Calculating Binding Free Energies in Model Host–Guest Systems with Unrestrained Advanced Sampling

Andrew V. Marquardt, Mohsen Farshad, and Jonathan K. Whitmer*

Department of Chemical and Biomolecular Engineering
University of Notre Dame
Notre Dame, Indiana 46556, United States

E-mail: jwhitme1@nd.edu

Abstract

Host–guest interactions are important to the design of pharmaceuticals, and more broadly to soft materials, as they can enable targeted, strong, and specific interactions between molecules. The binding process between host and guest may be classified as a "rare event" when viewing the system at atomic scales, such as those explored in molecular dynamics simulations. To obtain equilibrium binding conformations and dissociation constants from these simulations, it is essential to resolve such rare events. Advanced sampling methods such as Adaptive Biasing Force (ABF) promote the occurrence of less probable configurations in a system, therefore facilitating the sampling of essential collective variables (CVs) which characterize the host–guest interactions. Here, we present the application of ABF to a rod–cavitand coarse-grained (CG) model of host-guest systems to acquire the potential of mean force (PMF). We show that the employment of ABF enables the computation of configurational and thermodynamic properties of bound and unbound states, including the free energy landscape. Moreover, we identify important dynamical bottlenecks that limit sampling and discuss how these may be addressed in more general systems.

1 Introduction

Host–guest systems are a particular class of binding interactions where one molecule (the host) contains a cavity into which a second molecule (the guest) inserts.[1–5] The interactions can involve significant specificity in matching molecular shapes and charge distributions, and often lead to strong and selective interactions between the host and guest.[5–7] These host–guest interactions are crucial to a variety of biological[5,8] and chemical processes,[9] including catalysis,[10,11] sensor design,[12,13] and molecular sequestration from solution.[14,15]

It is crucial for the design of new materials and pharmaceuticals to have reliable knowledge of the degree of association between host and guest, as captured by standard binding free energies. Large, negative binding free energies ΔG° indicate thermodynamically stable host–guest associations. The binding affinity of the two molecules is typically expressed through the dissociation constant K_d in units of molarity (M).[16] The magnitude of this dissociation constant loosely correlates to the concentration in solution at which the host–guest complexes[17,18] will begin to dissociate. Small magnitudes of K_d correspond to strong binding and a favorable bound state, whereas large magnitudes of K_d imply the dissociation of host and guest and favor the unbound state. Formally, the dissociation constant is related to the standard binding free energy ΔG°, defined via Eq. 1 using the standard concentration, C° =1 M.[19,20]

$$\Delta G^\circ = k_{\mathrm{B}} T \log \left(K_d / C^\circ \right) \tag{1}$$

In particular, for pharmaceutical applications, knowledge of binding free energy is an essential component of molecular design.[19,21–24] Rather than undergo time-consuming processes to synthesize specific organic compounds and test their K_d using standard experimental techniques, researchers often begin by performing these calculations *in silico*, using atomistic molecular simulations with the best available accuracy, including polarization and quantum effects[25,26] when needed. For pharmaceutical applications, focus in molecular design is often on a combination of solubility and standard binding free energies. Recent efforts to determine dissociation constants for a wide variety of systems have utilized molecular dynamics in an effort to calculate the free energy of dissociation (from which a dissociation constant can be easily calculated, as in Eq. 1). These methods allow large search spaces to be explored through high-throughput molecular calculations, and minimize experimental cost, both in time and resources, enabling the focus of syntheses to be placed on strong candidate molecules.[27–31]

Historically, computational methods have utilized alchemical processes and thermodynamic integration.[32–34] In this technique, an unphysical, but computationally permissible, perturbation is made to the Hamiltonian describing the system. So long as the perturbations do not impose singularities on the Hamiltonian and its derivative, thermodynamic integration may capture the difference in free energy between bound molecules and individually solvated molecules, thus codifying the binding free energy at the system concentration.[35–37] A typical pathway involves the guest being completely removed from the complexed system by gradually turning off interactions between the guest molecule and the other molecules in the simulation, which results in an unbound host in solvent and a guest molecule in a hypothetical ideal gas state. Since both molecules must be solvated in the reaction equilibrium leading to Eq. 1, the guest molecule must be re-inserted into solvent to complete the thermodynamic path and obtain ΔG.[24,38] This can be corrected for volume to obtain ΔG°. However, one well-known difficulty with using this pathway for host–guest complexes involves ensuring the correct volume exploration for the guest molecule as it is deleted from the complex.[19] Additional problems include the sampling of so-called multivalent host molecules which could bind to a guest through multiple different mutual configurations, as each individual binding configuration must be independently calculated and properly weighted with statistical mechanics.

In the thermodynamic integration method, finite differences between intermediate alchemical states comprise a free energy landscape that is integrated to either a known state (such as the ideal gas or Einstein crystal) or a comparison state of interest, determining either an absolute free energy or a free energy change for the entire process.[39–42] The pathway is typically alchemical, meaning it does not correspond to physically accessible states and thus cannot obtain any information about reaction pathways that connect both states in configurational space or the dynamics of the system along them. However, advanced sampling methods,[37,43] including unrestrained biasing methods such as Funnel Metadynamics,[44,45] Steered Molecular Dynamics,[46,47] and Umbrella Sampling,[48,49] which apply biases to physical coordinates to reconstruct free energies along a collective variable (CV) pathway between states, *can* be useful for understanding transformations and dynamics, and have also proven effective for capturing host–guest free energies. These methods are attractive precisely because they are not alchemical; all CV-dependent pathways are defined as mappings of atomic coordinates, and though it is not required, it is often advantageous to make these variables continuous for use in molecular dynamics, and in interpreting the role of pathways and excitations in dynamic transformations. Properly designed, the CV captures all of the essential coordinates characterizing the reaction pathway, with a significantly reduced dimensionality relative to a full MD simulation. These often make use of biasing potentials or forces that accelerate sampling and aid in reconstructing the free energy landscape with restraining potentials often used to aid convergence and limit orthogonal exploration. However, this can be somewhat problematic in the context of host-guest interactions, as restraints typically require modified algorithms in order to capture the configurational entropy and multiple binding configurations present in these binding arrangements. This makes unrestrained methods, which naturally capture the influence of these effects on the ensemble of states, attractive for such studies.

It is our intent in this article to explore the behavior of biased simulations in a model context, to understand what features of a simulation can be used to determine reliability and convergence. Identifying characteristics of non-convergence in biased simulations will generally lead to more robust and accurate binding assays. Here, we employ the Adaptive Biasing Force (ABF) as an advanced sampling method in a collection of samples for free energy calculations of a coarse-grained host-guest system.[50,51] This host-guest model is inspired by the structure of the Cucurbit[n]uril (CB[n]) as the host molecule can bind with a compatible guest molecule containing hydrophobic groups.[31,52,53] Our chief interest is in understanding how the relative geometry of host and guest can affect binding, inspired by experimental observations of binding affinity in the CB[7]–Fentanyl complex.[15] This is a system which can potentially exhibit multiple binding conformations, but it is anticipated that the dominant mode of binding is insertion of the Fentanyl molecule lengthwise into the CB[7] (see Figure 1). The system in our study is entirely coarse-grained, but nonetheless models the process of insertion of a rod-like molecule into a cavitand, which we hypothesize is the primary binding configuration for this type of host–guest system. We seek to better characterize the ABF advanced sampling method and how it might be used to characterize binding free energies of molecules with complex geometries to a library of excipients and binding agents. The simplified geometry and coarse-graining allow us to focus on how these features influence the thermodynamics of host–guest binding, determine bottlenecks in molecular sampling, and overall affect the convergence of bias-based free energy methods. Specifically, the effects of altered guest geometry are explored to understand how it affects the overall system's behavior.

2 Model and Methods

The system consists of a single cavitand-shaped host structure and a single rod-shaped guest structure embedded in an implicit solvent. Both structures are composed of coarse-grained

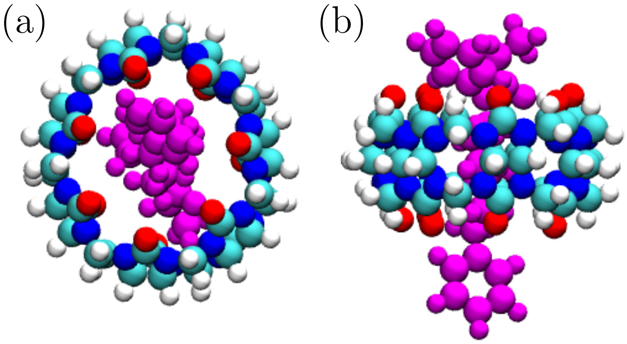

Figure 1: Hypothesized Cucurbit[7]uril (CB7)–Fentanyl binding complex from top (a) and side (b) perspectives. This molecular complex serves as an example of a real host–guest system which the model in this study approximates.[54]

Lennard-Jones (LJ) beads of identical size; this sets a natural unit of length within the system through the diameter (σ) used for all beads in the system. The host structure is a rigid body constructed from five stacked, concentric, identically-sized rings of particles containing two layers each at a radial spacing of 0.5σ. The resulting aggregated structure still closely approximates a cylindrical, hollowed cavitand molecule. The cavitand with a diameter of 6σ consists of 240 beads. Particles comprising the two layers are identical in diameter and mass, but are separated into two types of beads (1 and 2) depending on whether they are attractive (type 1) or repulsive (type 2) to the guest molecule. The inner ring is composed of type 1 particles, while the exterior ring is composed of type 2 particles to mitigate the tendency of the guest to adsorb to the exterior surface.

The guest structure is a linear rigid rod consisting of a variable number ($N = \{4, 6, 8$ or $10\}$) of type 1 particles, which results in rigid rods of effective length $L = N\sigma$. Figure 2 shows schematic examples of cavitand-rod interactions where the rod with a different number of beads has entered the pore and resides for a moment inside the cavitand. At the temperatures studied, the cavity–rod interaction, is sufficiently strong for all lengths to favor host–guest binding within the pocket of the cavitand. The differences in rod length allow for exploration of a key parameter in the host–guest binding of elongated molecules, notably the relative length L/D, where D is the diameter of the cavitand. It should be noted that

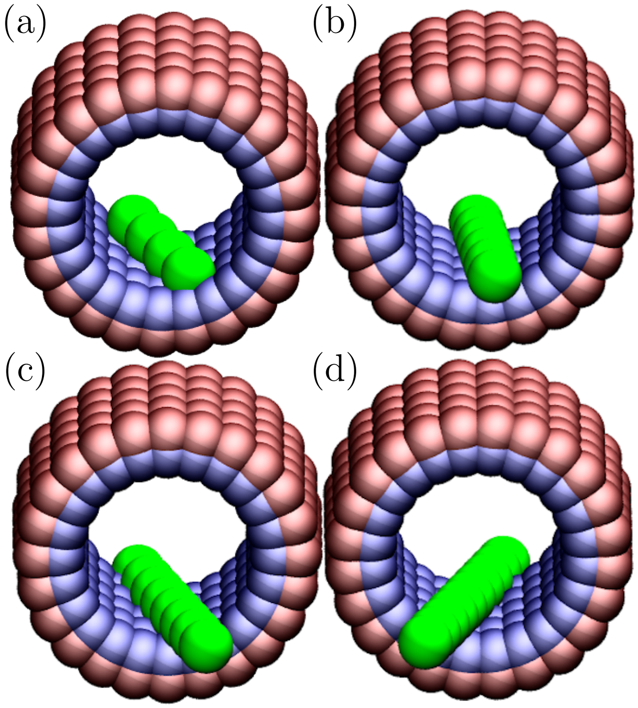

Figure 2: Rendered images of all four unique geometric configurations of the cavitand-rod system. Left to right: (a) 4σ, (b) 6σ, (c) 8σ, and (d) 10σ length rods. Each system is shown in a bound configuration at the interior binding pocket, though the exact orientation of the rod during binding is dependent on its geometry. The cavitand impermeability is shown clearly in all 4 images. Type 1 particles are depicted in pink, whereas type 2 particles are depicted in teal, demonstrating visually the favorable binding of the interior to the type 1 rod and unfavorable binding of the exterior to the same rod.

while the cavitand–rod interactions are more favorable within the interior binding pocket, weaker adhesive interactions are present between the rod and the exterior of the cavitand.

Interparticle interactions utilize the shifted–truncated Lennard-Jones (LJ) potential where the cutoff r_c is set to 2.5σ for both type 1 - type 1 and type 2 - type 2 interactions. On the other hand, the cutoff for type 1 - type 2 interactions is set to $r_c = 2^{1/6\sigma}$, resulting in a Weeks–Chandler–Anderson repulsive potential.[55] This is succinctly described by the following equation

$$U(r) = \begin{cases} 4\epsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6}] + U_0 & \text{if } r \leq r_\mathrm{c} \\ 0 & \text{if } r > r_\mathrm{c} \end{cases} . \tag{2}$$

In Eq. 2, r represents the distance between the host and guest molecules. The quantity U_0 is an added energy that makes the pair potential $U(r)$ continuous when evaluated at $r = r_\mathrm{c}$. With this formulation of the LJ potential, any foreign particle of type 1 will be strongly attracted to the interior of the cavitand composed of type 1 particles, but will only be weakly attracted to the exterior of the cavitand due to the mitigating repulsion of type 2 particles. As the host and guest are treated as rigid bodies, the only meaningful potential energy arises between host and guest. The remaining system parameters are nondimensionalized in the style typical for LJ-based systems,[55] with energy set by ϵ and masses by that of individual beads, m. The remaining system parameters may be obtained as derived units from combinations of σ, ϵ, and m.

We perform simulations using the open source code LAMMPS[56] (v.lammps-29Oct20) coupled to the SSAGES package (v.0.9.3) to perform advanced sampling. Example input files are provided in the supplementary information.[1] The host and guest are placed into a box with dimensions $V = L^3$ where $L = 10\sigma$. Particles are given randomized initial velocities, which imbue the rigid bodies with randomized initial center-of-mass and rotational motion, before equilibration using a Langevin thermostat with a damping coefficient of 1.0 and temperature of 1.5. The seed number varies from one simulation to another. A restraint is placed on the maximum distance between host and guest centers of mass ($\Delta r_{\mathrm{CM,max}}$) using SSAGES; this is set to 10σ for all systems studied. Utilizing the standard Lennard-Jones timescale, $\tau = \sqrt{\frac{m\sigma^2}{\epsilon}}$, where m refers to the mass of a single Lennard-Jones bead, timesteps were set to $\delta t = 10^{-4}\tau$. A set of trial simulations was performed focusing on identifying the proper simulation time length to ensure convergence; it was quickly observed that while some simulations converge quite quickly, convergence for all lengths L was unreliable for simulation times less than $10^4\tau$. Thus, a standard simulation of length $\tau_{\mathrm{sim}} = 10^5\tau$ was utilized for data gathering. Since all particles in this system participate in rigid body interactions, only cross-interactions between host and guest are calculated using the pair potential. Solvent is

---

[1] Electronic supplementary information for this article is posted at XXX.

treated implicitly through the Langevin thermostat and thus does not appear in the energy calculations.

Binding free energies are calculated using the SSAGES package, using a collective variable (CV) of the relative center of mass distance $\Delta\mathbf{r} = \sqrt{(\mathbf{r}_{\mathrm{CM,rod}} - \mathbf{r}_{\mathrm{CM,cav}})^2}$ of the rod and cavitand (cav). CV statistics are compiled over a range of 0σ to 10σ, and the resulting Potential of Mean Force (PMF) distribution, therefore, shares this range. No restraints are applied to the ABF CV range.

3 Results

An exemplary result for the free energy of a specific rod–cavitand system with $L = 10\sigma$ is given in Figure 3(a). It is illustrative to examine the approach to convergence, as represented by the trajectories in energy and CV space plotted in Figs. 3(b) and (c), respectively. The visualized dynamics in Figures 3(b) and (c) demonstrate that the rod is able to enter the interior region of the cavitand and interact with the adhesive beads there. However, in some cases, the rod can become stuck in a bound configuration, or through a combination of spatial configuration and sterics fail to explore the bound configurations at all. The restricted exploration limits the accuracy of the free energy computed on the other side of the transition state, visualized by the rightmost of the peaks in the PMF plotted in Figure 3(a) (occurring around $\Delta r_{\mathrm{CM}} = 6\sigma$). This peak represents the initial entry of the rod into the cavitand, and is an entropic barrier to binding which grows as the orientation becomes necessarily more constricted at the entry with growing N. The type of systematic undersampling we describe here introduces large errors at precisely this type of entropic barrier. Detailed analysis of the resulting free energy curves and trajectories in Figure 4(a) shows this is characteristic of the simulations which fail to converge.

The results in Figure 3(b) suggest that the system readily explores all possible host-guest distances in a properly converged system. This type of sampling does not imply convergence

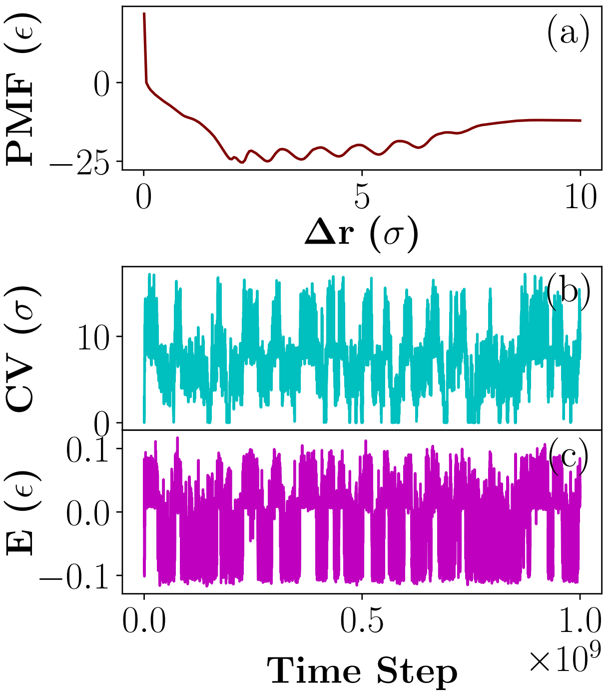

Figure 3: (a) PMF profile of a properly converged system. Example time series of the corresponding (b) CV and (c) total energy of the system and CV value. The wide fluctuations in CV and the total system's energy per atom is characteristic of all properly converged simulations, which is lacking in any improperly converged simulations.

on its own, but is a necessary requirement for convergence to occur, as it enables sufficient sampling of all relevant configurations to occur. The energy curve likewise explores a range of bound (predominately negative) and unbound (predominantly zero or positive) available energies as depicted in Fig. 3(c). This is in contrast to improperly converged systems where the CV in Figure 4(b) and total energy in Figrue 4(c) explore only a well-defined subset of rod–cavitand distances after initial evolution. In all observed cases this is due to the system becoming kinetically trapped within one region of CV space as a result of either entropic limitations or initially poor estimates of the mean force not allowing for relaxation over the timescale of simulations. Excluded states could manifest in one of two ways: either the CV became trapped beneath an artificial upper bound and explored only the bound basin or the CV became trapped above an artificial lower bound where dissociated host and guest keep trying, but failing, to enter the bound state. Geometrical analysis of the cavitand

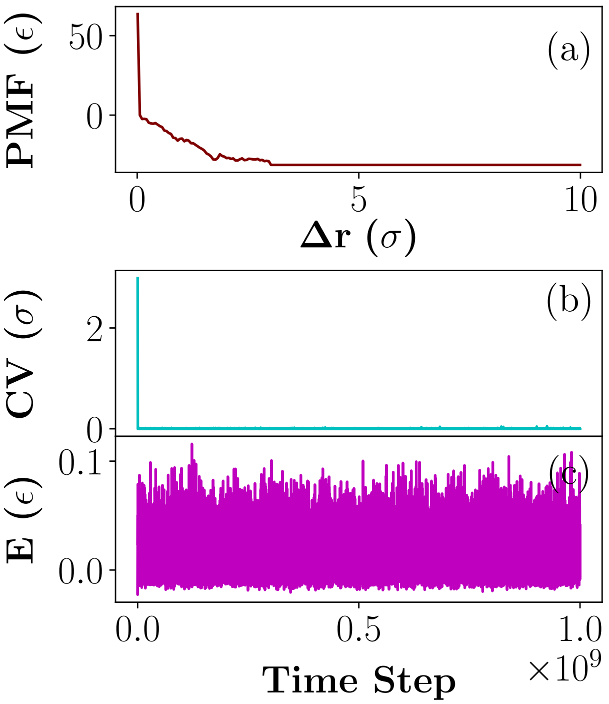

Figure 4: (a) PMF profile of an improperly converged system. Example time series of the corresponding (b) CV and (c) total energy of the system and CV value. The narrow fluctuations in CV and the total system's energy per atom is characteristic of an improperly converged simulation.

and rod suggests the value of this artificial bound corresponds roughly to the transition between the rod entering and exiting the interior binding pocket in all cases of unconverged simulations. We hypothesize that this is due to the entropic nature of the transition state in host–guest materials. If entropy on both sides of the barrier is similar and the rate-limiting step is merely an energetic excitation, ABF is likely to handle convergence well, as has been demonstrated for ABF and derived methods in a variety of one-dimensional systems. However, here, there is a significant entropic penalty to the binding event. As such, if the rod or cavitand approaches with incorrect orientation, the rigid nature of the molecules will not accommodate a binding event. This can lead to overestimation of the free energy of the transition state. Due to the additive nature of recorded forces in the ABF method, this will require extensive sampling to correct the mean force which may not be possible on tractable timescales.

To understand the correlation between the CV and total energy fluctuations in our simulations, we plot in Figure 5 the time-dependence of energy deviations in response to fluctuations in the CV. We use the equation

$$C_{\xi E}(t') = \frac{1}{N} \sum_{t=0}^{(N-1)\delta t} \frac{((\xi(t+t') - \langle \xi \rangle)(E(t) - \langle E \rangle)}{\sqrt{\langle \xi^2 \rangle \langle E^2 \rangle}} \tag{3}$$

where $\xi = \Delta r_{\mathrm{CM}}$ is the CV we use in our calculations. There is a clear correlation between fluctuations of CV and energy at the same points in time, and an oscillating response that results from the driving forces imparted by the ABF method.

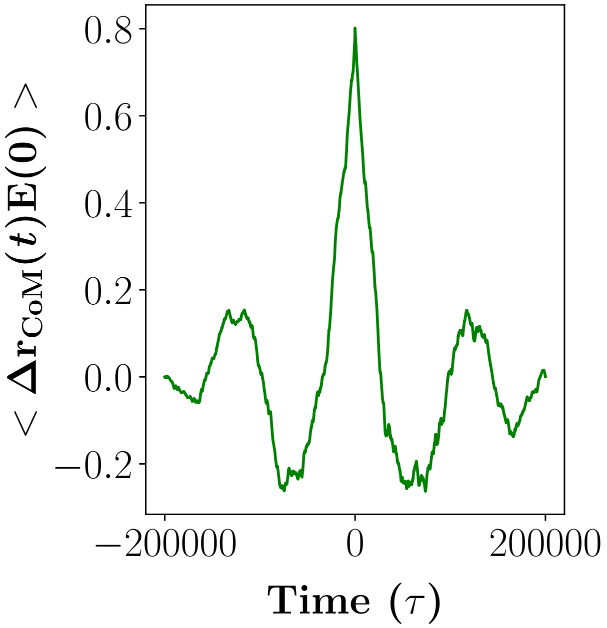

Figure 5: A plot of the time-correlations between the CV (Δr_{CM}) and the system's energy. A clear instantaneous correlation exists, followed by a periodic signature in the time correlations which likely results from application of the bias.

This is reflected in the extent of exploration for CVs and energy plotted in Figure 6, which represents calculations of the average variances of CV and scaled total energy in every individual simulation run. Figure 6 reveals a strong correlation between the average variances of total energy and CV for the host–guest system. The higher magnitude of variance indicates larger fluctuations in these variables which suggests a better coverage of the CV over the possibilities in the system. This further leads to an improvement in the convergence of PMF calculation. Note that as a result of the bias force driving sampling back from extremes of the CV domain, a periodic signature appears in the time series.

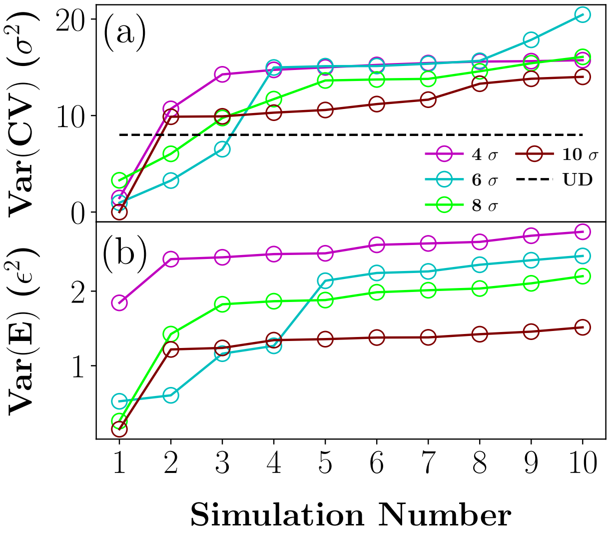

Figure 6: (a) The average variance of CV for 10 different simulations sorted from low to high versus simulation number for each system with different rod lengths of 4, 6, 8, 10 σ. Note that UD represents the variance of a uniform distribution within a range of CV $\in$ [0,10] (b) The variance of the scaled total energies. To make a convenient comparison between the total energy of four different systems we scaled the energy by multiplying values with the total number of atoms in a given system and then dividing it by the number of atoms in the rod. 4, 6, 8, 10 σ in the plots represents the size of the rod for the corresponding plots.

Properly converged simulations demonstrated a sufficiently comprehensive exploration of the entire CV range and demonstrated no apparent barrier to the rod's entry or exit of the interior binding pocket. Consequently, the energy varies within a wider possible range of values as shown from Figures 3 to 6. The trajectories of both energy and position as a function of time have the signature of a random walk; this type of behavior is often seen, for instance, in converged metadynamics simulations.[57] The random walk idea applies equally to simulations using ABF, though it is an imperfect analogy here, since ABF as applied does not regulate orientational entropy (and thus retains some system features within the CV dynamics). Since a uniform distribution results in the maximum variance of a hypothetical system, it is a logical reference to evaluate the convergence of our results.

We plot the expected uniform distribution of the CV in Figure 6(a), and find that simulations whose trajectories have a variance within 10% of this value are nearly always converged. Note that for simulation numbers greater than 4, these graphs demonstrate a plateau for each N studied; this behavior is representative of a simulation which is sampling the entire space and is converged. Also, it is worth noting that if this were a truly bimodal distribution with a bound state near $\Delta r_{\mathrm{CM}} = 0$ and unbound state located at $\Delta r_{\mathrm{CM}} = 10$, the maximum variance achievable is $25\sigma^2$; thus samples with much larger variances are possible which indicate sampling of both the minimum and unbound states with comparatively less time spent in between them. Long enough simulation runs which have converged will be expected to trend toward the uniform distribution (UD) value.

To extract a free energy landscape from our simulated ensembles, improperly converged samples were screened systematically and removed from the aggregated mean plots. Figure 7 presents the PMFs of our CG rod-cavitand systems where they vary in the number of beads in the rod as a guest interacting with the cavitand as the host.

The presence of a free energy well is evident for each system displayed in Figure 7 which corresponds to a binding event between the cavitand and rod. The primary well in the PMF widens as the rod length is increased. Additionally, the number of local minima in each well increases, suggesting that the unique binding events and conformations that contribute to local free energy minima are more numerous but less equivalent as rod length is increased. The absolute minima of these wells are roughly similar and independent of rod length. This corresponds to a CoM difference that is achievable only in an interior pocket bound state, demonstrating the most energetically favorable interaction. The longer rods also have the ability to bind in other metastable states due to the increased number of available binding locations (this is evidenced by the appearance of additional minima away from the primary minimum, spaced by $\approx \sigma$, but the best conformation is consistent across all geometries.

These concurrent phenomena display the two most observed challenges in achieving proper sampling: increasing rod length leading to longer convergence times and difficulty in

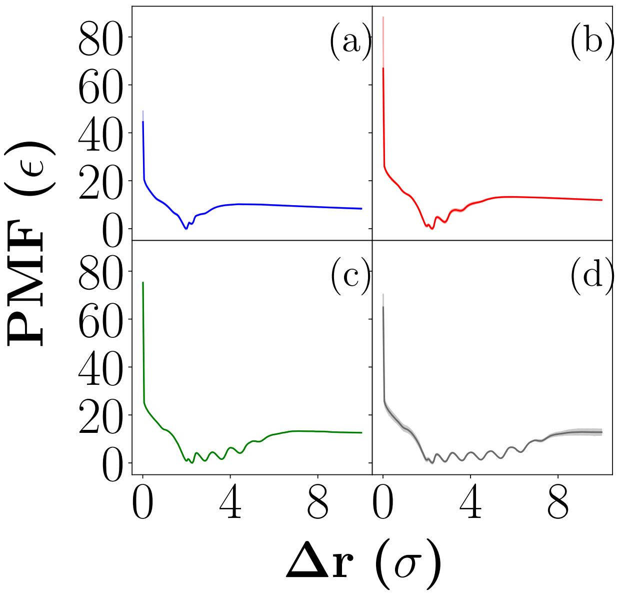

Figure 7: Mean ABF free energy curves for (a) 4σ, (b) 6σ, (c) 8σ, and (d) 10σ configurations. These curves are plotted along the ABF collective variable (difference between cavitand and rod CoM) and are the direct integrals of the ABF-generated PMF distributions. Free energy is in units of ϵ and normalized so that the minimum point on the curve corresponds to an energy value of 0.00, while CoM difference is in units of σ. Shaded regions represent the standard deviation regions of each curve, calculated from the included free energy curves of that configuration. Standard deviation is calculated independently at each point along the plotted CV.

achieving precise replicate measurements of the system in the bound state. Yet, the overall curve shape is not so drastically altered due to the increased error ranges that the result is confounded—both the binding pocket and the nonbinding region are sufficiently resolved when improperly converged runs are not included in the mean curve. Noting that our variances for the sampled CVs and energies (as plotted in Fig. 6) help to define systems with "good coverage," we anticipate this process may be automated in the future to ensure sufficient sampling is obtained within a reasonable number of overall free energy calculations, or the system is flagged for treatment using an alternative method (for instance, a partially restrained method such as funnel metadynamics).

4 Discussion

Adaptive biasing force methods are promising for exploring host-guest systems due to their tempered approach to initial free energy estimates, but they are not guaranteed to converge in finite-length simulations. Through the course of this model study, we have identified important signatures in the variance of CV and energy which indicate their state of convergence. Improperly converged runs exhibit smaller fluctuations in their CV and energy (whose fluctuations in turn are correlated) due to systematic undersampling. In extreme cases, these are several orders of magnitude less than those observed in converged calculations, which should trend toward a uniform distribution given sufficient sampling time. There is a distinct cause-and-effect relationship between this behavior in the energy profile and the inability to sample portions of the CV range corresponding to the interior or exterior of the cavitand. Therefore, the CV sampling exclusion, regardless of whether it excludes the range corresponding to the cavitand interior or exterior, has been identified as a significant system bottleneck. We reiterate that ABF should be capable of resolving these issues, given sufficient length of simulations, though it is impossible to determine precisely what additional length a simulation needs to run based on its current level of sampling. We can, however, use the fluctuations observed in simulation to determine if the extent of sampling achieved in a given simulation is sufficient by comparing the variance of the CV to the expected variance of a uniform distribution, as noted in the previous section.

We hypothesize this correlation is most likely due to the decrease in geometric obstructions to binding and exploration of conformations as rod length is decreased. Smaller rod lengths, especially those smaller than the cavitand diameter, are able to enter the interior of the cavitand, bind, unbind, and exit more freely due to an increase in available physical trajectories to reach the same end bound or unbound state. Larger rod lengths have more difficulty in entering and exiting the cavitand from the exterior space simply due to geometric steric hindrance, leading to more difficulty in executing these movements and sufficiently sampling the entire CV range. While later timesteps in the evolution of the system also incorporate a biasing force calculated from the PMF values obtained until that point in the simulation, the addition of this force is not necessarily sufficient to overcome the steric hindrances that discourage entrance and exit movements at longer rod lengths.

Adaptive sampling methods such as ABF could benefit greatly from designing protocols which identify or avoid such bottlenecks. Here, we only explored the relatively simple case of a single CV, though this issue could potentially be alleviated by biasing on both the separation distance and relative orientation of the host and guest. This streamlines resolution of the rate-limiting process, though it also requires significantly more sampling to be performed away from the bottleneck to resolve the full landscape in multiple dimensions. Multi-walker and multi-window–multi-walker simulations can be incorporated here, and there is the possibility for Monte Carlo protocols which connect the free energy landscape to resolve such bottlenecking issues.

5 Conclusions

We have shown, using a CG host-guest model, that ABF simulations can be utilized to obtain accurate PMFs in binding calculations. By systematically testing the ABF method on different rod lengths within our rod–cavitand CG model we were also able to identify the chief causes of bottlenecks which prevent the sampling of converged free energy landscapes, and discuss some strategies which may prove useful for mitigating similar undersampling issues. We anticipate these results to be useful to the computational community exploring host–guest interactions in materials or drug design contexts. As noted, the selection of this particular host-guest system was primarily motivated by a desire to study the binding between elongated guest molecules (with multiple potential binding configurations) and cavitands such as in the Fentanyl–CB7 system.[54] We anticipate this method can be extended to study other chemical and biological systems to earn knowledge about the kinetics and thermodynamics of host-guest binding and unbinding dynamics.

Acknowledgement

AVM and JKW acknowledge support for undergraduate research through the United States National Science Foundation (Award No. DMR-1751988). This work utilizes the SSAGES advanced sampling suite of codes, whose development in the group of JKW is supported by the MICCoM Center at Argonne National Laboratory under a Computational Materials Science center grant from the Department of Energy, Basic Energy Sciences Division. AVM, MF, and JKW acknowledge helpful discussions with A. C. Leonhard, and assistance from Logan Hennes in proofreading the manuscript.

Supporting Information Available

The SSAGES runfiles and other necessary scripts used in our simulations.

References

(1) Sayed, M.; Pal, H. An overview from simple host-guest systems to progressively complex supramolecular assemblies. *Phys. Chem. Chem. Phys.* **2021**, *23*, 26085–26107.

(2) Park, J.; Park, J.; Lee, J.; Lim, C.; Lee, D. W. Size compatibility and concentration dependent supramolecular host–guest interactions at interfaces. *Nat. Commun.* **2022**, *13*, 112.

(3) Yang, Y. W.; Sun, Y. L.; Song, N. Switchable host-guest systems on surfaces. *Acc. Chem. Res.* **2014**, *47*, 1950–1960.

(4) Khan, S. B.; Lee, S.-L. Supramolecular Chemistry: Host–Guest Molecular Complexes. *Molecules* **2021**, *26*, 3995.

(5) Ma, X.; Zhao, Y. Biomedical Applications of Supramolecular Systems Based on Host–Guest Interactions. *Chem. Rev.* **2015**, *115*, 7794–7839.

(6) Hunter, C. A. Meldola Lecture. The role of aromatic interactions in molecular recognition. *Chem. Soc. Rev.* **1994**, *23*, 101–109.

(7) Stoffelen, C.; Huskens, J. Soft Supramolecular Nanoparticles by Noncovalent and Host–Guest Interactions. *Small* **2016**, *12*, 96–119.

(8) Yang, H.; Yuan, B.; Zhang, X.; Scherman, O. A. Supramolecular chemistry at interfaces: Host-guest interactions for fabricating multifunctional biointerfaces. *Acc. Chem. Res.* **2014**, *47*, 2106–2115.

(9) Nassimbeni, L. R. Physicochemical aspects of host - Guest compounds. *Acc. Chem. Res.* **2003**, *36*, 631–637.

(10) Helm, M. P. V. D.; Li, G.; Hartono, M.; Eelkema, R. Transient Host-Guest Complexation To Control Catalytic Activity. *J. Am. Chem. Soc.* **2022**, *144*, 9465–9471.

(11) Li, Y.; Yu, J. Emerging applications of zeolites in catalysis, separation and host–guest assembly. *Nat. Rev. Mater.* **2021**, *6*, 1156–1174.

(12) Wang, H.; Fan, Y.; Hou, Y.; Chen, B.; Lei, J.; Yu, S.; Chen, X.; Hou, X. Host-guest liquid gating mechanism with specific recognition interface behavior for universal quantitative chemical detection. *Nat. Commun.* **2022**, *13*, 1906.

(13) Chen, J.; Hickey, B. L.; Wang, L.; Lee, J.; Gill, A. D.; Favero, A.; Pinalli, R.; Dalcanale, E.; Hooley, R. J.; Zhong, W. Selective discrimination and classification of G-quadruplex structures with a host–guest sensing array. *Nat. Chem.* **2021**, *13*, 488–495.

(14) Jonkergouw, C.; Beyeh, N. K.; Osmekhina, E.; Leskinen, K.; Taimoory, S. M.; Fedorov, D.; Anaya-Plaza, E.; Kostiainen, M. A.; Trant, J. F.; Ras, R. H. A.; Saavalainen, P.; Linder, M. B. Repurposing host-guest chemistry to sequester virulence and eradicate biofilms in multidrug resistant Pseudomonas aeruginosa and Acinetobacter baumannii. *Nat. Commun.* **2023**, *14*, 2141.

(15) Deng, C.-L.; Murkli, S. L.; Isaacs, L. D. Supramolecular hosts as *in vivo* sequestration agents for pharmaceuticals and toxins. *Chem. Soc. Rev.* **2020**, *49*, 7516–7532.

(16) Gumbart, J. C.; Roux, B.; Chipot, C. Efficient determination of protein-protein standard binding free energies from first principles. *J. Chem. Theory Comput.* **2013**, *9*, 3789–3798.

(17) Dill, K. A.; Bromberg, S. *Molecular driving forces: statistical thermodynamics in biology, chemistry, physics, and nanoscience*, 2nd ed.; Garland Science: London ; New York, 2011.

(18) Márquez, C.; Hudgins, R. R.; Nau, W. M. Mechanism of HostGuest Complexation by Cucurbituril. *J. Am. Chem. Soc.* **2004**, *126*, 5806–5816.

(19) Deng, Y.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113*, 2234–2246.

(20) Zhou, H.-X.; Gilson, M. K. Theory of Free Energy and Entropy in Noncovalent Binding. *Chem. Rev.* **2009**, *109*, 4092–4107.

(21) Boby, M. L.; Fearon, D.; Ferla, M.; Filep, M.; Koekemoer, L.; Robinson, M. C.; The COVID Moonshot Consortium‡; Chodera, J. D.; Lee, A. A.; London, N.; Von Delft, A.; Von Delft, F.; Achdout, H.; Aimon, A.; Alonzi, D. S.; Arbon, R.; Aschenbrenner, J. C.; Balcomb, B. H.; Bar-David, E.; Barr, H.; Ben-Shmuel, A.; Bennett, J.; Bilenko, V. A.; Borden, B.; Boulet, P.; Bowman, G. R.; Brewitz, L.; Brun, J.; Bvnbs, S.; Calmiano, M.; Carbery, A.; Carney, D. W.; Cattermole, E.; Chang, E.; Chernyshenko, E.; Clyde, A.; Coffland, J. E.; Cohen, G.; Cole, J. C.; Contini, A.; Cox, L.; Croll, T. I.; Cvitkovic, M.; De Jonghe, S.; Dias, A.; Donckers, K.; Dotson, D. L.; Douangamath, A.; Duberstein, S.; Dudgeon, T.; Dunnett, L. E.; Eastman, P.; Erez, N.; Eyermann, C. J.; Fairhead, M.; Fate, G.; Fedorov, O.; Fernandes, R. S.; Ferrins, L.; Foster, R.; Foster, H.; Fraisse, L.; Gabizon, R.; García-Sastre, A.; Gawriljuk, V. O.;

Gehrtz, P.; Gileadi, C.; Giroud, C.; Glass, W. G.; Glen, R. C.; Glinert, I.; Godoy, A. S.; Gorichko, M.; Gorrie-Stone, T.; Griffen, E. J.; Haneef, A.; Hassell Hart, S.; Heer, J.; Henry, M.; Hill, M.; Horrell, S.; Huang, Q. Y. J.; Huliak, V. D.; Hurley, M. F. D.; Israely, T.; Jajack, A.; Jansen, J.; Jnoff, E.; Jochmans, D.; John, T.; Kaminow, B.; Kang, L.; Kantsadi, A. L.; Kenny, P. W.; Kiappes, J. L.; Kinakh, S. O.; Kovar, B.; Krojer, T.; La, V. N. T.; Laghnimi-Hahn, S.; Lefker, B. A.; Levy, H.; Lithgo, R. M.; Logvinenko, I. G.; Lukacik, P.; Macdonald, H. B.; MacLean, E. M.; Makower, L. L.; Malla, T. R.; Marples, P. G.; Matviiuk, T.; McCorkindale, W.; McGovern, B. L.; Melamed, S.; Melnykov, K. P.; Michurin, O.; Miesen, P.; Mikolajek, H.; Milne, B. F.; Minh, D.; Morris, A.; Morris, G. M.; Morwitzer, M. J.; Moustakas, D.; Mowbray, C. E.; Nakamura, A. M.; Neto, J. B.; Neyts, J.; Nguyen, L.; Noske, G. D.; Oleinikovas, V.; Oliva, G.; Overheul, G. J.; Owen, C. D.; Pai, R.; Pan, J.; Paran, N.; Payne, A. M.; Perry, B.; Pingle, M.; Pinjari, J.; Politi, B.; Powell, A.; Pšenák, V.; Pulido, I.; Puni, R.; Rangel, V. L.; Reddi, R. N.; Rees, P.; Reid, S. P.; Reid, L.; Resnick, E.; Ripka, E. G.; Robinson, R. P.; Rodriguez-Guerra, J.; Rosales, R.; Rufa, D. A.; Saar, K.; Saikatendu, K. S.; Salah, E.; Schaller, D.; Scheen, J.; Schiffer, C. A.; Schofield, C. J.; Shafeev, M.; Shaikh, A.; Shaqra, A. M.; Shi, J.; Shurrush, K.; Singh, S.; Sittner, A.; Sjö, P.; Skyner, R.; Smalley, A.; Smeets, B.; Smilova, M. D.; Solmesky, L. J.; Spencer, J.; Strain-Damerell, C.; Swamy, V.; Tamir, H.; Taylor, J. C.; Tennant, R. E.; Thompson, W.; Thompson, A.; Tomásio, S.; Tomlinson, C. W. E.; Tsurupa, I. S.; Tumber, A.; Vakonakis, I.; Van Rij, R. P.; Vangeel, L.; Varghese, F. S.; Vaschetto, M.; Vitner, E. B.; Voelz, V.; Volkamer, A.; Walsh, M. A.; Ward, W.; Weatherall, C.; Weiss, S.; White, K. M.; Wild, C. F.; Witt, K. D.; Wittmann, M.; Wright, N.; Yahalom-Ronen, Y.; Yilmaz, N. K.; Zaidmann, D.; Zhang, I.; Zidane, H.; Zitzmann, N.; Zvornicanin, S. N. Open science discovery of potent noncovalent SARS-CoV-2 main protease inhibitors. *Science* **2023**, *382*, eabo7201.

(22) Jespers, W.; Åqvist, J.; Gutiérrez-de Terán, H. In *Protein-Ligand Interactions and Drug*

Design; Ballante, F., Ed.; Springer US: New York, NY, 2021; Vol. 2266; pp 203–226.

(23) Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017**, *57*, 2911–2937.

(24) Mobley, D. L.; Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annu. Rev. Biophys.* **2017**, *46*, 531–558.

(25) Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. Polarization effects in molecular mechanical force fields. *J. Condens. Matter Phys.* **2009**, *21*, 333102.

(26) Jorgensen, W. L. Special Issue on Polarization. *J. Chem. Theory Comput.* **2007**, *3*, 1877–1877.

(27) Wolf, S.; Lickert, B.; Bray, S.; Stock, G. Multisecond ligand dissociation dynamics from atomistic simulations. *Nat. Commun.* **2020**, *11*, 2918.

(28) Sandberg, R. B.; Banchelli, M.; Guardiani, C.; Menichetti, S.; Caminati, G.; Procacci, P. Efficient Nonequilibrium Method for Binding Free Energy Calculations in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2015**, *11*, 423–435.

(29) Hall, R.; Dixon, T.; Dickson, A. On Calculating Free Energy Differences Using Ensembles of Transition Paths. *Front. mol. biosci* **2020**, *7*, 106.

(30) Iida, S.; Tomoshi, K. Free energy and kinetic rate calculation via non-equilibrium molecular simulation: application to biomolecules. *Biophys. Rev.* **2022**, *14*, 1303–1314.

(31) Leonhard, A. C.; Whitmer, J. K. Accurate Determination of Cavitand Binding Free Energies via Unrestrained Advanced Sampling. *J. Chem. Theory Comput.* **2019**, *15*, 5761–5768.

(32) Straatsma, T. P.; McCammon, J. A. Computational Alchemy. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407–435.

(33) Shirts, M. R.; Mobley, D. L.; Chodera, J. D. *Annual Reports in Computational Chemistry*; Elsevier, 2007; Vol. 3; pp 41–59.

(34) Limongelli, V. Ligand binding free energy and kinetics calculation in 2020. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020**, *10*, e1455.

(35) Beutler, T. C.; Mark, A. E.; Van Schaik, R. C.; Gerber, P. R.; Van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529–539.

(36) Mobley, D. L.; Liu, S.; Cerutti, D. S.; Swope, W. C.; Rice, J. E. Alchemical prediction of hydration free energies for SAMPL. *J. Comput. Aided Mol. Des.* **2012**, *26*, 551–562.

(37) Amezcua, M.; Setiadi, J.; Ge, Y.; Mobley, D. L. An overview of the SAMPL8 host–guest binding challenge. *J. Comput. Aided Mol. Des.* **2022**, *36*, 707–734.

(38) Duarte Ramos Matos, G.; Kyu, D. Y.; Loeffler, H. H.; Chodera, J. D.; Shirts, M. R.; Mobley, D. L. Approaches for Calculating Solvation Free Energies and Enthalpies Demonstrated with an Update of the FreeSolv Database. *J. Chem. Eng. Data.* **2017**, *62*, 1559–1569.

(39) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245–268.

(40) Jorgensen, W. L. Free energy calculations: a breakthrough for modeling organic chemistry in solution. *Acc. Chem. Res.* **1989**, *22*, 184–189.

(41) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *Chem. Phys.* **2008**, *129*, 124105.

(42) Pohorille, A.; Jarzynski, C.; Chipot, C. Good practices in free-energy calculations. *J. Phys. Chem. B* **2010**, *114*, 10235–10253.

(43) Rizzi, A.; Jensen, T.; Slochower, D. R.; Aldeghi, M.; Gapsys, V.; Ntekoumes, D.; Bosisio, S.; Papadourakis, M.; Henriksen, N. M.; De Groot, B. L.; Cournia, Z.; Dickson, A.; Michel, J.; Gilson, M. K.; Shirts, M. R.; Mobley, D. L.; Chodera, J. D. The SAMPL6 SAMPLing challenge: assessing the reliability and efficiency of binding free energy calculations. *J. Comput. Aided Mol. Des.* **2020**, *34*, 601–633.

(44) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel metadynamics as accurate binding free-energy method. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 6358–6363.

(45) Raniolo, S.; Limongelli, V. Ligand binding free-energy calculations with funnel metadynamics. *Nat. Protoc.* **2020**, *15*, 2837–2866.

(46) Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *Chem. Phys.* **2003**, *119*, 3559–3566.

(47) Park, S.; Schulten, K. Calculating potentials of mean force from steered molecular dynamics simulations. *Chem. Phys.* **2004**, *120*, 5946–5961.

(48) Kästner, J. Umbrella sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 932–942.

(49) You, W.; Tang, Z.; Chang, C. E. A. Potential Mean Force from Umbrella Sampling Simulations: What Can We Learn and What Is Missed? *J. Chem. Theory Comput.* **2019**, *15*, 2433–2443.

(50) Darve, E.; Rodríguez-Gómez, D.; Pohorille, A. Adaptive biasing force method for scalar and vector free energy calculations. *Chem. Phys.* **2008**, *128*, 144120.

(51) Comer, J.; Gumbart, J. C.; Hénin, J.; Lelievre, T.; Pohorille, A.; Chipot, C. The adaptive biasing force method: Everything you always wanted to know but were afraid to ask. *J. Phys. Chem. B* **2015**, *119*, 1129–1151.

(52) Isaacs, L. Cucurbit[n]urils: from mechanism to structure and function. *Chem. Commun.* **2009**, 619–629.

(53) Das, D.; Assaf, K. I.; Nau, W. M. Applications of Cucurbiturils in Medicinal Chemistry and Chemical Biology. *Front. Chem.* **2019**, *7*, 619.

(54) Cheng, M.; Isaacs, L. Acyclic cucurbituril featuring pendant cyclodextrins. *Supramolecular Chemistry* **2021**, *33*, 53–62.

(55) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*, second edition ed.; Oxford University Press: Oxford, United Kingdom, 2017.

(56) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.

(57) Zheng, L.; Yang, W. Essential energy space random walks to accelerate molecular dynamics simulations: Convergence improvements via an adaptive-length self-healing strategy. *Chem. Phys.* **2008**, *129*, 014105.

TOC Graphic