Published in final edited form as:

J Phys Chem B. 2018 May 31; 122(21): 5579–5598. doi:10.1021/acs.jpcb.7b11820.

Binding Modes of Ligands Using Enhanced Sampling (BLUES): Rapid Decorrelation of Ligand Binding Modes via Nonequilibrium Candidate Monte Carlo

Samuel C. Gill[†], Nathan M. Lim[‡], Patrick B. Grinaway^{¶,§}, Ariën S. Rustenburg^{§,¶}, Josh Fass^{¶,§}, Gregory A. Ross[§], John D. Chodera[⊥], and David L. Mobley^{*,‡,†}

[†]Department of Chemistry, University of California, Irvine

[‡]Department of Pharmaceutical Sciences, University of California, Irvine

[¶]Graduate Program in Physiology, Biophysics, and Systems Biology, Weill Cornell Medical College, New York, NY 10065

[§]Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065

[¶]Tri-Institutional PhD Program in Computational Biology and Medicine, New York, NY 10065

[⊥]Memorial Sloan Kettering Cancer Center, New York NY 10065

Abstract

Accurately predicting protein-ligand binding affinities and binding modes is a major goal in computational chemistry, but even the prediction of ligand binding modes in proteins poses major challenges. Here, we focus on solving the binding mode prediction problem for rigid fragments. That is, we focus on computing the dominant placement, conformation, and orientations of a relatively rigid, fragment-like ligand in a receptor, and the populations of the multiple binding modes which may be relevant. This problem is important in its own right, but is even more timely given the recent success of alchemical free energy calculations. Alchemical calculations are increasingly used to predict binding free energies of ligands to receptors. However, the accuracy of these calculations is dependent on proper sampling of the relevant ligand binding modes. Unfortunately, ligand binding modes may often be uncertain, hard to predict, and/or slow to interconvert on simulation timescales, so proper sampling with current techniques can require prohibitively long simulations. We need new methods which dramatically improve sampling of

dmobley@mobleylab.org, Phone: 949-824-6383.

Supporting Information Available

The Supporting Information is available free of charge on the ACS Publications website at (details) and includes a PDF file containing SI Figure 1 (showing work distributions for rotating toluene in lysozyme as a function of the amount of NCMC relaxation), Figure 2 (showing the work standard deviation for toluene in lysozyme as a function of the amount of switching), Figure 3 (showing the dihedral progress coordinate used for 3-iodotoluene), Figure 4 (showing the estimated MSM transition matrix for toluene in lysozyme), Figure 5 (showing acceptance of NCMC vs standard MC move proposals as a function of dihedral angle/binding mode, given a fixed ensemble of MD snapshots), and Figure 6 (showing the acceptance probability for toluene in water as a function of the amount of switching); a `.tar.gz` file containing a set of scripts for running MD and MD/MC and MD/NCMC, simulation run input files for the GROMACS simulations described here, scripts for the OpenMM simulations described, input topology and coordinate files for all simulations, a README.md file detailing organization, and scripts for MSM construction and PCCA analysis as noted in Methods, as well as a copy of the BLUES code used to generate the data presented here.

ligand binding modes. Here, we develop and apply a nonequilibrium candidate Monte Carlo (NCCMC) method to improve sampling of ligand binding modes. In this technique, the ligand is rotated and subsequently allowed to relax in its new position through alchemical perturbation before accepting or rejecting the rotation and relaxation as a nonequilibrium Monte Carlo move. When applied to a T4 lysozyme model binding system, this NCCMC method shows over two orders of magnitude improvement in binding mode sampling efficiency compared to a brute force molecular dynamics simulation. This is a first step towards applying this methodology to pharmaceutically-relevant binding of fragments and, eventually, drug-like molecules. We are making this approach available via our new Binding Modes of Ligands using Enhanced Sampling (BLUES) package which is freely available on GitHub.

TOC image

1 Introduction

1.1 Ligand binding modes are important, but difficult to predict

One of the motivations of computer aided drug design is to understand and predict what factors increase ligand binding affinity to allow for better design of new ligands for further drug development. Successfully predicting binding affinity depends on many factors, including the determination of the most favorable or relevant binding mode, or modes, of the ligand. Medicinal chemists often use knowledge of the likely binding mode or binding modes to attempt rational improvements upon the scaffold, as well as giving insight into the important interactions driving binding. The binding mode or binding modes also provide a fundamental input for many calculations that can predict binding affinities, such as free energy calculations.¹

As important as binding modes are, actually determining them can be difficult. The standard experimental techniques for binding mode determination, X-ray crystallography and nuclear magnetic resonance, can be time-consuming, difficult, or costly, and are not suitable for all targets (membrane proteins can be particularly challenging, for instance). Additionally, experimental methods do not always clearly resolve the binding mode. For example, in the case of fragment-based drug discovery, small, relatively rigid ligands can often have some ambiguity in their binding modes because of internal pseudosymmetry, or other issues.^{2,3} Additionally, methods to make X-ray diffraction data easier to collect—such as cryocooling crystals—potentially stabilize binding modes that are not observed under the conditions of interest.⁴ Multiple binding modes may also contribute substantially to a ligand's affinity,^{5–9}

therefore knowledge of a single experimental binding mode may be misleading or provide an incomplete picture.

Computationally determining binding modes is similarly difficult. One of the most widely used computational methods for binding mode determination is docking, which scores a variety of ligand poses in the binding site.^{10,11} Docking has been shown to perform well in generating candidate binding poses from the top scoring poses; however, the top scoring poses from docking tend not to be the ones found experimentally.^{12,13} This is partly because docking was designed to have a particularly low computational cost (usually seconds or less per molecule) in order to allow application to large databases.^{14,15} Thus, docking necessarily makes many approximations in order to achieve this speed.

In a recent D3R challenge, which consisted of predicting binding modes of HSP90 ligands, different docking studies had varying levels of accuracy—even within submissions using the same docking software—but human screening of the structures seemed to help identify the correct binding mode.¹⁶ Four of the 11 top scoring methods used visual inspection of the computationally predicted poses, while the less successful methods did not, indicating how it remains extremely challenging to predict binding modes.¹⁶ Another study by Warren *et al.* looked at how well different docking programs performed across a variety of different protein targets.¹² They found that docking methods could explore the conformational space of the ligand sufficiently, but the top scoring pose often did not correspond to the observed crystallographic pose.¹²

In fact, humans tend to outperform automated methods at predicting binding modes in blind challenges,¹⁷ further showing that there are still many aspects of binding mode prediction that could benefit from improvement. In the SAMPL4 HIV integrase challenge, for example, determining the binding mode or even binding site of a set of ligands to HIV integrase was shown to be very difficult for many methods.¹³ A human expert with more than 10 years working on the target provided the best submission, in large part guided by his expertise.¹⁷ The best purely computational method in this challenge used docking followed by free energy calculations to predict whether compounds would bind to HIV integrase. In this study, the majority of false negative binding results used starting poses vastly dissimilar from crystallographic observations,¹⁸ indicating that many of the failures resulted from incorrect binding modes.

An alternative to docking which is more rigorous, but computationally expensive, is to apply free energy calculations based on molecular simulations to predict populations of possible stable binding modes. For example, the “confine and release” approach allows multiple binding modes to be treated separately by distinct free energy calculations, and then subsequently combining the individual binding free energies to yield a total binding free energy¹⁹ (Figure 1(a)). Specifically, the overall binding free energy of a ligand to a protein can be decomposed into a particular type of average over the effective binding free energies of different metastable binding modes.²⁰ As long as these metastable binding modes are defined a way that they cover the full bound state including all the relevant binding modes, and they do not overlap in phase space, this approach is rigorous. However, the number of required binding free energy calculations scales linearly with the number of binding modes

for this already computationally demanding approach, making it unappealing to consider multiple candidate binding modes separately in this manner.¹ This approach has been exemplified recently in binding free energy studies on bromodomains; multiple binding modes were considered separately. The crystallographic binding mode was always found to have a higher affinity than other metastable binding modes.²¹ A similar approach has also been used with empirical free energy methods²²

1.2 Other approaches exist to determine binding modes

Another option is to sample over the binding modes within a given binding free energy calculation, as reviewed elsewhere.¹ Many binding free energy calculations use *alchemical* techniques^{23–25} where binding free energies are computed by turning off interactions between the ligand and receptor (controlled by an alchemical parameter λ), taking the ligand through a nonphysical pathway that allows it to be moved from the binding site to solution, yielding the binding free energy. The Binding Energy Distribution Analysis Method²⁶ (BEDAM) is one such alchemical method which includes multiple binding modes by allowing the non-interacting or weakly-binding ligand to rearrange and reorient in the binding site before turning back on interactions, thus allowing relatively easy interchange between binding modes in a single set of simulations.

A similar approach is taken by Wang *et al.* in the application of Hamiltonian replica exchange molecular dynamics to ligand binding.⁸ In their work, multiple replicas of a protein-ligand system were simulated in implicit solvent with varying λ couplings of the steric and electrostatics between replicas. To enhance conformational sampling, translational and rotational Monte Carlo moves were applied before exchange attempts. This potentially allows efficient sampling across binding modes in a single binding free energy calculation, though the use of implicit solvent was critical to the success of their instantaneous Monte Carlo moves. The POPFEP work of Jayachandran *et al.* proposed an alternative approach, correcting for poorly mixing sampling that resulted in highly erroneous binding pose populations by decomposing the sampled configurations into distinct poses with a Markov state model and independently computing alchemical binding free energies with respect to a common noninteracting state.²⁷

1.3 Efficiently sampling binding modes in a simulation would greatly increase free energy calculation performance

Our goal in this paper is a computational method which can reproduce equilibrium binding mode populations with much less computational time than treating binding modes separately. Specifically, each alchemical binding free energy calculation requires simulation at N different alchemical intermediate states (where N is typically at least 12–20^{8,20,26}), where each alchemical state is associated with a vector of alchemical parameters λ (which we will refer to as λ values). If we consider M different binding modes, the total cost of a binding free energy calculation that covers all binding modes separately is $MN \times$ where x is the cost of a single simulation (Figure 1(A) and (C)). This becomes impractical as the number of potential binding modes grows. Instead, the approach we envision is one where we calculate an absolute binding free energy for a single, reasonably populated binding mode and, in a separate calculation that can be run concurrently, efficiently determine the

relative free energies (or equilibrium populations) of all M potential binding modes (Figure 1(B)). Then, we can combine the populations of the individual binding modes and the free energy estimate for a single binding mode into a binding free energy that includes all of the possible binding modes.¹ Thus this approach would have a computational cost of $Nx+y$, where y is the simulation time to determine the binding mode populations.

Such an approach could work as outlined in Ref.,¹ providing a way to compute interconversion free energies between different metastable binding modes. This would have implications for both absolute and relative binding free energy calculations. For absolute calculations, depending on how many binding modes are being considered, such an approach could drastically reduce the total amount of simulation time, as long as $y \ll Nx$ (see Figure 1(C)) (and the wallclock time as long as $y \ll x$ since the N independent alchemical simulations could be run in parallel). This is not currently feasible because we have no suitable, general-purpose method for efficiently sampling binding modes, and thus the cost of these calculations (y) is far too expensive in terms of both human effort and computational time. Our focus here is on developing a method for obtaining binding mode populations which has a cost y which is relatively favorable compared to Nx , or ideally even x so that a parallel calculation could complete in at most x wall-clock time.

A method that allows efficient sampling of binding modes would have broad uses in free energy applications, but would also aid in predictions of binding modes for structure-based design, fragment-based discovery, and other applications.^{28–30} Here, our primary focus is on different “binding modes”: defined as different metastable conformations of a fragment-like ligand within a single relatively rigid protein cavity. Metastable binding modes are thus those which are slow to interconvert on a simulation timescale z . Generally, if binding modes interconvert at a timescale slower than $z/10$ then proper sampling is a major concern; different metastable conformations may have different binding free energies but will be sampled in incorrect proportions, resulting in highly biased results. Moreover, the concept of a binding mode can include multiple ligand conformations in the same site, binding to in multiple sites, or even multiple protein conformations,¹ though we do not specifically address enhanced sampling of protein motion here.

2 Theory and computational methods

2.1 Various sampling methods can be applied

There are a number of common sampling methods which can be used so that simulations sample the equilibrium distribution of populations. The efficiency of these methods can vary dramatically depending on which particular system or class of problem they are applied to, and a method that works best for one class of problem is not necessarily most suitable for another class. Thus it is often nontrivial to determine which sampling method is best suited to a particular problem, or whether there is even a suitable method. Here, our particular interest is in accelerating sampling across ligand binding modes while still sampling the correct distribution of populations. Our goal is to develop a general method that can efficiently determine binding mode populations, in part by reducing the time it takes for simulations to switch between binding modes relative to other methods. This section will

discuss some common sampling methods and the difficulties they encounter when applied to the ligand binding mode sampling problem.

2.1.1 Molecular dynamics (MD) is limited by the metastability of ligand binding modes

—MD is typically used to simulate the dynamics of biomolecular systems by application of a force field which gives the forces between the atoms in the system as a function of their positions. With enough simulation time, MD should sample different metastable states with populations that are correct for a given choice of force field and ensemble, assuming that other simulation details—such as the integrator used to propagate dynamics—do not introduce errors. However, in practice sampling transitions between binding modes using MD is typically inefficient because of large energy barriers (and hence slow timescales) separating binding modes.^{1,20,31,32}

Some free energy calculations attempt to get around this problem by assuming that similar ligands will have similar binding modes, so if a bound structure of a related ligand is available, it is assumed that new related ligands will share the same binding mode. However, this is not necessarily the case – even closely related ligands can have disparate binding modes that are slow to interconvert^{1,7,32} Perhaps this is one reason why the accuracy of relative free energy calculations based on MD still falls short of what is desired for pharmaceutical applications,³³ therefore, adequate binding mode sampling via direct MD simulation requires considerable computational expense and can necessitate specialized simulation hardware.³⁴ This inefficiency is compounded further in free energy calculations, as detailed above, where it is often necessary to adequately sample all relevant binding modes at each λ value to obtain correct binding free energies. In some cases, it is possible to sample long enough at the physical end states to cover all binding modes and then apply restraints to restrict the space treated at intermediate λ values, then compute the free energy of imposing and removing the restraints at the end states. This can improve efficiency modestly,²⁰ but still requires simulations on timescales substantially longer than the timescales of the relevant motions.

2.1.2 Markov State Models (MSMs) can predict long timescale behavior efficiently, but are not ideally suited to our problem

—The MSM approach assumes that a trajectory is generated from a Markov process. This assumption allows a statistical interpretation of MD trajectories. Specifically, a Markov State Model (MSM) is a matrix containing the transition probabilities between defined microstates, which can be used to predict the long timescale behavior of a system. The resulting model approximates the temporally coarse-grained dynamics with a Markovian surrogate model, which has certain properties that can be used to predict the kinetics and equilibrium populations of each state.³⁵ Because a MSM is concerned only with the transitions between states, multiple simulations can be used to generate the model, leading to more efficient use of computational resources. Specifically, rather than running a single very long simulation to adequately sample all binding modes, many shorter simulations can be used with substantially less wall-clock time, at least if parallel resources are available.³⁶

The MSM framework also works to predict the long-timescale behavior of a system even before global equilibrium is reached, as long as local equilibrium is achieved, allowing a

smaller total amount of simulation time to be used to estimate the equilibrium populations of all states rather than having to fully converge to the global equilibrium.³⁷ Sampling at different thermodynamic states can also be employed with MSMs to improve transition estimates and sampling.^{27,38}

However, the mathematics and assumptions behind MSMs unfortunately make this method difficult to use without expert knowledge and considerable care, and a maximum increase in efficiency is obtained only with prior knowledge of all potential binding modes. A sufficient number of transitions between states is necessary to properly estimate equilibrium populations from the MSM robustly. It is difficult to know *a priori* how much simulation data will be required to reach this stage, and it can require careful checking to know when this has been achieved.³⁹ There are also many parameter choices (order parameters, lagtimes, clustering methods) which make constructing MSMs difficult to generalize, although recent developments such as the generalized matrix Raleigh quotient⁴⁰ and time-lagged independent component analysis (tICA)^{41,42} help to reduce dependence on parameter choices.

2.1.3 Effective Monte Carlo proposals can accelerate sampling, but are difficult to construct for condensed-phase systems—In some cases, sampling can be dramatically accelerated by introducing Monte Carlo proposals informed by prior information, but this becomes increasingly difficult in condensed phase systems. As a running example, consider a bistable dimer.⁴³ If we know approximately the relative locations of free energy minima (i.e. how far apart are the minima of the bond-length term), we might construct proposals that instantaneously hop from the vicinity of one minimum to the vicinity of the other. In a vacuum, this will dramatically accelerate mixing between the two metastable states of the dimer.⁴⁴ However, in a densely solvated environment it can be difficult to construct nontrivial proposals that avoid having near-universal rejection since instantaneously perturbing the coordinate of interest is likely to introduce clashes with solvent. While high acceptance rates could be achieved with extremely small perturbations, the long correlation times resulting from these small perturbations leads to highly inefficient sampling. Thus, while MC techniques have seen substantial use for biomolecular systems,^{45,46} much of the field has moved towards using MD as a more general sampling engine and MC has to some extent fallen out of favor, partly because naïve MC moves in densely packed systems tend to overwhelmingly be rejected.

However, there have been some successes at combining MC and MD. For example the common replica⁴⁷ and Hamiltonian replica⁴⁸ exchange approaches use MC moves (involving swaps between replica simulations run under different conditions) to allow increased sampling in a variety of systems. MD itself can also be used as a Markov Chain Monte Carlo (MCMC) proposal move as in hybrid Monte Carlo.⁴⁹ Additionally, in prior work in the YANK package, MC rotational and translational moves have been combined with MD to help with rapid ligand positional/orientational decorrelation while doing binding free energy calculations in implicit solvent.^{8,50} In general, however, designing MC moves that fully exploit available knowledge (to make non-local proposals) while retaining reasonable acceptance rates is difficult in the condensed-phase.

2.1.4 Nonequilibrium Candidate Monte Carlo (NCMC) couples MD with the benefits of MC and yields more efficient sampling—

Nonequilibrium Candidate Monte Carlo (NCMC) provides a framework for translating insight about the system (in the form of a naïve Monte Carlo proposals) into practical algorithms⁴⁴ that retain some of the advantages of Monte Carlo while providing dramatically higher acceptance. The motivation is that it can be easier to construct a *finite-time proposal process* (a nonequilibrium “protocol”) that achieves high acceptance rates with short correlation times than to construct a successful *instantaneous proposal*. In the dimer example above, instead of instantaneously proposing a single large dimer extension move, we may construct a nonequilibrium process including a sequence of small dimer extension increments. If, after every incremental “perturbation,” the rest of the system is allowed to “relax”/“propagate,” then we might end up with an acceptable proposal that has crossed a free energy barrier.

NCMC itself has already seen some use in a variety of capacities, including driving torsion angles to increase sampling,^{51,52} performing constant pH simulations^{53–56} and creating an osmostat to simulate under macroscopic salt conditions.⁵⁷

NCMC was originally presented in a very general setting, where (1) the target distribution is an *expanded ensemble* of configurations and thermodynamic states, (2) the protocols may mix arbitrary sequences of steps, and (3) each proposal is drawn from a distribution over protocols. Here, we consider a special case where we have only a single thermodynamic state, a single time-symmetric protocol, and a simple “perturbation kernel,” so many of these terms cancel out and leave a simpler exact expression for the acceptance criterion. We make a further approximation in the acceptance criterion to improve performance, as we discuss further below.

NCMC permits nonequilibrium relaxation of most of the system while part of the system is being driven over (or around) kinetic or energetic barriers prior to acceptance or rejection of the NCMC move. Instead of proposing large instantaneous perturbations to the system, NCMC divides a target large perturbation into a series of steps consisting of smaller instantaneous perturbations, where each perturbation is followed by propagated dynamics. After this series of perturbation and propagation steps, the whole sequence is accepted or rejected as an NCMC move. The intermediate states are always discarded and do not count towards any equilibrium averages or other properties as they are transiently out of equilibrium.

The NCMC procedure is performed via a protocol Λ , which utilizes a sequence of *perturbation* kernels $a(\mathbf{x}, \mathbf{y})$ and *propagation* kernels $K(\mathbf{x}, \mathbf{y})$, where x and y are microstates of the system. By “kernels” we mean conditional probability distributions, p , that we can evaluate pointwise and draw samples from. Furthermore, each kernel p must satisfy the requirement that if $p(\mathbf{x}, \mathbf{y}) > 0$ then $p(\mathbf{y}, \mathbf{x}) > 0$, for all pairs $\mathbf{x}, \mathbf{y}$.

Here, we use a symmetric protocol consisting of T steps, where the perturbation and propagation steps are alternated with either a or K appearing at both the beginning and the end, so that $\Lambda = (a_1, K_1, a_2, K_2, \dots, K_T, a_{T+1}) = \tilde{\Lambda}$, where $\tilde{\Lambda}$ is the reverse protocol. This protocol produces a trajectory $X \equiv (\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T)$. To generate the appropriate acceptance

for an NCMC move to maintain detailed balance, we also have to consider the probability of observing a time-reversed trajectory $\tilde{X} \equiv (\tilde{\mathbf{x}}_T, \dots, \tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_0)$ under the reverse protocol

$\tilde{\Lambda} \equiv (a_{T+1}, K_T, a_T, K_{T-1}, a_{T-1}, \dots, K_1, a_1)$, where $\tilde{\mathbf{x}}_t$ is the microstate $\mathbf{x}_t$ with reversed momenta. Because we are considering a symmetric protocol, however, the forward and reverse protocol are identical, thus simplifying the acceptance criterion.

The protocol used in our package, Binding modes of Ligands Using Enhanced Sampling (BLUES) is thus a symmetric sequence of perturbation and propagation events, starting and ending with perturbation. The perturbation typically consists “thermodynamic perturbation” — modifying the potential energy function to change interactions between the ligand and the protein. However, the central perturbation step in each NCMC cycle is an instantaneous perturbation of the ligand coordinates. These perturbation (thermodynamic or instantaneous) events are interspersed with propagation via Langevin dynamics.

For *perturbation*, we alchemically modify the potential energy function (described in detail below) to slowly annihilate and then restore ligand interactions with the environment, resulting in a sequence of reduced potentials u_t that incorporate the time-dependent interactions of the ligand with its environment. In the middle of the protocol, when the ligand is no longer interacting with the environment, we rotate the ligand into a new orientation in an operation that does not change the potential energy of the system. This is done by translating the center of mass of the ligand to the origin, applying a rotation matrix to its coordinates, and reversing the translation to restore the ligand’s original center of mass. The rotation matrix is drawn uniformly over the space of all rotations using a quaternion approach, in which a random quaternion is generated uniformly over a 4D hypersphere and converted to a rotation matrix. This ensures the probabilities of generating a particular rotation matrix and its inverse are equal so that the overall proposals are time-symmetric.

For *propagation*, we use a Langevin integrator with specific properties (described below).

A variety of acceptance criteria $A[X]$ applied at this point would restore the system to equilibrium. For the case of symmetric protocols $\Lambda = \tilde{\Lambda}$ where all perturbation operations are symplectic (preserve phase space volume), the acceptance probabilities for a proposed NCMC trajectory X given protocol Λ must satisfy

$$\frac{A[X]}{A[\tilde{X}]} \equiv e^{-\Delta S[X]} = e^{-w[X]}, \quad (1)$$

where $\Delta S(X)$ is the conditional path action difference, which is equivalent to the total work $w[X]$. The total work includes both protocol work and “shadow work”,⁵⁸

$$w[X] \equiv w_{\text{protocol}}[X] + w_{\text{shadow}}[X] \quad (2)$$

where the protocol work depends on the changes in reduced potential energy during each of the perturbation steps,

$$w_{\text{protocol}}[X] \equiv \sum_{t=1}^T [u_t(\mathbf{x}_t) - u_{t-1}(\mathbf{x}_t)] \quad (3)$$

By contrast, the $w_{\text{shadow}}(X)$ depends on internal details of the propagation scheme used.⁵⁸

While neglect of the shadow work can lead to large errors in general,⁵⁸ we select a specific Langevin integrator that preserves the configurational distribution to very high accuracy, the BAOAB integrator of Matthews and Leimkuhler,^{59,60} allowing us to neglect this contribution without introducing large error. We justify this approximation by observing that the sequence of Langevin propagation kernels are *nearly exact* Markov kernels, each preserving the distribution $\pi_t(\mathbf{x}) \propto e^{-u_t(\mathbf{x})}$ with high fidelity.⁶¹ Recall that, due to discretization error, the invariant distribution $\rho_t(\mathbf{x})$ sampled by a numerical algorithm for Langevin dynamics will differ slightly from the target (i.e. $\rho_t \approx \pi_t$), and the magnitude of this difference increases with the integrator step size. This may introduce bias. We neglect this bias, since the specific integrator employed here is thought to preserve the configurational distribution to very high accuracy.^{59,60} This conclusion is based on analytical results showing that the integrator approximates configurational averages to fourth-order in the timestep (as opposed to second-order for competing integrators), and extensive numerical evidence examining particular biomolecular observables.^{59,60} Note that we would also use this criterion if each propagation step were a reversible MCMC move.

In practice, using an exact MCMC kernel (such as Generalized Hamiltonian Monte Carlo) for propagation would substantially increase the computational expense of a protocol by (a) introducing costly energy evaluations during accept-reject steps, (b) reducing the feasible integration timestep, and (c) dramatically increasing correlation times if the acceptance rate is even slightly less than 1.⁶² Including the shadow work contribution would also substantially reduce the acceptance rate of long protocols. In principle, though, one could simply use the total energy difference between the start and end states in the acceptance criteria; however, this would both lower the acceptance rate and lose the exact contributions of the protocol work and shadow work to the total work. Knowing their exact contributions is important to know how much the shadow Hamiltonian differs from the true Hamiltonian. In future work, we will examine the bias vs. efficiency trade-offs of our neglect of the shadow work, and the extent to which these can be mitigated by choice of Langevin integrator, or by using reduced-momentum-flipping variants of Hamiltonian Monte Carlo.^{62–64} While it has been argued more generically that the contribution of “shadow work” in nonequilibrium simulations can be neglected without introducing much bias,⁵³ this is likely highly dependent on the specific choice of integrator used, so we recommend caution if other integrators are considered.

2.1.5 We combine NCMC with random ligand rotational moves—Here, we provide details of our NCMC move proposals for ligand binding mode sampling. We combine thermodynamic perturbation of the ligand (alchemically changing its interactions

with the protein) with uniform random rotation around the ligand center of mass (COM). Specifically, we scale λ over a series of n NCMC steps until the ligand no longer interacts with the protein (removing its steric and electrostatic interactions). λ scales the interactions between the ligand and the rest of the system; to elaborate further, λ controls the strength of interactions between the ligand and its environment, with $\lambda = 1$ corresponding to the fully interacting state, and $\lambda = 0$ corresponding to the non-interacting state (as discussed in 2.5). The ligand is then randomly rotated to a new orientation in the binding site around its center of mass. Then its interactions are turned back on by scaling λ over a series of another n NCMC steps, as conceptually shown in Figure 2. Finally, we use the analogue of the Metropolis-Hastings acceptance criteria⁶⁵ that satisfies Eq. 4 to accept or reject the resulting move.

$$A[X] = \min \left\{ 1, e^{-w_{\text{protocol}}(X)} \right\} \quad (4)$$

Ligand rotation does not strictly need to be around the COM; it could be around a randomly selected heavy atom, or a point chosen within a Gaussian distribution around the COM, or various other options; we use the COM here for simplicity.

Figure 2 shows a cartoon of how these NCMC moves can work for exploring ligand binding modes. The ligand starts fully interacting (Figure 2(A)) and its interactions with the rest of the system are slowly turned off through alchemical λ coupling over a series of NCMC steps (Figure 2(B,C)). When the ligand is fully non-interacting, a random rotation (see Section 2.1.4) around the ligand's center of mass is performed (Figure 2(D)). Then the ligand's interactions are subsequently turned back on until it is once again fully interacting, potentially allowing it to find an alternate favorable orientation in the binding site (Figure 2(E,F)). We then accept or reject the move based on the acceptance criteria in Equation 4. In order to preserve detailed balance, the momenta of the system must be reversed after acceptance or rejection of proposed moves, or the momenta must be reassigned randomly from a Boltzmann distribution after the move;^{62,66,67} in this work, we take the latter approach and randomly assign the momenta of the of the entire system after a NCMC move attempt.

2.2 We study a T4 lysozyme cavity mutant which binds simple ligands

Here, we test several methods, including our new NCMC rotational method, on a T4 lysozyme cavity mutant which binds simple ligands. The T4 lysozyme L99A cavity mutant studied here has a buried binding site which readily binds non-polar molecules, making it a common model system for free energy calculations.⁶⁸

Toluene, a T4 lysozyme L99A binder, was chosen as the initial ligand for testing this method for a number of reasons. One is that toluene's symmetry allows for a convenient check of correctness; symmetric binding modes should have equivalent populations with adequate sampling. Also, the different potential binding modes for toluene differ primarily based on rigid body rotation of the ligand in the binding site, so rotational moves should increase sampling of the relevant binding mode(s). In addition, previous conventional MD

simulations we ran of toluene bound to T4 lysozyme suggest two distinct stable binding modes are present. Adequate sampling of even these two simple binding modes poses significant challenges for conventional MD.²⁰

After testing our NCMC rotational method on toluene, we also explored its capacities on 3-iodotoluene, a more bulky ligand. 3-iodotoluene does not have the same symmetry as toluene, meaning that we cannot take advantage of symmetry as a convenient check for convergence of populations. However, it is still valuable as a test for efficiency on bulkier molecules.

The T4 lysozyme binding site is relatively simple, and most ligands have only a small number of stable/metastable binding modes (e.g. here, M in Figure 1 is typically two to a few). This limits any efficiency gains our method might provide in this site relative to the efficiency which could be expected compared to conducting separate binding free energy calculations for each stable binding mode. However, biological binding sites can have far more metastable binding modes (e.g.⁶⁹). Thus, our main focus in this work is to compare the efficiency of different approaches for sampling ligand binding modes in lysozyme, and not to benchmark of our method to absolute binding free energy calculations.

2.3 System preparation

2.3.1 T4 lysozyme/toluene system setup—The T4 lysozyme L99A structure with toluene bound was taken from the 4W53 protein structure from the Protein Data Bank. Hydrogens were added to the protein using `tLeap` from AmberTools14.^{70,71} Hydrogens were added to toluene using Maestro and parameterized using GAFF version 1.7⁷² and AM1-BCC charges.⁷³ Hydrogens and missing atoms of the protein were added using `tLeap` in AmberTools14, and parameterized using `ff99sbuildn`.⁷⁴ A TIP3P rectangular box was added with 10Å padding from the protein to the nearest box edge, and Cl^- atoms were added to neutralize the charge of the system. The resulting `.prmtop` and `.inpcrd` files were converted to the equivalent GROMACS formats using `ACPYPE`.⁷⁵

The system was then minimized using steepest descent running for 2500 steps. The system was then equilibrated in GROMACS 5.1 for 25000 2 fs steps with constant volume, then equilibrated under constant pressure for the same number of steps using a Parrinello-Rahman barostat to maintain a pressure of 1 atm. Long range electrostatics were calculated using Particle Mesh Ewald.⁷⁶ Long range dispersion corrections were used for calculating the energy and pressure. These preparatory simulations were performed at 300 K and `v-rescale` with `tau_t` set to 0.1 ps was used to perform temperature coupling. `v-rescale` is the GROMACS option that provides temperature coupling using velocity rescaling with a stochastic term; `tau_t` specifies the average time between velocity rescalings. Full details of the simulation setup can be found in the `.mdp` files in the Supporting Information (SI).

2.3.2 Toluene/water system setup—Toluene was parameterized and created the same way as Section 2.3.1, except without any protein present. The system's energy was then locally minimized with a tolerance of 10 kJ/mol. The system was subsequently equilibrated

in OpenMM under constant pressure at 0.987 atm with and 300 K using a Monte Carlo barostat for 25 ns using a Langevin integrator with a 1 ps timestep and 1/ps collision rate.

2.3.3 T4 lysozyme/3-iodotoluene system setup—The T4 lysozyme L99A structure was taken from the 4W53 protein structure from the Protein Data Bank, with the toluene ligand removed. 3-iodotoluene was then docked using OpenEye's FRED (ver 3.2.0.2) and we retained the top scoring generated pose. Preparation of the system using `tleap` was done the same as in 2.3.1. The system's energy was then locally minimized with a tolerance of 10 kJ/mol. The system was subsequently equilibrated in OpenMM under constant pressure at 0.987 atm with and 300 K using a Monte Carlo barostat for 25 ns using a Langevin integrator with a 1 ps timestep and 1/ps collision rate.

2.3.4 Setup for OpenMM NCMC simulations—OpenMM 7.1.0 was used.⁷⁷ The OpenMM simulations used the same systems loaded from the `.prmtop` and `.inpcrd` files as prepared in Section 2.3.1. For the MD portions of the protocol a Langevin integrator was used with a 2fs timestep and 1/ps collision rate. No barostat was used for these simulations after equilibration (and thus simulations were done in the NVT ensemble). Long range electrostatics were calculated using Particle Mesh Ewald.

2.3.5 Setup for Yank binding free energy calculations—Yank⁵⁰ (ver 0.20.1) was used to perform binding free energy calculations to compare the free energy differences between the observed toluene binding modes. Three simulations were performed in total. All simulations were run in the NPT ensemble with a pressure 1 atm at 300K. The λ coupling for the electrostatics and sterics for each simulation was determined using Yank's `auto` setting. The first Yank simulation had harmonic restraints in place whose λ coupling was also determined by the `auto` setting. The second and third Yank simulations each used restraints to restrict the toluene ligand to a single binding mode. The restraints used were a set of one bond, two angle, and three dihedral restraints.^{20,78} The spring constant for the bond restraint was set at 10 kcal/(mol*Å²) while the angle and dihedral restraints were set at 10 kcal/(mol*radians²). These restraints were constant throughout the λ windows. All other system parameters were the same as Section 2.3.4. Full details of the Yank simulations are included in the `.yaml` files in the SI.

The first and second simulation were started using a representative crystallographic pose from the NCMC simulations (labeled A in this paper), while the third simulation was started from a representative pose corresponding to the other binding mode (labeled B).

2.4 We built Markov state models of toluene binding to lysozyme

The T4 lysozyme system with toluene bound (as described in 2.3.1) was minimized in GROMACS 5.1^{79,80} via steepest-descent, followed by 1 ns of NVT simulation and then 5 ns of NPT simulation at 1 atm and 300 K for equilibration. The leapfrog integrator was used with a 2 fs timestep and the bonds involving hydrogen constrained with LINCS.⁸¹ The system was then simulated for a total of 806 ns under the same NPT conditions, saving configurations to a trajectory file every 30ps. tICA was performed on the pairwise-distances of the toluene heavy-atoms and the alpha carbons of the binding site of the trajectory with a

lagtime of 0.6ns to generate order parameters for MSM construction. Of the 210 initial dimensions, 22 dimensions were retained— enough to account for 95% of the kinetic variance in the data, and were scaled by the kinetic map scheme. These large number of initial dimensions were used to help ensure all relevant binding modes were separated. An initial MSM was estimated from the order parameters computed from the trajectory using PyEMMA,⁸² using 1000 microstates generated from k-means clustering and a lagtime of 6 ns. The MSM was coarse-grained into four macrostates using Perron-Cluster Cluster Analysis ++(PCCA+);^{83–85} full details are available in scripts deposited in the SI. Two random trajectory frames from each of the four macrostates were then used as the starting point for new simulations to further sample each identified binding mode and potentially generate additional transitions. These eight additional simulations were each run for 60 ns and combined with the longer run above to re-estimate a MSM, following the same sequence of steps, with these additional simulations added to better explore transitions out of each macrostate. The total amount of aggregated simulation time used for the final MSM was 1.286 μ s spread across nine trajectories. Additional simulation details can be found in the SI.

2.5 We use Nonequilibrium candidate Monte Carlo (NMC) to study toluene binding to lysozyme

2.5.1 Our NCMC procedure uses random rotational moves to enhance binding mode sampling—Here we use NCMC to enhance sampling of ligand binding modes in the T4 lysozyme binding site. As discussed in Section 2.1.5, coupling thermodynamic perturbation with rotational move proposals can allow the ligand to cross energy barriers between binding modes while allowing some amount of relaxation to improve the acceptance of proposed moves. In our procedure (Figure 2) interactions between the protein and ligand are on at the beginning of a move proposal. Then the interactions are turned off by scaling λ from 1 to 0 (where 1 corresponds to full interactions and 0 corresponds to no interactions) over a series of n steps, following the scheme shown in Figure 3. Soft core potentials were used to avoid numerical instabilities related to scaling the steric and electrostatic interactions, with a 1-1-6 potential with $\alpha = 0.5$.⁸⁶ As the NCMC protocol progresses, we first turn off the electrostatics of the ligand by scaling its potential energy contribution linearly with lambda so that the electrostatics are completely removed as we go from $\lambda = 0$ to $\lambda = 0.2$. Then we decouple the Lennard-Jones interactions from $\lambda = 0.2$ to $\lambda = 0.5$ using soft core potentials so that the ligand is now fully non-interacting with the rest of the system (although its intramolecular Lennard-Jones interactions are still in place).

Then a random rotation of the ligand is performed (as described in Sec 2.1.4), with the random quaternion generated using `mdtraj`.⁸⁸ Finally, the interactions are subsequently turned back on via a reverse of the original procedure.

The work done during this process is accumulated and used to accept or reject the move (consisting of the full decoupling, rotation, and recoupling procedure) via Equation 4.

After the NCMC move is accepted or rejected, velocities are randomized by drawing from the Maxwell-Boltzmann distribution appropriate for the temperature and then a phase of conventional MD is performed to better sample the other (protein/solvent) degrees of freedom. This full procedure consisting of NCMC moves plus MD steps is then repeated

many times until convergence, and the populations can be then estimated from clustering the resulting trajectory and computing the time spent in each state.

2.5.2 We implemented this approach via our BLUES package for binding mode sampling—

We constructed a package called Binding modes of Ligands Using Enhanced Sampling (BLUES) to facilitate the use of NCMC to enhance ligand sampling. BLUES implements the approach outlined in Section 2.5.1 and switches between sampling the system via normal MD and NCMC alchemical perturbation. The BLUES package allows straightforward control of the number of MD steps between each NCMC move, the number of alchemical steps used within each NCMC move, and the total simulation time and number of MD/NCMC cycles.

In BLUES the `alchemicalfactory` module of `openmmtools`⁸⁹ version 0.11.2 was used to allow annihilation and restoration of toluene's steric and electrostatic interactions. The MD portions of these simulations used OpenMM's Langevin integrator. The NCMC portion of the OpenMM simulations used an implementation of the BAOAB integrator of Langevin dynamics⁶⁰ During the NCMC propagation phase, we also froze the positions of protein residues further than 5 Å from the ligand, and solvent molecules (which in this system were always more distant than 5 Å from the ligand, to reduce the variance in work over the course of the NCMC protocol, thereby improving acceptance probabilities.⁵² The selection of the frozen water and protein residues was not updated during the simulation; this was appropriate in this case as the binding site is a buried, non-polar binding site with no water nearby and the ligands remain stably in the binding site, so no updates were needed. The long range dispersion correction was turned off during the NCMC integration steps due to computational costs recalculating the correction while scaling λ , but was accounted for by taking into account the differences in energy between the alchemical and normal systems at the initial and final states.

Full details of the implementation and a class diagram are available on GitHub at the link below.

BLUES is also an extensible framework in that it allows general MC moves to be performed during the NCMC portion. Here we consider only random rigid-body rotations around the ligand center of mass as described in Section 2.5.1. However, other moves which might be of interest to explore later could include translations of subsets of a given system, moves involving ligand internal coordinates, or sidechain MC moves.

BLUES is freely available on GitHub under the MIT license at <https://github.com/MobleyLab/blues>. We used BLUES version 0.1.0 to obtain the data found in this paper. The same systems from Sections 2.3.1–2.3.3 were used.

2.5.3 Variations in the NCMC protocol dramatically impact move acceptance—

With NCMC in BLUES, we can vary the number of perturbation and propagation steps relative to the amount of standard MD in order to allow an adequate amount of relaxation in order to ensure reasonable acceptance without using so much relaxation that the approach becomes tremendously inefficient. Here, we tested this by measuring the acceptance ratio as

a function of the amount of relaxation (Figure 4), and found that the acceptance probability increases rapidly from around 100 NCMC switching steps up to 10000 NCMC switching steps, then begins increasing more slowly with further relaxation. Thus, here, we selected 10000 NCMC steps and 10000 MD steps per cycle as a reasonable choice in order to determine how much enhancement in sampling (and therefore efficiency) NCMC can provide relative to standard MD or MD with MC.

Corresponding work distributions and standard deviations of work distributions are shown in SI Figures 1–2. For trials using 1000 NCMC switching steps and more, the uncertainty was calculated based on blocking.^{90,91} The number of blocks used was the amount that maximized the standard deviations of the acceptance rate across blocks. For trials using fewer than 1000 NCMC switching steps, accepted moves were rare enough that we took the standard deviation across four trials and computed the standard error from that.

One notable feature of Figure 4 is that in the region of shorter switching times (>1000 switching steps) there is a decrease of acceptance compared to instantaneous MC. This has been observed previously with NCMC,^{44,51} and in our case might arise from the asymmetric work distributions associated with removing and restoring the ligand's interactions⁹²

2.6 For reference, we compare NCMC with conventional MD and MD/MC

To compare the efficiency in sampling with NCMC versus more traditional forms of sampling, we also ran normal MD and MD with MC rotational moves on toluene with the same T4 lysozyme/toluene system. In order to make a fair comparison between methods we kept the number of energy evaluations consistent across methods.

For MD, we ran 20000 integration steps of MD during one iteration to mimic the number of NCMC energy evaluations per iteration. For MD with MC we ran 20000 integration steps followed by 10 MC random rigid body ligand rotations using the Metropolis criteria for each iteration. We ran each of these methods for 5000 iterations and then compared the trajectories and binding mode populations found in our NCMC simulation. The code used to perform these calculations can be found in the SI.

We observed very few transitions between binding modes in both the normal MD and MC with MD simulations. Because there are so few transitions in these cases we cannot expect the binding mode populations in these simulations to be converged to the equilibrium populations.

2.7 We analyzed our binding mode sampling using a dihedral angle which discriminates between the stable binding modes

Originally, we used many pairwise-distances as an order parameter when constructing the MSM to identify toluene's binding modes (Sec 2.4). Once those binding modes were identified however, further analysis found that a simple 1-dimensional progress coordinate could separate and identify them. To monitor the binding mode of toluene in the cavity, we picked a dihedral angle which clearly discriminates between toluene's four distinct binding modes. Specifically, toluene's binding mode was monitored via calculation of the dihedral formed by the alpha carbon of ARG118 and the C1, C5, and C7 atoms of toluene (Figure 5).

These angles were then used for construction of histograms to monitor populations of the observed binding modes and the number of transitions between binding modes. The populations of the different binding modes were monitored by using the following different bin boundaries $[-\pi, -1.5)$, $[-1.5, 0)$, $[0, 1.5)$, $[1.5, \pi)$. We assigned the following state labels according to the bin locations: $[-\pi, -1.5)$ is B1, $[-1.5, 0)$ is A1, $[0, 1.5)$ is B2, and $[1.5, \pi)$ is A2, where the A labels correspond to the crystallographic binding mode and B labels correspond to an in-plane rotated binding mode. To determine the uncertainty in the computed population as a function of simulation time, the populations of each binding mode were determined using fractions of the total simulation data in a blocking approach.^{90,91} The uncertainty in the computed populations were determined based on breaking each 10% of the simulation into a set of smaller blocks. The number of blocks used was the amount that maximized the standard deviations of the populations between blocks. For NCMC/MD this was 8 blocks. For MD and MC/MD, the standard deviation in the mean across blocks via bootstrapping failed to reach a maximum even with only one block per 10% of the simulation. The error bars in the plots of the MD and MC/MD simulations are shown based on one block per 10% of the data, but are likely to severely underestimate the true error.

2.8 We generated synthetic data to compare MD and NCMC transition efficiency

To get a better sense of the efficiency gains of NCMC compared to standard MD we constructed statistical models from the data of the MSM and the NCMC simulations. For the MSM we took the estimated MSM transition matrix directly after clustering the 1.286 μ s into four clusters corresponding to the four binding modes. For the NCMC simulation, we assigned each iteration to a particular macrostate using the dihedral order parameter as defined in Sec 2.7. We used those state assignments from those 5000 iterations to generate a transition matrix. To generate the synthetic data, we started from state A2 and iteratively applied the transition matrices to get trajectories of states. A new state could be the same as the previous, depending on the transition probabilities given by that particular transition matrix. This process was repeated, and the total state populations at each iteration were recorded. We performed 5000 propagation iterations for the MD and NCMC transition matrices with 1000 trials each. This allowed us to cheaply estimate uncertainties in the populations of each state at each time point by taking the standard deviation in the estimated population across all trials. Overall, this analysis facilitated an assessment of the rate of convergence of the populations.

2.9 We examined rotational distributions and added the case of 3-iodotoluene as an example of a bulkier ligand

To better compare the performance of NCMC and standard MC move proposals (as discussed in Results 3.2.3), we chose the bulky lysozyme ligand 3-iodotoluene and compared the efficiency of running a large number of standard MC move proposals with the efficiency of running many NCMC move proposals. Because our BLUES tool is not designed for efficient MC performance (since it has additional overhead to facilitate relaxation with NCMC) we did this test outside of BLUES, instead running standard MD simulations and then selecting snapshots from these to compare acceptance of MC and NCMC move proposals. Thus this test is not a benchmark of NCMC against MC, but an

assessment of the performance of NCMC and MC move proposals starting from the same ensemble of MD snapshots.

As preparation, we simulated the T4 lysozyme system with 3-iodotoluene for 90 ns, saving the positions every 0.2 ns, thereby saving a total of 450 trajectory snapshots. We used these snapshots to facilitate our comparison of the acceptance of MC moves and NCMC moves. For 3-iodotoluene, our goal is to assess overall acceptance, and see whether substantial rotational moves are being accepted with reasonable frequency – in contrast to our work on toluene (Section 2.6) where we were interested in estimating populations in order to ensure that our approach converges to the correct populations. Thus at the start of each MC or NCMC move attempt we randomly pick a starting trajectory snapshot to use as a starting point for a new move proposal. This allowed us to have a variety of starting points for our move proposals, while also ensuring that we were assessing the performance of move proposals with equivalent environments. For MC we performed 10 trials, each consisting of 2,000,000 move attempts where each move is instantaneous. For NCMC we performed 7 trials, each consisting of 2000 move attempts; each move consisted of 6500 NCMC switching steps.

Since we were interested in comparing not just acceptance rate but how these moves fared at *substantially* decorrelating ligand orientation within the binding site by sampling across different binding modes, we also monitored the angle by which each move rotated the ligand. Specifically, when a move was accepted, we calculated the angle of rotation by first calculating the rotation matrix needed to generate the final ligand positions from the initial ligand positions,⁹³ then calculating the angle of rotation θ by Eq 5, where R is the rotation matrix.

$$\theta = \arccos\left(\frac{\text{Tr}(R) - 1}{2}\right) \quad (5)$$

We also performed a similar routine to determine the rotation distributions in the toluene case, except we performed MD/MC or MD/NCMC sampling as previously described in Sec 2.6. For toluene, we ran 5 MD/MC trials consisting of 10000 iterations, with each iteration consisting of 10 MC moves attempts followed by 1000 steps of MD. For MD/NCMC we ran one trial consisting of 10000 iterations, with each iteration consisting of a NCMC move of 10000 NCMC switching steps and 1000 steps of MD.

In all cases, the resulting rotational distributions were histogrammed using 8 bins of 22.5 degrees; the error for each bin was determined using the standard error in the mean estimated across trials, except for the MD/NCMC T4 lysozyme/3-iodotoluene simulation, in which we estimated the error by dividing up the accepted frames of the trajectory into 8 blocks, which maximized the standard deviation, then computing the standard error in the mean by bootstrapping over the accepted blocks.

To monitor the binding mode, we found a dihedral order parameter that separated the 3-iodotoluene binding modes observed during the simulations (see SI); this involved the C1, C5, and I8 atoms of 3-iodotoluene and the alpha carbon of VAL111 of T4 lysozyme. This

was used to monitor the overall orientation of the ligand in the binding site, e.g. in SI Figure 3.

Results are given in Section 3.2.3.

3 Results and Discussion

3.1 Kinetics and populations of binding modes through MD and Markov State Modeling

We constructed a MSM from approximately 1 μ s of simulation data on the T4 lysozyme/toluene system (see Sec 2.4) to estimate equilibrium populations of the binding modes and timescales of interconversion. From the implied timescales of the MSM (Figure 6) we identified 4 kinetically separated binding modes of toluene as expected from the gaps between the third and fourth timescales.

Perron cluster cluster analysis+ (PCCA+) was then used to separate the trajectory frames of the MSM into 4 clusters. Visual inspection of the resulting macrostate clusters from PCCA+ revealed that there were two clusters, each with a symmetry equivalent partner (Figure 7). The populations estimated from the MSM show the populations of the symmetric states to be roughly equal, with $32 \pm 8\%$ and $26 \pm 6\%$ for the two symmetric binding modes corresponding to the crystallographic binding mode. The other binding mode showed $18 \pm 5\%$ and $23 \pm 6\%$ populations for the symmetric equivalents (SI Figure 4).

We find that timescales for binding mode interconversion are extremely slow, both from analyzing our long conventional MD simulation directly, and from the implied timescales of the MSM. Direct analysis of our long single 806 ns trajectory (Figure 8) showed that certain binding mode transitions are quite rare.

Additionally, the slow kinetics involved in sampling are highlighted by the implied timescales of the MSM (Figure 6). The slowest transition—switching between symmetric binding modes—occurs on a timescale of 100 ns, while the faster transitions occur approximately every 4ns.

It is important to note that even with the simplicity of both this binding site and toluene (which is a rather small ligand compared to the size of the site), slow transitions are still observed, consistent with earlier observations that binding mode interconversion is quite slow in the buried lysozyme binding site.⁶

Given the timescale of the slowest binding mode transitions observed here, obtaining accurate ligand binding mode populations from brute force MD or even MSMs seems particularly costly in this case. Specifically, to generate an accurate representation of the populations either approach will need to observe multiple transitions between binding modes. Especially in the MD case, this would require simulations which are at least 10 times longer than the 100ns timescale for the slower binding mode interconversion events—equivalent to at least 1 μ s of simulation. While toluene's symmetry could be used to obtain correct populations without adequately sampling the symmetric ring flip, any new ligand differing by a substitution breaking this symmetry (and there are many such ligands which bind in this site, such as the 3-iodotoluene case considered later in this work⁵) would require

adequate sampling of these previously symmetry-equivalent binding modes. In other words, adequate binding mode population estimates would likely require multiple microseconds of simulation time.

The populations calculated from this MSM agree well with values obtained from calculated absolute binding free energies for the different binding modes. We performed absolute binding free energy calculations from the two non-symmetric binding modes (see Section 2.3.5). The binding free energy for the crystallographic binding mode was $-7.1 \pm 0.2 k_B T$, while the non-crystallographic binding mode was calculated to be $-6.5 \pm 0.2 k_B T$. The populations estimated from the MSM would translate to a free energy difference of approximately $0.41 k_B T$, which is within error of the difference in absolute binding free energies, $0.6 \pm 0.3 k_B T$.

3.2 BLUES rapidly samples binding modes

3.2.1 Initial simulations indicate BLUES samples more rapidly than MD—Here, to compare the efficiency of our NCMC approach, BLUES, to that of brute force molecular dynamics and MSMs, we applied BLUES to the T4 lysozyme/toluene system. We applied the NCMC protocol of thermodynamic perturbation with random rigid-body ligand rotations (as described in Section 2.1.4) to observe the protocol's efficiency in sampling binding mode interconversions. The NCMC protocol was applied over 5000 iterations, each consisting of 10000 MD steps separated by NCMC move proposals consisting of turning off and restoring ligand interactions over 10000 steps, with random rotations while the ligand is noninteracting.

For reference, we also performed standard MD and MD/MC simulations using the same total number of energy evaluations. In the MD case, this meant 5000 iterations of 20000 MD steps. And in the MD/MC case, this meant 5000 iterations of 20000 MD steps interspersed by 10 conventional MC move proposals involving random ligand rotation. The number of energy evaluations at each iteration was kept constant between methods. Thus, with a 2fs timestep we simulated for an equivalent of 200ns total with each method.

Figure 9 shows the dihedral angle (indicating the binding mode) sampled versus time over each method's iterations, along with the resulting histogram of the binding mode populations. Compared to MD or MD coupled with traditional MC moves, this NCMC method allows rapid interconversion between all four binding modes.

This allows BLUES to reproduce the correct equilibrium populations (Figure 9(C), right), and on a force evaluation basis, NCMC was approximately 17 times more efficient than brute force MD with Markov State Modeling.

This is also evidenced by the fact that over the same number of iterations, BLUES converges rapidly to the correct equilibrium populations within 2000 iterations (Figure 10), whereas MD and MD/MC still have significant errors. For MD after 5000 iterations the major binding mode populations differ from the equilibrium populations by as much as 45%; for MC they differs by as much as 26%.

Although this NCMC implementation is more efficient on a force evaluation basis compared to MD, there is some computational overhead for alchemically modifying the sterics and electrostatics during integration in OpenMM for GPU simulations, causing the wall-clock time per NCMC iteration to be about three times longer than a MD iteration. Specifically, our calculations shown in Figure 10 took 2249 minutes for MD, 2189 minutes for MD/MC, and 5413 minutes MD/NCMC. Convergence to within uncertainty of the correct population, and to within 5% of the correct population, appears to occur for this system (Figure 10) well before 40% of the total simulation time (80ns); with a factor of three in additional cost, this takes about as long to run as 240ns of conventional MD simulation. Thus the savings of NCMC in terms overall wall-clock time is still about a factor of five compared to the MSM approach for this system, which required roughly $1.3\mu\text{s}$ of aggregate simulation data.

3.2.2 Statistical analysis confirms the efficiency of BLUES compared to MD—

To validate and further assess the relative efficiency of BLUES compared to MD, we built a statistical model of convergence of populations in these two cases. Specifically, we wanted to use transition matrices between the four states (in both cases) to propagate the populations over a long time in order to analyze convergence properties. We obtained the MSM transition matrix for the MD case. For the NCMC case we constructed a transition matrix from our BLUES simulation (as described in Section 2.8). We then built a model of convergence using these two matrices as a starting point. In each case, we started a hypothetical simulation from binding mode A1 and used the transition matrix to propagate the populations, at each timestep choosing a new state to transition based on the probabilities in the transition matrix. In the MD case the transition matrix was constructed using a lagtime of 6 ns so our simulation timesteps corresponded to 6 ns, whereas in the BLUES case the transition matrix was constructed for a 40 ps MD/NCMC iteration so timesteps were 40 ps. We then repeated 1000 such simulations for each case and examined the mean population as a function of time and the standard deviation over trials. These are shown in Figure 11. In this case we find that NCMC converges much more quickly than MD, specifically, for MD it takes approximately 12000 nanoseconds for the standard deviation in the slowest converging population to drop below 5%, indicating that typical simulations would have a population error of no more than 5%. In contrast, for NCMC the standard deviation drops below 5% for the slowest converging population by 60 nanoseconds, a factor of 200 reduction in total simulation time compared to MD.

It is important to note that in this model the transition matrices are only estimates of the true transition matrices, so populations will eventually converge to a stationary distribution as seen in Figure 11, but the final populations will have some amount of error. Here we are more interested in examining the rate of convergence than the populations as our goal is to measure the relative efficiency of both techniques.

Ultimately, the difference in performance between MD/NCMC and the benchmark MD and MD/MC approaches is fairly simple to understand. Conventional MD cannot cross kinetic barriers any faster than their inherent timescales, so, since timescales for interconversion between the slowest binding modes here are around 100ns (Section 3.1), convergence in conventional MD will necessarily take many times longer than 100ns. The MD/MC approach here couples conventional MD with occasional random ligand rotational moves

which are accepted or rejected via conventional Monte Carlo, but because the binding site is relatively densely packed—even though it is not solvent exposed—the vast majority of these are rejected for toluene (giving an acceptance rates of $0.091 \pm 0.004\%$). Thus, this approach performs almost equivalently to standard MD. Our NCMC approach implemented in BLUES converges much more rapidly because ligand rotational move proposals can relax before being accepted or rejected, thus dramatically enhancing the acceptance rate to approximately 11%. These acceptance rates are reflected in the transitions between states, (Figure 12), which show that MD/NCMC produced 497 transitions. This is more than twice as many transitions than the other methods employed; during the same number of iterations MD produced 242 transitions and MD/MC produced 230 transitions. Also, MD/NCMC produced high transition counts from any given binding mode to any other binding mode, while the other methods primarily produced transitions from a given binding mode to a subset of all the binding modes. Specifically, the other methods had the most transitions between in-plane binding modes (A1-B1 or A2-B2 transitions, Figure 12(a) and (b)) which are relatively fast to interconvert in normal dynamics, where BLUES had significant numbers of transitions between all binding modes, even the out-of-plane flip, which has a characteristic timescale of roughly 100 ns for conventional MD (Section 3.1). For example, the A1 to A2 transition occurred only twice in standard MD, and once in MD/MC, but 48 times in BLUES. This is also clearly apparent from Figure 9 (left panels), where the MD and MD/MC approaches have few transitions between the top pair of states and the bottom pair of states, but BLUES has a very large number.

3.2.3 NCMC does not compare as favorably to MC for toluene, but NCMC moves have clear advantages for bulkier iodotoluene—While BLUES compares favorably to standard MD in the case of toluene bound to lysozyme, and has a better acceptance ratio ($10 \pm 1\%$) than standard MC ($0.091 \pm 0.004\%$), the difference in acceptance ratio between BLUES and standard MC is not actually enough to justify the additional computational expense of the NCMC relaxation. Specifically, instead of doing 10000 NCMC steps (as in BLUES) to achieve a reasonable acceptance rate, we could simply do a very large number of MC trials (e.g. 10000, for a similar computational expense) with the low acceptance rate and still see a reasonable number of moves accepted. We believe this relative advantage of standard MC may be a peculiarity of toluene in this particular binding site (which is relatively large compared to the size of toluene, and known to be especially rigid^{5,68,94,95}), and not representative of typical MC performance in condensed phase systems.⁴⁴

To further test the relative performance of MC and NCMC, we examined performance of NCMC versus MC move proposals for a larger ligand in the lysozyme binding site, 3-iodotoluene; we find that the presence of the bulky iodo substituent dramatically impairs acceptance of MC moves, presumably due to the larger size of the ligand relative to the size of the binding site (see Section 2.9 for methods). 3-iodotoluene is another known binder in the lysozyme L99A site; however, due to its lack of symmetry, we are unable to take advantage of ligand symmetry to provide a simple metric for convergence of binding mode populations. It is nevertheless useful here as a good example of a larger ligand which should have several different metastable binding modes in this site.

For 3-iodotoluene, our comparison focuses just on acceptance of MC and NCMC *move proposals* given a fixed ensemble of MD snapshots as a starting point, and we find that the acceptance rate of standard MC is $(1.2 \pm 0.2) \times 10^{-2} \%$ while for NCMC, it is $0.8 \pm 0.1 \%$. Thus, standard MC results in an order of magnitude lower acceptance for 3-iodotoluene than toluene, meaning that standard MC is closer to NCMC's performance on 3-iodotoluene rather than outperforming it (in terms of acceptance rate) as in the case of toluene.

However, the overall acceptance probability is not the only consideration – we are also interested in how each technique improves the acceptance of *substantial* moves that significantly alter the ligand binding mode. After all, when proposing random ligand rotational moves, a rotation of zero, or of very nearly zero, is a valid move proposal, so potentially many of the moves being accepted are in fact very small ligand rearrangements. To examine this, we determined the fraction of rotational moves accepted as a function of amount of rotation (Figure 13), both for toluene and for 3-iodotoluene.

For toluene, due to its relatively small size, MC and NCMC result in relatively similar acceptance profiles as a function of the amount of rotation (Figure 13(a) and (b)), except perhaps that NCMC yields improved acceptance of intermediate amounts of rotation between 0 and 180 degrees. However, for 3-iodotoluene (Figure 13(b) and (c)), standard MC results in virtually no acceptance of moves larger than 20 degrees $((5 \pm 2) \times 10^{-5} \%)$, whereas NCMC retains good acceptance of such moves $(0.68 \pm 0.07 \%)$. (For the MC case on the 3-iodotoluene system, we performed a total of ten trials of 2,000,000 MC attempts and in each trial, typically saw at most one or two accepted moves consisting of significant rotations.) Accounting for the 6500 relaxation steps used in NCMC, MC would have a $(3 \pm 1) \times 10^{-1} \%$ acceptance of significant rotations indicating that NCMC is still approximately twice as efficient than MC in this case.

This is further supported by SI Figure 5, where we examine how effective MC versus NCMC moves are at sampling new binding modes not represented in the MD trajectory providing the starting points for our move attempts. For MC, only very few moves outside the starting region are accepted, whereas NCMC is quite effective at sampling new binding modes. Additionally, MC gives the apparently false impression that the initial set of binding modes is by far the most favorable (because it is so difficult to find a combination of ligand orientation and protein conformation which can allow a rotational move to be accepted with no relaxation), whereas NCMC suggests that there are alternate binding modes that may be considerably more favorable.

Thus, the acceptance ratio only gives a small part of the overall picture; NCMC does dramatically better than MC at sampling significant binding mode transitions, enough so that even in this relatively simple 3-iodotoluene test system, NCMC outperforms simply performing a very large number of MC trials (with an equivalent number of energy evaluations) by roughly a factor of 2.

We also examined the performance of NCMC move proposals for 3-iodotoluene as a function of the amount of relaxation, as shown in Figure 14. In keeping with the analysis just prior, we find that NCMC move proposals perform nearly as well for substantial rotations as

for all rotations, whereas MC move proposals do not. Additionally we performed simulations analogous to Sec. 9 for 3-iodotoluene, comparing the sampling of MC to NCMC (SI Figure 6) and found that NCMC was able to sample alternate binding modes much more consistently.

Overall, our tests on 3-iodotoluene indicate that for larger ligands and/or more confined environments, standard MC move proposals perform dramatically worse than for toluene, in keeping with what might be expected for large moves in condensed-phase systems.⁴⁴ This, combined with the overall flexibility of the NCMC approach in combining some of the advantages of MD with those of MC, indicates that this approach may be a good general strategy for ligand binding mode sampling. Additionally, this work highlights how important it is not just to monitor the overall acceptance rate of moves, but how the acceptance rate of moves is coupled to the size of moves; here, NCMC results in high acceptance of large moves, while MC does not.

4 Conclusions

4.1 Summary

Overall, we find that NCMC with random ligand rotational moves dramatically enhances sampling of ligand binding modes compared to the other more conventional methods employed here.

Particularly, we have shown that NCMC can greatly enhance move acceptance for exploring ligand binding modes by allowing for relaxation during attempts. NCMC also allows dramatically faster sampling than standard MD because of its ability to cross steric barriers; advantages over standard MC are less clear but grow with the size of the ligand relative to the amount of space it has in the binding site. The generality of this method is particularly appealing. We did not use any prior information about the binding modes in generating our move proposals, which involve random rigid-body rotations, thus this type of move shows promise in broadly sampling different potential binding modes without any prior knowledge. Extensions of this approach however, could potentially make use of other information—for instance from docking—to perform guided rotations targeting specific binding modes. Although NCMC rotational moves can help sample potentially slow binding mode transitions, there are some factors which can pose challenges for this approach. The acceptance rate will likely decrease as the ligand size grows, since a larger percentage of possible random rotations will lead to particularly significant clashes that cannot relax in the span of the move, and favorable binding modes will become correspondingly harder to find by random exploration. Additionally, rotational moves alone will not cover all binding mode possibilities in some cases, but the addition of other Monte Carlo moves (such as translation) could perhaps help address this. Also, rigid body random rotations of a ligand will likely not be as effective for flexible ligands, whose binding modes can be dependent on changes to the internal degrees of freedom such as torsional rotations.

While toluene and iodotoluene binding to T4 lysozyme might not seem to be particularly relevant to drug discovery problems, the problem confronted here actually has considerable similarity to problems encountered in fragment based drug discovery (FBDD). FBDD

attempts to find promising leads for early stage drug discovery by studying the binding of very small, often relatively rigid, ligands.^{96–98} These ligands can in fact be of relatively similar size and rigidity to toluene in some cases.³ Thus, prediction of binding modes of small rigid fragments is in fact of considerable interest. Additionally, even when structural data is available for the binding of fragments, the X-ray crystal structures obtained from FBDD campaigns sometimes have ambiguous electron density for the ligand, making the binding mode difficult to determine.²⁸ Applying this NCMC method to cases involving rigid ligands could help determine the major binding mode(s) and/or resolve ambiguity in experimental structural data.

4.2 Future Work

Future work will focus on exploring other degrees of freedom not just of the ligand, but also the protein. For example, previous studies of T4 lysozyme with p-xylene have shown the VAL101 sidechain orientation greatly impacts which of p-xylene's binding modes are favorable. That valine sidechain is, however, slow to sample and thus would make an excellent test case for NCMC sidechain rotational sampling.

We are also interested in exploring the internal degrees of freedom of the ligand. Performing random rotations of ligand rotatable bonds might be one way to explore the internal degrees of freedom. T4 lysozyme with n-propylbenzene might be suitable for such a test, as the crystal structure shows multiple binding modes due to rotations of the ligand's alkyl tail. Also, rings within a molecule can often be pseudosymmetric, thus necessitating sampling of each ring conformation. These ring flips can be similarly treated by rotating the internal bonds of the molecule.

We would also like to investigate what effect the binding site has on acceptance rates of these NCMC moves. The T4 lysozyme binding cavity we tested on is fairly rigid; however, it could be more difficult to achieve high acceptance rates with this approach in systems with solvent-exposed or flexible binding sites, due to protein or water rearrangements as the ligand is turned off. Preliminary tests with BLUES rotational moves of toluene in solution (without protein) show reasonable acceptance rates (SI Figure 7), suggesting that water rearrangements, at least, would not cause major issues with this methodology. Another potential concern is that the ligand could more easily leave the binding site when its interactions are turned off. We have not observed such an effect here, probably in part because our moves are purely rotational and the time spent non-interacting is minimal. However, if this were to occur, modest restraints could perhaps be used to prevent this.

The NCMC framework in BLUES has been written to allow straightforward extension to other types of move proposals, such as protein sidechain or ligand torsion rotations as noted above. Even more ambitious move types may be of interest as well. For example, techniques like smart darting⁹⁹ could potentially be used to allow ligand hops between different candidate binding sites or binding modes that have been determined in advance.

One more avenue we would like to pursue is to use the nonequilibrium work generated from the BLUES protocol to enhance our binding mode population comparisons. During our protocol in BLUES we are essentially performing two nonequilibrium simulations—a ligand

deletion followed by an insertion. By separating the work done during these two phases we should then be able to use those work values to alternatively estimate the relative free energies of the different binding modes using the Jazinski equality.^{100,101} The resulting free energies could provide a check of the binding mode populations during our BLUES simulations, potentially providing a way to recover correct populations from partial sampling.

4.3 BLUES

The BLUES package introduced in this work as available free and open-source at <https://github.com/MobleyLab/blues>. We believe this approach shows considerable promise for enhanced sampling of protein and ligand motion and will be useful for a wide range of applications.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

DLM and SCG appreciate the financial support from the National Science Foundation (CHE 1352608) and the National Institutes of Health (1R01GM108889-01) and computing support from the UCI GreenPlanet cluster, supported in part by NSF Grant CHE-0840513. JDC appreciates support from the Sloan Kettering Institute and NIH grant P30 758 CA008748. We appreciate helpful discussions with Christopher I. Bayly (OpenEye Scientific Software) and Ioan Andricioaei (UC Irvine). We also appreciate the anonymous reviewer who suggested that our protocol work values could be useful to obtain free energy estimates, not just as part of the acceptance criterion.

References

1. Mobley DL, Klimovich PV. Perspective: Alchemical Free Energy Calculations for Drug Discovery. *J Chem Phys.* 2012; 137:230901. [PubMed: 23267463]
2. Pearce NM, Krojer T, Bradley AR, Collins P, Nowak RP, Talon R, Marsden BD, Kelm S, Shi J, Deane CM, et al. A Multi-Crystal Method for Extracting Obscured Crystallographic States from Conventionally Uninterpretable Electron Density. *Nat Commun.* 2017; 8:15123. [PubMed: 28436492]
3. Nair PC, Malde AK, Drinkwater N, Mark AE. Missing Fragments: Detecting Co-operative Binding in Fragment-Based Drug Design. *ACS Med Chem Lett.* 2012; 3:322–326. [PubMed: 24900472]
4. Fischer M, Shoichet BK, Fraser JS. One Crystal, Two Temperatures: Cryocooling Penalties Alter Ligand Binding to Transient Protein Sites. *Chembiochem.* 2015; 16:1560–1564. [PubMed: 26032594]
5. Mobley DL, Graves AP, Chodera JD, McReynolds AC, Shoichet BK, Dill KA. Predicting Absolute Ligand Binding Free Energies to a Simple Model Site. *J Mol Biol.* 2007; 371:1118–1134. [PubMed: 17599350]
6. Boyce SE, Mobley DL, Rocklin GJ, Graves AP, Dill KA, Shoichet BK. Predicting Ligand Binding Affinity with Alchemical Free Energy Methods in a Polar Model Binding Site. *J Mol Biol.* 2009; 394:747–763. [PubMed: 19782087]
7. Mobley DL, Dill KA. Binding of Small-Molecule Ligands to Proteins: “What You See” Is Not Always “What You Get”. *Structure.* 2009; 17:489–498. [PubMed: 19368882]
8. Wang K, Chodera JD, Yang Y, Shirts MR. Identifying Ligand Binding Sites and Poses Using GPU-Accelerated Hamiltonian Replica Exchange Molecular Dynamics. *J Comput Aided Mol Des.* 2013; 27:989–1007. [PubMed: 24297454]

9. Kaus JW, Harder E, Lin T, Abel R, McCammon JA, Wang L. How To Deal with Multiple Binding Poses in Alchemical Relative Protein-Ligand Binding Free Energy Calculations. *J Chem Theory Comput.* 2015; 11:2670–2679. [PubMed: 26085821]
10. Taylor RD, Jewsbury PJ, Essex JW. A Review of Protein-Small Molecule Docking Methods. *J Comput Aided Mol Des.* 2002; 16:151–166. [PubMed: 12363215]
11. Leach AR, Shoichet BK, Peishoff CE. Prediction of Protein-Ligand Interactions. Docking and Scoring: Successes and Gaps. *J Med Chem.* 2006; 49:5851–5855. [PubMed: 17004700]
12. Warren GL, Andrews CW, Capelli A-M, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, et al. A Critical Assessment of Docking Programs and Scoring Functions. *J Med Chem.* 2006; 49:5912–5931. [PubMed: 17004707]
13. Mobley DL, Liu S, Lim NM, Wymer KL, Perryman AL, Forli S, Deng N, Su J, Branson K, Olson AJ. Blind Prediction of HIV Integrase Binding from the SAMPL4 Challenge. *J Comput Aided Mol Des.* 2014; 28:327–345. [PubMed: 24595873]
14. Coleman RG, Carchia M, Sterling T, Irwin JJ, Shoichet BK. Ligand Pose and Orientational Sampling in Molecular Docking. *PLoS ONE.* 2013; 8:e75992. [PubMed: 24098414]
15. Rueda M, Abayan R. Best Practices in Docking and Activity Prediction. *bioRxiv.* 2016
16. Gathiaka S, Liu S, Chiu M, Yang H, Stuckey JA, Kang YN, Delproposto J, Kubish G, Dunbar JB, Carlson HA, et al. D3R Grand Challenge 2015: Evaluation of Protein-Ligand Pose and Affinity Predictions. *J Comput Aided Mol Des.* 2016; 30:651–668. [PubMed: 27696240]
17. Voet ARD, Kumar A, Berenger F, Zhang KYJ. Combining in Silico and in Cerebro Approaches for Virtual Screening and Pose Prediction in SAMPL4. *J Comput Aided Mol Des.* 2014; 28:363–373. [PubMed: 24446075]
18. Gallicchio E, Deng N, He P, Wickstrom L, Perryman AL, Santiago DN, Forli S, Olson AJ, Levy RM. Virtual Screening of Integrase Inhibitors by Large Scale Binding Free Energy Calculations: The SAMPL4 Challenge. *J Comput Aided Mol Des.* 2014; 28:475–490. [PubMed: 24504704]
19. Mobley DL, Chodera JD, Dill KA. Confine-and-Release Method: Obtaining Correct Binding Free Energies in the Presence of Protein Conformational Change. *J Chem Theory Comput.* 2007; 3:1231–1235. [PubMed: 18843379]
20. Mobley DL, Chodera JD, Dill KA. On the Use of Orientational Restraints and Symmetry Corrections in Alchemical Free Energy Calculations. *J Chem Phys.* 2006; 125:084902. [PubMed: 16965052]
21. Aldeghi M, Heifetz A, Bodkin MJ, Knapp S, Biggin PC. Predictions of Ligand Selectivity from Absolute Binding Free Energy Calculations. *J Am Chem Soc.* 2017; 139:946–957. [PubMed: 28009512]
22. Stjernschantz E, Oostenbrink C. Improved Ligand-Protein Binding Affinity Predictions Using Multiple Binding Modes. *Biophys J.* 2010; 98:2682–2691. [PubMed: 20513413]
23. Tembe BL, McCammon JA. Ligand Receptor Interactions. *Comput Chem.* 1984; 8:281–283.
24. Gilson M, Given J, Bush B, McCammon J. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys J.* 1997; 72:1047–1069. [PubMed: 9138555]
25. Kollman P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem Rev.* 1993; 93:2395–2417.
26. Gallicchio E, Lapelosa M, Levy RM. Binding Energy Distribution Analysis Method (BEDAM) for Estimation of Protein-Ligand Binding Affinities. *J Chem Theory Comput.* 2010; 6:2961–2977. [PubMed: 21116484]
27. Jayachandran G, Shirts MR, Park S, Pande VS. Parallelized-over-Parts Computation of Absolute Binding Free Energy with Docking and Molecular Dynamics. *J Chem Phys.* 2006; 125:084901. [PubMed: 16965051]
28. Georgiou C, McNae I, Wear M, Ioannidis H, Michel J, Walkinshaw M. Pushing the Limits of Detection of Weak Binding Using Fragment-Based Drug Discovery: Identification of New Cyclophilin Binders. *J Mol Biol.* 2017; 429:2556–2570. [PubMed: 28673552]
29. Rees DC, Congreve M, Murray CW, Carr R. Fragment-Based Lead Discovery. *Nat Rev Drug Discov.* 2004; 3:660–672. [PubMed: 15286733]
30. Murray CW, Verdonk ML, Rees DC. Experiences in Fragment-Based Drug Discovery. *Trends Pharmacol Sci.* 2012; 33:224–232. [PubMed: 22459076]

31. Clark AJ, Tiwary P, Borrelli K, Feng S, Miller EB, Abel R, Friesner RA, Berne BJ. Prediction of Protein-Ligand Binding Poses via a Combination of Induced Fit Docking and Metadynamics Simulations. *J Chem Theory Comput.* 2016; 12:2990–2998. [PubMed: 27145262]
32. Heinzelmann G, Henriksen NM, Gilson MK. Attach-Pull-Release Calculations of Ligand Binding and Conformational Changes on the First BRD4 Bromodomain. *J Chem Theory Comput.* 2017; 13:3260–3275. [PubMed: 28564537]
33. Sherborne B, Shanmugasundaram V, Cheng AC, Christ CD, DesJarlais RL, Duca JS, Lewis RA, Loughney DA, Manas ES, McGaughey GB, et al. Collaborating to Improve the Use of Free-Energy and Other Quantitative Methods in Drug Discovery. *J Comput Aided Mol Des.* 2016; 30:1139–1141. [PubMed: 28013427]
34. Shan Y, Kim ET, Eastwood MP, Dror RO, Seeliger MA, Shaw DE. How Does a Drug Molecule Find Its Target Binding Site? *J Am Chem Soc.* 2011; 133:9181–9183. [PubMed: 21545110]
35. Prinz J-H, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schütte C, Noé F. Markov Models of Molecular Kinetics: Generation and Validation. *J Chem Phys.* 2011; 134:174105. [PubMed: 21548671]
36. Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR. Constructing the Equilibrium Ensemble of Folding Pathways from Short Off-Equilibrium Simulations. *Proc Natl Acad Sci USA.* 2009; 106:19011–19016. [PubMed: 19887634]
37. Pande VS, Beauchamp K, Bowman GR. Everything You Wanted to Know about Markov State Models but Were Afraid to Ask. *Methods.* 2010; 52:99–105. [PubMed: 20570730]
38. Wu H, Mey ASJS, Rosta E, Noé F. Statistically Optimal Analysis of State-Discretized Trajectory Data from Multiple Thermodynamic States. *J Chem Phys.* 2014; 141:214106. [PubMed: 25481128]
39. Chodera JD, Noé F. Markov State Models of Biomolecular Conformational Dynamics. *Curr Opin Struct Biol.* 2014; 25:135–144. [PubMed: 24836551]
40. McGibbon RT, Pande VS. Variational Cross-Validation of Slow Dynamical Modes in Molecular Kinetics. *J Chem Phys.* 2015; 142:124105. [PubMed: 25833563]
41. Pérez-Hernández G, Paul F, Giorgino T, De Fabritiis G, Noé F. Identification of Slow Molecular Order Parameters for Markov Model Construction. *J Chem Phys.* 2013; 139:015102. [PubMed: 23822324]
42. Schwantes CR, Pande VS. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J Chem Theory Comput.* 2013; 9:2000–2009. [PubMed: 23750122]
43. Lelièvre, T., Rousset, M., Stoltz, G. *Free Energy Computations: A Mathematical Perspective.* Imperial College Press; 2010. p. 149–258.
44. Nilmeier JP, Crooks GE, Minh DDL, Chodera JD. Nonequilibrium Candidate Monte Carlo Is an Efficient Tool for Equilibrium Simulation. *Proc Natl Acad Sci USA.* 2011; 108:E1009–E1018. [PubMed: 22025687]
45. Jorgensen WL, Tirado-Rives J. Molecular Modeling of Organic and Biomolecular Systems Using BOSS and MCPRO. *J Comput Chem.* 2005; 26:1689–1700. [PubMed: 16200637]
46. Michel J, Taylor RD, Essex JW. Efficient Generalized Born Models for Monte Carlo Simulations. *J Chem Theory Comput.* 2006; 2:732–739. [PubMed: 26626678]
47. Marinari E, Parisi G. Simulated Tempering: A New Monte Carlo Scheme. *EPL.* 1992; 19:451–458.
48. Sugita Y, Okamoto Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem Phys Lett.* 1999; 314:141–151.
49. Duane S, Kennedy AD, Pendleton BJ, Roweth D. Hybrid Monte Carlo. *Phys Lett B.* 1987; 195:216–222.
50. Chodera, JD. Yank. <https://github.com/choderalab/yank>
51. Kurut A, Fonseca R, Boomsma W. Driving Structural Transitions in Molecular Simulations Using the Nonequilibrium Candidate Monte Carlo. *J Phys Chem B.* 2018; 122:1195–1204. [PubMed: 29260565]
52. Giovannelli E, Gellini C, Pietraperzia G, Cardini G, Chelli R. Nonequilibrium Candidate Monte Carlo Simulations with Configurational Freezing Schemes. *J Chem Theory Comput.* 2014; 10:4273–4283. [PubMed: 26588124]

53. Chen Y, Roux B. Constant-pH Hybrid Nonequilibrium Molecular Dynamics-Monte Carlo Simulation Method. *J Chem Theory Comput.* 2015; 11:3919–3931. [PubMed: 26300709]
54. Chen Y, Roux B. Enhanced Sampling of an Atomic Model with Hybrid Nonequilibrium Molecular Dynamics-Monte Carlo Simulations Guided by a Coarse-Grained Model. *J Chem Theory Comput.* 2015; 11:3572–3583. [PubMed: 26574442]
55. Chen Y, Roux B. Generalized Metropolis Acceptance Criterion for Hybrid Non-Equilibrium Molecular Dynamics-Monte Carlo Simulations. *J Chem Phys.* 2015; 142:024101. [PubMed: 25591332]
56. Chen Y, Roux B. Efficient Hybrid Non-Equilibrium Molecular Dynamics - Monte Carlo Simulations with Symmetric Momentum Reversal. *J Chem Phys.* 2014; 141:114107. [PubMed: 25240345]
57. Ross G, Rustenburg AS, Grinaway PB, Fass J, Chodera JD. Biomolecular Simulations under Realistic Macroscopic Salt Conditions. *bioRxiv.* :226001.
58. Sivak DA, Chodera JD, Crooks GE. Using Nonequilibrium Fluctuation Theorems to Understand and Correct Errors in Equilibrium and Nonequilibrium Simulations of Discrete Langevin Dynamics. *Phys Rev X.* 2013; 3:011007.
59. Leimkuhler B, Matthews C. Rational Construction of Stochastic Numerical Methods for Molecular Sampling. *Appl Math Res Express.* 2013; 2013:34–56.
60. Leimkuhler B, Matthews C. Robust and Efficient Configurational Molecular Sampling via Langevin Dynamics. *J Chem Phys.* 2013; 138:174102. [PubMed: 23656109]
61. Fass, Josh, Sivak, DA., Crooks, GE., Beauchamp, KA., Leimkuhler, B., Chodera, JD. Quantifying Configuration-Sampling Error in Langevin Simulations of Complex Molecular Systems. *bioRxiv.*
62. Wagoner JA, Pande VS. Reducing the Effect of Metropolization on Mixing Times in Molecular Dynamics Simulations. *J Chem Phys.* 2012; 137:214105. [PubMed: 23231215]
63. Sohl-Dickstein J, Mudigonda M, De-Weese MR. Hamiltonian Monte Carlo Without Detailed Balance. *Proceedings of the 31st International Conference on International Conference on Machine Learning.* 2014; 32:I-719–I-726.
64. Campos CM, Sanz-Serna J. Extra Chance Generalized Hybrid Monte Carlo. *J Comput Phys.* 2015; 281:365–374.
65. Hastings WK. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika.* 1970; 57:97–109.
66. Horowitz A. A Generalized Guided Monte Carlo Algorithm. *Phys Lett B.* 1991; 268:247–252.
67. Kennedy AD, Edwards R, Mino H, Pendleton B. Tuning the Generalized Hybrid Monte Carlo Algorithm. *Nucl Phys B Proc Suppl.* 1996; 47:781–784.
68. Mobley DL, Gilson MK. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annu Rev Biophys.* 2017; 46:531–558. [PubMed: 28399632]
69. Evoli S, Mobley DL, Guzzi R, Rizzuti B. Multiple Binding Modes of Ibuprofen in Human Serum Albumin Identified by Absolute Binding Free Energy Calculations. *Phys Chem Chem Phys.* 2016; 18:32358–32368. [PubMed: 27854368]
70. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber Biomolecular Simulation Programs. *J Comput Chem.* 2005; 26:1668–1688. [PubMed: 16200636]
71. Case, DA., Babin, V., Berryman, JT., Betz, RM., Cai, Q., Cerutti, DS., Cheatham, TE., Darden, TA., Duke, RE., Gohlke, H., et al. *Amber*. University of California; San Francisco: 2014. p. 14
72. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and Testing of a General Amber Force Field. *J Comput Chem.* 2004; 25:1157–1174. [PubMed: 15116359]
73. Jakalian A, Jack DB, Bayly CI. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J Comput Chem.* 2002; 23:1623–1641. [PubMed: 12395429]
74. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins.* 2010; 78:1950–1958. [PubMed: 20408171]

75. da Silva AWS, Vranken WF. ACPYPE-Antechamber Python Parser Interface. BMC research notes. 2012; 5:367. [PubMed: 22824207]
76. Darden T, York D, Pedersen L. Particle Mesh Ewald: An $N\log(N)$ Method for Ewald Sums in Large Systems. J Chem Phys. 1993; 98:10089–10092.
77. Eastman P, Friedrichs MS, Chodera JD, Radmer RJ, Bruns CM, Ku JP, Beauchamp KA, Lane TJ, Wang L-P, Shukla D, et al. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. J Chem Theory Comput. 2013; 9:461–469. [PubMed: 23316124]
78. Boresch S, Tettinger F, Leitgeb M, Karplus M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. J Phys Chem B. 2003; 107:9535–9551.
79. Berendsen HJC, Van Der Spoel D, van Drunen R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. Comput Phys Comm. 1995; 91:43–56.
80. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. SoftwareX. 2015; (1–2):19–25.
81. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A Linear Constraint Solver for Molecular Simulations. J Comput Chem. 1997; 18:1463–1472.
82. Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, Wehmeyer C, Prinz J-H, Noé F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. J Chem Theory Comput. 2015; 11:5525–5542. [PubMed: 26574340]
83. Schutte C, Fischer A, Huisinga W, Deuffhard P. A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo. J Comput Phys. 1999; 151:146–168.
84. Deuffhard P, Weber M. Robust Perron Cluster Analysis in Conformation Dynamics. Linear Algebra Appl. 2005; 398:161–184.
85. Röblitz S, Weber M. Fuzzy Spectral Clustering by PCCA+: Application to Markov State Models and Data Classification. Adv Data Anal Classif. 2013; 7:147–179.
86. Beutler TC, Mark AE, van Schaik RC, Gerber PR, van Gunsteren WF. Avoiding Singularities and Numerical Instabilities in Free Energy Calculations Based on Molecular Simulations. Chem Phys Lett. 1994; 222:529–539.
87. Wang J, Deng Y, Roux B. Absolute Binding Free Energy Calculations Using Molecular Dynamics Simulations with Restraining Potentials. Biophys J. 2006; 91:2798–2814. [PubMed: 16844742]
88. McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, Schwantes CR, Wang L-P, Lane TJ, Pande VS. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. Biophys J. 2015; 109:1528–1532. [PubMed: 26488642]
89. Chodera, JD., Rizzi, A. Openmmtools. <https://github.com/choderalab/openmmtools/blob/master/openmmtools/>
90. Grossfield A, Zuckerman DM. Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations. Annu Rep Comput Chem. 2009; 5:23–48. [PubMed: 20454547]
91. Flyvbjerg H, Petersen HG. Error Estimates on Averages of Correlated Data. J Chem Phys. 1989; 91:461–466.
92. Lu N, Kofke DA. Accuracy of Free-Energy Perturbation Calculations in Molecular Simulation. I. Modeling. J Chem Phys. 2001; 114:7303–7311.
93. Besl PJ, McKay ND. A Method for Registration of 3-D Shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1992; 14:239–256.
94. Wang L, Berne BJ, Friesner RA. On Achieving High Accuracy and Reliability in the Calculation of Relative Protein-Ligand Binding Affinities. Proc Natl Acad Sci USA. 2012; 109:1937–1942. [PubMed: 22308365]
95. Morton A, Baase WA, Matthews BW. Energetic Origins of Specificity of Ligand Binding in an Interior Nonpolar Cavity of T4 Lysozyme. Biochemistry. 1995; 34:8564–8575. [PubMed: 7612598]
96. Hajduk PJ, Greer J. A Decade of Fragment-Based Drug Design: Strategic Advances and Lessons Learned. Nat Rev Drug Discov. 2007; 6:211–219. [PubMed: 17290284]

97. Murray CW, Rees DC. The Rise of Fragment-Based Drug Discovery. *Nat Chem*. 2009; 1:187–192. [PubMed: 21378847]
98. Erlanson DA, McDowell RS, O'Brien T. Fragment-Based Drug Discovery. *J Med Chem*. 2004; 47:3463–3482. [PubMed: 15214773]
99. Andricioaei I, Straub JE, Voter AF. Smart Darting Monte Carlo. *J Chem Phys*. 2001; 114:6994–7000.
100. Jarzynski C. Equilibrium Free-Energy Differences from Nonequilibrium Measurements: A Master-Equation Approach. *Phys Rev E*. 1997; 56:5018.
101. Hendrix DA, Jarzynski C. A “Fast Growth” Method of Computing Free Energy Differences. *J Chem Phys*. 2001; 114:5974–5981.

Figure 1. Potential free energy efficiency gains using binding mode populations

(A) shows calculations of M different effective binding free energy values ΔG_i° for each different metastable binding mode of a ligand in a receptor; these effective binding free energies can be rigorously combined to recover the total binding free energy,²⁰ with a total computational cost (C) of MNx where N is the number of intermediate alchemical states used and x is the length of the simulation used at each alchemical state. Alternatively, (B) shows how, if relative populations (p_i) of different metastable binding modes can be recovered from end state simulations (colored circles, top; each circle represents an amount of simulation time spent in the binding mode, so the populations can be determined from counting time in each mode, with binding modes separated by clustering techniques or any reasonable decomposition of state space¹), then the full binding free energy can be recovered from the calculation of a single effective binding free energy (here, ΔG_3° is selected for convenience) and the populations of the different binding modes. This approach has a computational cost (C) of $Nx + y$, where y is the cost of determining the binding mode populations.

Figure 2. NCMC moves for ligand binding modes

The blue circles represent the atoms in the binding site, black circles represent the fully interacting ligand, white circles represent the fully non-interacting ligand, and gray circles indicate intermediate levels of interaction. A) The ligand is fully interacting in the binding site. B) The ligand's interactions are partially off, allowing the protein to modestly relax the binding site. C) The ligand's interactions are fully turned off. D) The ligand is randomly rotated around its center of mass; its interactions remain off. E) The ligand's interactions are partially turned on and the propagation steps of NCMC allow relaxation of the rotated binding mode to resolve clashes. F) At the end of the NCMC protocol the ligand is again fully interacting in a new orientation. The NCMC move is then accepted or rejected based on the work performed via Equation 4.

Figure 3. Lambda scaling over the course of our NCMC steps

The ligand's electrostatic interactions are first turned off, followed by the sterics, until the halfway point. The interactions are then turned on in reverse order. This protocol resembles what is typically done for efficient alchemical free energy calculations, such as binding free energy calculations. In particular, the electrostatics are the first to turn off and the last to turn on because having electrostatic interactions present without first turning off the steric interactions can lead to numerical instabilities.⁸⁷

Figure 4. Acceptance probability for toluene as a function of the amount of NCMC relaxation
The acceptance probability—also referred to as the acceptance rate—is shown on a log scale as a function of the total number of NCMC switching steps per cycle, for toluene in the L99A site of T4 lysozyme. The acceptance probability increases dramatically up to 10000 NCMC switching steps per cycle, then increases more slowly, so much of our work in this study focuses on comparing efficiency with other approaches at 10000 steps per cycle. The red dashed line marks the acceptance probability of instantaneous conventional MC rotations without an accompanying NCMC protocol, *i.e.* standard MC for ligand rotation. Error bars are the standard error in the acceptance rate (Sec. 2.5.3).

Figure 5. Order parameter used for identifying binding modes of toluene

The dihedral angle order parameter which we monitor to differentiate toluene's binding modes is defined by the alpha carbon of ARG118 and the C1, C5, and C7 toluene atoms, shown in orange in CPK representation. In the image, the atoms involved in the dihedral are connected by a purple line, and the dihedral angle measures rotation around the central dashed purple line. The protein is shown in a blue ribbon representation, and toluene is shown in cyan.

Figure 6. Implied timescales of binding mode transitions

The implied timescales shown here were calculated from an MSM utilizing all of our MD simulations of toluene in T4 lysozyme L99A. The black line denotes when the lagtime is equal to the implied timescale; timescales below this line (gray) have already relaxed and cannot be estimated accurately; shown here are the 10 slowest implied timescales. Overall, this shows that the slowest timescale in this system (in this case the out-of-plane flip of the ring) has an implied timescale of roughly 100 ns.

Figure 7. Toluene binding modes

Toluene exhibits four binding modes. The toluene molecule shown in orange corresponds to the crystallographic binding mode, while toluene in blue corresponds to another binding mode. The other two binding modes come about from the symmetric equivalents of these two binding modes, where the molecule is flipped in the plane of the ring.

Figure 8. Toluene binding mode populations from a long trajectory

(a) Dihedral angle (corresponding to binding modes) observed in the initial long trajectory as a function of simulation time (see Sec 2.7). (b) A histogram plot of the selected dihedral order parameter computed from the trajectory (as shown in Figure 5). Labels A1 and A2 correspond to the two different, but symmetry-equivalent populations of the more favorable binding mode; B1 and B2 to the symmetry-equivalent populations of the less favorable binding mode. The legend shows computed populations; with enough simulation time the symmetric binding modes should have equivalent populations, which is not the case after over 800 ns of simulation, partly because out-of-plane flips between symmetry equivalent modes are so rarely observed. Thus, A2 and B2 end up underpopulated relative to their symmetry equivalent partners A1 and B1.

Figure 9. Binding mode sampling of toluene in T4 lysozyme with various methods over 5000 iterations

This compares the performance of various methods for sampling the four binding modes of toluene in T4 lysozyme over a comparable number of iterations; each iteration corresponds to the same number of energy evaluations (20000) for each method. The dihedral angle plotted (on the vertical axis in the left column) discriminates between binding modes. (A,C,E) The dihedral order parameter plotted as a function of iteration number (loosely, simulation time). The slow out-of-plane flip of toluene results in a transition between the top two states and the bottom two states; relatively few such transitions can be seen in (A) and

(C), though more can be seen in (E). (*B,D,F*) Histogram plots of dihedral angles observed in the trajectories, colored by binding mode. Each binding mode's fractional population is indicated in the legend. Labels A1 and A2 correspond to the two different, but symmetry-equivalent populations of the more favorable binding mode; B1 and B2 to the symmetry-equivalent populations of the less favorable binding mode. (*A,B*) MD sampling of toluene in T4 lysozyme. (*C,D*) MC with MD sampling of toluene in T4 lysozyme. (*E,F*) NCMC with MD sampling of toluene in T4 lysozyme. Overall, the MD/NCMC approach leads to dramatically faster transitions between binding modes and apparently better converged populations; for example, the symmetry-equivalent A1–A2 pair has dramatically different populations in (B), as does the B1–B2 pair.

Figure 10. Convergence of binding mode populations for toluene

Shown are binding mode populations over 5000 iterations (200ns) for toluene in T4 lysozyme L99A. Labels A1 and A2 correspond to the two different, but symmetry-equivalent populations of the more favorable binding mode; each should converge to 0.30, marked by the dashed blue line. Labels B1 and B2 correspond to the populations of the less favorable binding mode; each should converge to 0.20, marked by the dashed red line. The populations computed by BLUES are within uncertainty of the true result well before 10% of the total simulation time, whereas with MD and MC the populations are not until much later if at all.

(a) MD

(b) MD/NCMC

Figure 11. A model of the convergence of binding mode populations for toluene in T4 lysozyme L99A

The transition matrices from the MSM and MD/NCMC simulation were used to estimate populations as a function of time for a hypothetical simulation starting in state A1. We ran 1000 trials in each case. For each trial we propagated the transition matrix by selecting a new state to transition to at each timestep with probabilities given by the transition matrix as described in the text. Heavy lines show the mean population estimated over the trials, and the lighter shaded regions give the standard deviation over trials, indicating the region within which a typical single simulation would usually fall. Vertical bars denote the point at which the standard deviation of each estimated population first falls below 5%. (a) The statistical model estimated from the MSM which shows that it takes approximately 12000 ns for the standard deviation in the slowest converging population to drop below 5%. (b) For the MD/NCMC simulation, it takes approximately 60 ns for the standard deviation in the slowest converging population to drop below 5%. In both cases, because the transition matrices were

estimated from relatively short simulations, the populations converge to a steady state but have some error due to the underlying transition matrices. Together, (a) and (b) demonstrate that MD/NCMC results in dramatically faster convergence of populations as a function of simulation time compared to MD alone.

Figure 12. Binding mode transitions for toluene

Shown is the transition matrix counting the number of transitions between binding modes for toluene in T4 lysozyme L99A over 5000 iterations (200ns), for the different sampling methods. Labels A1 and A2 correspond to the more favorable binding mode; B1 and B2 to the less favorable binding mode. A1 and A2 comprise a symmetry-equivalent pair, as do B1 and B2, but to transition between states in a symmetry-equivalent pair (A1 to A2, or B1 to B2) requires an out-of-plane flip. Transition counts to the same binding mode (the main diagonal of the matrix) are omitted for clarity. Here, hotter colors are better as they indicate more transitions between binding modes. (a) Transitions of the MD simulation. The total number of transitions is 242. (b) Transitions of the MD/MC simulation. The total number of transitions is 230. (c) Transitions of the MD/NCMC simulation. The total number of transitions is 497. Here, it can be seen that in the MD case, only the A2 to B2 and B2 to A2 cases have more than 30 transitions, because the simulation mostly remained stuck in these two states without flipping out-of-plane (Figure 9) and a similar effect happened in the

MD/MC case but for A1 to B1. In contrast, in the NCMC case, all transitions occur more than 30 times because out-of-plane transitions are also relatively frequent.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 13. Rotational distribution of accepted moves for toluene and 3-iodotoluene in T4 lysozyme

Shown are the probabilities of accepting rotational moves of a particular size with standard Monte Carlo and with NCMC, for toluene (top) and the bulkier iodotoluene (bottom). Results come from 10000 MC iterations of 10 attempts each (a and c) or 10000 NCMC iterations (b and d). Iodotoluene is substantially bulkier compared to toluene, and it is difficult to rotate it in the binding site without at least some amount of relaxation, so the acceptance rate for MC moves is lower (Sec. 3.2.3) and the number of *significant* rotations is

dramatically lower (c), with virtually no rotations larger than 22.5 degrees observed; for panel (c) we use a log scale so it is clear that *some* significant rotations were observed.

Figure 14. Acceptance probability for 3-iodotoluene as a function of the amount of NCMC relaxation

The acceptance rate shown here is analogous to that in Figure 4 except that this test uses a fixed set of MD snapshots as a basis for move proposals, as described in the text. Overall acceptance (black line) increases dramatically up to 10000 NCMC switching steps per cycle, then increases more slowly. The black dashed line marks the acceptance probability of instantaneous MC rotations, given the same set of MD snapshots as starting points. The solid blue line denotes the acceptance probability of *substantial* rotations, those larger than 45 degrees, and the dashed blue line indicates the overall acceptance probability of instantaneous MC rotations from the same set of snapshots. Thus, NCMC does only modestly worse at sampling substantial rearrangements than sampling all rearrangements, whereas MC has orders of magnitude lower acceptance of substantial rearrangements.