Spectral gap optimization of order parameters for sampling complex molecular systems

Pratyush Tiwary[a] and B. J. Berne[a,1]

[a]Department of Chemistry, Columbia University, New York, NY 10027

Contributed by B. J. Berne, January 20, 2016 (sent for review November 4, 2015; reviewed by Peter G. Bolhuis, Ken A. Dill, and Attila Szabo)

In modern-day simulations of many-body systems, much of the computational complexity is shifted to the identification of slowly changing molecular order parameters called collective variables (CVs) or reaction coordinates. A vast array of enhanced-sampling methods are based on the identification and biasing of these low-dimensional order parameters, whose fluctuations are important in driving rare events of interest. Here, we describe a new algorithm for finding optimal low-dimensional CVs for use in enhanced-sampling biasing methods like umbrella sampling, metadynamics, and related methods, when limited prior static and dynamic information is known about the system, and a much larger set of candidate CVs is specified. The algorithm involves estimating the best combination of these candidate CVs, as quantified by a maximum path entropy estimate of the spectral gap for dynamics viewed as a function of that CV. The algorithm is called spectral gap optimization of order parameters (SGOOP). Through multiple practical examples, we show how this postprocessing procedure can lead to optimization of CV and several orders of magnitude improvement in the convergence of the free energy calculated through metadynamics, essentially giving the ability to extract useful information even from unsuccessful metadynamics runs.

collective variables | timescale separation | spectral gap | caliber | enhanced sampling

With the advent of increasingly accurate force fields and powerful computers, molecular-dynamics (MD) simulations have become a ubiquitous tool for studying the static and dynamic properties of systems across disciplines. However, most realistic systems of interest are characterized by deep, multiple free-energy basins separated by high barriers. The timescales associated with escaping such barriers can be formidably high compared with what is accessible with straightforward MD even with the most powerful computing resources. Thus, to accurately characterize such landscapes with atomistic simulations, a large number of enhanced-sampling schemes have become popular, starting with the pioneering works of Torrie, Valleau, Bennett, and others (1–13). Many of these schemes involve probing the probability distribution along selected low-dimensional collective variables (CVs), either through a static preexisting bias or through a bias constructed on-the-fly, that enhances the sampling of hard-to-access but important regions in the configuration space.

The quality, reliability, and usefulness of the sampled distribution is in the end deeply dependent on the quality of the chosen CV. Specifically, one key assumption inherent in several enhanced-sampling methods is that of timescale separation (14): for efficient and accurate sampling, the chosen CV should encode all of the relevant slow dynamics in the system, and any dynamics not captured by the CV should be relatively fast. For most practical applications, there are a large number of possible CVs that could be chosen, and it is not at all obvious how to construct the best low-dimensional CV or CVs for biasing from these various possible options. Success in enhanced-sampling simulations has traditionally relied on an apt use of physical intuition to construct such low-dimensional CVs. Identification of good low-dimensional CVs is in fact useful not just for enhanced-sampling simulations such as umbrella sampling and metadynamics but also for distributed computing techniques like Markov state models (MSMs) (15), allowing one to significantly improve the quality and reliability of the constructed kinetic models. Last but not the least, having an optimal low-dimensional CV can also help in the building of Brownian dynamics-type models (16, 17). Indeed, given the importance of this problem, there exists a range of methods that have been proposed to solve it (18–25).

In this communication, we report a new and computationally efficient algorithm for designing good low-dimensional slow CVs. We suggest that the best CV is one with the maximum separation of timescales between visible slow and hidden fast processes (14, 26). This timescale separation is calculated as the spectral gap between the slow and fast eigenvalues of the transition probability matrix (see *Theory* for a rigorous definition and implementation of the spectral gap as used in this work). The method is named spectral gap optimization of order parameters (SGOOP). Note that, in this work, henceforth we refer to the best CV in the singular, without loss of any generality in the treatment. The notion of such a timescale separation and spectral gap is at the core of not just enhanced-sampling methods but also coarse-grained, multiscale, MSM, and projection operator methods (15, 27–29).

Our algorithm involves learning the best linear or nonlinear combination of given candidate CVs, as quantified by a maximum path entropy (30) estimate of the spectral gap for the dynamics of that CV. The input to the algorithm is any available information about the static and dynamic properties of the system, accumulated through (*i*) a biased simulation performed along a suboptimal trial CV, possibly (but not necessarily) complemented by (*ii*) short bursts of unbiased MD runs, or (*iii*) by knowledge of experimental observables. Any type of biased simulation could be used in *i*, as long as it allows projecting the stationary probability density estimate on generic CVs without

Significance

Molecular-dynamics (MD) simulations have become a versatile tool for exploration of complex molecular systems. However, they are limited in the timescales that can be reached. Thus, over the years, a suite of enhanced-sampling algorithms have been proposed that assist MD to transcend the timescale limitation, with diverse applications across physical and life sciences. A continuing grand challenge in the success of many such sampling methods pertains to a judicious choice of order parameters. In this work, we propose a new method for designing order parameters that minimizes the role played by human intuition and makes the progress significantly more automated than before. We expect this algorithm to be of great use in furthering the success of enhanced sampling.

Author contributions: P.T. and B.J.B. designed research; P.T. and B.J.B. performed research; P.T. analyzed data; and P.T. and B.J.B. wrote the paper.

Reviewers: P.G.B., University of Amsterdam; K.A.D., Stony Brook University; and A.S., National Institutes of Health.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. Email: bb8@columbia.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1600917113/-/DCSupplemental.

having to repeat the simulation. Metadynamics (31) provides this functionality in a straightforward manner, and hence it is our method of choice here. Given this information, we use the principle of maximum caliber (30) to set up an unbiased master equation for the dynamics of various trial CVs. Through a simple postprocessing optimization procedure, we then find the CV with the maximal spectral gap of the associated transfer matrix. For instance, this optimization can be performed through a simulated annealing approach that maximizes the spectral gap by performing a robust global search in the space of trial CVs.

Through three practical examples, we show how our postprocessing procedure can lead to better choices of CVs, and to several orders of magnitude improvement in the convergence of the free energy calculated through the popular enhanced-sampling technique metadynamics. Furthermore, the algorithm is generally applicable irrespective of the number of stable basins. Our algorithm essentially provides the much needed ability to extract useful information about relevant CVs even from unsuccessful metadynamics runs. In addition to use in free-energy sampling methods, the optimized CV can then also be used in other methods that provide kinetic rate constants (32, 33). We expect this algorithm to be of widespread use in designing CVs for biasing during enhanced-sampling simulations, making the process significantly more automatic and far less reliant on human intuition.

Theory

Let us consider a molecular system with N atoms at temperature T. We assume there exists a large number d of available order parameters with $1 \ll d \ll N$, collectively referred to as $\{\Theta\}$, such that the dynamics in this d-dimensional space is Markovian. These could be intermolecular distances (18), torsional angles, solvation states, nucleus size/shape (34), bond order parameters (35), etc. The identification of such order parameters poses another complicated problem, but as routinely done in other methods aimed at optimizing CVs (15, 18, 24), we assume such order parameters are a priori known.

There are several available biasing techniques that can sample the probability distribution of the space $\{\Theta\}$, and even calculate the rate constants for escape from stable states in this space (32). All of these techniques are feasible only for a very small number of CVs whose number is much smaller than d—typically one to three. These are the order parameters whose fluctuations are deemed to be most important for the system or process being studied, and by building a fixed or time-dependent bias of these CVs, one should be able to determine the true unbiased probability distribution of the full space $\{\Theta\}$. However, how does one decide what is an optimal low-dimensional subset or combination of the available order parameters? This dimensionality reduction is of central importance to methods such as umbrella sampling, metadynamics, and others, the answer to which decides the speed of convergence of the biased simulation, or if it will even ever converge within practically useful simulation times.

The key idea in the current work is to perform enhanced sampling (e.g., metadynamics) with a choice of trial CVs, complemented by information gathered from short bursts of unbiased MD simulations and experimental observables when available, to iteratively improve the CVs. The maximum caliber framework (30, 36, 37), which is a dynamical generalization of the hugely popular maximum entropy framework (38), provides a method for accomplishing this, which is now used in fields as diverse as biology, signal processing, and image reconstruction. In this, given certain information about the system at hand, one builds a model that is consistent with our ignorance of unknown or missing information. The maximum caliber approach (30) is a generalization of this approach to dynamics, with similar underlying ideas.

We start by choosing a trial CV given by $f\{\Theta\}$, where f maps the space $\{\Theta\}$ onto a lower-dimensional space. The space along this trial CV $f\{\Theta\}$ is then discretized in grids labeled n. This CV could be multidimensional, with n then indexing the multidimensional grids. Let $p_n(t)$ denote the instantaneous probability of the system being found in grid n. For the sake of clarity, we assume that f is a linear combination of $\{\Theta\}$, i.e., $f = c_1\Theta_1 + c_2\Theta_2 + \ldots + c_d\Theta_d$. The treatment developed here applies to nonlinear combinations as well, which we show in the examples. Then, for a fixed Δt, we write a master equation:

$$\frac{\Delta p_n(t)}{\Delta t} = \Sigma_m k_{mn} p_m(t) - \Sigma_m k_{nm} p_n(t) \equiv \Sigma_m \mathbf{K}_{nm} p_m(t), \qquad [1]$$

where k_{nm} is the rate of transition from grid n to m per unit time (39). The matrix $\mathbf{K}$, where $\mathbf{K}_{nm} = k_{mn}$, is the entirety of all these rates. If the dynamics of $f\{\Theta\}$ is Markovian, then the matrix $\mathbf{\Omega}$ of transition probabilities is given for small Δt by the following:

$$\mathbf{\Omega} = \exp(\mathbf{K}\Delta t) \approx \mathbf{I} + \mathbf{K}\Delta t, \qquad [2]$$

and should not depend on the value of Δt used in Eq. **1**. This provides a self-consistency check of whether or not the CV so generated is Markovian. Similar to $\mathbf{K}$, the matrix Ω has terms $\mathbf{\Omega}_{nm} = \omega_{mn}$, where $\omega_{ab} = k_{ab}\Delta t$ for $a \neq b$ and the normalization $\Sigma_b \omega_{ab} = 1$ is satisfied. In the maximum caliber approach, one uses all available stationary state and dynamical information to construct probabilities of micropaths. Instead of defining the entropy as a function of microstate probabilities as in information theory and statistical thermodynamics (38), one now defines an entropy S as a functional of the probabilities of micropaths, which is essentially a path integral. For the Markovian process of Eq. **1** (40):

$$S = -\Sigma_{ab} p_a \omega_{ab} \log \omega_{ab}. \qquad [3]$$

Note that ω_{ab} are not rate constants but transition probabilities of a Markov model that is discrete in both space and time. Path ensemble averages of time-dependent quantities A_{ab} can now be calculated as follows (30), where the subscripts a,b denote grid indices:

$$\langle A \rangle = \Sigma_{ab} p_a \omega_{ab} A_{ab}. \qquad [4]$$

The path entropy of Eq. **3** incremented by quantities accounting for constraints placed by our knowledge of observables $\langle A^n \rangle$, where n runs over the number of known observables, and some other constraints such as detailed balance, is collectively called caliber (30). As derived for instance in ref. 37, maximizing the caliber is then equivalent to being least committal about missing dynamic and static information, with the end result being that one obtains a relation between the grid-to-grid rates and the stationary probabilities as follows:

$$\omega_{ab} = \sqrt{\frac{p_b}{p_a}} e^{-\Sigma_i \rho_i A^i_{ab}}. \qquad [5]$$

Here, i runs over the number of available dynamical pieces of information, and ρ_i is the Lagrange multiplier for the associated constraint. As a special case, consider when the only observable at hand is the mean number of transitions $\langle N \rangle$ in observation interval Δt over the entire gridded CV (37). $\langle N \rangle$ would be a measure of the total number of jumps in the time Δt between any two points on the gridded CV. In this case, the above equation takes a particularly simple and useful form:

$$\omega_{ab} = \sqrt{\frac{p_b}{p_a}} e^{-\rho}. \qquad [6]$$

Eqs. **5** and **6** are the two central equations in this work upon which the estimation of the spectral gap of the dynamics is based.

Interestingly, an equation similar to Eq. **6** has been previously derived by Bicout and Szabo by assuming a constant position-dependent diffusivity (41).

Spectral Gap. Our method involves calculating for various trial CVs the spectral gap of the associated transition probability matrix $\mathbf{\Omega}$. Let $\{\lambda\}$ denote the set of eigenvalues of $\mathbf{\Omega}$, with $\lambda_0 \equiv 1 > \lambda_1 \geq \lambda_2 \ldots$. The size of this set depends on the discretization interval of the trial CV f—for the purposes of improving CVs, we found very little sensitivity to the details of the discretization. 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 $\gtrsim k_B T$). Estimating the Lagrange multiplier is computationally expensive, so a first estimate for maximizing the spectral gap is performed using Eq. **6** where the Lagrange multiplier ρ need not be computed, because it sets only the overall timescale but does not influence the spectral gap. Also note that, in the limit of small Δt, the matrix $\mathbf{\Omega}$ will be diagonally dominated (42), and to estimate the spectral gap one needs only an accurate estimate of relative local free energies.

There is a wide scope for creativity in choosing the dynamic observables to be used to constrain the caliber for calculating the spectral gap. For instance, one could consider the average number of transitions per unit time not on the entire grid as we do here, but separately in different parts of the configuration space. One could even include experimental observables such as correlation functions from scattering experiments. More static or dynamical information (41, 43–47) simply introduces additional Lagrange multipliers and can be treated through Eq. **5**. This can be done if the intention is to calculate an accurate kinetic model with correct estimates of the dominant eigenvalues and not just the spectral gap. For detailed balance to be satisfied through Eq. **5**, the observable must be symmetric or be symmetrized on the grid, i.e., $A_{ab} = A_{ba}$.

Algorithm. We are now in a position to describe the actual algorithm. It comprises the following two steps in a sequential manner, and can be improved by iterating:

i) Perform metadynamics along a trial CV $f = c_1\Theta_1 + c_2\Theta_2 + \ldots + c_d\Theta_d$ to get a crude estimate of the stationary density.

ii) As postprocessing, perform optimization in the space of mixing coefficients $\{c_1, c_2 \ldots c_d\}$ to identify the CV with the maximal spectral gap. The reweighting functionality (31) of metadynamics allows projection of free-energy estimates on different CVs with minimal computational effort, and is used to calculate the $\mathbf{\Omega}$ matrix through Eq. **6** (see ref. 31 and *Supporting Information* for a summary of reweighting in metadynamics). We elaborate on the optimization procedure details in the next section (*Illustrative Examples*).

The optimization procedure gives the best CV as the one with highest spectral gap, given the information at hand. As in any maximum entropy framework (38), the better the quality of this information, the more accurate will be the spectral gap. However, even with very poor quality information, as we show in the examples, the algorithm still leads to significant improvements in the CV. Furthermore, whether or not the CV is Markovian can also be checked by repeating step *ii* for different time intervals Δt of observation and determining whether the spectral gap is independent of the value of Δt.

It is natural to compare our approach with MSM. The similarity between these two approaches begins and ends with the construction of the master equation Eq. **1**. A MSM builds this equation by constructing extensive unbiased simulations and attempts to obtain all relevant eigenvalues. Ours is a maximum path entropy-based approach that uses biased and unbiased simulations to obtain the difference between the slow and fast eigenvalues rather than the exact spectrum of eigenvalues.

Illustrative Examples

Model 2D Landscapes: The De Leon–Berne Potential. The first illustrative example for SGOOP is a model two-state potential introduced by De Leon and Berne (48). To sample this landscape at temperature $k_B T = 0.1$, we perform metadynamics with path CVs, a class of widely used CVs that can capture nonlocal and nonlinear fluctuations (see *Supporting Information* and ref. 49 for details). Path CVs require specification of a series of landmarks between two points in configuration space, where the landmarks can be described in terms of generic order parameters. Fluctuations in the system can then be enhanced in the direction along and perpendicular to these landmarks, leading to efficient exploration of the space. In Fig. 1*A*, we show the 2D potential along with several possible path CVs imposed on it. We first perform a short trial metadynamics run biasing the y coordinate. By postprocessing this, we generate the spectral gaps for various paths using Eq. **6** (Fig. 1*B*). By comparing Fig. 1*A* against Fig. 1*B*, it is clear how the path with maximum spectral gap is the minimum energy pathway passing through the saddle point. In this case, although this result could have simply been obtained through nudged elastic band-type calculations (50), the point is to use this example to develop intuition for the method. Also note that moving around the best path to others that are a bit distant from it, does not lead to much change in the spectral gap. This is consistent with the observation that, in several enhanced-sampling

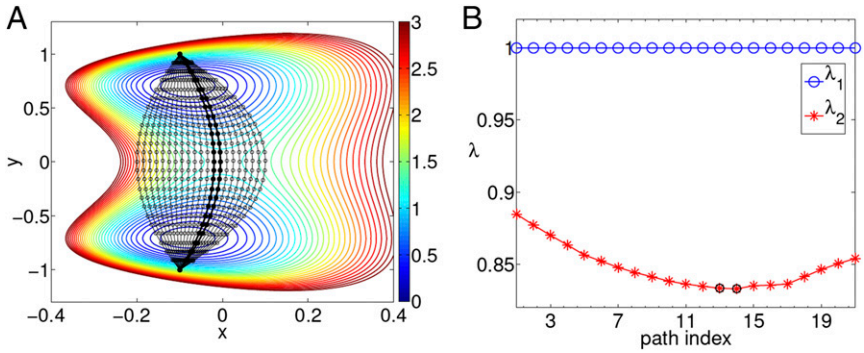

Fig. 1. In *A*, we provide the 2D De Leon–Berne potential (48) with several candidate path CVs imposed on it. Black circles denote the corresponding landmarks (49). See *Supporting Information* for further details of path CVs. In *B*, the corresponding eigenvalues λ_1 and λ_2 (i.e., excluding the stationary eigenvalue λ_0) are shown for each of these paths. As per the spectral gap given here by $\lambda_1 - \lambda_2$, we identify two possible good paths marked with black circles in *B* and correspondingly with thicker black lines in *A*. Energy is in absolute units and $k_B T = 0.1$.

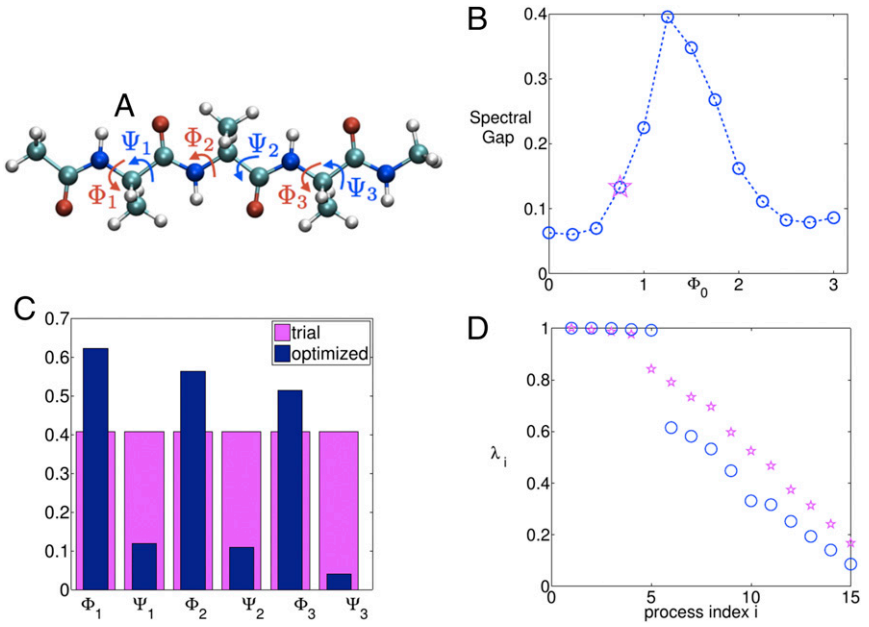

Fig. 2. (*A*) The five-residue peptide studied in this work. The six dihedral angles are marked. (*B*) The output of the simulated annealing algorithm run separately for different θ_0 values (blue circles). The starting value with the trial choice of CV is marked with a magenta-colored star. (*C*) The trial (magenta) and optimized (blue) mixing coefficients $\{c\}$ for the six dihedrals. (*D*) The spectrum of eigenvalues for dynamics projected on the trial (magenta) and optimized (blue) CVs. A distinct improvement can be seen in the spectral gap. Process index *i* refers to the *i*th index in the transition matrix.

methods such as metadynamics or umbrella sampling (3, 7, 8), the CV need not be precisely the true reaction coordinate, as long as it has a sufficient overlap with it (49, 51).

In *Supporting Information*, we provide a similar analysis on another 2D model potential but with three states (Fig. S1). The conclusions are similar.

Five-Residue Peptide. Now, we move to a more complex system, which has also been considered as a test case for new enhanced-sampling methods (52) to establish their usefulness. This is the five-residue peptide Ace–Ala_3–Nme in vacuum (Fig. 2*A*), where there are six possibly relevant dihedral torsion angles. Here, we ask the question: what is the best possible 1D linear combination of these six dihedrals that we could bias but still maximally enhance exploration of the 6D space comprising all of the dihedrals?

In this problem, for periodicity-related numerical reasons, we bias a reference cosine defined by $\cos(\theta-\theta_0)$, where θ is one of the six dihedral angles, and θ_0 is some reference value whose optimal choice we do not know a priori. Through our algorithm we then seek to identify:

i) The best choice of mixing coefficients $\{c\}$ to use in trial CV $f=c_1\Phi_1'+c_2\Psi_1'+c_3\Phi_2'+c_4\Psi_2'+c_5\Phi_3'+c_6\Psi_3'$, where we keep the Euclidean norm of $\{c\}=1$, and for any angle θ the prime denotes the transformation $\theta\mapsto 0.5+\cos(\theta-\theta_0)$;

ii) The best choice of θ_0, kept same for all six dihedrals.

We start with the trial CV where all members of $\{c\}$ are the same subject to Euclidean norm of $\{c\}$ = 1, and an arbitrary choice of $\theta_0=0.75$ radians is taken. A short metadynamics run is performed biasing this trial CV. See *Supporting Information* for details of the metadynamics and MD parameters (53–55), and Fig. 3*A* for the metadynamics trajectory used for spectral gap optimization. Based on the free-energy estimate generated from this run, a simulated annealing procedure is performed in the space $\{c\}$ for various θ_0 values. Starting from the spectral gap estimated using Eq. **6** for the trial CV, this involves executing Metropolis moves in the $\{c\}$ space with an attempt to find the global maxima of the spectral gap. In Fig. 2, *B–D*, respectively, we show how the spectral gap is increased by the simulated annealing procedure, and the corresponding best estimate of $\{c,\theta_0\}$. The algorithm suggests the minimal role of the angles

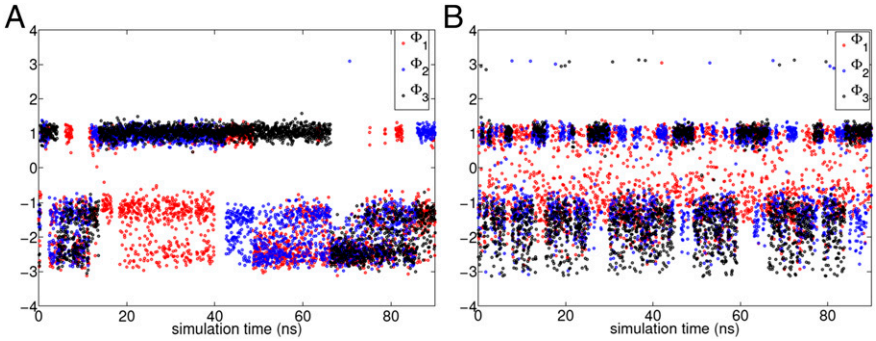

Fig. 3. *A* and *B* show trajectories obtained from metadynamics biasing the trial CV and the optimized CV, respectively. The first 20 ns of the trajectory shown in *A* was used to generate the optimized CV for *B*. A very pronounced improvement in the enhancement of sampling can be seen with the optimized CV.

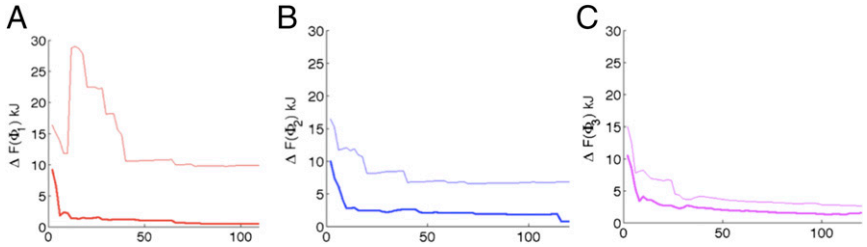

Fig. 4. Errors in the 1D free energies for three dihedrals in kilojoules calculated with respect to respective reference free energies (52, 61, 62) using the error metric from ref. 56. Thin and thick lines denote values using the trial and optimized CVs, respectively. *A*, *B*, and *C* denote error in the free energies for the dihedrals Φ_1, Φ_2, and Φ_3, respectively.

Ψ_1, Ψ_2, Ψ_3 as can be seen through their relatively low weights (52) (Fig. 2*C*). The spectrum of eigenvalues for dynamics projected on the trial (magenta) and optimized (blue) CVs, along with respective spectral gaps is provided in Fig. 2*D*. Fig. 3, *A* and *B*, shows the metadynamics trajectories for the three dihedral angles Φ_1, Φ_2, Φ_3, with the trial and the optimized CVs, respectively. A very pronounced improvement in the quality of sampling can be seen. Fig. 4 *A*–*C* shows the rate of convergence of the error of the estimated free energy (31) with respect to reference values from other approaches (52), through metadynamics runs performed with each of the trial and optimized CVs, respectively. The error metric is the same as in refs. 52 and 56, and is calculated for all points within 25 kJ of the global minimum in the respective 1D free energy. The behavior is robust with respect to the choice of this threshold value. As can be seen, the optimized CV, even though it was obtained on the basis of a very poorly converged and short (20-ns) metadynamics run, leads to several orders of magnitudes improvement in the rate at which the free energies converge. Interestingly, iterating the algorithm with the improved 1D CV did not lead to much improvement in the sampling, reflecting that the optimized coefficients $\{c\}$ are close to the best that can be achieved with a 1D CV for this problem.

Discussion

To conclude, we have introduced a new approach named SGOOP for improving the choice of low-dimensional CVs for biasing in enhanced sampling in complex systems. This is accomplished through the use of a maximum caliber-based approach, where we build kinetic models for different CVs. For each CV, we separate out the slow motions that involve crossing barriers, from hidden or orthogonal motions. Through a spectral gap maximization, we make the orthogonal fluctuations as fast as possible, compared with the slow motions apparent in the CV. The algorithm is iterative in spirit and attempts to learn how to improve CVs based on available stationary and dynamic data. We also provide several proof-of-concept practical examples to establish the potential usefulness of the method. For model 2D potentials, the algorithm was shown to yield the minimum energy pathway. For a small peptide, we found very significant improvement in determining the best 1D CV from six possible functions with no ad hoc or intuition-based tuning. Future work will use this algorithm to treat a range of problems, especially involving protein–ligand unbinding. For instance, the displacement of water molecules and protein flexibility are often slowly varying order parameters in unbinding (33, 51, 57, 58), but do we really need to bias one or both of these for the purpose of sampling? Another issue to be considered in future work is whether we can use these optimized CVs to obtain reliable dynamical information from metadynamics (25, 32), including the very important off-rate for ligand unbinding (51, 59).

One central limitation of this algorithm is having to specify possibly a large number of order parameters that may be important. However, for many physical problems, one does have a sense of which order parameters could be at work, and this is where we expect this algorithm to be of tremendous use. Another obvious limitation is with systems devoid of a timescale separation (60)—for example, in glassy systems where there is an effectively continuous spectrum of eigenvalues with no discernible timescale separation. However, the dynamics of many complex and real-world molecular systems does thankfully show a timescale separation between few relevant slow modes and remaining fast ones (62), and we expect our algorithm to be of help in unraveling the thermodynamics and dynamics in such systems.

ACKNOWLEDGMENTS. We thank Purushottam Dixit for helpful discussions regarding caliber, Omar Valsson for providing system setup and reference free energies for the peptide, and Jed Brown for originally suggesting a spectral gap approach. This work was supported by National Institutes of Health Grant NIH-GM4330 and Extreme Science and Engineering Discovery Environment Grant TG-MCA08X002.

1. Bolhuis PG, Chandler D, Dellago C, Geissler PL (2002) Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu Rev Phys Chem* 53(1):291–318.
2. Valsson O, Tiwary P, Parrinello M (2016) Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint. *Annu Rev Phys Chem* 67(1):1–27.
3. Torrie GM, Valleau JP (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J Comput Phys* 23(2):187–199.
4. Carter E, Ciccotti G, Hynes JT, Kapral R (1989) Constrained reaction coordinate dynamics for the simulation of rare events. *Chem Phys Lett* 156(5):472–477.
5. Hansmann UH, Okamoto Y (1993) Prediction of peptide conformation by multicanonical algorithm: New approach to the multiple-minima problem. *J Comput Chem* 14(11):1333–1338.
6. Voter AF (1997) Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Phys Rev Lett* 78(2):3908–3911.
7. Laio A, Parrinello M (2002) Escaping free-energy minima. *Proc Natl Acad Sci USA* 99(20):12562–12566.
8. Barducci A, Bussi G, Parrinello M (2008) Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys Rev Lett* 100(2):020603–020606.
9. Darve E, Rodríguez-Gómez D, Pohorille A (2008) Adaptive biasing force method for scalar and vector free energy calculations. *J Chem Phys* 128(14):144120.
10. Abrams CF, Vanden-Eijnden E (2010) Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics. *Proc Natl Acad Sci USA* 107(11):4961–4966.
11. Zheng L, Chen M, Yang W (2008) Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc Natl Acad Sci USA* 105(51):20227–20232.
12. Tiwary P, van de Walle A (2013) Accelerated molecular dynamics through stochastic iterations and collective variable based basin identification. *Phys Rev B* 87(9):094304–094307.
13. Faradjian AK, Elber R (2004) Computing time scales from reaction coordinates by milestoning. *J Chem Phys* 120(23):10880–10889.
14. Berezhkovskii A, Szabo A (2011) Time scale separation leads to position-dependent diffusion along a slow coordinate. *J Chem Phys* 135(7):074108.
15. Pérez-Hernández G, Paul F, Giorgino T, De Fabritiis G, Noé F (2013) Identification of slow molecular order parameters for Markov model construction. *J Chem Phys* 139(1):015102.
16. Ermak DL, McCammon J (1978) Brownian dynamics with hydrodynamic interactions. *J Chem Phys* 69(4):1352–1360.
17. Morrone JA, Li J, Berne BJ (2012) Interplay between hydrodynamics and the free energy surface in the assembly of nanoscale hydrophobes. *J Phys Chem B* 116(1):378–389.
18. Best RB, Hummer G (2005) Reaction coordinates and rates from transition paths. *Proc Natl Acad Sci USA* 102(19):6732–6737.
19. Coifman RR, et al. (2005) Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proc Natl Acad Sci USA* 102(21):7426–7431.

20. Peters B, Trout BL (2006) Obtaining reaction coordinates by likelihood maximization. *J Chem Phys* 125(5):054108.
21. Ma A, Dinner AR (2005) Automatic method for identifying reaction coordinates in complex systems. *J Phys Chem B* 109(14):6769–6779.
22. Rohrdanz MA, Zheng W, Maggioni M, Clementi C (2011) Determination of reaction coordinates via locally scaled diffusion map. *J Chem Phys* 134(12):124116.
23. Ceriotti M, Tribello GA, Parrinello M (2011) From the Cover: Simplifying the representation of complex free-energy landscapes using sketch-map. *Proc Natl Acad Sci USA* 108(32):13023–13028.
24. Chen M, Yu TQ, Tuckerman ME (2015) Locating landmarks on high-dimensional free energy surfaces. *Proc Natl Acad Sci USA* 112(11):3235–3240.
25. Salvalaglio M, Tiwary P, Parrinello M (2014) Assessing the reliability of the dynamics reconstructed from metadynamics. *J Chem Theory Comput* 10(4):1420–1425.
26. Coifman RR, Kevrekidis IG, Lafon S, Maggioni M, Nadler B (2008) Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems. *Mult Mod Sim* 7(2):842–864.
27. Berne BJ, Pecora R (2000) *Dynamic Light Scattering* (Dover Publications, Inc., Mineola, NY).
28. Car R, Parrinello M (1985) Unified approach for molecular dynamics and density-functional theory. *Phys Rev Lett* 55(22):2471–2474.
29. Kevrekidis IG, et al. (2003) Equation-free, coarse-grained multiscale computation. *Commun Math Sci* 1(4):715–762.
30. Pressé S, Ghosh K, Lee J, Dill KA (2013) Principles of maximum entropy and maximum caliber in statistical physics. *Rev Mod Phys* 85:1115.
31. Tiwary P, Parrinello M (2015) A time-independent free energy estimator for metadynamics. *J Phys Chem B* 119(3):736–742.
32. Tiwary P, Parrinello M (2013) From metadynamics to dynamics. *Phys Rev Lett* 111(23): 230602–230606.
33. Tiwary P, Mondal J, Morrone JA, Berne BJ (2015) Role of water and steric constraints in the kinetics of cavity-ligand unbinding. *Proc Natl Acad Sci USA* 112(39): 12015–12019.
34. ten Wolde PR, Ruiz-Montero MJ, Frenkel D (1999) Numerical calculation of the rate of homogeneous gas–liquid nucleation in a Lennard-Jones system. *J Chem Phys* 110(3): 1591–1599.
35. Steinhardt PJ, Nelson DR, Ronchetti M (1983) Bond-orientational order in liquids and glasses. *Phys Rev B* 28(2):784–805.
36. Jaynes ET (1980) The minimum entropy production principle. *Annu Rev Phys Chem* 31(1):579–601.
37. Dixit PD, Jain A, Stock G, Dill KA (2015) Inferring transition rates of networks from populations in continuous-time markov processes. *J Chem Theory Comput* 11(11): 5464–5472.
38. Jaynes ET (1957) Information theory and statistical mechanics. *Phys Rev* 106(4): 620–630.
39. Zwanzig R (2001) *Nonequilibrium Statistical Mechanics* (Oxford Univ Press, New York).
40. Filyukov A, Karpov VY (1967) Method of the most probable path of evolution in the theory of stationary irreversible processes. *J Eng Phys Thermophys* 13(6):416–419.
41. Bicout D, Szabo A (1998) Electron transfer reaction dynamics in non-Debye solvents. *J Chem Phys* 109(6):2325–2338.
42. Rosta E, Hummer G (2015) Free energies from dynamic weighted histogram analysis using unbiased Markov state model. *J Chem Theory Comput* 11(1):276–285.
43. Hummer G (2005) Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New J Phys* 7(1):34–48.
44. Marinelli F, Pietrucci F, Laio A, Piana S (2009) A kinetic model of trp-cage folding from multiple biased molecular dynamics simulations. *PLoS Comput Biol* 5(8):e1000452.
45. Berne B, Pechukas P, Harp G (1968) Molecular reorientation in liquids and gases. *J Chem Phys* 49(7):3125–3129.
46. Granata D, Camilloni C, Vendruscolo M, Laio A (2013) Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics. *Proc Natl Acad Sci USA* 110(17):6817–6822.
47. Bonomi M, Camilloni C, Cavalli A, Vendruscolo M (2015) Metainference: A Bayesian inference method for heterogeneous systems. *Science Advances* 2(1):e1501177.
48. De Leon N, Berne B (1981) Intramolecular rate process: Isomerization dynamics and the transition to chaos. *J Chem Phys* 75(7):3495–3510.
49. Branduardi D, Gervasio FL, Parrinello M (2007) From A to B in free energy space. *J Chem Phys* 126(5):054103–054112.
50. Henkelman G, Uberuaga BP, Jónsson H (2000) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J Chem Phys* 113(22): 9901–9904.
51. Tiwary P, Limongelli V, Salvalaglio M, Parrinello M (2015) Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proc Natl Acad Sci USA* 112(5):E386–E391.
52. Valsson O, Parrinello M (2014) Variational approach to enhanced sampling and free energy calculations. *Phys Rev Lett* 113(9):090601–090605.
53. Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G (2014) Plumed 2: New feathers for an old bird. *Comput Phys Commun* 185(2):604–613.
54. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4(3):435–447.
55. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126(1):014101.
56. Branduardi D, Bussi G, Parrinello M (2012) Metadynamics with adaptive Gaussians. *J Chem Theory Comput* 8(7):2247–2254.
57. Ladbury JE (1996) Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem Biol* 3(12): 973–980.
58. Berne BJ, Weeks JD, Zhou R (2009) Dewetting and hydrophobic interaction in physical and biological systems. *Annu Rev Phys Chem* 60(60):85–103.
59. Copeland RA, Pompliano DL, Meek TD (2006) Drug-target residence time and its implications for lead optimization. *Nat Rev Drug Discov* 5(9):730–739.
60. Zwanzig R (1990) Rate processes with dynamical disorder. *Acc Chem Res* 23(5): 148–152.
61. Valsson O, Parrinello M (2015) Well-tempered variational approach to enhanced sampling. *J Chem Theor Comput* 11(5):1996–2002.
62. Machta BB, Chachra R, Transtrum MK, Sethna JP (2013) Parameter space compression underlies emergent theories and predictive models. *Science* 342(6158):604–607.

Supporting Information

Tiwary and Berne 10.1073/pnas.1600917113

Metadynamics

Here, we briefly describe metadynamics and the related concepts that are used in the present work.

Reweighting. The reweighting operation in metadynamics is central to this work, as it allows projecting probability densities on arbitrary collective variables (CVs) without having to repeat the simulation. A more detailed discussion can be found in ref. 31. In metadynamics, one constructs a time-dependent bias V (**s**,t) as a function of some low-dimensional CV **s** (**R**), where **R** denotes the configurational coordinates. At time t, the biased probability distribution for **R** can then be written as follows:

$$P(\mathbf{R},t)=\frac{e^{-\beta[U(\mathbf{R})+V(\mathbf{s}(\mathbf{R}),t)]}}{\int d\mathbf{R}\, e^{-\beta[U(\mathbf{R})+V(\mathbf{s}(\mathbf{R}),t)]}}, \qquad [S1]$$

where $U(\mathbf{R})$ is the potential energy of the system (31). This in turn can be rewritten as follows:

$$P(\mathbf{R},t)=P_0(\mathbf{R})\, e^{-\beta[V(\mathbf{s}(\mathbf{R}),t)-c(t)]}, \qquad [S2]$$

where $P_0(\mathbf{R})$ is the unbiased Boltzmann probability density and the function $c(t)$ is defined as follows:

$$c(t)=\frac{1}{\beta}\log\frac{\int ds e^{-\beta F(s)}}{\int ds e^{-\beta(F(s)+V(s,t))}}. \qquad [S3]$$

β is the inverse of the temperature multiplied by the Boltzmann constant k_B. The time-dependent function $c(t)$ is an estimator for the reversible work done by the bias. As shown in ref. (31), it can be calculated by substituting the running estimate of $F(s)$ (31) into Eq. **S3** as follows:

$$c(t)=\frac{1}{\beta}\log\frac{\int d\mathbf{s}\exp\left[\frac{\gamma}{\gamma-1}\beta V(\mathbf{s},t)\right]}{\int d\mathbf{s}\exp\left[\frac{1}{\gamma-1}\beta V(\mathbf{s},t)\right]}. \qquad [S4]$$

Here, γ is the bias factor in well-tempered metadynamics that modulates the decay of hill height each time a point is revisited (8). Using Eq. **S4** in Eq. **S2**, we can then easily calculate the distribution of any generic observable $O(\mathbf{R})$ over the unbiased ensemble from the metadynamics trajectory through the following:

$$\langle O(\mathbf{R})\rangle_0=\left\langle O(\mathbf{R})e^{\beta[V(\mathbf{s}(\mathbf{R}),t)-c(t)]}\right\rangle. \qquad [S5]$$

Path CVs. In the path CV framework (2, 49), one assumes that initial and final states A and B are known. One then specifies a series of landmarks between these two points, which can be described in terms of generic order parameters in some high-dimensional space **R**. This series of landmarks then denotes a trial path connecting the initial and final states in the space **R**, which we call $\mathbf{S}_0(t)$. Two variables s and z are then introduced, defined as follows, that, for a given series of landmarks, respectively denote distances along and perpendicular to the trial path:

$$s(\mathbf{R})=\lim_{\lambda\to\infty}\frac{\int_0^1 dt\, t\, e^{-\lambda\|\mathbf{S}(\mathbf{R})-\mathbf{S}_0(t)\|^2}}{\int_0^1 dt\, e^{-\lambda\|\mathbf{S}(\mathbf{R})-\mathbf{S}_0(t)\|^2}}, \qquad [S6]$$

$$z(\mathbf{R})=\lim_{\lambda\to\infty}\left(-\frac{1}{\lambda}\log\int_0^1 dt\, e^{-\lambda\|\mathbf{S}(\mathbf{R})-\mathbf{S}_0(t)\|^2}\right). \qquad [S7]$$

In practice, the paths are discretized (i.e., a finite number of landmarks are chosen), and λ is taken as the inverse distance between points in the path (49).

Simulation Setup for Metadynamics Calculations

All peptide simulations are performed with the GROMACS 4.5.4 MD package (54), patched with the PLUMED 2.2 plugin (53). The production runs were NVT (constant number, volume, temperature) with a temperature of 300 K implemented with the stochastic velocity rescaling thermostat (55). An integration time step of 2 fs was used for all runs. The model potentials were simulated in an in-house code.

For De Leon–Berne potential (main text), Gaussian hills were added every 50,000 integration time steps, with a starting height of 0.1 k_BT, width of 0.1 unit, and tempering factor (8) of 6.

For the five-residue peptide (main text), Gaussian hills were added every 400 integration time steps, with a starting height of 1.7 kJ/mol, width of 0.03 units, and tempering factor (8) of 15.

For the three-state potential (*SGOOP Optimization Details*), Gaussian hills were added every 50,000 integration time steps, with a starting height of 0.4 k_BT, width of 0.2 unit, and tempering factor (8) of 15.

SGOOP Optimization Details

For the model potentials, the maximum spectral gap was ascertained by tabulation against candidate CVs. For the peptide, simulated annealing was performed with negative of the spectral gap as the objective function. A starting temperature of 2.5 units was used for the Metropolis moves, with a geometric cooling schedule, where at each step the temperature was reduce by a factor of 0.995.

SGOOP on a 2D Three-State Potential

Similar to the De Leon–Berne potential described in the main text, we tested SGOOP as a proof of principle on another 2D potential but with three states at temperature $k_BT=0.15$. This potential can be seen in Fig. S1*A*, along with five candidate path CVs imposed on it (49). The functional form of this potential is given by the following:

$$V(x,y)=-3.0e^{-(x+2.8)^2-(y-2.5)^2}-3.7e^{-(x+0.1)^2-(y-3.5)^2}-3.7e^{-(x+1.4)^2-(y-0.3)^2}+0.005\left((x+1)^6+(y-1)^6\right).$$

We first perform short trial metadynamics run biasing the y coordinate. By applying SGOOP, we obtain spectrum of eigenvalues corresponding to trial paths, and the corresponding spectral gaps (Fig. S1*B*). As can be seen, the maximal spectral gap so obtained is approximately for the minimum energy pathways on this landscape filtering out the bad paths.

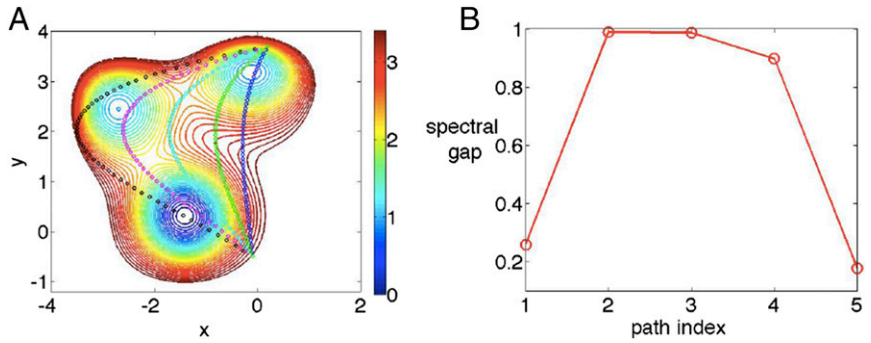

Fig. S1. (*A*) Three-state potential with five trial path CVs imposed on it, and (*B*) corresponding spectral gaps. Energies are in absolute units and simulation temperature was $k_B T = 0.15$. In both figures, paths are to be counted in the same order, and the second and third paths counting from the left have the maximal spectral gaps. These can be seen to be roughly the minimum energy pathways.