Estimation of binding rates and affinities from multiensemble Markov models and ligand decoupling

Yunhui Ge[1, a)] and Vincent A. Voelz[2, b)]

[1)]*Department of Pharmaceutical Sciences, University of California, Irvine, CA 92697, USA*

[2)]*Department of Chemistry, Temple University, Philadelphia, PA 19122, USA*

Accurate and efficient simulation of the thermodynamics and kinetics of protein-ligand interactions is crucial for computational drug discovery. Multiensemble Markov Model (MEMM) estimators can provide estimates of both binding rates and affinities from collections of short trajectories, but have not been systematically explored for situations when a ligand is decoupled through scaling of non-bonded interactions. In this work, we compare the performance of two MEMM approaches for estimating ligand binding affinities and rates: (1) the transition-based reweighting analysis method (TRAM) and (2) a Maximum Caliber (MaxCal) based method. As a test system, we construct a small host-guest system where the ligand is a single uncharged Lennard-Jones (LJ) particle, and the receptor is an 11-particle icosahedral pocket made from the same atom type. To realistically mimic a protein-ligand binding system, the LJ ϵ parameter was tuned, and the system placed in a periodic box with 860 TIP3P water molecules. A benchmark was performed using over 80 μs of unbiased simulation, and an 18-state Markov state model used to estimate reference binding affinities and rates. We then tested the performance of TRAM and MaxCal when challenged with limited data. Both TRAM and MaxCal approaches perform better than conventional MSMs, with TRAM showing better convergence and accuracy. We find that subsampling of trajectories to remove time correlation improves the accuracy of both TRAM and MaxCal, and that in most cases only a single biased ensemble to enhance sampled transitions is required to make accurate estimates.

Keywords: Markov Model, molecular kinetics, multiensemble, Maximum Caliber, TRAM, MBAR, free energy estimators

INTRODUCTION

Estimating both the thermodynamics and kinetics of protein-ligand binding is essential for understanding biological function and for the rational design of therapeutics. In the last decade, alchemical free energy perturbation—a technique which relies on sampling from multiple thermodynamic ensembles—has emerged as the main tool to accurately estimate binding thermodynamics from molecular simulations.[1–8] Along with the growing recognition of the importance of binding rates in drug discovery (in particular, the residence time[9–11]), there has also been increased interest in estimating the kinetics of protein-ligand binding using molecular simulation. Although different methods to estimate on- and off-rates of ligand-binding have been developed,[12–21] most of them estimate rates separately from (or at the expense of) thermodynamics.

Markov State Models (MSMs) of molecular kinetics, which describe conformational dynamics as a network of transitions between metastable states,[22–26] can provide combined estimates of the thermodynamics and kinetics of ligand binding from large ensembles of short trajectories.[27–31] In practice, however, estimates of slow dissociation rates are limited by the sampling of rare

events, because a key assumption is that the trajectories are sampled at equilibrium.[32,33] Although rare events can be sampled in ultra-long trajectories using specialized hardware,[34] a more general approach would be desired.

A great improvement to this situation has come from the introduction of so-called multi-ensemble Markov Model (MEMM) estimators, which use trajectory data collected from multiple thermodynamic ensembles to make MSM estimates of thermodynamics and kinetics.[35–39] The essential idea, as applied to ligand binding, is to collect energy snapshots and metastable state transitions in biased thermodynamic ensembles that are *not* limited by rare event sampling, in order to make more statistically significant estimates of rates and affinities for the unbiased ensemble. One of these estimators, the transition-based reweighting analysis method (TRAM) of Wu et al.,[37] has been used to estimate the slow dissociation of small-molecule and peptide ligands using harmonic bias potentials.[40,41] Less explored has been the use of TRAM with ensembles of scaled nonbonded protein-ligand interactions, as is common in alchemical free energy methods.

Another MSM approach that utilizes multiple thermodynamic ensembles to infer binding affinities and rates is the Maximium Caliber (MaxCal) based method described in Wan et al.[42] In this approach, the principle of maximum path entropy (caliber) is used to infer changes in inter-state transition rates from changes in the equilibrium populations.[43–49] Like TRAM, the MaxCal approach relies on the ability to collect many transitions

————————

a)Electronic mail: yunhuig2@uci.edu
b)Electronic mail: voelz@temple.edu

in a biased thermodynamic ensemble, enabling more statistically significant estimation of transition rates for an unbiased ensemble. As with TRAM, the use of MaxCal with ensembles where non-bonded protein-ligand interactions are scaled has not been systematically explored.

In this manuscript, our goal is to test the performance of TRAM and MaxCal vs. conventional MSM approaches in estimating ligand binding rates and affinities, for a set of thermodynamic ensembles with decoupled ligand interactions, when challenged with limited sampling. To do this, we first constructed a toy host-guest system that realistically mimics protein-ligand association, and performed thorough sampling to obtain a reference benchmark for comparison. We then test the performance of TRAM and MaxCal estimators using as input a collection of short trajectories sampled in each thermodynamic ensemble.

I. METHODS

A. Simulation system

In our toy model of ligand-receptor binding, the receptor is an 11-particle icosahedral binding pocket composed of atoms with the same Lennard-Jones (LJ) parameters as a CT carbon atom in the AMBER force fields: $\sigma = 0.339967$ nm, $\epsilon = 0.457730$ kJ mol^{-1} (Figure 1). Equilibrium bond lengths, bond force constants, and dihedral potentials were kept the same as a CT-CT AMBER bond type, with a soft harmonic angle potential ($\theta_0 = 180.0°, k_\theta = 150.0$ kJ mol^{-1} rad^{-2}) to enforce the icosahedral shape. The ligand is a single uncharged LJ particle of the same type, where we varied the ϵ parameter from 0.0 to 10.0 kJ mol^{-1} in increments of 0.5 kJ mol^{-1}.

All simulations used the GROMACS 5.1.4 molecular dynamics package.[50] The system was solvated in a cubic periodic box and equilibrated in the NPT ensemble at 300 K and 1 atm using a Berendsen thermostat with 860 TIP3P waters, two Na$^+$ ions, and two Cl$^+$ ions to determine the box volume $(3.00859$ nm$)^3$ to be used for production runs in the NVT ensemble. For the production run, the stochastic dynamics thermostat was used to control the temperature (300 K) in the simulations with a 2 fs time step, hydrogen bonds constrained using LINCS, and Particle Mesh Ewald (PME) electrostatics. System topologies and simulation parameters can be found at https://github.com/yunhuige/toy_model_paper.

To choose the value of ϵ used in all subsequent tests, we simulated 10 trajectories of length 1 μs for each of the 21 different values of ϵ, resulting in 210 μs of aggregate trajectory data. Snapshots were recorded every 100 ps. After inspecting the apparent residence times of the ligand for each of these simulations (Figure 2), we chose an ϵ value of 2.5 kJ mol^{-1} for our reference system, resulting in a bound-state population of $\pi_{bound} = 81.9\%$ and residence times in the range of tens of nanoseconds.

After establishing the value of ϵ to use in our reference system, we performed further sampling to generate an exhaustive set of trajectory data to be used for mak-
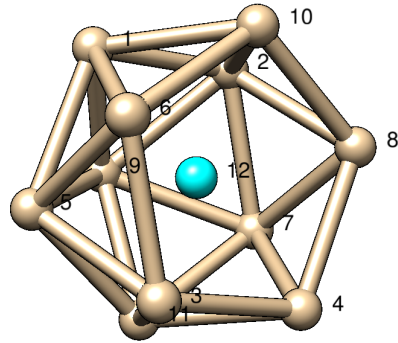

FIG. 1. A toy host-guest system that mimics protein-ligand binding. The receptor is an icosahedral binding pocket of Lennard-Jones (LJ) particles (beige) that binds a single LJ particle ligand (cyan) in its central cavity. This system is solvated in a periodic box with counterions. Particles are labeled with their atom indices.
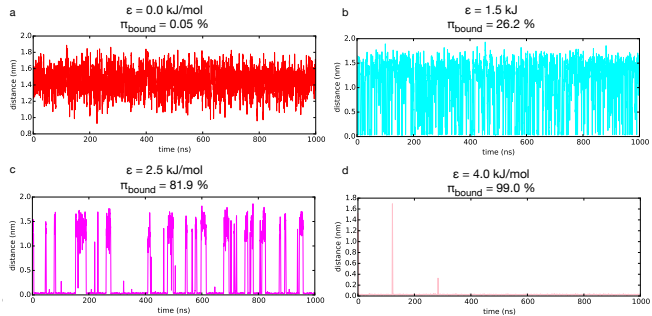

FIG. 2. Trajectory traces of ligand distance from the center of the binding pocket, shown for selected ϵ values: (a) 0.0 kJ mol^{-1}, (b) 1.5 kJ mol^{-1}, (c) 2.5 kJ mol^{-1}, and (d) 4.0 kJ mol^{-1}, labeled with the apparent bound-state population π_{bound} in each 1-μs trajectory. The value $\epsilon = 2.5$ kJ mol^{-1} was chosen for all subsequent tests,

ing high-quality reference estimates of binding rates and affinities. Twenty trajectories were generated, each of length 4 μs, for an aggregate of 80 μs. Coordinates were saved at a frequency of 1 ps to preserve kinetic information at high temporal resolution.

1. Ligand decoupling simulations

The free energy perturbation (FEP) functionality of GROMACS was used to perform simulations in which the ligand was decoupled by scaling non-bonded interactions (in this case, van der Waals interactions, because the ligand is uncharged). Separate simulations were performed for scaling constants $\lambda = 0.5, 0.6, 0.7, 0.8, 0.9$, and 1.0, where $\lambda = 1.0$ represents the unbiased thermodynamic ensemble with full vdW interactions. Two starting configurations were used to initialize the simulations: one where the ligand is in the bound state, and another where the ligand is unbound in solution. For each of the

two starting configurations, trajectories of length 400 ns were generated, with snapshots recorded every 1 ps. Soft-core interactions were used for non-bonded interactions (`sc-alpha = 0.5`, `sc-power = 1`, and `sc-sigma = 0.3`). In total, 4.8 μs ($6 \times 2 \times 400$ ns) of trajectory data were collected—800 ns data for each λ value. Velocity-verlet integration was performed using a velocity-rescaling thermostat.

B. Multiensemble dynamical estimators

1. Conventional Markov State Models

PyEMMA software package[51] was used to construct conventional MSMs. This process involves first assigning trajectory snapshots to discrete metastable states. Here, the toy binding system is simple enough that we can manually assign metastable states. From inspection of the distribution of sampled conformations (Figure 3), we identify 18 metastable states that partition conformational space: a *bound* state (0) where the ligand is strongly bound inside the receptor, *encounter-complex* states (1-10) where the ligand is weakly bound to one of the faces of the icosahedron, a state where the ligand associates with *entrance* of the binding site (16), and an *unbound* state (17).

State indices i are assigned based on the ligand position vector $\vec{x}$ where the origin is the center of the binding site (computed as the mean position of the receptor atoms, excluding atom 9 in Figure 1), and its distance from the origin $|\vec{x}|$. If $|\vec{x}| > 0.85$ nm, it is considered unbound (state 17); if $|\vec{x}| < 0.2$ nm it is considered bound (state 0). Otherwise, it is assigned to the state i having the smallest angular distance to the vectors $\vec{f}_i$ normal to each isocahedral face (the entrance is also considered a face): $\arg\max_{i}[(\vec{x}/|\vec{x}|) \cdot (\vec{f}_i/|\vec{f}_i|)]$ (Figure 4).

We also examined a two-state decomposition bound vs. unbound states, defined by a radial cutoff of 0.2 nm (see Results). This coarse grained state decomposition provided an opportunity to test state decomposition effect on kinetic and thermodynamic predictions.

Using the state assignments described above, a conventional MSM is constructed by first compiling the matrix of the observed transition counts c_{ij}^{τ}, the number of times the system is in state i at time t and in state j at time $t + \tau$) for a suitable lag time τ. To infer the matrix of transition probabilities p_{ij}, we use the standard maximum-likelihood estimator (MLE) enforcing a constraint on detailed balance. The transition matrix contains full information about the thermodynamics and kinetics. Equilibrium populations π_i are obtained as the stationary eigenvector of the transition matrix. Inverse mean first passage times between bound and unbound states are used to compute binding (k_{on}) and unbinding (k_{off}) rates.
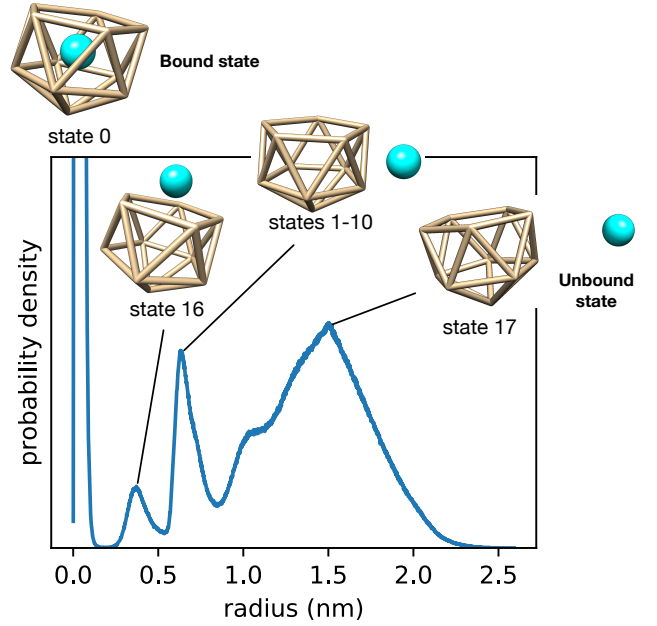

FIG. 3. The sampled distribution of the distance between the ligand and the center of mass of the binding pocket. Peaks in the probability density correspond to particular metastable states, representative snapshots of which are shown.
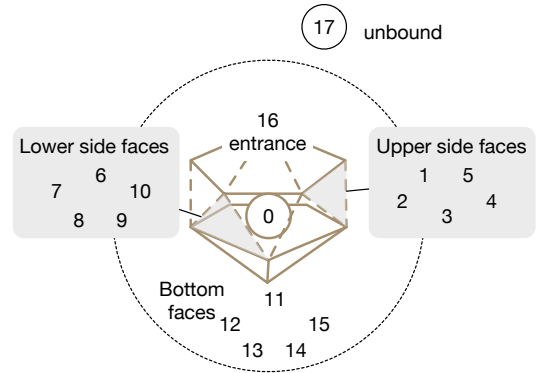

FIG. 4. Metastable state indices used in the construction of an 18-state MSM.

2. Transition-based reweighting analysis method (TRAM)

The TRAM estimator[37] as implemented in the PyEMMA software package[51] was used to construct multiensemble Markov Models from the simulation trajectory data. TRAM attempts to simultaneously optimize (1) the transitions p_{ij} between state i and j for given free energies f_i^k for all conformational states i at each thermodynamic ensemble k and (2) the free energies f_i^k for all thermodynamic ensembles at each conformational state i. This is achieved by maximizing a joint likelihood function L_{TRAM} that is the product of a reversible MSM estimator likelihood function $L_{MSM}^{(k)}$ and a free energy estimator likelihood function $L_{LEQ}^{(k)}$:

$$L_{\mathrm{TRAM}} = \prod_{k=1}^{K} \underbrace{\left(\prod_{i,j} \left(p_{ij}^{(k)}\right)^{c_{ij}^{(k)}}\right)}_{L_{\mathrm{MSM}}^{(k)}} \underbrace{\left(\prod_{i=1}^{m} \prod_{x \in X_i^{(k)}} \mu(x) e^{f_i^{(k)} - b^{(k)}(x)}\right)}_{L_{\mathrm{LEQ}}^{(k)}}$$

(1)

where $p_{ij}^{(k)}$ is the transition probability between state i and j given observed transition counts $c_{ij}^{(k)}$ in ensemble k, $\mu(x)$ is the normalized equilibrium distribution (such that $\sum_x \mu(x) = 1$) of samples x assigned to state i in the kth ensemble ($x \in X_i^{(k)}$), f_i^k is the local free energy of state i in ensemble k and $b^{(k)}(x)$ is the bias potential. The TRAM solution is identical to the reversible MSM estimator when only a single thermodynamic ensemble is available. In the limit of infinite sampling, TRAM is equivalent to multistate Bennett acceptance ratio (MBAR) estimator[37,52].

3. Maximum-caliber (MaxCal) method

In the Maximum-caliber (MaxCal) method of Wan et al.[42], a conventional MSM is first constructed from trajectory data sampled in thermodynamic ensemble k, to obtain equilibrium populations $\pi_i^{(k)}$ and transition probabilities $p_{ij}^{(k)}$. Given estimates of perturbed equilibrium populations $\pi_i^{(l)}$ in thermodynamic ensemble l, the MaxCal method infers transition probabilities $p_{ij}^{(l)}$ by maximizing the path entropy, or caliber $\mathcal{C}$,

$$\mathcal{C} = \sum_{i,j} -\pi_i p_{ij}^{(l)} \ln\left(\frac{p_{ij}^{(l)}}{p_{ij}^{(k)}}\right),$$

(2)

with a restraint on detailed balance. The solution to this maximization problem is the self-consistent iteration of two equations involving a set of weights w_i:

$$p_{ij}^{(l)} \leftarrow w_i p_{ij}^{(k)} \left(\frac{\pi_j^{(l)} p_{ji}^{(k)} w_j}{\pi_i^{(l)} p_{ij}^{(k)} w_i}\right)^{\frac{1}{2}},$$

(3)

$$w_i \leftarrow w_i / \sum_j p_{ij}^{(l)},$$

(4)

(Note that that the solution of Wan et al.[42] applies to discrete-time kinetic networks; a more general solution for continuous rate equations is provided by Dixit et al.[53])

In our context, thermodynamic ensemble k is one in which the ligand is sufficiently decoupled so as to observe many transitions, and ensemble l is the unbiased thermodynamic ensemble. To obtain estimates of $\pi_i^{(l)}$, we use the MBAR free energy estimator[52] as implemented in the PyEMMA software package.[51]

	$k_{\mathrm{on}}(\times 10^9 \text{ M}^{-1}\text{s}^{-1})$	$k_{\mathrm{off}}(\times 10^7 \text{ s}^{-1})$	$\pi_{\mathrm{bound}}(\%)$	$K_D (\mathrm{mM})$
18-state	3.165 ± 0.488	3.294 ± 0.588	84.4 ± 0.4	10.66 ± 2.61
2-state	3.269 ± 0.516	3.672 ± 0.640	84.4 ± 0.4	11.54 ± 2.87

TABLE I. Reference values of ligand binding rates and affinities estimated from conventional MSMs.

II. RESULTS

A. Determination of reference rates and affinities for the toy ligand binding system.

After choosing the value $\epsilon = 2.5$ kJ mol^{-1} to be used in our tests (see Methods), we performed a series of long simulations (aggregate trajectory data of 80 μs) to precisely determine reference values for the bound-state equilibrium population π_{bound}, k_{on}, and k_{off} rates, via the construction of a conventional MSM.

To determine a suitable lag time τ for the construction of the MSM, a series of MSMs were constructed for lag times up to 4 ns, and the slowest implied timescale $t_2 = -\tau/(\ln \mu_2)$ computed, where $\mu_2 < 1$ is the second-largest eigenvalue of the transition probability matrix. This timescale corresponds to the slowest dynamical motion observed in the simulations, which in this case corresponds to ligand binding and unbinding. Thus, t_2 very closely approximates $(k_{\mathrm{on}} + k_{\mathrm{off}})^{-1}$.

The implied timescale plots of t_2 versus the lag time τ plateaus after a few 100 ps, indicating the dynamics is sufficiently Markovian (Figure 5a). Based on this result, we chose a lag time of $\tau = 1$ ns to construct an 18-state MSM to compute reference rates and affinities (Table I). Association and dissociation rates k_{off} and k_{on} were computed from inverse mean first passage times, and bound-state populations π_{bound} were estimated from the equilibrium populations of the MSM, with uncertainties estimated using the standard deviation of the results across 20 independent trajectories. The dissociation constant K_D was estimated from the computed rates as $k_{\mathrm{off}}/k_{\mathrm{on}}$, with uncertainty estimates from 10000 trials with normally distributed error. We also tried constructing a two-state MSM (bound and unbound states, Table I and Figure 5b), and found that both k_{on} and k_{off} were larger than the 18-state MSM estimates, likely due to discretization error.[24]
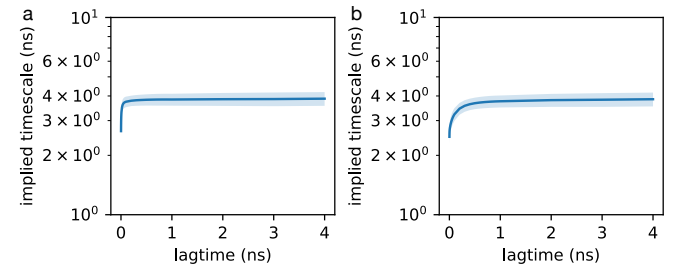

FIG. 5. Slowest MSM implied timescale t_2 versus MSM lag time for (a) an 18-state MSM, and (b) a 2-state MSM.

To explore how these (18-state MSM) estimates depend on the amount of trajectory data sampled, we di-

vided the trajectories into 10 blocks and constructed MSMs using cumulatively increasing amounts of data (e.g. 10%, 20%, ..., 100%). The estimated bound state populations in Figure 6 shows 20% data (800 ns for each independent run) is necessary for reasonably accurate estimates, with converged values obtained after about 60% of the trajectory. The estimated binding and unbinding rates, however, converge faster, requiring only 25% of the trajectory data. (Similar results were found for the two-state MSM, data not shown.)

These results illustrate the difficulty of estimating both thermodynamics and kinetics of ligand binding. Even in this simple model with a residence time of $(k_{\text{off}})^{-1} = 30.3$ ns, a large amount of trajectory data (1.6 μs) is required for converged estimation using MSMs. When it comes to realistic all-atom simulations of ligand binding to protein receptors proteins, the amount of total data required will surely exceed the limits of typical available computing time. For our purposes in this paper, we will keep in mind the value of (1.6 μs) as an upper bound for the trajectory dataset required for an unbiased simulation, which we will compare to the trajectory datasets needed for accurate multiensemble estimations, as described below.
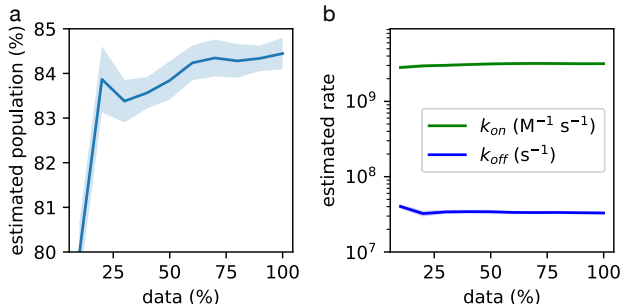
FIG. 6. Estimated bound-state populations (a) and rates (b) as a function of the amount of input data.

B. A single additional λ-scaled ensemble can efficiently estimate binding affinities and rates

With converged estimates of ligand binding rates and affinities in hand for our reference system, we now can assess the performance of multiensemble estimators. First, we examine the accuracy of TRAM and MaxCal estimators in the case where we are allowed to include one additional biased thermodynamic ensemble. Whereas our reference calculation used 80 μs of aggregate trajectory data (20 trajectories of 4 μs each), in this test we limit the trajectory data from the unbiased ensemble (i.e., $\lambda = 1.0$) to only 1% of that used for the reference calculations (2 trajectories of 400 ns each). Combined with the one additional biased thermodynamic ensemble, the total data we used here (2 × 2 × 400 ns) is only 2% of the reference calculations (80 μs) and 10% of the data required for converged results from the reference calculations (1.6 μs).

	$\lambda = 0.5$	$\lambda = 0.6$	$\lambda = 0.7$	$\lambda = 0.8$	$\lambda = 0.9$	$\lambda = 1.0$
all	400001	400001	400001	400001	400001	400001
subsampled	3602	1602	802	802	802	802

TABLE II. Number of samples used in the calculation before and after subsampling. The maximum amount of subsampling is for a stride equal to the lag time, resulting in 802 subsamples.

We used data from $\lambda = 1.0$ simulation as the unbiased ensemble and include one more ensemble with λ value taken from the set (0.5, 0.6, 0.7, 0.8, 0.9). For comparison, we also built MEMMs using trajectory data from all λ values. For the MaxCal estimator, our protocol was to first build a conventional MSM using only unbiased ($\lambda=1.0$) trajectories to estimate equilibrium state populations $\pi_i^{(k)}$ and transition rates $p_{ij}^{(k)}$. Then, we used the same data as used in MEMM construction (unbiased + one biased ensemble, or all ensembles) as input to MBAR to make improved estimates of equilibrium state populations $\pi_i^{(l)}$ in the $\lambda=1.0$ ensemble. Using these new populations, MaxCal was used to infer improved estimates of the transition rates $p_{ij}^{(l)}$.

It is well-known that subsampling of trajectory data to remove time-correlated data is important to make accurate estimates of free energies and their uncertainties.[8,52,54,55] We suspected that such considerations are similarly important for MEMM estimators TRAM and MaxCal. To remove time-correlation in our trajectory data, we subsampled trajectories generated in each thermodynamic ensemble at a time interval of $(2\tau_c + 1)$ steps, where τ_c is an estimate of the integrated correlation time calculated from the distance of the ligand from the origin (the center of the pocket, see Methods) over time.

After subsampling, we find that the number of effectively independent samples decreases dramatically (Table II). For thermodynamic ensembles $\lambda \geq 0.7$, correlation times are so long that subsampling would be at intervals longer than the MSM lag time of 1 ns. Therefore, we limit subsampling of trajectories from these ensembles to a maximum interval of 1 ns.

The performance of different estimators using subsampled and non-subsampled data are shown in Figure 7. These results clearly show the success of using simulations performed in λ-scaled thermodynamic ensembles to make multiensemble estimates of binding affinities and rates. Whereas conventional MSM estimates (using only the unbiased $\lambda=1$ ensemble) underestimate bound-state populations and off-rates, multiensemble estimates using only one additional thermodynamic ensemble more accurately predict the reference values.

In particular, thermodynamic ensembles with the most decoupling ($\lambda=0.5$ and $\lambda=0.6$) are best at reproducing populations and dissociation rates. It is in these ensembles that more frequent binding and unbinding transitions are observed, contributing most favorably to the statistical precision of affinity and rate estimates. Specifically, the use of trajectory data from $\lambda=0.6$ ensembles appears to give the best results, which, apart from finite sampling issues, may be a consequence of fewer re-
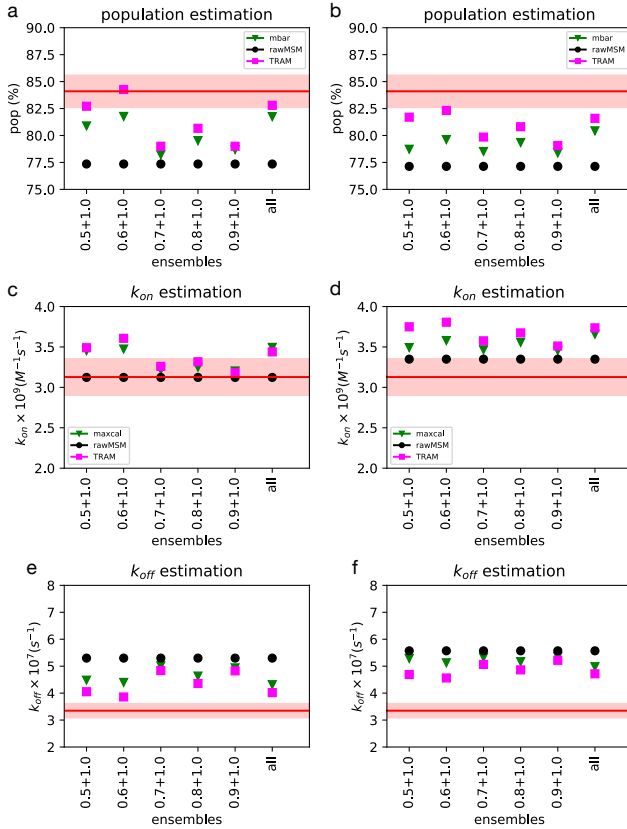
FIG. 7. Estimated rates and affinities from conventional (MSM) vs. multiensemble (TRAM and MaxCal) estimators applied to various trajectory data sets. Each data set consists of two 400-ns trajectories biased using a particular value of λ (e.g. λ=0.5) and two 400-ns unbiased trajectories (λ=1.0). Also considered were data sets in which trajectories from all λ values were included. Shown are estimated bound-state populations from (a) subsampled and (b) non-subsampled data, estimated values of k_{on} from (c) subsampled and (d) non-subsampled data, and estimated values of k_{off} from (e) subsampled and (f) non-subsampled data. The red and line and shaded region represent the values computed from our reference simulations, and their uncertainties, respectively. In all panels, the estimated uncertainties are smaller than the marker size.

binding events in the λ=0.5 ensemble. Indeed, simulations at small λ values spend the majority of time in the unbound state, which is less useful for rate estimation (data not shown).

While MaxCal produces improved estimates of affinities and rates compared to those of a conventional MSM, estimates from TRAM consistently outperform MaxCal. TRAM is able to make particularly accurate predictions of k_{off}. A potential explanation is that while free energy estimated from TRAM are informed by both sampled energies and transition counts, MaxCal free energy estimates (calculated by MBAR) are only informed by sampled energies.

For both TRAM and MaxCal estimates, inclusion of trajectory data from all thermodynamic ensembles results in only modest improvement in estimated quantities, rivaled by the results when only trajectory data from

λ=0.6 and λ=1.0 ensembles are used. This suggests that bias from finite sampling is quite important; including more trajectory data from ensembles where few transitions are observed is less effective than including data where many transitions are observed. A similar conclusion can be made from the comparison of subsampled vs. non-subsampled data. Despite the severe reduction in the number of samples used in the calculation, predictions using subsampled trajectory data consistently outperform predictions made using non-subsampled data.

C. Convergence of multiensemble estimates with increasing amounts of trajectory data

To assess the robustness of conventional and multiensemble estimators in dealing with limited input data, we performed tests in which different subsets of data (after subsampling) are used in the estimations. In these tests, we randomly select a given extent of λ=0.6 and λ=1.0 trajectory data (e.g. 10%, 20%, ..., 90% of the full trajectory) as a block segment, to be used as input to conventional and multiensemble estimators. This operation is repeated 10 times, resulting in 10 independent trials for statistical analysis.

As stated in the original TRAM paper[37], unlike MBAR, a global equilibrium sampling is not required for TRAM estimation, because of its efficient use of simulation data. Our results, which suggest that multiensemble esitmates from TRAM converges faster than the other estimators we evalutated, corroborate this statement. Using only 20% of the trajectory from ensembles λ=0.6 and λ=1.0, TRAM can accurately estimate bound-state populations, with faster convergence than MaxCal (Figure 8).

To highlight the reduction of uncertainty as the amount of trajectory data is increased, the magnitude of the uncertainty estimates (error bars in Figures 8a,c,e) are plotted separately (Figures 8b,d,f). It is clear that TRAM estimates have the smallest uncertainties among all estimators studied, and the fastest convergence with increasing amounts of trajectory data. Using only 10% of the trajectory data, the uncertainties of MaxCal estimates are large, becoming comparable with TRAM uncertainties when at least 30% of the trajectory data is used.

III. DISCUSSION

We have constructed a novel toy binding system that has several advantages. The system is small enough (2596 atoms) that it can be used to generate large amounts of trajectory data. Despite its simplicity, the system realistically mimics protein-ligand binding, with a well-defined binding site, encounter-complex conformations, and realistic solvation through the use of explicit water with counterions. It is straightforward to apply any number of enhanced sampling techniques to this system (e.g. umbrella simulations, free energy perturbation, metadynamics, etc), and very easy to tune the LJ ϵ parameter to
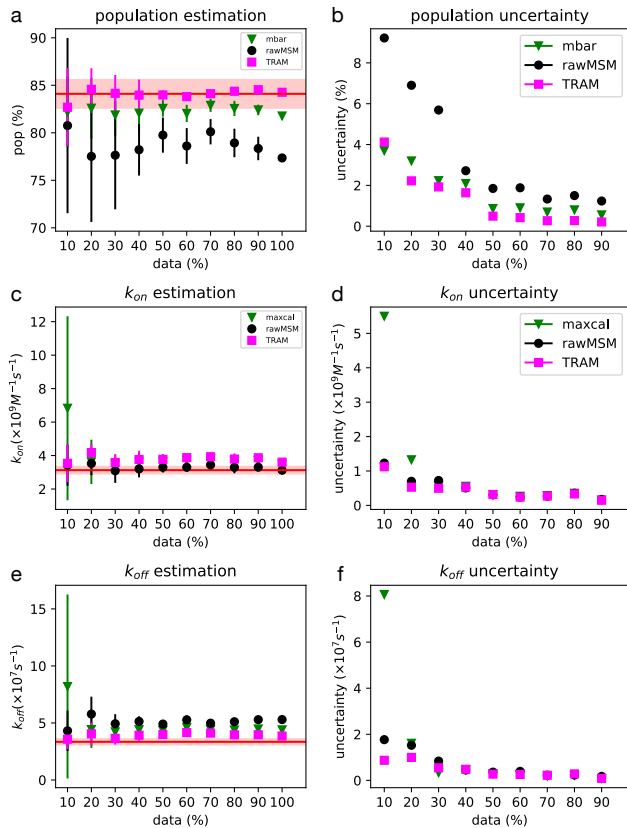
FIG. 8. Estimated bound-state populations (a), k_{on} (c), and k_{off} (e) as a function of increasing amounts of subsampled input data for ensembles ($\lambda = 0.6, 1.0$). Uncertainties in bound-state populations (b), k_{on} (d), and k_{off} (f) (the error bars shown in panels a,c,e), estimated by computing standard deviations from 10 independent trials.

increase or decrease the ligand affinity to suit particular problems of interest. The metastable state decomposition is well-defined and does not require any specialized dimensionality reduction or featurization. Thus, we expect this toy binding system to find many uses in future studies.

In this work, we specifically address how multiensemble estimators of ligand binding rates and affinities perform when used with thermodynamic ensembles where the ligand is decoupled by scaling its non-bonded interactions. To our knowledge, this study is the first application of TRAM and MaxCal to this problem, and the first to directly compare TRAM with MaxCal. The use of ligand decoupling as a bias potential has several advantages over umbrella sampling, a biasing technique that has has been used in the many of first applications of TRAM.[37,40,41] For one, ligand decoupling can take advantage of vast array of existing simulation tools to perform these calculations (for example, soft-core potentials). Moreover, whereas umbrella sampling is good at focusing sampling on a certain region, ligand-decoupling in effect serves to "de-focus" sampling away from the bound state, preserving the ability to sample transitions between many metastable states, a key requirement for the success of multiensemble estimators.

In our tests, we find that multi-ensemble estimators TRAM and MaxCal are superior to conventional MSM estimators, and that TRAM yields the most accurate estimates when faced with limited trajectory data. MaxCal takes a close second place to TRAM, achieving comparable results (Figure 7). In some applications, however, a MaxCal approach might be preferred to TRAM. One of the limitations of the TRAM estimator is that it is somewhat computationally demanding, requiring that key states be sampled in the unbiased ensemble for the calculation to be performed. Depending on the nature of the trajectory data being analyzed, this requirement may be too stringent, especially when there is poor phase space overlap between thermodynamic ensembles. In contrast, MBAR and MaxCal approaches are not limited by this requirement.

As is known to be case for free energy estimators, we find that TRAM and MaxCal estimates are sensitive to presence of time-correlated input data, and that these estimates can significantly improve when subsampled to remove this correlated samples. This is not surprising, since TRAM is itself a free energy estimator, and the MaxCal approach we use relies on free energy estimates from MBAR. Still, this finding underscores the need for careful curation of trajectory data to achieve accurate results.

IV. CONCLUSION

In this work, we have introduced a novel toy model that realistically mimics a protein-ligand binding system. After tuning and thoroughly benchmarking the affinities and binding rates in this model, we used it to study the performance of multiensemble MSM estimators when used with ligand decoupling. The key idea in this work is that many more binding and unbinding events can be observed in decoupled thermodynamic ensembles, enabling better statistical estimation of rates and affinities. Indeed, we find that TRAM and MaxCal estimators outperform conventional MSM estimators at this task, and the TRAM is the most accurate and efficient overall. We find that accuracy is improved when subsampling is performed to remove time-correlation.

These results suggest that multiensemble methods used with ligand decoupling simulations may be highly valuable to simultaneously obtain accurate predictions of ligand binding affinities and rates. In the future, we anticipate the growing use of these methods in computational drug discovery.

[1] J. D. Chodera, D. L. Mobley, M. R. Shirts, R. W. Dixon, K. Branson, and V. S. Pande, Current opinion in structural biology **21**, 150 (2011).

[2] D. L. Mobley and P. V. Klimovich, The Journal of Chemical Physics **137**, 230901 (2012).

[3] L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M. K. Dahlgren, J. Greenwood, *et al.*, Journal of the American Chemical Society **137**, 2695 (2015).

[4] M. Aldeghi, A. Heifetz, M. J. Bodkin, S. Knapp, and P. C. Biggin, Chemical science **7**, 207 (2016).

[5] Z. Cournia, B. Allen, and W. Sherman, Journal of Chemical Information and Modeling **57**, 2911 (2017).

[6] Z. Cournia, B. K. Allen, T. Beuming, D. A. Pearlman, B. K. Radak, and W. Sherman, Journal of Chemical Information and Modeling , acs.jcim.0c00116 (2020).

[7] L. F. Song and K. M. Merz Jr, Journal of Chemical Information and Modeling **60**, 5308 (2020).

[8] A. S. Mey, B. K. Allen, H. E. B. Macdonald, J. D. Chodera, D. F. Hahn, M. Kuhn, J. Michel, D. L. Mobley, L. N. Naden, S. Prasad, *et al.*, Living Journal of Computational Molecular Science **2**, 18378 (2020).

[9] R. A. Copeland, D. L. Pompliano, and T. D. Meek, Nature Reviews Drug Discovery **5**, 730 (2006).

[10] R. A. Copeland, Nature Reviews Drug Discovery **15**, 87 (2016).

[11] D. A. Schuetz, W. E. A. de Witte, Y. C. Wong, B. Knasmueller, L. Richter, D. B. Kokh, S. K. Sadiq, R. Bosma, I. Nederpelt, L. H. Heitman, E. Segala, M. Amaral, D. Guo, D. Andres, V. Georgi, L. A. Stoddart, S. Hill, R. M. Cooke, C. De Graaf, R. Leurs, M. Frech, R. C. Wade, E. C. M. de Lange, A. P. IJzerman, A. Müller-Fahrnow, and G. F. Ecker, Drug discovery today **22**, 896 (2017).

[12] L. Mollica, I. Theret, M. Antoine, F. Perron-Sierra, Y. Charton, J.-M. Fourquez, M. Wierzbicki, J. A. Boutin, G. Ferry, S. Decherchi, *et al.*, Journal of medicinal chemistry **59**, 7167 (2016).

[13] A. Dickson, P. Tiwary, and H. Vashisth, Current topics in medicinal chemistry **17**, 2626 (2017).

[14] D. B. Kokh, M. Amaral, J. Bomke, U. Gradler, D. Musil, H.-P. Buchstaller, M. K. Dreyer, M. Frech, M. Lowinski, F. Vallee, *et al.*, Journal of chemical theory and computation **14**, 3859 (2018).

[15] M. Bernetti, M. Masetti, W. Rocchia, and A. Cavalli, Annual review of physical chemistry **70**, 143 (2019).

[16] A. Nunes-Alves, D. B. Kokh, and R. C. Wade, Current Opinion in Structural Biology **64**, 126 (2020).

[17] S. Decherchi and A. Cavalli, Chemical Reviews **120**, 12788 (2020).

[18] S. D. Lotz and A. Dickson, Journal of the American Chemical Society **140**, 618 (2018).

[19] R. Hall, T. Dixon, and A. Dickson, Frontiers in molecular biosciences **7**, 106 (2020).

[20] P. Tiwary, V. Limongelli, M. Salvalaglio, and M. Parrinello, Proceedings of the National Academy of Sciences of the United States of America **112**, E386 (2015).

[21] B. R. Jagger, C. T. Lee, and R. E. Amaro, The journal of physical chemistry letters **9**, 4941 (2018).

[22] F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, Proceedings of the National Academy of Sciences **106**, 19011 (2009).

[23] V. a. Voelz, G. R. Bowman, K. Beauchamp, and V. S. Pande, Journal of the American Chemical Society **132**, 1526 (2010).

[24] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, The Journal of chemical physics **134**, 174105 (2011).

[25] G. Zhou and V. A. Voelz, The Journal of Physical Chemistry B **120**, 926 (2016).

[26] B. E. Husic and V. S. Pande, Journal of the American Chemical Society **140**, 2386 (2018).

[27] N. Plattner and F. Noé, Nature communications **6**, 1 (2015).

[28] G. Zhou, G. a. Pantelopulos, S. Mukherjee, and V. a. Voelz, Biophysj **113**, 785 (2017).

[29] N. Plattner, S. Doerr, G. De Fabritiis, and F. Noé, Nature chemistry **9**, 1005 (2017).

[30] Y. Ge, E. Borne, S. Stewart, M. R. Hansen, E. C. Arturo, E. K. Jaffe, and V. a. Voelz, Journal of Biological Chemistry **293**, 19532 (2018).

[31] Y. Ge and V. A. Voelz, "Markov state models to elucidate ligand binding mechanism," in *Protein-Ligand Interactions and Drug Design*, edited by F. Ballante (Springer US, New York, NY, 2021) pp. 239–259.

[32] H. Wan and V. A. Voelz, The Journal of chemical physics **152**, 024103 (2020).

[33] E. Suárez, R. P. Wiewiora, C. Wehmeyer, F. Noé, J. D. Chodera, and D. M. Zuckerman, bioRxiv (2020).

[34] A. C. Pan, H. Xu, T. Palpant, and D. E. Shaw, Journal of chemical theory and computation **13**, 3372 (2017).

[35] H. Wu, A. S. J. S. Mey, E. Rosta, and F. Noé, The Journal of Chemical Physics **141**, 214106 (2014).

[36] E. Rosta and G. Hummer, Journal of chemical theory and computation **11**, 276 (2015).

[37] H. Wu, F. Paul, C. Wehmeyer, and F. Noé, Proceedings of the National Academy of Sciences of the United States of America **113**, E3221 (2016).

[38] L. S. Stelzl, A. Kells, E. Rosta, and G. Hummer, Journal of chemical theory and computation **13**, 6328 (2017).

[39] S. Kieninger, L. Donati, and B. G. Keller, Current opinion in structural biology **61**, 124 (2020).

[40] F. Paul, C. Wehmeyer, E. T. Abualrous, H. Wu, M. D. Crabtree, J. Schöneberg, J. Clarke, C. Freund, T. R. Weikl, and F. Noé, Nature Communications **8**, 1095 (2017).

[41] Y. Ge, S. Zhang, M. Erdelyi, and V. Voelz, Journal of Chemical Information and Modeling **61**, 2353– (2021).

[42] H. Wan, G. Zhou, and V. a. Voelz, Journal of Chemical Theory and Computation **12**, 5768 (2016).

[43] P. D. Dixit, A. Jain, G. Stock, and K. A. Dill, Journal of chemical theory and computation **11**, 5464 (2015).

[44] P. D. Dixit, J. Wagoner, C. Weistuch, S. Pressé, K. Ghosh, and K. A. Dill, The Journal of Chemical Physics **148**, 010901 (2018).

[45] P. D. Dixit and K. A. Dill, Journal of chemical theory and computation **14**, 1111 (2018).

[46] D. Meral, D. Provasi, and M. Filizola, The Journal of chemical physics **149**, 224101 (2018).

[47] D. Meral, D. Provasi, D. Prada-Gracia, J. Möller, K. Marino, M. J. Lohse, and M. Filizola, Scientific reports **8**, 1 (2018).

[48] X. Hu, Y. Wang, A. Hunkele, D. Provasi, G. W. Pasternak, and M. Filizola, PLoS computational biology **15**, e1006689 (2019).

[49] K. Ghosh, P. D. Dixit, L. Agozzino, and K. A. Dill, Annual review of physical chemistry **71**, 213 (2020).

[50] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, SoftwareX **1**, 19 (2015).

[51] M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz, and F. Noé, Journal of Chemical Theory and Computation **11**, 5525 (2015).

[52] M. R. Shirts and J. D. Chodera, The Journal of Chemical Physics **129**, 124105 (2008).

[53] P. D. Dixit, The Journal of Chemical Physics **148**, 091101 (2018), https://doi.org/10.1063/1.5023232.

[54] J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok, and K. A. Dill, Journal of Chemical Theory and Computation **3**, 26 (2007).

[55] A. Grossfield, P. N. Patrone, D. R. Roe, A. J. Schultz, D. W. Siderius, and D. M. Zuckerman, Living journal of computational molecular science **1** (2018).