

How wet should be the reaction coordinate for ligand unbinding?

Pratyush Tiwary^{a)} and B. J. Berne^{b)}

Department of Chemistry, Columbia University, New York, New York 10027, USA

(Received 23 May 2016; accepted 8 July 2016; published online 3 August 2016)

We use a recently proposed method called Spectral Gap Optimization of Order Parameters (SGOOP) [P. Tiwary and B. J. Berne, *Proc. Natl. Acad. Sci. U. S. A.* **113**, 2839 (2016)], to determine an optimal 1-dimensional reaction coordinate (RC) for the unbinding of a bucky-ball from a pocket in explicit water. This RC is estimated as a linear combination of the multiple available order parameters that collectively can be used to distinguish the various stable states relevant for unbinding. We pay special attention to determining and quantifying the degree to which water molecules should be included in the RC. Using SGOOP with under-sampled biased simulations, we predict that water plays a distinct role in the reaction coordinate for unbinding in the case when the ligand is sterically constrained to move along an axis of symmetry. This prediction is validated through extensive calculations of the unbinding times through metadynamics and by comparison through detailed balance with unbiased molecular dynamics estimate of the binding time. However when the steric constraint is removed, we find that the role of water in the reaction coordinate diminishes. Here instead SGOOP identifies a good one-dimensional RC involving various motional degrees of freedom. *Published by AIP Publishing.* [<http://dx.doi.org/10.1063/1.4959969>]

I. INTRODUCTION

The unbinding of ligand-substrate systems is a problem of great theoretical and practical relevance. To take an example from the biological sciences, there is now an emerging view that the pharmacological efficacy of a drug depends not just on its thermodynamic affinity for the host protein, but also, and perhaps even more so, on when and how it unbinds from the protein.^{1,2} While a variety of experimental techniques can provide unbinding rate constants, glean a clear molecular scale understanding from such experiments into the dynamics of unbinding is difficult, and at best indirect. This makes it in principle very attractive to use atomistic molecular dynamics (MD) simulations to study the unbinding process. However, most successful drugs unbind at time scales much longer than milliseconds.^{1,2} Even with the fastest available supercomputers, this makes it virtually impossible to use MD simulations to obtain statistically reliable insight into unbinding dynamics.

This time scale limitation makes it crucial to complement MD with enhanced sampling techniques. These techniques accelerate the movement between metastable states separated by high ($\gg k_B T$) barriers but still allow recovering the unbiased thermodynamics and kinetics. While in principle one could construct Markov State Models (MSM)³ to study the unbinding dynamics from multiple short, unbiased simulations without any enhanced sampling, the associated high barriers typical for unbinding make this extremely difficult. As such, reported applications of MSM to such problems have been indirect, and instead of directly studying unbinding, these studies^{4,5} have actually looked at the drug binding problem

where the barriers tend to be smaller. To directly simulate the unbinding process, it thus becomes unavoidable to use enhanced sampling methods.⁶

On the other hand, the use of enhanced sampling methods to study high barrier systems has its own caveats. Many such methods involve controlling the probability distribution along a low-dimensional reaction coordinate (RC), which best captures all the relevant slow degrees of freedom. Typically many such order parameters or collective variables (CVs) are available that can distinguish between various metastable states of the system at hand. For ligand unbinding, these CVs could include ligand-host relative displacement, their conformations, and their hydration states. However, often the fluctuations in these CVs can be coupled in a non-trivial manner, and it can be tricky to select a RC without having a prescience of the CVs whose fluctuations matter the most for driving the process of interest.

In this work, we aim to answer the following question: given a certain choice of order parameters (or collective variables) for a ligand-host system, what is the optimal 1-dimensional RC for unbinding that can be expressed as a linear combination of these collective variables? We are especially interested in determining how wet this RC is. Wetness here denotes the weight ascribed to the descriptor of the solvation state of the binding site, relative to other descriptors contributing to the RC. This will indicate how important biasing water density fluctuations in the host binding pocket is to the kinetics of ligand unbinding. While it is well-known through various theoretical, simulation, and experimental studies that collective water motion into/out of binding pockets is correlated with unbinding/binding, respectively,⁶⁻¹² we wish to have a quantitative measure of the utility of biasing these water fluctuations in the sampling of ligand unbinding.

^{a)}Electronic mail: pt2399@columbia.edu

^{b)}Electronic mail: bb8@columbia.edu

Here, we investigate this question for ligand unbinding in a much studied model hydrophobic ligand-host system (Fig. 1) interacting through Lennard-Jones potential in an aqueous environment made of explicit TIP4P water molecules.^{9,11,13} Many excellent methods exist for the purpose of RC optimization.^{14–22} However, the energy barrier for unbinding in this system as reported through previous studies is as high as $30\text{--}35 k_B T$, making it crucial for the purpose of RC optimization to use a method that does not rely on accurate sampling of rare reactive unbinding trajectories. For this reason, we use a recently proposed method SGOOP (spectral gap optimization of order parameters)²³ that enables us to determine an optimal RC through relatively short biased simulations performed using a trial RC (see Fig. 2 and Sec. II A for details of SGOOP).

We consider two different scenarios in this work, both of which are expected to arise in the context of ligand unbinding. In the first scenario, we sterically constrain the system so that the ligand can move only along the centro-symmetric axis z (see Fig. 1). In the second, we lift this steric constraint. We find that in the presence of the steric constraint, water density fluctuations in the host cavity must be part of the optimal RC. This is in excellent agreement with the previous work on this and related systems^{9,11,24–28} where for a sterically constrained setup, there is a bimodal water distribution at a critical ligand-cavity separation, around which the unbinding pathway involves moving from dry to wet states. However we find that when the steric constraint is removed and the ligand is free to move in any direction, the role of water in the optimal RC is minimal to none. In this case, water is less of a driving variable for unbinding, but more of a driven variable that follows the movement of the ligand. Here SGOOP identifies how the optimal RC is distorted from the z -axis (Fig. 1), which turns out to be the minimum free energy pathway for this system as reported in a previous work.¹¹

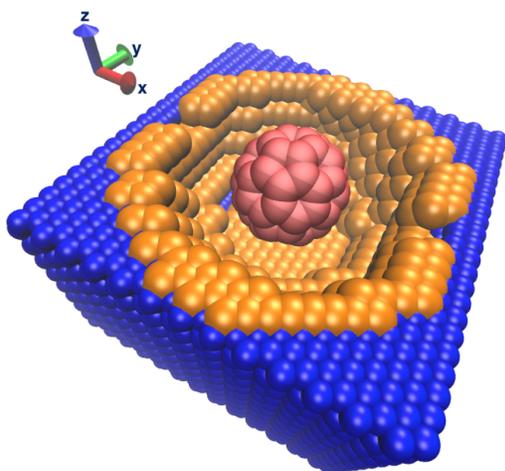


FIG. 1. Cavity-ligand system in explicit water with axes marked. Red: fullerene shaped ligand atoms. Orange: cavity atoms that interact with the ligand and with water molecules. Blue: wall atoms. The water molecules are not shown for clarity. See Section III and the supplementary material for corresponding interaction potentials and further details.²⁹

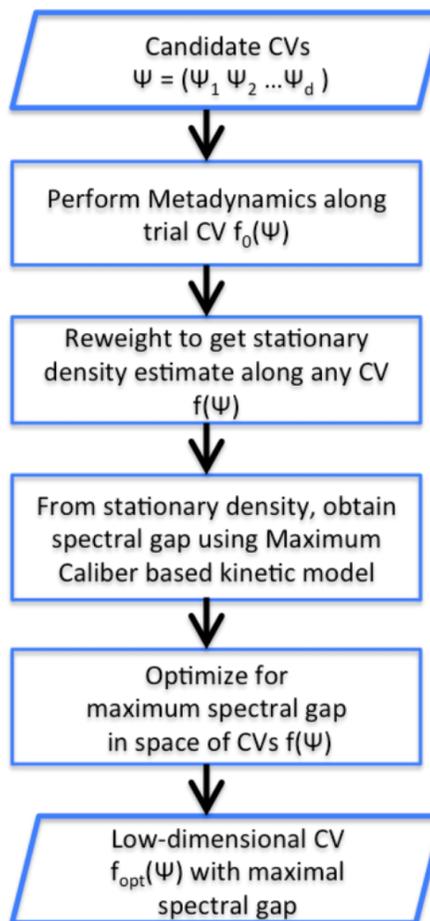


FIG. 2. Flowchart summarizing the various key steps in SGOOP.²³ The whole process can in principle be iterated between the second and the last steps to further improve the sampling.

We validate our results through extensive calculations of unbinding time statistics for the sterically constrained ligand using the infrequent metadynamics approach³⁰ and find that the optimal RC is indeed wet to some extent. In addition, because the analogous barrier for ligand binding is much smaller than for unbinding, we use unbiased MD estimates of the binding time and validate that detailed balance is satisfied between unbinding and binding rates. We perform the infrequent metadynamics calculations using the optimal RC as per SGOOP and two other sub-optimal RCs with no water content and more than optimal water content, respectively. Our findings clearly demonstrate the improvement in the quality and accuracy of the unbinding time statistics by using the optimized RC predicted through SGOOP. With the optimized CV, the unbinding time statistics gives a superior agreement with the binding time statistics obtained through unbiased MD. Furthermore, it also gives a much improved Poisson fit for the cumulative distribution function of unbinding times, as quantified through the Kolmogorov-Smirnov test proposed in Ref. 31. This shows that the optimized RC predicted through SGOOP indeed does a better job of capturing the slow dynamics of the system. Previous applications of

SGOOP²³ were restricted to using the optimized RC for faster convergence of the free energy. The results reported in this work comprise the first demonstration of improving kinetics calculations using SGOOP and mark a step further towards systematic high-throughput studies of unbinding dynamics.

II. THEORY

In this section, we summarize the key methods^{23,30,31} used in this work and their underlying principles.

A. Spectral gap optimization of order parameters (SGOOP)

SGOOP²³ is a method to optimize low-dimensional order parameters or collective variables for use in enhanced sampling biasing methods like umbrella sampling and metadynamics, when only limited prior information is known about the system (see Fig. 2 for a flowchart summarizing the key steps in SGOOP). This optimization is done from a much larger set of candidate CVs $\Psi = (\Psi_1, \Psi_2, \dots, \Psi_d)$, which are assumed to be known *a priori*. SGOOP is based on the idea that the best order parameter, which we call the reaction coordinate (RC), is one with the maximum separation of time scales between visible slow and hidden fast processes. This time scale separation is calculated as the spectral gap between the slow and fast eigenvalues of the transition probability matrix on a grid along any CV.²³ The transition probability matrix is calculated in SGOOP using an approximate kinetic model that can be derived, for example, through the principle of maximum caliber.^{23,32,33} Let $\{\lambda\}$ denote this set of eigenvalues, with $\lambda_0 \equiv 1 > \lambda_1 \geq \lambda_2 \dots$. The spectral gap is then defined as $\lambda_s - \lambda_{s+1}$, where s is the number of barriers apparent from the free energy estimate projected on the CV at hand that are higher than a user-defined threshold (typically $\gtrsim k_B T$). In this case, assuming overdamped dynamics, the eigenvalues beyond the first $s + 1$ correspond to relaxation times in each of the individual wells,^{34–36} which for an optimal RC should be much smaller than the escape times from the wells.

The key input to SGOOP as used in this work is an estimate of the stationary probability density (or equivalently the free energy) of the system, accumulated through a biased simulation performed along a sub-optimal trial RC given by some linear or non-linear function $f_0(\Psi)$, where Ψ denotes the larger set of candidate CVs. Any type of biased simulation could be used for this purpose, as long as it allows projecting the stationary probability density estimate on generic combinations of CVs without having to repeat the simulation. Metadynamics³⁷ provides this functionality in a straightforward manner and hence we use it here. Given this information, we use the principle of maximum caliber²³ to set up an unbiased master equation for the dynamics of various trial CVs $f(\Psi)$. Through a post-processing optimization procedure, we then find the optimal RC as the $f(\Psi)$ which gives the maximal spectral gap of the associated transfer matrix. We refer to Ref. 23 for details of the master equation and the maximum caliber expression that relates the transfer

matrix to stationary probabilities and facilitates calculation of the eigenvalues and hence the spectral gap.

As described in the Introduction, for the problem of ligand unbinding in this work, we take this larger set of CVs to be the various components of the separation between the ligand and the host, and the solvation state of the host pocket (Fig. 1). In more complex systems, further members could be added to this set. Since counting the number of barriers in a projected free energy profile could be affected by sampling noise, we smooth the free energy by averaging over bins. To ensure that the calculated spectral gaps are robust with respect to the amount of smoothening, we perform an averaged estimate of the spectral gaps using different amounts of smoothing (see the supplementary material for details).²⁹

Note that the approximate kinetic model used here in SGOOP is equivalent to the Smoluchowski equation whereby (i) the dynamics of any CV is described by a forced diffusion process and (ii) the diffusion constant along this CV is independent of position. This kinetic model is used in SGOOP to improve the choice of the RC that should be biased given limited information starting with a trial RC. The calculation of rates is then done with this improved RC. It is important to note that the infrequent metadynamics method for calculating rate constants³⁰ *does not* assume Smoluchowski dynamics or constant diffusivity (see Sec. II B for details).

B. Dynamics from infrequent metadynamics

The infrequent metadynamics approach^{30,31} is a recently proposed method which has been used to obtain rate constants in various molecular systems.^{11,38} It involves time-dependent biasing of a few selected (typically one to three) order parameters or collective variables (CVs) out of the many available, in order to hasten the escape from metastable free energy basins.³⁹ By periodically adding repulsive bias (typically in the form of Gaussians) in the regions of CV space as they are visited, the system is encouraged to escape stable free energy basins where they would normally be trapped for long periods of time. The central idea in infrequent metadynamics is to deposit bias rarely enough compared to the time spent in the transition state regions so that dynamics in the saddle region is very rarely perturbed. Through this approach, one then increases the likelihood of not corrupting the transition states and preserves the sequence of transitions between stable states. The acceleration of transition rates achieved through biasing can then be calculated by appealing to generalized transition state theory,⁴⁰ which yields the following simple running average for the acceleration:³⁰

$$\alpha = \langle e^{\beta V(s,t)} \rangle_t, \quad (1)$$

where s is the collective variable being biased, $\beta = 1/k_B T$ is the inverse temperature, $V(s,t)$ is the bias experienced at time t , and the subscript t indicates averaging under the time-dependent potential. This approach is expected to work best in the diffusion controlled regime.⁴¹

The infrequent metadynamics method requires a good and small set of slow collective variables demarcating all relevant stable states of interest. Whether this is the case or not can be verified *a posteriori* by checking if the cumulative distribution

function for the transition times out of each stable state is Poissonian,³¹ as quantified through the Kolmogorov-Smirnov test described in detail in Ref. 31. While metadynamics can still be performed with two, three, or more biasing CVs, the computational gain obtained by compressing the slow dynamics into an optimized 1-dimensional RC is immense, especially given the infrequent nature of biasing (see the supplementary material for detailed simulation parameters such as the frequency of biasing used in this work).²⁹ Using SGOOP (Sec. II A) allows us to select a good 1-dimensional RC as a function of the many available choices of CVs, as we show in this work. This choice increases the probability of passing the test of Ref. 31 once the relatively expensive infrequent metadynamics runs are performed.

III. SYSTEM DETAILS

The model ligand used in this work is a C_{60} fullerene and the binding pocket is an ellipsoidal cavity carved from a hydrophobic slab, all interacting via Lennard-Jones potentials and enclosed by a periodic box with explicit water and cubic edge length 5.96 nm. This system was introduced previously in works such as Refs. 9 and 11. The pocket sites are fixed and interact with the model ligand with a Lennard-Jones site-site potential having $\sigma = 0.4152$ nm, which is kept the same for all interactions. The pocket itself comprises 2 types of atomic species, interacting with the ligand atoms (color red in Fig. 1) as described below. The system comprised a total of 34 296 atoms, with the total number of ligand, cavity, and solvent atoms equaling 60, 9020, and 25 216, respectively.

1. Cavity atoms (color orange in Fig. 1), with Lennard-Jones $\epsilon = 0.008$ kJ/mol.
2. Wall atoms (color blue in Fig. 1), with Lennard-Jones $\epsilon = 0.0024$ kJ/mol.

The solute-solvent interactions are represented by the geometric mean of the respective water and solute parameters, in accordance with the Optimized Potential for Liquid Simulations (OPLS) formalism.⁴² All simulations are performed in explicit TIP4P water¹³ with the GROMACS 4.5.4 MD package⁴³ patched with the PLUMED plugin.⁴⁴ During the equilibration stage, temperature and pressure are controlled with the stochastic velocity rescaling thermostat⁴⁵ and Berendsen barostat.⁴⁶ The production runs were NVT (constant number, volume, temperature) with a temperature of 300 K. The PLUMED plugin⁴⁴ was used to carry out metadynamics calculations. An integration time step of 2 fs was used for all runs. All other relevant simulation details are provided in the supplementary material.²⁹

IV. RESULTS AND DISCUSSION

A. Ligand constrained to move along one direction

In the first investigated case, the system dynamics is sterically constrained so that the ligand can move only along the centro-symmetric axis z (Fig. 1). This system and constraint have already been investigated in studies aimed at understanding hydrophobic interactions.^{9,11,24–26} Here we

consider two descriptors; the z -component of the ligand-cavity separation and the number of water molecules in the host cavity, denoted w . The number of water molecules is computed using a sigmoidal function which makes w continuous and differentiable (see the supplementary material²⁹ for details including precise definition of w) as implemented in the enhanced sampling plugin PLUMED.⁴⁴ We then seek the best 1-d RC f of the following form:

$$f(z, w) = \{z + m_w w; m_w \geq 0\}. \quad (2)$$

Throughout this paper m_w is a measure of the wetness of the RC, with $m_w = 0$ corresponding to a completely dry RC, and higher values denoting increasingly wetter RCs.

We first perform a short metadynamics simulation by biasing $f_0 = z$. This starting run is performed with frequent biasing since the objective here is to get a sense of the free energy, and not the kinetics (see the supplementary material for various biasing frequencies and other parameters).²⁹ Through this, we can obtain an estimate of the stationary probability density along any $f(z, w)$ by using the reweighting functionality of metadynamics.³⁷ By using SGOOP, we then get an estimate of the optimal $m_w \approx 0.075$ in Eq. (2) which maximizes the spectral gap. This is shown in Fig. 3(a) where an estimate of the spectral gap versus m_w for different lengths of the starting metadynamics trajectory is provided. Other trajectories used in SGOOP shown in Figs. 3(b) and 3(c) are of length 10 ns, 15 ns, and 20 ns, respectively. The results are extremely robust with respect to simulation time and parameters. Fig. 3(a) has 3 different curves calculated which for all practical purposes collapse into one, indicating that the spectral gaps estimated with trajectories of three different simulation times are virtually indistinguishable and well-

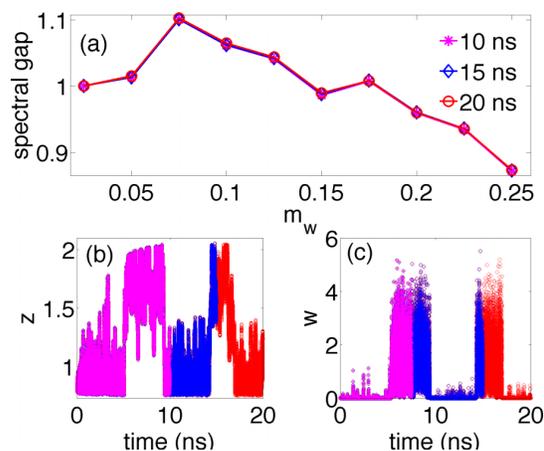


FIG. 3. (a) Spectral gap versus the amount of wetness of the RC, m_w (see Eq. (2)) for the case when the ligand constrained to move along a line. The optimal RC can be clearly seen to be at $m_w \approx 0.075$. Three different profiles are provided, which were calculated by using the starting metadynamics trajectories of different lengths as indicated in legends, performed with biasing CV z . The spectral gap is normalized so that its value for $m_w = 0$ is 1. (b) and (c) are the corresponding trajectories for the distance z and the number of pocket waters w . See Sec. IV A and the supplementary material for precise definition of w .²⁹ In all sub-figures here, magenta stars, blue diamonds, and red circles denote results for trajectories of lengths 10 ns, 15 ns, and 20 ns, respectively.

converged. Furthermore, using an entirely different starting metadynamics trajectory generated using different Gaussian width gives the same optimal wetness of the RC (see the supplementary material).²⁹

The optimal wetness of the RC in Eq. (2) given by $m_w \approx 0.075$ is validated by performing extensive multiple independent unbinding simulations using infrequent metadynamics, starting from the bound pose $z = 0$ (Sec. II B). The unbinding time is calculated as the time taken to reach $z = 1.4$ nm for the first time.¹¹ We perform three independent sets of 24 simulations (totaling 72 simulations) for (1) $m_w = 0$, a dry RC, (2) $m_w = 0.075$, the RC with optimal wetness found from SGOOP, and (3) $m_w = 0.15$, the RC with more than optimal wetness. The empirical and fitted cumulative distribution functions for the unbinding time statistics using the three different RCs with varying amounts of wetness are shown in Figs. 4(b)-4(d), along with the respective p-values for fits to the ideal Poisson distributions, quantified using the Kolmogorov-Smirnov test from Ref. 31, and mean times $\log(2)$ divided by median ratio for each case. An ideal fit to the Poisson distribution would result if both these numbers would be close to 1, and this would suggest that the accelerated time scales found using metadynamics are reliable. The RC with

optimal water coefficient $m_w = 0.075$ obtained using SGOOP gives Poisson metrics closest to 1. Fig. 4(a) shows the mean unbinding times obtained using the three RCs with different values of the wetness parameter m_w and these are compared with the corresponding estimate provided in the literature⁹ calculated from accurate free energy calculations together with the principle of detailed balance. While it must be said that the completely dry RC does a reasonable job in terms of the p-value and order of magnitude agreement with unbiased MD, it is very clear from this plot as well that the RC with optimal wetness gives the best performance as per various metrics shown in Fig. 4. Thus to summarize, the optimal RC for this case indeed has a small but distinct amount of wetness.

B. Ligand free to move in any direction

In this case, we remove the steric constraint forcing the system to move along z and allow the ligand to freely move in any direction (see Fig. 1). Because the system is axially symmetric, we consider 3 order parameters, namely, the z -component of the ligand-cavity separation, $\rho = \sqrt{x^2 + y^2}$, and the number of water molecules in the host cavity, denoted w .

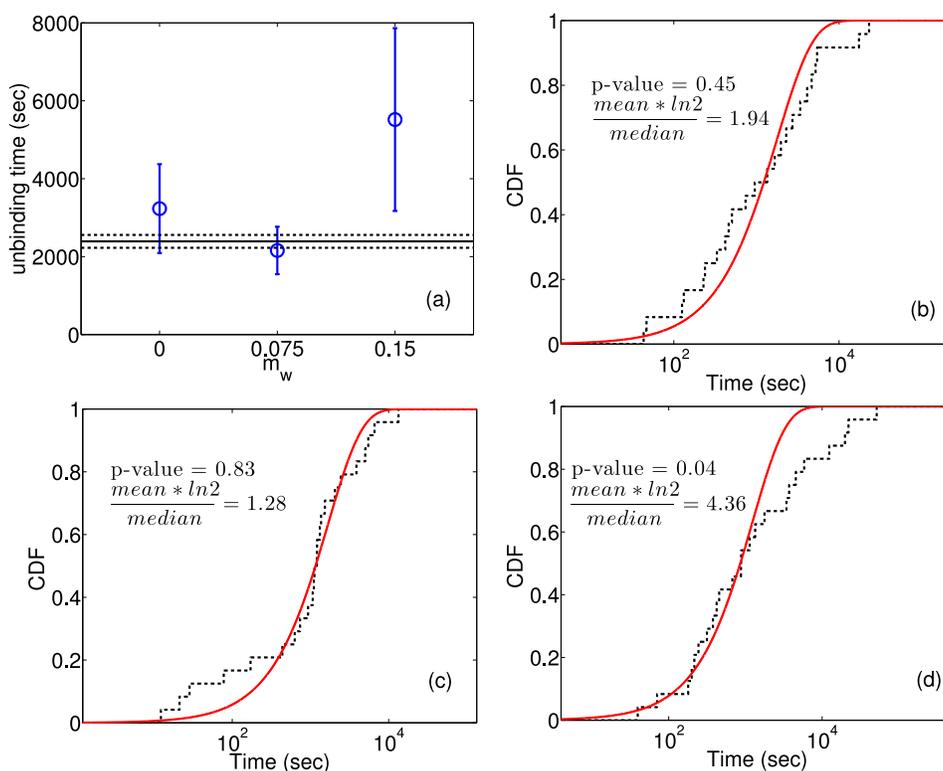


FIG. 4. Unbinding times for the sterically constrained ligand using different simulation protocols. In (a) the mean unbinding times as obtained through the three RCs with different water coefficients are plotted along with error bars (blue circles). Also plotted is the corresponding estimate of mean unbinding time (solid black line) with errors (dashed black line) by using the principle of detailed balance with the unbiased estimate of binding time. All error bars correspond to \pm standard deviation intervals. (b) to (d) give the empirical (black dashed line) and fitted (solid red line) cumulative distribution functions (CDF) for unbinding time statistics using different RCs with varying amounts of wetness m_w (see Eq. (2)). From (b) to (d), respectively, m_w is 0, 0.075 and 0.15. Also indicated are respective p-values for fit to the ideal Poisson distribution, quantified using the Kolmogorov-Smirnov test from Ref. 31, and mean times $\log(2)$ divided by median ratio for each case. The closer are both these values to 1, the more ideal is the Poisson distribution fit illustrating the reliability of the dynamics generated from metadynamics. As can be seen from these figures, the RC with optimal water coefficient of 0.075 as obtained from SGOOP gives the best Poisson metric as per both these criteria.

We then seek the best 1-d RC f of the following form:

$$f(z, \rho, w) = \{z + m_\rho \rho + m_w w; m_\rho \geq 0, m_w \geq 0\}. \quad (3)$$

We first perform a short metadynamics simulation by biasing with $f_0 = z$, a purely dry RC. As before, this starting run is performed with frequent biasing since the objective here is to get a sense of the free energy, and not the kinetics. This gives an estimate of the stationary probability density along any $f(z, \rho, w)$ by applying the reweighting functionality of metadynamics.³⁷ We then use SGOOP to obtain an estimate of the optimal values as $m_\rho \approx 0.6$, $m_w \approx 0.0$ in Eq. (3). These values maximize the spectral gap.

Fig. 5 gives an estimate of the spectral gap versus (m_ρ, m_w) based on an initial metadynamics trajectory of duration 20 ns biasing z . The results are again extremely robust with respect to how long the simulation was run. In this case as well, using an entirely different starting metadynamics trajectory generated using different Gaussian width gives the same optimal value of the RC (see the supplementary material).²⁹

As can be seen by comparing Fig. 5 to Fig. 3, the wetness of the optimal RC in the case of unconstrained motion is closer to 0. In a sense, the water fluctuations in the cavity appear to be caused or driven by the unbinding, rather than being a driving variable for unbinding as it is in the constrained case. The primary reaction coordinate depends on z and ρ , the displacement variables of the ligand with respect to the cavity. Indeed SGOOP finds $m_\rho \approx 0.6$, which gives the distortion of the reaction path from the z -axis (see Fig. 1). This is the same as the slope of the minimum free energy pathway in (z, ρ) space reported in the previous work.¹¹ As described in detail in Ref. 11, removing the steric constraint causes the bound ligand to roll over and take a slightly more stable off-center ground state and therefore a different binding pose. This leads to a slight difference in the binding free energies of the two setups considered in this work.

Since the optimal wetness of the RC in this case is close to 0, we do not perform any kinetics calculations. Instead we refer to the results from Ref. 11, where infrequent metadynamics with a similar completely dry RC for this setup gave very

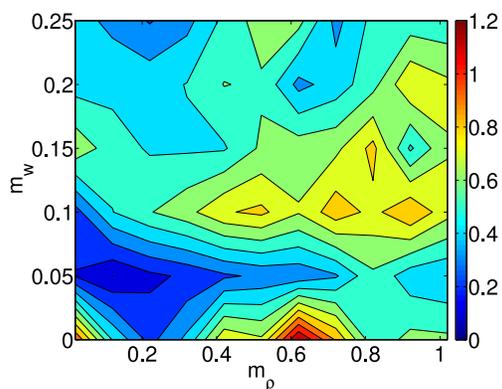


FIG. 5. Contour plot of spectral gap versus (m_ρ, m_w) with the starting metadynamics trajectory of duration 20 ns used in SGOOP. The optimal RC can be clearly seen to be at $(m_\rho, m_w) \approx (0.6, 0.0)$. The spectral gap is normalized so that its value for $(m_\rho, m_w) = (0, 0)$ is 1.

good agreement through detailed balance with the unbiased MD estimate of the binding time.

In the supplementary material,²⁹ we also provide illustrative free energy profiles for both setups along a variety of RCs.

V. DISCUSSION AND CONCLUSIONS

In this work, we have applied the recently proposed method SGOOP²³ to the problem of determining the reaction coordinate for ligand unbinding in a model system in explicit water. By using short biased metadynamics simulations performed using a sub-optimal reaction coordinate, we find that the true reaction coordinate involves water in the case when the system is sterically constrained to move along an axis of symmetry. In the case when this constraint is lifted, the role of water in the optimal RC is reduced. Our predictions of the optimal RC are validated by extensive calculations of the unbinding rate constant using metadynamics with infrequent biasing^{30,31} with different RCs. We believe that the application of SGOOP to optimize the choice of RC for ligand unbinding, combined with the approach of Refs. 30 and 31, provides an important step in the quest to invent methods useful for systematic and possibly high throughput calculations of the unbinding rate constant in more complex and realistic protein-ligand systems, a quantity extremely difficult to compute without careful enhanced sampling based approaches.^{38,47} The hope is that this approach will contribute a step toward the success of computational drug discovery programs that take drug unbinding dynamics into account. It would also be interesting to see how closely the RC and associated unbinding pathways identified through SGOOP correlate with realistic reaction paths obtained from single molecule experiments on dissociation of biomolecular complexes. We also think that the current work is a demonstration of how SGOOP may be used to answer similar questions in systems other than drug unbinding where the role of water density fluctuations in driving the dynamics is believed to play a role but which is hard to quantify.

Using the model system in this work allows us to study an unbinding problem involving solvation and steric related complexities, yet where we can perform extensive simulations of the reverse binding process. Undoubtedly more realistic systems will be harder to tackle than the model system of the current work, possibly involving a much larger set of trial collective variables than used here and requiring more care in coming up with this trial set to begin with. For instance, it might be crucial to incorporate the conformational fluctuations in the receptor or the ligand molecule or both into the trial set of collective variables. In principle, this should be possible with SGOOP and is the subject of current investigations. As long as the system's intrinsic dynamics displays a time scale separation between few slow and remaining fast processes, and hence possesses an associated spectral gap, we expect SGOOP to be useful in obtaining a sense of fluctuations that matter for driving the dynamics in rare event systems.

We would like to emphasize that the systems considered in this work, in spite of their model nature, are in fact quite challenging test cases. This is due to the enormous

barrier height involved (around 30–35 $k_B T$), and the relative insignificance of the barrier in the dewetting related bimodal distribution (around 1–2 $k_B T$)⁹ relative to the main barrier. As such, even the trial RC that excludes wetness entirely, considered in this work and in Ref. 11, does a remarkably decent job when used with metadynamics.^{30,31} Yet SGOOP does very well in picking up signals in the right directions for improving the RC towards ideal. This demonstration makes us optimistic that in more complex systems where the barrier associated to movement of water is expected to be higher,^{38,48–50} the algorithm will be even more useful. Some such studies are already underway and will be the subject of future publications.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (Grant No. NIH-GM4330) and the Extreme Science and Engineering Discovery Environment (XSEDE) (Grant No. TG-MCA08X002).

- ¹R. A. Copeland, D. L. Pompliano, and T. D. Meek, *Nat. Rev. Drug Discovery* **5**, 730 (2006).
- ²R. A. Copeland, *Nat. Rev. Drug Discovery* **15**, 87 (2015).
- ³G. R. Bowman, K. A. Beauchamp, G. Boxer, and V. S. Pande, *J. Chem. Phys.* **131**, 124101 (2009).
- ⁴I. Buch, T. Giorgino, and G. De Fabritiis, *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10184 (2011).
- ⁵N. Plattner and F. Noé, *Nat. Commun.* **6**, 7653 (2015).
- ⁶A. C. Pan, D. W. Borhani, R. O. Dror, and D. E. Shaw, *Drug Discovery Today* **18**, 667 (2013).
- ⁷R. Baron and J. A. McCammon, *Annu. Rev. Phys. Chem.* **64**, 151 (2013).
- ⁸P. Setny, R. Baron, P. M. Kekenus-Huskey, J. A. McCammon, and J. Dzubiella, *Proc. Natl. Acad. Sci. U. S. A.* **110**, 1197 (2013).
- ⁹J. Mondal, J. A. Morrone, and B. J. Berne, *Proc. Natl. Acad. Sci. U. S. A.* **110**, 13277 (2013).
- ¹⁰J. Mondal, R. A. Friesner, and B. Berne, *J. Chem. Theory Comput.* **10**, 5696 (2014).
- ¹¹P. Tiwary, J. Mondal, J. A. Morrone, and B. J. Berne, *Proc. Natl. Acad. Sci. U. S. A.* **112**, 12015 (2015).
- ¹²R. G. Weiß, P. Setny, and J. Dzubiella, “Solvent fluctuations induce non-Markovian kinetics in hydrophobic pocket-ligand binding,” *J. Phys. Chem. B* (to be published).
- ¹³W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).
- ¹⁴R. B. Best and G. Hummer, *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6732 (2005).
- ¹⁵R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7426 (2005).

- ¹⁶B. Peters and B. L. Trout, *J. Chem. Phys.* **125**, 054108 (2006).
- ¹⁷A. Ma and A. R. Dinner, *J. Phys. Chem. B* **109**, 6769 (2005).
- ¹⁸M. A. Rohrdanz, W. Zheng, M. Maggioni, and C. Clementi, *J. Chem. Phys.* **134**, 124116 (2011).
- ¹⁹G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, and F. Noé, *J. Chem. Phys.* **139**, 015102 (2013).
- ²⁰M. Ceriotti, G. A. Tribello, and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13023 (2011).
- ²¹M. Chen, T.-Q. Yu, and M. E. Tuckerman, *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3235 (2015).
- ²²G. Hummer and A. Szabo, *J. Phys. Chem. B* **119**, 9029 (2014).
- ²³P. Tiwary and B. J. Berne, *Proc. Natl. Acad. Sci. U. S. A.* **113**, 2839 (2016).
- ²⁴J. A. Morrone, J. Li, and B. J. Berne, *J. Phys. Chem. B* **116**, 378 (2012).
- ²⁵J. Li, J. A. Morrone, and B. Berne, *J. Phys. Chem. B* **116**, 11537 (2012).
- ²⁶P. G. Bolhuis and D. Chandler, *J. Chem. Phys.* **113**, 8154 (2000).
- ²⁷A. J. Patel, P. Varilly, and D. Chandler, *J. Phys. Chem. B* **114**, 1632 (2010).
- ²⁸G. Hummer and S. Garde, *Phys. Rev. Lett.* **80**, 4193 (1998).
- ²⁹See supplementary material at <http://dx.doi.org/10.1063/1.4959969> for further simulation details.
- ³⁰P. Tiwary and M. Parrinello, *Phys. Rev. Lett.* **111**, 230602 (2013).
- ³¹M. Salvalaglio, P. Tiwary, and M. Parrinello, *J. Chem. Theory Comput.* **10**, 1420 (2014).
- ³²S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, *Rev. Mod. Phys.* **85**, 1115 (2013).
- ³³P. D. Dixit, A. Jain, G. Stock, and K. A. Dill, *J. Chem. Theory Comput.* **11**, 5464 (2015).
- ³⁴R. R. Coifman, I. G. Kevrekidis, S. Lafon, M. Maggioni, and B. Nadler, *Multiscale Model. Simul.* **7**, 842 (2008).
- ³⁵B. Matkowsky and Z. Schuss, *SIAM J. Appl. Math.* **40**, 242 (1981).
- ³⁶H. Risken, *Fokker-Planck Equation* (Springer, 1984).
- ³⁷P. Tiwary and M. Parrinello, *J. Phys. Chem. B* **119**, 736 (2014).
- ³⁸P. Tiwary, V. Limongelli, M. Salvalaglio, and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.* **112**, E386 (2015).
- ³⁹O. Valsson, P. Tiwary, and M. Parrinello, *Annu. Rev. Phys. Chem.* **67**, 159 (2016).
- ⁴⁰B. J. Berne, M. Borkovec, and J. E. Straub, *J. Phys. Chem.* **92**, 3711 (1988).
- ⁴¹P. Tiwary and B. J. Berne, *J. Chem. Phys.* **144**, 134103 (2016).
- ⁴²W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, *J. Am. Chem. Soc.* **118**, 11225 (1996).
- ⁴³B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl, *J. Chem. Theory Comput.* **4**, 435 (2008).
- ⁴⁴G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, *Comput. Phys. Commun.* **185**, 604 (2014).
- ⁴⁵G. Bussi, D. Donadio, and M. Parrinello, *J. Chem. Phys.* **126**, 014101 (2007).
- ⁴⁶H. J. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. Haak, *J. Chem. Phys.* **81**, 3684 (1984).
- ⁴⁷I. Teo, C. G. Mayne, K. Schulten, and T. Lelièvre, *J. Chem. Theory Comput.* **12**, 2983 (2016).
- ⁴⁸P. Liu, X. Huang, R. Zhou, and B. J. Berne, *Nature* **437**, 159 (2005).
- ⁴⁹Y. Shan, E. T. Kim, M. P. Eastwood, R. O. Dror, M. A. Seeliger, and D. E. Shaw, *J. Am. Chem. Soc.* **133**, 9181 (2011).
- ⁵⁰M. Ø. Jensen, V. Jogini, D. W. Borhani, A. E. Leffler, R. O. Dror, and D. E. Shaw, *Science* **336**, 229 (2012).