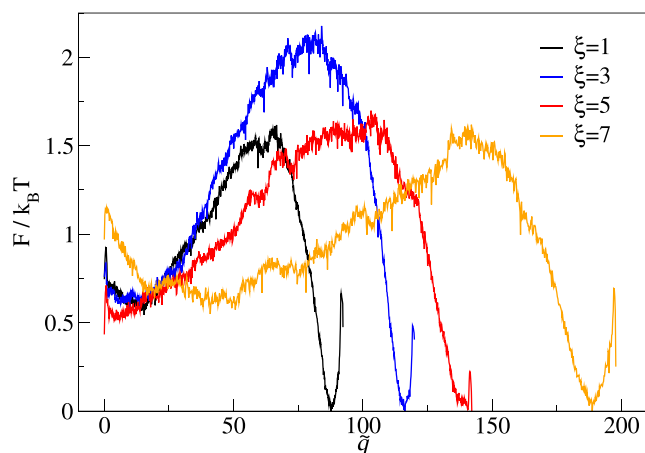


Figure 2. Free-energy landscapes $F(\tilde{q})/k_B T$ along the rescaled committor for different ξ .

Finally, in Figure 2 we show the free-energy profiles $F(\tilde{q})/k_B T$ obtained for systems with different anisotropies ξ along the rescaled committor RC $\tilde{q}$. Interestingly, just as for the energy-based profiles,⁴² the free-energy barriers display a nonmonotonic behavior with anisotropy. Even so, as the values included in Table 1 indicate, anisotropy slows down overall kinetics, as measured by $\tau_{B \rightarrow A}$ and N_{AB} . This is explained by the monotonic decrease of diffusivity with increasing ξ , that is, the range $\Delta\tilde{q} = \tilde{q}_B - \tilde{q}_A$ between the two basins in Figure 2 increase monotonically with ξ .

To the best of our knowledge, the computation of committor q and its landscape for the entire aggregation process in finite yet large systems (i.e., with $N \sim 10^2$ molecules) is not available in the literature. Studies including committor analysis are generally restricted to the first steps of aggregation with only a handful of aggregating molecules (see e.g., refs 45, 46 where more sophisticated models of protein aggregation were considered). By considering a nonparametric variational approach³⁹ supplied with a proposed set of permutationally invariant CVs, we confirmed that the diffusive model along the committor can be used to reproduce kinetic properties exactly. In addition, our results indicate that usual physically motivated RCs like n may not lead to accurate kinetic description at arbitrary time scales due to strong non-Markovian effects. When dealing with such suboptimal coordinates, more sophisticated stochastic models with a memory kernel are needed,¹³ in which case the free-energy profile loses its simple interpretation, or one needs to employ a suitable separation of time scales, as we did recently for the energy coordinate.⁴⁰ The use of the committor as RC preserves the appealing simple picture of diffusion on the free-energy profile while avoiding the need for time scale separation. Such an accurate model of the dynamics can also be used to compute rigorously and in a direct manner, the pre-exponential factor,^{23,34,47} a key determinant of aggregation kinetics. Finally, it is worth mentioning that our approach is general and its implementation for the analysis of aggregation kinetics in more realistic multimolecular systems^{1–5,45,46} is straightforward. Our method is also of interest to the machine learning community, which is seeking methods to determine the committor for longitudinal data sets.^{25–29} In particular, the stringent $Z_{C,1}$ validation test can be used to assess the accuracy of committor estimates obtained by various methods, including the many neural network architectures currently available, thereby greatly simplifying direct comparison between approaches.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.5c01427>.

It contains a practical example of how the permutationally invariant CVs are computed; additional methodological details for implementing the variational approach, defining boundary states, and computing kinetic quantities from diffusive models; additional results for larger lattice systems and a Lennard-Jones system; additional references (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Lair F. Trugilho – Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, United Kingdom; Departamento de Física, Universidade Federal de Viçosa (UFV), 36570-900 Viçosa, Brazil; orcid.org/0000-0002-6704-9744; Email: lair.trugilho@ufv.br

Sergei V. Krivov – Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, United Kingdom; Astbury Center for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, United Kingdom; orcid.org/0000-0002-3493-0068; Email: s.krivov@leeds.ac.uk

Authors

Stefan Auer – School of Chemistry, University of Leeds, Leeds LS2 9JT, United Kingdom; orcid.org/0000-0003-0418-1745

Leandro G. Rizzi – Departamento de Física, Universidade Federal de Viçosa (UFV), 36570-900 Viçosa, Brazil; orcid.org/0000-0002-3890-6022

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jctc.5c01427>

Funding

The Article Processing Charge for the publication of this research was funded by the Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior (CAPES), Brazil (ROR identifier: 00x0ma614).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Lair F. Trugilho thanks FAPEMIG and CAPES (PSDE program) for the studentships. Leandro G. Rizzi acknowledges the support from CNPq (Grants No. 312999/2021-6 and No. 308285/2025-5) and FAPEMIG (Process APQ-01462-24). The authors also thank the computational resources made available by GISC-UFV.

REFERENCES

- (1) Cho, W. J.; Kim, J.; Lee, J.; Keyes, T.; Straub, J. E.; Kim, K. S. Limit of Metastability for Liquid and Vapor Phases of Water. *Phys. Rev. Lett.* **2014**, *112*, 157802.
- (2) Sosso, G. C.; Chen, J.; Cox, S. J.; Fitzner, M.; Pedevilla, P.; Zen, A.; Michaelides, A. Crystal Nucleation in Liquids: Open Questions and Future Challenges in Molecular Dynamics Simulations. *Chem. Rev.* **2016**, *116*, 7078.
- (3) Auer, S.; Kashchiev, D. Phase Diagram of α -Helical and β -Sheet Forming Peptides. *Phys. Rev. Lett.* **2010**, *104*, 168105.
- (4) Wang, Y.; Bunce, S. J.; Radford, S. E.; Wilson, A. J.; Auer, S.; Hall, C. K. Thermodynamic phase diagram of amyloid- β (16–22) peptide. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 2091.
- (5) Duan, C.; Wang, R. Electrostatics-Induced Nucleated Conformational Transition of Protein Aggregation. *Phys. Rev. Lett.* **2023**, *130*, 158401.
- (6) Nadler, W.; Hansmann, U. H. E. Generalized ensemble and tempering simulations: A unified view. *Phys. Rev. E* **2007**, *75*, 026109.
- (7) Balakrishnan, V. *Elements of Nonequilibrium Statistical Mechanics*; Springer, 2021.
- (8) Blow, K. E.; Quigley, D.; Sosso, G. C. The seven deadly sins: When computing crystal nucleation rates, the devil is in the details. *J. Chem. Phys.* **2021**, *155*, 040901.
- (9) Chahine, J.; Oliveira, R. J.; Leite, V. B. P.; Wang, J. Configuration-dependent Diffusion Can Shift the Kinetic Transition State and Barrier Height of Protein Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 14646–14651.
- (10) Best, R. B.; Hummer, G. Coordinate-dependent Diffusion in Protein Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 1088–1093.
- (11) Trugilho, L. F.; Rizzi, L. G. Shape-free theory for the self-assembly kinetics in macromolecular systems. *EPL* **2022**, *137*, 57001.
- (12) Kuhnhold, A.; Meyer, H.; Amati, G.; Pelagejcev, P.; Schilling, T. Derivation of an exact, nonequilibrium framework for nucleation: Nucleation is a priori neither diffusive nor Markovian. *Phys. Rev. E* **2019**, *100*, 052140.
- (13) Tepper, L.; Dalton, B.; Netz, R. R. Accurate Memory Kernel Extraction from Discretized Time-Series Data. *J. Chem. Theory Comput.* **2024**, *20*, 3061–3068.
- (14) Banushkina, P. V.; Krivov, S. V. Optimal reaction coordinates. *WIREs Comput. Mol. Sci.* **2016**, *6*, 748–763.
- (15) Krivov, S. V. On Reaction Coordinate Optimality. *J. Chem. Theory Comput.* **2013**, *9*, 135–146.
- (16) Peters, B. Reaction Coordinates and Mechanistic Hypothesis Tests. *Annu. Rev. Phys. Chem.* **2016**, *67*, 669–690.
- (17) Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. S. On the transition coordinate for protein folding. *J. Chem. Phys.* **1998**, *108*, 334–350.
- (18) Berezhkovskii, A. M.; Szabo, A. Diffusion along the Splitting/Commitment Probability Reaction Coordinate. *J. Phys. Chem. B* **2013**, *117*, 13115–13119.
- (19) Vanden-Eijnden, E.; Venturoli, M.; Ciccotti, G.; Elber, R. On the assumptions underlying milestoning. *J. Chem. Phys.* **2008**, *129*, 174102.
- (20) Lu, J.; Vanden-Eijnden, E. Exact dynamical coarse-graining without time-scale separation. *J. Chem. Phys.* **2014**, *141*, 044109.
- (21) 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.
- (22) Metzner, P.; Schütte, C.; Vanden-Eijnden, E. Illustration of transition path theory on a collection of simple examples. *J. Chem. Phys.* **2006**, *125*, 084110.
- (23) Krivov, S. V. Protein Folding Free Energy Landscape along the Committor - the Optimal Folding Coordinate. *J. Chem. Theory Comput.* **2018**, *14*, 3418–3427.
- (24) Krivov, S. V. Blind Analysis of Molecular Dynamics Click to copy article link. *J. Chem. Theory Comput.* **2021**, *17*, 2725–2736.
- (25) Khoo, Y.; Lu, J.; Ying, L. Solving for high-dimensional committor functions using artificial neural networks. *Res. Math. Sci.* **2019**, *6*, 1.
- (26) Li, Q.; Lin, B.; Ren, W. Computing committor functions for the study of rare events using deep learning. *J. Chem. Phys.* **2019**, *151*, 054112.
- (27) Chen, H.; Roux, B.; Chipot, C. Discovering Reaction Pathways, Slow Variables, and Committor Probabilities with Machine Learning. *J. Chem. Theory Comput.* **2023**, *19*, 4414–4426.
- (28) Jung, H.; Covino, R.; Arjun, A.; Leitold, C.; Dellago, C.; Bolhuis, P. G.; Hummer, G. Machine-guided path sampling to discover mechanisms of molecular self-organization. *Nat. Comput. Sci.* **2023**, *3*, 334–345.
- (29) Kang, P.; Trizio, E.; Parrinello, M. Computing the committor with the committor to study the transition state ensemble. *Nat. Comput. Sci.* **2024**, *4*, 451–460.
- (30) Krivov, S. V. Nonparametric Analysis of Nonequilibrium Simulations. *J. Chem. Theory Comput.* **2021**, *17*, 5466–5481.
- (31) Vroylandt, H. On the derivation of the generalized Langevin equation and the fluctuation-dissipation theorem. *EPL* **2022**, *140*, 62003.
- (32) Krivov, S. V.; Karplus, M. Diffusive reaction dynamics on invariant free energy profiles. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13841–13846.
- (33) Krivov, S. V. Is Protein Folding Sub-Diffusive? *PloS Comput. Biol.* **2010**, *6*, e1000921.
- (34) Krivov, S. V. The Free Energy Landscape Analysis of Protein (FIP35) Folding Dynamics. *J. Phys. Chem. B* **2011**, *115*, 12315–12324.
- (35) Krivov, S. V. Numerical Construction of the p-fold (Committer) Reaction Coordinate for a Markov Process. *J. Phys. Chem. B* **2011**, *115*, 11382–11388.
- (36) Best, R. B.; Hummer, G. Reaction coordinates and rates from transition paths. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6732–6737.
- (37) Peters, B.; Trout, B. L. Obtaining reaction coordinates by likelihood maximization. *J. Chem. Phys.* **2006**, *125*, 054108.
- (38) Roux, B. String Method with Swarms-of-Trajectories, Mean Drifts, Lag Time, and Committor. *J. Phys. Chem. A* **2021**, *125*, 7558–7571.
- (39) Banushkina, P. V.; Krivov, S. V. Nonparametric variational optimization of reaction coordinates. *J. Chem. Phys.* **2015**, *143*, 184108.

- (40) Trugilho, L. F.; Auer, S.; Rizzi, L. G. A density of states-based approach to determine temperature-dependent aggregation rates. *J. Chem. Phys.* **2024**, *161*, 051101.
- (41) Herringer, N. S. M.; Dasetty, S.; Gandhi, D.; Lee, J.; Ferguson, A. L. Permutationally Invariant Networks for Enhanced Sampling (PINES): Discovery of Multimolecular and Solvent-Inclusive Collective Variables. *J. Chem. Theory Comput.* **2024**, *20*, 178–198.
- (42) Trugilho, L. F.; Rizzi, L. G. Microcanonical thermostatics of aggregation transition in a system with anisotropically interacting molecules. *J. Phys. Conf. Ser.* **2020**, *1483*, 012011.
- (43) Kashchiev, D. *Nucleation: Basic Theory With Applications*; Butterworth-Heinemann, 2000.
- (44) Rhee, Y. M.; Pande, V. S. One-Dimensional Reaction Coordinate and the Corresponding Potential of Mean Force from Commitment Probability Distribution. *J. Chem. Phys.* **2005**, *109*, 6780–6786.
- (45) Yang, Y. I.; Gao, Y. Q. Computer simulation studies of $A\beta_{37-42}$ aggregation thermodynamics and kinetics in water and salt solutions. *J. Phys. Chem. B* **2015**, *119*, 662–670.
- (46) Phan, T. M.; Schmit, J. D. Conformational entropy limits the transition from nucleation to elongation in amyloid aggregation. *Biophys. J.* **2022**, *121*, 2931–2939.
- (47) Banushkina, P. V.; Krivov, S. V. High-Resolution Free-Energy Landscape Analysis of α -Helical Protein Folding: HP35 and Its Double Mutant. *J. Chem. Theory Comput.* **2013**, *9*, S257–S266.