TEMPLE
UNIVERSITY®

Efficient estimation of binding kinetics using scaled non-bonded interactions and harmonic restraints

Yunhui Ge and Vincent A. Voelz

Department of Chemistry, Temple University, Philadelphia, PA

ABSTRACT

Understanding the thermodynamics and kinetics of protein-ligand interactions is crucial for computational drug discovery. In this project, we simulate a small host-guest system where the ligand is a single, uncharged Lennard-Jones (LJ) particle, and the binding site is an 11-particle icosahedral pocket made from the same atom type; the system is simulated in a $(3.00859 \text{ nm})^3$ periodic box filled with 860 TIP3P waters. We first tuned the LJ parameter ϵ to construct a system that realistically mimics a protein-ligand binding system. Using over 80 μs of unbiased simulation, we found a value of $\epsilon = 2.5 \text{ kJ/mol}$ produced a suitably challenging test case, with a dissociation constant of $K_d = 0.19 \pm 0.04 \text{ M}$, and an implied timescale of $3.84 \pm 0.27 \text{ ns}$, estimated using an 18-state Markov state model (MSM). We then tested the accuracy of different estimators when less data is used as input to the model, but sampled from biased ensembles using a combination of harmonic restraints and scaled non-bonded interactions. We compare results for the thermodynamic estimator MBAR and multi-ensemble MSM estimators dTRAM and TRAM to estimate unbiased thermodynamic and kinetic properties from the biased simulations.

A toy host-guest system

Parameterizing a toy host-guest system. As a toy system to test different estimators, we used an 11-particle icosahedral pocket of Lennard-Jones (LJ) particles as a model receptor, with a single uncharged LJ particle as ligand, inside a 27.23 nm^3 cubic periodic box filled with 860 TIP3P water molecules, 2 Na^+ ions, and 2 Cl^- ions, comprising a total of 2596 particles (**Figure 1**). All LJ particles had $\sigma = 3.39967 \text{ \AA}$, modeled after an ffAMBER CT carbon, and a tunable ϵ parameter. Bond parameters for pocket atoms were copied from CT atom types, and bond angle restraints of 108° were enforced with a force constant of $150.0 \text{ kJ mol}^{-1} \text{ rad}^{-2}$. To determine a good value of ϵ , ten stochastic dynamics (Langevin) trajectories of length 1 μs were generated at 300 K, each for ten values of ϵ between 0.0 and 10.0 kJ mol^{-1} . A value was $\epsilon = 2.5 \text{ kJ mol}^{-1}$ was chosen, resulting in a reasonably large bound-state populations with 1–10 ns dwell times.

Figure 1. The toy binding site is an 11-particle icosahedral pocket.

Figure 2. Trajectory traces using the chosen value of $\epsilon = 2.5 \text{ kJ mol}^{-1}$.

Markov State Model (MSM) construction and estimation of reference populations and rates. 18 metastable states were defined using criteria determined from the observed binding landscape (**Figure 3**). The bound state (index 0) is assigned when the radius $r < 0.15 \text{ nm}$, the unbound state (index 17) when $r > 0.85 \text{ nm}$, and states with intermediate radii assigned to state indices 1–16 according to the nearest icosahedral face or entrance face (**Figure 4**). To determine a set of reference populations and rates, an aggregate of 80 μs ($20 \times 4\text{-}\mu\text{s}$ trajectories) of stochastic dynamics was performed. MSMs transition matrices estimated using a 2 ns lag time yielded reference values for future comparisons (**Table 1**).

Figure 3. The distribution of the distance between the center of mass of the binding pocket and the ligand. Representative snapshots of conformational states used to build the MSM are shown.

Figure 4. Indices of metastable states.

Table 1. Estimated reference values.

$\pi_{\text{bound}} (\%)$	84.1 $\pm$ 1.5
$K_{\text{on}} (\times 10^9 \text{ M}^{-1} \text{ s}^{-1})$	3.128 $\pm$ 0.226
$k_{\text{off}} (\times 10^7 \text{ s}^{-1})$	3.348 $\pm$ 0.265

Can biased simulations be used to efficiently estimate rates and populations for our toy system? We used non-bonded scaling and harmonic bias potentials to sample a variety of thermodynamic ensembles.

Scaling of non-bonded interactions

- $\lambda = 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$

Umbrella-sampling simulations

- force constants $k = 5, 10, 15, 20, 25, 40, 50, 75, 100 \text{ kJ mol}^{-1} \text{ nm}^{-2}$
- equilibrium distances $r_0 = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 \text{ nm}$

We generated a test set consisting of 800 ns of trajectory data for each unique set of λ , k and r_0 values (a single trajectory of 400 ns each for unbound and bound starting states). The PyEMMA software package was used to construct MSMs using the TRAM, dTRAM, and MBAR estimators. Rates were estimated using inverse mean first passage times (MFPTs) between states 0 and 17.

METHODS

Multiensemble dynamical estimators

MBAR/MaxCal. In this method, a conventional MSM is built from unbiased trajectory data. The MBAR free energy estimator (Shirts and Chodera, 2008) is then used to update the estimates of unbiased populations from biased-ensemble samples. From these updated populations, the Maximum Caliber (MaxCal) approach of Wan and Voelz (2016) is then used to infer updated MSM transition rates.

TRAM (Wu et al. 2016) and **dTRAM** (Wu et al. 2014). These are transition reweighting estimators that use multi-ensemble trajectory data to simultaneously estimate unbiased free energies and MSM transitions.

RESULTS

Figure 5. Estimated thermodynamic and kinetic properties from λ -scaling non-bonded interactions simulations using (a) all data, and (b) sub-sampling. Error bars are smaller than the marker size.

Only a single λ -scaled trajectory ($\lambda=0.6$) + an unbiased trajectory ($\lambda=1.0$) is enough to efficiently estimate thermodynamics and kinetics. Estimated properties using subsampled $\lambda=0.6$ and $\lambda=1.0$ trajectories, calculated as a function of varying input data, are shown in **Figure 6**. Even using only 10% of the trajectory data ($4 \times 40 \text{ ns}$, 960 samples), thermodynamic and kinetic properties are accurate. This accuracy improves as more data is included in the calculation.

Figure 6. (a) Estimated thermodynamic and kinetic properties by using varying amounts of subsampled input data for ensembles ($\lambda = 0.6, 1.0$). (b) Uncertainties estimated by a bootstrap procedure.

RESULTS

The accuracy of multi-ensemble estimates are more sensitive to umbrella positions than force constants. **Figure 7a** shows estimates incorporating unscaled ($\lambda=1.0$) trajectory data from all umbrella positions, where only data from a single force constant used (in addition to unbiased data). Only dTRAM estimates are strongly sensitive to force constant, likely because of the inaccuracy of estimating discrete-state bias free energies at large force constants. **Figure 7b** shows estimates incorporating unscaled trajectory data where only a single force constant ($k = 20 \text{ kJ mol}^{-1} \text{ nm}^{-2}$, and only a single umbrella position is used. Here, the most accurate estimate comes from trajectories using umbrellas placed at $r_0 = 0.3 \text{ nm}$, near the entrance of the pocket. This umbrella location encourages more binding and unbinding events than other biased ensembles.

Figure 7.

Application: cyclic β -hairpin binding to MDM2

Figure 8. The slowest dynamics (along tIC1) corresponds to the association of a designed cyclic peptide to MDM2. The next-slowest dynamics corresponds to the ligand binding with different poses other than the crystal structure.

Figure 9. Superposition of the crystal structure of designed cyclic peptide (magenta) and p53 TAD domain (green) bound with MDM2 (red).

Table 3. Estimated kinetic and thermodynamic properties using different estimators.

	$\pi_{\text{bound}} (\%)$	$k_{\text{on}} (\text{s}^{-1} \text{M}^{-1})$	$k_{\text{off}} (\text{s}^{-1})$
Experiment	99.997	$8.30 (\pm 0.02) \times 10^5$	0.1174 ± 0.0001
MSM	51.74 ± 9.48	$7.34 (\pm 2.13) \times 10^6$	795.07 ± 220.38
TRAM	96.12 ± 4.25	$8.21 (\pm 5.89) \times 10^6$	40.11 ± 39.74

FUTURE WORK

Here, we study estimator performance in a simple toy model. Future research will focus on applying these methods to efficiently estimate kinetics and thermodynamics in larger and more biologically relevant systems. We anticipate MSM-based approaches using all-atom molecular dynamic simulations will be especially useful to future drug discovery efforts.

Acknowledgements

This research was supported in part by the National Institutes of Health through grant 1S10OD020095-01. Temple University's HPC resources are supported in part by the National Science Foundation through major research instrumentation grant number 1625061 and by the US Army Research Laboratory under contract number W911NF-16-2-0189. Y.G. and V.V. were supported by National Institutes of Health grant 1R01GM123296-01.

DISCUSSION

1. The toy host-guest system is useful for testing different estimation methods for ligand binding properties.
2. The TRAM estimator works with λ -scaling of non-bonded interactions, requiring only a few biased ensembles for accurate kinetic and thermodynamic estimation.
3. TRAM estimation converges fast and is robust in dealing with limited input data.
4. Subsampling to remove time correlation improves the accuracy and efficiency of different estimators.
5. Estimates from harmonic biased simulations are most sensitive to umbrella location.
6. More thermodynamic ensembles of umbrella sampling are required for accurate estimation of properties than λ -scaling simulations.

REFERENCES

1. Wu, H., Paul, F., Wehmeyer, C., & Noé, F. (2016). *PNAS*, 113(23), E3221–30.
2. Wu, H., Mey, A. S. J. S., Rosta, E., & Noé, F. (2014). *J. Chem. Phys.*, 141(21), 214106–12.
3. Shirts, M. R., & Chodera, J. D. (2008). *J. Chem. Phys.*, 129(12), 124105–11.
4. Scherer, M. K., Trendelkamp-Schroer, B., Paul, F., Pérez-Hernández, G., Hoffmann, M., Plattner, N., et al. (2015). *J. Chem. Theor. Comput.*, 11(11), 5525–5542.