Multiensemble Markov models of molecular thermodynamics and kinetics

Hao Wu[a,1], Fabian Paul[a,1], Christoph Wehmeyer[a], and Frank Noé[a,2]

[a]Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany

Edited by David Baker, University of Washington, Seattle, WA, and approved April 22, 2016 (received for review December 21, 2015)

We introduce the general transition-based reweighting analysis method (TRAM), a statistically optimal approach to integrate both unbiased and biased molecular dynamics simulations, such as umbrella sampling or replica exchange. TRAM estimates a multiensemble Markov model (MEMM) with full thermodynamic and kinetic information at all ensembles. The approach combines the benefits of Markov state models—clustering of high-dimensional spaces and modeling of complex many-state systems—with those of the multistate Bennett acceptance ratio of exploiting biased or high-temperature ensembles to accelerate rare-event sampling. TRAM does not depend on any rate model in addition to the widely used Markov state model approximation, but uses only fundamental relations such as detailed balance and binless reweighting of configurations between ensembles. Previous methods, including the multistate Bennett acceptance ratio, discrete TRAM, and Markov state models are special cases and can be derived from the TRAM equations. TRAM is demonstrated by efficiently computing MEMMs in cases where other estimators break down, including the full thermodynamics and rare-event kinetics from high-dimensional simulation data of an all-atom protein–ligand binding model.

molecular dynamics | enhanced sampling | Markov state models | transition-based reweighting analysis method | Bennett acceptance ratio

Computer simulations have become important tools in the investigation of biomolecular processes, including transmembrane transport (1–4), ligand reception and receptor activation (5–7), and endocytosis (8–10). Unbiased atomistic molecular dynamics (MD) simulations have recently reached the ability to extensively sample biomolecular processes on timescales up to milliseconds, including protein folding (11), conformational changes (5, 12), and protein–ligand association and dissociation (13–15). In addition to breakthroughs in computer hardware (16), simulation software (17–19), and distributed computing (20, 21), a key technology to reconcile swarms of individually short simulations to long-time kinetics are kinetic models, such as Markov state models (MSMs) (22–28). A key advantage of Markov state modeling over many other approaches is that it integrates well with dimension reduction and clustering techniques (29–31) that can process high-dimensional data, and can thus treat complex kinetics that are not well described by few states or reaction coordinates (13, 32–34). A limitation of MSMs is that they rely on the rare events being reversibly sampled in the underlying MD simulation data.

However, important biological processes are still out of reach for unbiased MD simulation. For example, although downhill processes such as protein–ligand association to the bound pose can now be spontaneously sampled (6, 13, 15, 34), the dissociation of stable inhibitory complexes can involve timescales of hours or longer (35). Enhanced sampling methods such as umbrella sampling (US) (36, 37), parallel or simulated tempering (38–40), metadynamics (41), and others (42–44) use simulations at different ensembles in which bias potentials or higher temperatures are used to speed up events that are rare in the physical ensemble (45, 46). Reweighting methods such as the weighted histogram analysis method (WHAM) (37, 47, 48) and binless WHAM, also known as multistate Bennett acceptance ratio (MBAR) (49–51), can combine multiple-ensemble simulation data to estimates of the unbiased thermodynamics (free energies or probabilities). These methods treat their input data as uncorrelated samples of the ensemble distribution and are therefore not suitable for simulation data with long correlation times in some variables, as it is common for unbiased MD simulations and biased simulations with slow unbiased coordinates (32, 52).

To overcome individual limitations of MSMs and enhanced sampling techniques, we propose to integrate simulation data from multiple ensembles in a multiensemble Markov model (MEMM) (Fig. 1), in such a way as to (*i*) work with high-dimensional data and coarse state-space discretizations, (*ii*) use unbiased MD simulations from nonequilibrium starting points but avoid rate models beyond MSMs, and (*iii*) optimally combine data to full thermodynamics and kinetics at all simulated ensembles.

Here, we develop the generic transition-based reweighting analysis method (TRAM), an estimation method for MEMMs that combines the above features as follows. (*i*) Statistical weights of sampled configurations are reweighted between ensembles in a binless manner, a key property for working in high-dimensional spaces. (*ii*) Conditional transition statistics are used in an MSM-based likelihood, and thus simulations only need to be in local but not in global equilibrium. TRAM only relies on the MSM approximation and detailed balance relations to predict rare-event kinetics (53) and avoids the use of additional rate models. (*iii*) TRAM provides a maximum-likelihood MEMM with full thermodynamic and kinetic information at all ensembles. In summary, TRAM goes significantly beyond previously proposed transition-based reweighting methods (54–57) and other methods to estimate thermodynamics and kinetics

Significance

Molecular dynamics simulations can provide mechanistic understanding of biomolecular processes. However, direct simulation of slow transitions such as protein conformational transitions or protein–ligand dissociation are unfeasible with commonly available computational resources. Two typical strategies are (*i*) conducting large ensembles of short simulations and estimating the long-term kinetics with a Markov state model, and (*ii*) speeding up rare events by bias potentials or higher temperatures and estimating the unbiased thermodynamics with reweighting estimators. In this work, we introduce the transition-based reweighting analysis method (TRAM), a statistically optimal approach that combines the best of both worlds and estimates a multiensemble Markov model (MEMM) with full thermodynamic and kinetic information at all simulated ensembles.

Author contributions: H.W., F.P., and F.N. designed research; H.W., F.P., and F.N. performed research; H.W., F.P., and C.W. analyzed data; H.W., F.P., and C.W. wrote software; and H.W., F.P., and F.N. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]H.W. and F.P. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: frank.noe@fu-berlin.de.

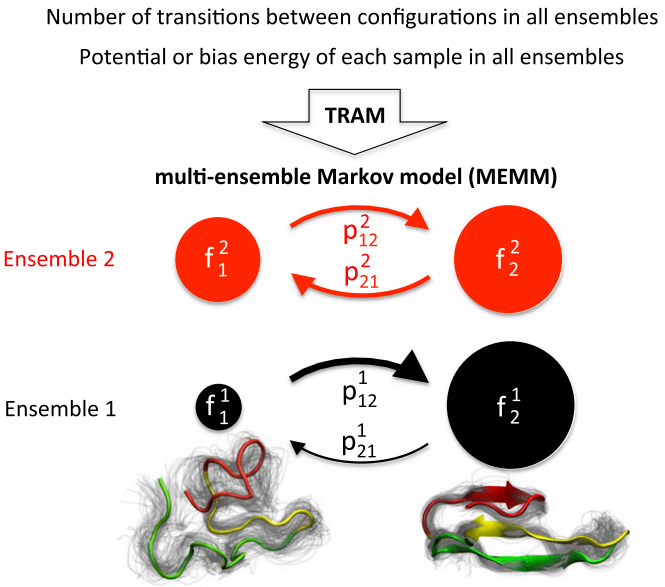

Fig. 1. Multiensemble Markov models (MEMMs) contain thermodynamic (e.g., free energies) and kinetic (e.g., transition probabilities) information among all configuration states (subscript index) and ensembles (superscript index). TRAM estimates MEMMs by combining transitions observed between configuration states, and the statistical weights/reduced energies of samples in all ensembles (reweighting).

from multiensemble data (53, 58–61), which offer some but not all of the above properties (for more detailed discussion, see below). TRAM is a formal generalization to WHAM, MBAR, reversible MSMs, and discrete TRAM that can all be derived from TRAM.

We apply TRAM on two benchmark systems and an all-atom model of the trypsin protein with the benzamidine inhibitor. We illustrate that TRAM can estimate the thermodynamics at ensembles more accurately and with less simulation data than previous estimation methods, and that additionally unbiased models of the kinetics can be built. We demonstrate that our MEMM approach offers a systematic treatment of the common problem of slow unbiased coordinates in US simulations and provides efficient estimates of rare-event kinetics, such as protein–ligand dissociation.

TRAM

Basics. Let us consider a molecular system in a reference ensemble with configuration x in a configuration space and dimensionless potential function $u(x)$. $u(x)$ has units of thermal energy $k_BT=\beta^{-1}$, where T is the temperature. $u(x)$ is a sum of terms, including $\beta U(x)$ with the potential energy function U, and pressure–volume or chemical potential terms, depending on the ensembles under consideration (49). The system has an equilibrium distribution as follows:

$$\mu(x)=\mathrm{e}^{f-u(x)}, \qquad [1]$$

where the free energy f is the negative logarithm of the partition function and has the role to normalize $\mu(x)$.

Suppose we are given simulations from K different ensembles (indexed by the superscript k), which may comprise an arbitrary combination of the unbiased ensemble, simulations with biased energy functions, or different temperatures. We can formally relate any ensemble with dimensionless potential $u^k(x)$ to the reference ensemble by introducing a bias potential $b^k(x)$ such that $u^k(x)=u(x)+b^k(x)$. The corresponding equilibrium distribution $\mu^k(x)$ of the kth ensemble can be expressed as follows:

$$\mu^k(x)=\mathrm{e}^{f^k-b^k(x)}\mu(x), \qquad [2]$$

where the relative free energy f^k of ensemble k is chosen such that $\mu^k(x)$ is normalized. Consider the following examples to see how $b^k(x)$ must be chosen to model commonly used enhanced sampling methods:

i) In US, the potential energy function of each simulation is $U(x)+B^k(x)$, where $B^k(x)$ is the kth umbrella potential. The bias potential is as follows:

$$b^k(x)=\beta B^k(x). \qquad [3]$$

ii) Replica exchange or parallel tempering simulations are performed at different temperatures $T^1,\ldots,T^K$, the bias of the kth temperature with respect to the reference ensemble (e.g., the lowest temperature) is as follows:

$$b^k(x)=U(x)\left(\beta^k-\beta\right). \qquad [4]$$

MSMs for Molecular Kinetics. An MSM at ensemble k consists of a partition of the molecular configuration space into m discrete and nonoverlapping configuration states $S_1,\ldots,S_m$ and the conditional transition probabilities $p_{ij}^k(\tau)$ that a system that is in state S_i at time t will be found in state S_j at time $t+\tau$.

We first define the local free energy f_i^k of configuration state S_i in ensemble k. The exponential of f_i^k is proportional to the statistical weight of this state:

$$\mathrm{e}^{-f_i^k}=\mathrm{e}^{-f^k}\int_{S_i}\mu^k(x)dx, \qquad [5]$$

where the integral evaluates to the equilibrium probability of the system to be in state S_i when simulated in ensemble k.

For given simulation data from ensemble k that contains c_{ij}^k transitions from state S_i at time t and to state S_j at time $t+\tau$, the likelihood of an MSM with transition matrix $\mathbf{P}^k=[p_{ij}^k]$ is as follows:

$$L_{\mathrm{MSM}}^k=\prod_{i=1}^m\prod_{j=1}^m\left(p_{ij}^k\right)^{c_{ij}^k}. \qquad [6]$$

When simulations are conducted at thermal equilibrium (i.e., without adding or removing energy to the system) in ensemble k, equilibrium and transition probabilities are related by the detailed balance equations $\mathrm{e}^{-f_i^k}p_{ij}^k=\mathrm{e}^{-f_j^k}p_{ji}^k$, and the Markov model is said to be reversible. With detailed balance constraints, the maximum likelihood of Eq. **6** has no closed-form solution but can be iteratively solved (28, 62, 63).

Local Equilibrium Model. If simulations sample from multiple ensembles, a central problem is to infer the equilibrium distribution $\mu(x)$ at a reference ensemble, given the simulation data at all ensembles. The principle behind such inference is that we can reweight the equilibrium probability of a sample x between different ensembles by means of Eq. **2**.

A widely used estimator is the binless WHAM method, (50, 51), also called MBAR (49), which provides an optimal estimate of $\mu(x)$ under the assumption that at each ensemble k, the samples x are drawn independently from their global equilibrium distribution $\mu^k(x)$. MBAR can be derived by maximizing a likelihood that is simply given by the product of $\mu^k(x)$ over all samples x and all ensembles k (49–51).

However, we do not want to depend on the global equilibrium assumption. Hence we define the local equilibrium distribution for each configuration state S_i:

$$\mu_i^k(x) = \begin{cases} e^{f_i^k - f^k}\mu^k(x) & x \in S_i \\ 0 & \text{else.} \end{cases} \quad [7]$$

We assume that simulations are sampling from these local equilibrium distributions, but they do not need to be in equilibrium between configuration states, which is key for invoking the MSM framework. We obtain the following likelihood of generating the simulation data for a given sequence of discrete states:

$$L_{\text{LEQ}}^k = \prod_{i=1}^{m} \prod_{x \in X_i^k} \mu_i^k(x), \quad [8]$$

where X_i^k denotes the set of all samples generated from the kth ensemble and in configuration state S_i. As $\mu_i^k(x)$ can be related to $\mu(x)$ via Eqs. **2** and **7**, the local equilibrium model is key to reweight samples between different ensembles.

TRAM Likelihood. We develop the TRAM estimator. The TRAM likelihood combines the MSM likelihood [**6**] and local equilibrium likelihood [8]. Inserting Eqs. **2** and **7**, we obtain the following:

$$L_{\text{TRAM}} = \prod_{k=1}^{K} \underbrace{\left(\prod_{i,j} \left(p_{ij}^k\right)^{c_{ij}^k} \right)}_{L_{\text{MSM}}^k} \underbrace{\left(\prod_{i=1}^{m} \prod_{x \in X_i^k} \mu(x) e^{f_i^k - b^k(x)} \right)}_{L_{\text{LEQ}}^k}. \quad [9]$$

This likelihood expresses the probability that a given set of trajectories sampling from different ensembles has visited a particular sequence of discrete states (L_{MSM}^k) and has sampled the local configurations inside these discrete states (L_{LEQ}^k). The structure of the TRAM likelihood is similar to that of a hidden Markov model (64).

The trajectory statistics include the bias potentials $b^k(x)$ that are defined by the simulation protocol [e.g., US or replica exchange molecular dynamics (REMD)], and the number of observed transitions c_{ij}^k. The unknown variables in the TRAM likelihood are the point densities $\mu(x)$, the local free energies f_i^k, and the transition probabilities p_{ij}^k. The TRAM problem is to maximize the likelihood [**9**] in the variable space subject to the following constraints:

$$e^{-f_i^k} p_{ij}^k = e^{-f_j^k} p_{ji}^k, \quad \text{for all } i,j,k \quad [10]$$

$$\sum_j p_{ij}^k = 1, \quad \text{for all } i,k \quad [11]$$

$$\sum_{x \in X} \mu(x) = 1, \quad [12]$$

where [**11**] and [**12**] are simple normalization constraints, and $\mu(x)$ is considered as a discrete distribution on the set of all samples, X. The detailed balance condition [**10**] couples the dynamical (MSM) part to the local equilibrium part. Unfortunately, the detailed balance constraints make the above problem very hard to solve.

Maximum-Likelihood Solution. The TRAM problem contains $(m^2K + |X|)$ unknowns, $(mK + 1)$ linear equality constraints (normalization), and $Km(m-1)/2$ nonlinear equality constraints (detailed balance), so finding the optimal solution by directly using gradient- or Newton-type methods is difficult even for systems with only few configuration states or ensembles. Fortunately, we can transform the TRAM problem into a more tractable system of nonlinear algebraic equations and solve the resulting system by an iterative algorithm. By using the Lagrange duality theory (*Appendix*), it can be proved that the maximum of the TRAM likelihood satisfies the following equations:

$$\sum_j \frac{c_{ij}^k + c_{ji}^k}{\exp\left[f_j^k - f_i^k\right] v_j^k + v_i^k} = 1, \quad \text{for all } i,k \quad [13]$$

$$\sum_{x \in X_i} \frac{\exp\left(f_i^k - b^k(x)\right)}{\sum_l R_i^l \exp\left[f_i^l - b^l(x)\right]} = 1, \quad \text{for all } i,k \quad [14]$$

where v_i^k are Lagrange multipliers that can be interpreted as counts (for infinite statistics $v_i^k = \sum_j c_{ij}^k$, see *Appendix*). X_i is the set of all samples in configuration state S_i, no matter from which ensemble. The factor R_i^l is given by the following:

$$R_i^k = \sum_j \frac{\left(c_{ij}^k + c_{ji}^k\right) v_j^k}{v_j^k + \exp\left[f_i^k - f_j^k\right] v_i^k} + N_i^k - \sum_j c_{ji}^k, \quad [15]$$

where N_i^k is the number of samples in X_i^k. R_i^k are effective state counts (see below). When [**9**–**12**] are fulfilled and in the limit of infinite statistics, $R_i^k = N_i^k$.

In contrast to Eqs. **9**–**12**, the formulation in [**13**] and [**14**] only contains the $2mK$ unknowns v_i^k and f_i^k and does not involve $\mathbf{P}^k$ and $\mu(x)$ explicitly. Given the solution of Eqs. **13** and **14**, we can compute all MSM transition matrices by the following:

$$p_{ij}^k = \frac{c_{ij}^k + c_{ji}^k}{\exp\left[f_j^k - f_i^k\right] v_j^k + v_i^k}, \quad [16]$$

and the unbiased statistical weights of all samples by the following:

$$\mu(x) = \frac{1}{\sum_k R_{i(x)}^k \exp\left[f_{i(x)}^k - b^k(x)\right]}, \quad [17]$$

where we have defined $i(x)$ such that $i(x) = j$ when $x \in X_j$. The TRAM estimator defined by Eqs. **13** and **14** is statistically optimal, asymptotically correct, i.e., converges to the correct results of f_i^k, p_{ij}^k, and $\mu(x)$ as the length or number of simulation trajectories increases, and the most general multiensemble Markov model estimator (see below and *Appendix*).

Eqs. **13** and **14** are reminiscent of other estimators: Eq. **13** arises when optimizing an MSM transition matrix with given stationary weights $\exp(-f_i^k)$ (63). Eq. **14** has the same form as the self-consistent MBAR equation (49) for the ensemble free energies f_i^k of a single configuration state S_i, but instead of the number of samples in that state N_i^k, the modified counts R_i^k are used (detailed interpretation in *Appendix*). The TRAM equations can therefore be thought of expressing two optimization problems simultaneously: (*i*) at each ensemble k, the optimization of the MSMs for given free energies, f_i^k for all configurations S_i. (*ii*) At each configuration S_i, the optimization of the free energies, f_i^k for all ensembles.

Optimization Algorithm. The TRAM equations [**13**] and [**14**] are coupled and can only be solved numerically. Here, we transform

them into a simple fixed-point problem, in which the following equations need to be iterated until convergence:

$$v_i^{k,\mathrm{new}} := v_i^k \sum_j \frac{c_{ij}^k + c_{ji}^k}{\exp\left[f_j^k - f_i^k\right] v_j^k + v_i^k}, \qquad [18]$$

$$f_i^{k,\mathrm{new}} := -\ln \sum_{x \in X_i} \frac{\exp\left[-b^k(x)\right]}{\sum_l R_i^l \exp\left[f_i^l - b^l(x)\right]}. \qquad [19]$$

More implementation details of this algorithm, including initialization, termination, and convergence acceleration are given in *Appendix*. Note that, instead of a fixed-point iteration, we could attempt a Newton-based (65, 66) or stochastic optimization method (67, 68).

Thermodynamics and Kinetics from TRAM.

Thermodynamics. The correct calculation of stationary (thermodynamic) properties does not rely on Markovianity, but only requires the unbiased estimation of free energies f_i^k or the stationary density $\mu^k(k)$ at the chosen lag time τ. However, it is required that the simulations are in local equilibrium within the configuration states, and violations of local equilibrium can be compensated by using longer lag times τ. The robustness of TRAM estimates should therefore be tested as a function of lag time (see results, Fig. 3).

Kinetics. Asymptotic correctness of all p_{ij}^k at the selected lag time τ does not imply that powers of the matrix $\mathbf{P}^k$ are a good prediction of the transition probabilities at longer lag times. Whether the multiensemble Markov model is able to predict long-term kinetics depends on the quality of the discretization and on τ being sufficiently large, as usual for MSMs (28, 56) (see results, Fig. 3*D*). Note that this behavior does not change if a rate matrix is used instead of a transition matrix.

Generality of TRAM. TRAM is a generalization of discrete TRAM, binless WHAM/MBAR, binned WHAM, and reversible MSMs (Fig. 2). These specialized estimators can be derived from TRAM by adding the specific assumptions made by them. MBAR can be derived from TRAM by assuming that samples are drawn from the global equilibrium distribution of each ensemble. Discrete TRAM can be derived by assuming that the bias energies are piecewise constant and pointwise reweighting can be replaced by histogram reweighting. WHAM is derived using a combination of both assumptions. Finally, if we have only a single ensemble, the TRAM solution is identical to the reversible MSM estimator (derivations in *Appendix*).

TRAM has the Markovianity assumption in its likelihood model, but otherwise only uses fundamental relations such as detailed balance and pointwise reweighting. It is therefore the most general MSM-based estimator for simulation data from multiple ensembles. Other transition-based reweighting methods are related as follows: trajectory reweighting techniques (58–60) are applicable without any state space discretization, but assume the trajectory starting points to emerge from a global equilibrium distribution. The dynamic weighted histogram analysis method (DHAM) (57) uses a kinetic reweighting scheme that can predict kinetics at ensembles not simulated from, but is based on a rate model, uses histogram binning and does not optimize with respect to detailed balance. An advantage in not enforcing the detailed balance constraint is that DHAM is a single-shot estimator, while TRAM and dTRAM are estimators that need to be iterated to solution. xTRAM (55) is a bin-less TRAM method, but in contrast to the present method not statistically optimal for finite data (*Applications*).

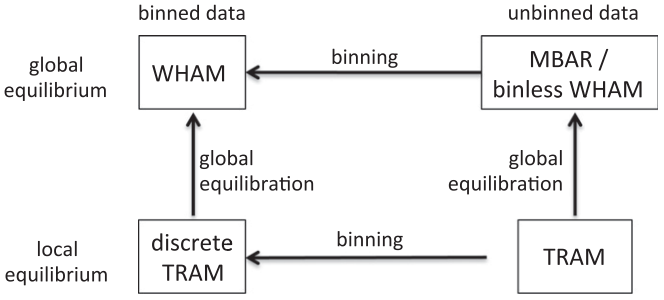

Fig. 2. Relationship of different statistically optimal reweighting estimators. TRAM is the most general method considered here, it can be specialized to MBAR, discrete TRAM, and WHAM by adding the assumption of global equilibrium or performing a binning of sampled configurations.

Applications

Multitemperature Replica-Exchange. We compare the performance of MBAR and TRAM for REMD simulations of solvated alanine dipeptide using 33 exponentially spaced temperatures in the range of 300–600 K (Fig. 3*A*) (see ref. 55 for simulation protocol). In multitemperature simulations, the bias potentials between ensembles depend on the potential energy (Eq. **4**). To analyze such data with histogram-based methods, one would have to bin the potential energy axis in addition to the coordinate(s) of interest (69). For many-body systems such as solvated macromolecules, the large range of potential energies sampled and the required resolution to approximate the bias energies (Eq. **4**) disables binned estimators such as WHAM, discrete TRAM, and DHAM, and instead require binless methods such as MBAR and TRAM.

The configuration space of alanine dipeptide is partitioned into 20 discrete states using k-means clustering in the space of the coordinates $\{\cos\phi, \sin\phi, \cos\psi, \sin\psi\}$ (Fig. 3*A*). The equilibrium probabilities on the sets I–IV are compared between estimators. Within statistical error, both MBAR and TRAM converge to the same values, whereas TRAM converges significantly faster (Fig. 3*B*). TRAM outperforms MBAR because TRAM relies only on local rather than global equilibria. As a result, TRAM does not suffer from the fact that initial structures are not sampled from a global equilibrium distribution, and the REMD simulation must first relax to sample from global equilibrium. Even after this relaxation phase, TRAM uses the data more effectively because a smaller number of simulation steps is required to generate an uncorrelated sample from local equilibrium compared with global equilibrium.

Next, we test the robustness of TRAM estimates as a function of the lag time (Fig. 3 *C* and *D*). It is seen that the stationary probabilities, and thus the results in Fig. 3*B*, are independent of the lag time, demonstrating that the Markov property is not required to get correct estimates of the stationary properties. Unbiased estimates of equilibrium properties only require that the simulations are in local equilibrium. For REMD simulations with a good state space discretization, this is fulfilled even at short lag times τ. In contrast, unbiased estimates of the kinetic properties require sufficiently long lag times for the Markov property to be valid. The estimated relaxation timescale at temperature 366 K is constant above $\tau = 10$ ps (Fig. 3*D*), which was used for all TRAM estimates in Fig. 3. Only trajectory segments in which no temperature swap was executed for 10 ps or longer were used for this estimate.

Fig. 3 *E* and *F* show thermodynamics and kinetics obtained from the multiensemble Markov model as a function of the temperature. The probabilities of metastable states become more similar with increasing temperatures, but the temperature dependence is very weak, indicating that entropy differences play a minor role (Fig. 3*E*). The mean first passage times (inverse transition rates) from I/II to III/IV and back decrease strongly with temperature (Fig. 3*F*). The decrease is exponential (Arrhenius-like) up to 450 K, but shows a weaker temperature dependence

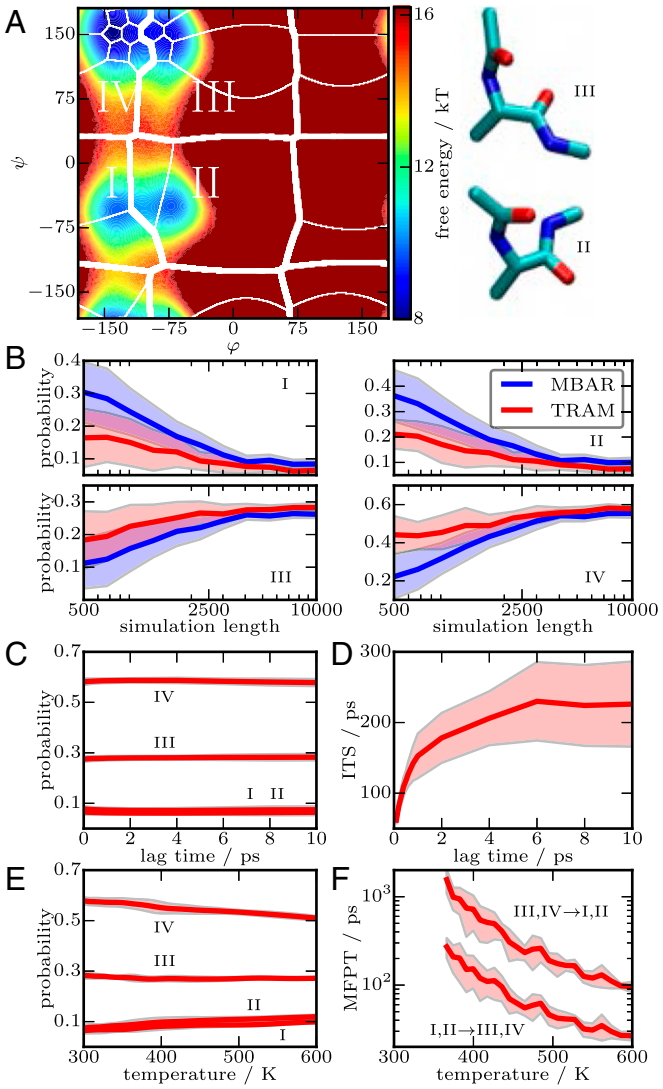

Fig. 3. MBAR versus TRAM using REMD simulations of alanine dipeptide; mean and SD over seven independent simulations are shown. (*A*) Histogram-based free-energy landscape in the backbone torsions ϕ, ψ at a temperature of 300 K. Thin lines: borders of 20 clusters used to partition the space $\{\cos\phi, \sin\phi, \cos\psi, \sin\psi\}$ (hence the nonlinear boundaries). Thick lines: borders of four metastable sets analyzed in *B*. (*B*) Convergence of equilibrium probability estimates of sets I–IV at 300 K, as a function of simulation length. (*C–F*) TRAM estimates. (*C*) Equilibrium probabilities as a function of the lag time τ. (*D*) Slowest relaxation timescale at 366 K as a function of the lag time. (*E*) Equilibrium probabilities of sets I–IV as a function of temperature. (*F*) Mean first passage times at temperatures where transitions were found.

for higher temperatures, indicating that the kinetics are limited by diffusion rather than barrier crossing in this range.

Biased Simulations with Slow Orthogonal Degrees of Freedom. Simulations in which sampling is enhanced along predefined reaction coordinates (e.g., using bias potentials) are often hampered by unforeseen rare transitions in other coordinates (52). For illustration, we use a pathological 2D toy potential with three wells (Fig. 4*A*). US simulations are conducted using only the x coordinate as bias coordinate (details in *Appendix*).

As the potential wells I and III cannot be separated on the x axis, it takes a long time to converge to the global equilibrium when the simulations are confined to values of $x < 15$. Especially simulations with the second umbrella potential centered at $x = 8.33$ exhibit rare-event transitions along the y axis (Fig. 4*B*), characterized by an autocorrelation time of 500 steps (Fig. 4*C*) (70). In contrast, the largest value of the autocorrelation time along the x axis is only about 22 steps.

Rare events in nonenhanced coordinates are a common problem in enhanced sampling simulations and cause major problems in their analysis. For estimation methods relying on global equilibrium sampling such as MBAR, the statistically correlated samples should be discarded before running the estimator (49). In umbrella simulation 2, this would result in retaining only one effective sample for each 500 samples in the simulation, resulting in the loss of almost all data and requiring very long simulation times.

TRAM does not require global equilibrium sampling and can therefore use simulation data much more efficiently. We discretize the configuration space Ω into 20 states as shown in Fig. 4*A*, and then use TRAM with lag time $\tau = 1$ to estimate the unbiased equilibrium distribution from the US data. Fig. 4*D* summarizes estimation errors of TRAM for different lengths of each simulation trajectory and compares them with those of MBAR and the previously described xTRAM method (55). Here, the error is evaluated as the Kullback–Leibler divergence between the estimated probability distributions of the three macrostates I, II, and III and the true reference. In contrast to MBAR, TRAM can effectively overcome the influence of the nonequilibrium distribution of the data through Markov state modeling and achieve accurate estimates even in the case of trajectory length smaller than autocorrelation times $\tau_{\mathrm{ess}}(y)$ of some biased simulations. Furthermore, TRAM also significantly outperforms xTRAM, which is a consistent estimator under the MSM assumption but not statistically optimal for finite data.

Protein–Ligand Binding and Kinetics. Finally, we demonstrate that TRAM can help to resolve the problem of rare events in orthogonal degrees of freedom and provides efficient estimates of rare-event kinetics in all-atom, explicit-solvent simulations of the serine protease trypsin and its inhibitor benzamidine (see ref. 34 for detailed setup). This illustrates the usefulness of the estimator

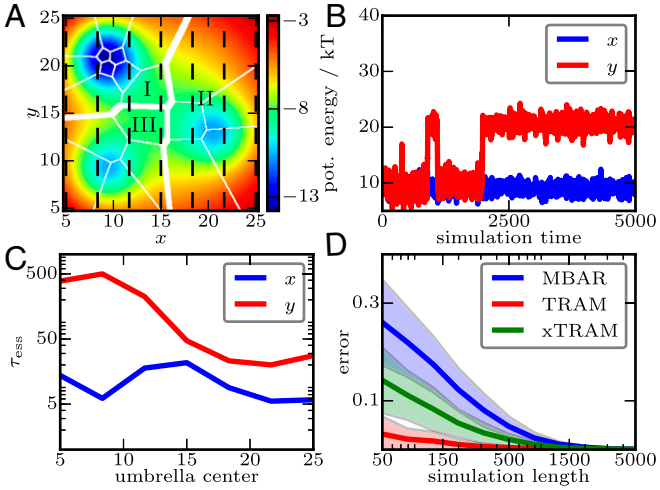

Fig. 4. Comparison of MBAR, xTRAM, and TRAM for estimating the equilibrium distribution of a three-well potential from US data. (*A*) Potential function $u(x, y)$, where thin white lines represent the borders of 20 discrete states, thick white lines represent the borders of the three potential wells, and the dashed black lines indicate the umbrella centers. (*B*) A simulation trajectory with the bias potential centered at $x = 8.33$. (*C*) Autocorrelation times $\tau_{\mathrm{ess}}(x)$ and $\tau_{\mathrm{ess}}(y)$ with respect to x axis and y axis for different umbrella centers. (*D*) Average estimation errors and their SDs of MBAR, xTRAM, and TRAM for different simulation trajectory lengths over 30 independent realizations of US.

in high-dimensional spaces where binning of all relevant coordinates is not an option.

We first analyze pure US simulations with 150 umbrella windows used to sample the position of benzamidine between the bound pose and a prebinding site (Fig. 5*A*, structures *i–iv*; details in *Appendix*).

To detect rare events in the unbiased coordinates, time-lagged independent component analysis (TICA) (29, 30) was used with the Cartesian coordinates of residues around the binding site. The first independent component (IC) is strongly correlated with the US coordinate. From the remaining ICs, two had timescales implied by the TICA eigenvalues larger than the trajectory length, indicating undesirable metastable transitions orthogonal to the umbrella coordinate. The second IC corresponds to closing of the binding pocket by the Trp 215 side chain (Fig. 5*D*, structures *i* and *iii*). The third IC corresponds to an isomerization of the disulfide bond between Cys 191 and Cys 220. An analysis using MBAR or WHAM is thus unfeasible or inefficient, as the global equilibrium assumption is strongly violated.

One strategy to deal with this very common problem is to restrain coordinates orthogonal to the umbrella coordinate, to avoid undesirable degrees of freedom from switching (45). Although this approach is useful for computing energy differences between end states, it may change or restrain the transition mechanism and artificially increase free-energy barriers along the pathway. With TICA and TRAM, we now have the possibility to allow these orthogonal dynamics to happen, and to treat these events explicitly.

The space spanned by the US coordinate and the second IC was discretized into 100 Voronoi cells with the *k*-means algorithm (Fig. 5*A*). This number of states is far smaller than the number of bins that would be required with a binned estimator such as WHAM or discrete TRAM. A count matrix c_{ij}^k was estimated for every umbrella at a lag time of 11 ns, and the largest strongly connected component $\mathcal{S}$ of the summed count matrix $c_{ij} = \sum_k c_{ij}^k$ was determined. The initial set was strongly disconnected, and we therefore adaptively started new umbrella simulations in nine rounds, to improve the connectivity (*Appendix*). In the complete dataset, some clusters are still disconnected (red clusters in Fig. 5*A*).

In particular, these disconnected states include structures in which the binding site is occluded by a tryptophan side chain, while benzamidine is still inside, and structures in which the binding site attempts to close during the exit pathway. TRAM is applied on the connected subset of states (white clusters in Fig. 5*A*). The TRAM results show that the Trp-occluded conformation is a local minimum in the free-energy landscape (Fig. 5*C*). This is confirmed by refs. 13 and 61 where the Trp-occluded conformation is shown to be a metastable conformation of the protein. In contrast, this local minimum is not found by MBAR, and several disconnected minima are spuriously estimated (boxes in Fig. 5*B*).

To analyze the full high-dimensional binding mechanism and estimate unbiased kinetics, we must go beyond US simulations. We therefore used TRAM to combine the US data with up to 49.1 μs of unbiased MD data (details in *Appendix*). The unbiased trajectories started in the unbound state, such that many binding

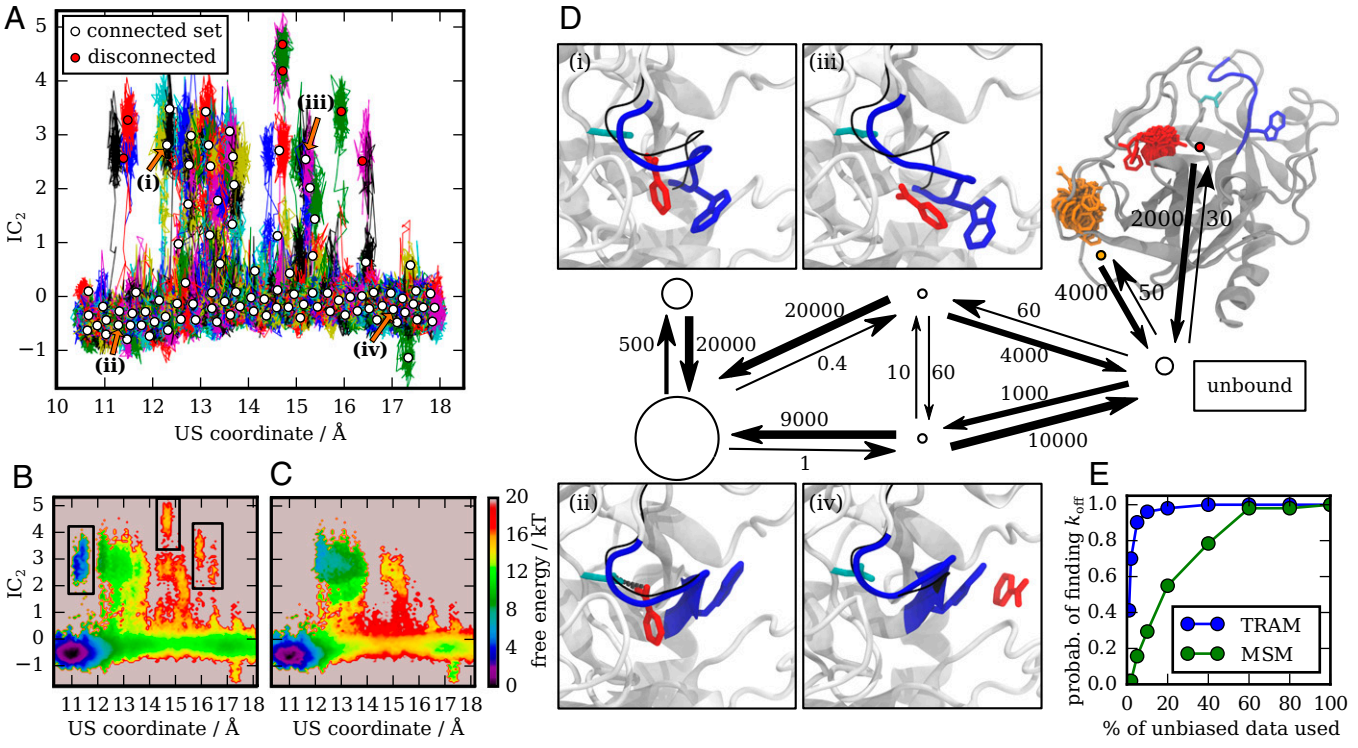

Fig. 5. Thermodynamics and kinetics of all-atom protein–ligand binding model for trypsin–benzamidine. (*A–C*) US simulations. (*D* and *E*) MEMM using both unbiased and US simulations. (*A*) Trajectories projected on the space of the umbrella sampling (US) coordinate and the second independent component (IC). The US coordinate describes a transition from benzamidine bound to Asp-189 to benzamidine located outside the binding pocket on the surface of trypsin. The second IC corresponds to concerted opening of loop (Trp-215-Gln-221) and flipping of Trp-215. The Voronoi centers of the Markov states are shows as disks. Markov states that are irreversibly connected to the data set are shown as red disks and are excluded from the MEMM. (*B*) Potential of mean force (PMF) in the same coordinate space computed with MBAR; (*C*) PMF computed with TRAM. Besides a higher barrier along the US coordinate, the TRAM-PMF gives the Trp-occluded conformation a lower free energy compared with the MBAR result. (*D*) Coarse-grained kinetic network of the MEMM. Structures (*i*, *ii*, *iii*, and *iv*) are found in the four quadrants of *A*. The largest transition rates (where at least one direction exceeds 1/ms) between these macrostates, the unbound state and two alternatively bound states are shown as arrows. Units are events per millisecond. (*E*) Efficiency of TRAM in the estimation of unbinding kinetics compared with an MSM built from the same unbiased data. Shown is the probability that $\log k_{\text{off}}$ calculated from a bootstrap sample falls into the interval $[0.5 \log k_{\text{off}}^{\text{all}}, 2 \log k_{\text{off}}^{\text{all}}]$ where $k_{\text{off}}^{\text{all}}$ is the TRAM estimate calculated using all data.

events are present. Individual steps of dissociation events are found in some trajectories, but no complete dissociation event is found in any single trajectory. By combining the free-energy information inherent in the biased trajectories with the binding kinetics from the unbiased trajectories, the full unbinding kinetics can be estimated with TRAM. TRAM gives the estimate $k_{\mathrm{off}}^{\mathrm{TRAM}} = 1{,}170\ \mathrm{s}^{-1}$, with 95% confidence intervals of [617 s^{-1}, 2,120 s^{-1}]. For comparison, the MSM estimated from the unbiased simulation data only, using the same state definition and lag time as for TRAM, provides an estimate of $k_{\mathrm{off}}^{\mathrm{MSM}} = 1{,}863\ \mathrm{s}^{-1}$ with a larger uncertainty of [876 s^{-1}, 4,816 s^{-1}] [all errors estimated using bootstrap, experimental dissociation rate 600 s^{-1} (71)].

To assess the data efficiency of TRAM and the MSM, we varied the amount of unbiased MD data that was used for the estimation. With TRAM, only 5–10% of the unbiased MD data are needed compared with an MSM to reliably estimate k_{off} (Fig. 5*E*).

Fig. 5*D* shows a kinetic network of the binding/dissociation events at the unbiased ensemble of the multiensemble Markov model. The kinetic data include association to several secondary binding sites; two of them are shown in Fig. 5*D*. At the lag time that we chose (30 ns), the prebound states *iii* and *iv* are not metastable and indirect transitions where these states are skipped during binding/unbinding appear in the transition matrix (not shown in the figure for clarity).

Conclusions

We have derived the TRAM for estimating MEMMs from simulation data comprising arbitrary combinations of unbiased MD, biased enhanced sampling simulations such as US, or multitemperature simulations such as REMD. TRAM does not require binning of the bias energies and is therefore suitable for the analysis of multitemperature simulations and of high-dimensional state spaces. TRAM is a Markov modeling method as it only requires local equilibrium and uses conditional transition statistics to estimate the MEMM—i.e., it can use short trajectories whose starting points were not sampled from global equilibrium. Even when just being used for estimating thermodynamics, e.g., the equilibrium distribution at the unbiased ensemble, or temperature-dependent free energies, TRAM is superior to global equilibrium-based estimators such as WHAM or MBAR.

In an application to US simulations of protein–ligand binding, we have used MSM concepts of finding slow coordinates and detecting a connected set of states to define a meaningful subspace for computing a ligand dissociation pathway and its free-energy profile. We have also sketched an approach to identify sampling bottlenecks and extend simulations in nonconverged umbrella windows to adaptively improve the convergence of the umbrella simulation, in line with other adaptive approaches (52, 72). In this example, combining US with replica exchange simulations may also have improved the sampling (73, 74).

We demonstrated that TRAM can be used to compute an unbiased estimate of protein–ligand dissociation kinetics on the order of a millisecond by using only a few microseconds of simulation data. Beyond the simple two-state rate, TRAM is ideally suited to estimate the full multistate kinetics that was found in refs. 13 and 46 with rate models or much more simulation data. TRAM significantly expands the power of the MSM framework by allowing to integrate the full power of enhanced sampling simulations. The TRAM estimator is included in PyEMMA (75) as of version 2.2. Tutorials can be found under pyemma.org.

Materials and Methods

Three-Well Potential Setup. The potential shown in Fig. 4*A* is defined by a sum of four Gaussians $u(x,y) = -\sum_{i=1}^{4} a_i g_{h_1,h_2,\sigma_1,\sigma_2}(x,y)$ with parameters (8,15,15,10,10), (4.8, 9,9,2.5, 2.5), (8,9,21,2.5, 2.5), and (4,21,13,2.5, 2.5), and $g_{h_1^i,h_2^i,\sigma_1^i,\sigma_2^i}(x,y) = \exp[-(x-h_1)^2/2\sigma_1^2 - (y-h_2)^2/2\sigma_1^2]$ on a square $[5,25]\times[5,25]$ and ∞ outside. US simulations are conducted using bias potentials $b^k(x,y) = (x-\bar{x}^k)^2/5$ for $k = 1, \ldots, 7$ with umbrella centers $\{\bar{x}^k\}$ positioned at $\bar{x}^k = (10k+5)/3$. We generate 20 independent simulation trajectories $\{(x_t^k, y_t^k)\}$ for each biased potential using the Metropolis sampling algorithm, where $x_0^k = \bar{x}^k$, y_0^k is randomly drawn from [5,25] and the candidate sample follows the uniform distribution on $[x_t^k - 3, x_t^k + 3]\times[y_t^k - 3, y_t^k + 3]$ for a given (x_t^k, y_t^k). The autocorrelation time of $\{(x_t^k, y_t^k)\}$ in x is given by $\tau_{\mathrm{ess}} = 1 + 2\sum_{s=1}^{\infty}\rho_s(x)$, and likewise in y (70), where $\rho_s(h)$ denotes the autocorrelation at lag s of h. We compute τ_{ess} from a long trajectory with 10^6 steps as described in ref. 76.

Trypsin–Benzamidine Setup.

US. The US coordinate is defined as the distance from the center of mass of all backbone atoms of Asp 189 and Pro 161 to the center of mass of the benzamidine ring atoms. The setup consists of 150 harmonic umbrellas with uniform force constant of 100 kcal·mol^{-1}·Å^{-2} and umbrella centers positioned along the US coordinate according to $x_{\mathrm{rest},i} = 10.5\text{Å} + i\cdot 0.05\text{Å}$ for $i = 0 \ldots 149$. For each umbrella, multiple independent runs were generated all starting from the same initial conditions that come from an unbiased binding trajectory. In total, 459 trajectories each having a length of 20 ns were generated adaptively in nine rounds of restarts. After an initial exploratory round, eight additional rounds were started to increase the overlap $o_{k,k+1} = \sum_{i\in S}\min(N_i^k, N_i^{k+1})$ between ensembles. Restarts were done in ensembles k, $k+1$ where $o_{k,k+1} < 100$. For the analysis, TICA (29) was used at a lag time 5.5 ns on the Cartesian coordinates of all heavy atoms within a 15Å radius around Asp 189 in the Protein Data Bank structure.

Molecular dynamics. A total of 491 unbiased MD simulation trajectories of length 100 ns each (data from ref. 34) were discretized by selecting the nearest neighbor heavy-atom contacts between benzamidine and all trypsin residues as input features (75). The features were transformed by TICA (lag time, 5 ns) to a kinetic map preserving 95% of the kinetic variance (31), resulting in a 31-dimensional transformed space. Discretization of all data in the joint space of 31 ICs and the two coordinates shown in Fig. 5*A* was done with the k-means algorithm, using $k = 500$. Microstates were grouped into seven macrostates. Four macrostates correspond to the quadrants of Fig. 5*A*, splitting microstates near the binding site at the US coordinate 14.5Å and the TICA coordinate $\mathrm{IC}_2 = 1$. Nearness to the binding site is defined by an US coordinate <18.2Å and being inside a binding funnel, defined by $\cos\gamma \le 0.74$, where γ measures the angle between the vectors connecting centers of mass of benzamidine with Pro 161 and Trp 215 with Pro 161. The remaining microstates were grouped in three macrostates: the unbound state and two alternatively bound states where benzamidine binds to secondary binding sites of trypsin, found with PCCA++ (77).

Kinetics. Using TRAM, MEMMs were estimated combining the US data and the unbiased MD data. The MEMM lag times were chosen as 30 ns for the unbiased data and as 10 ns for the US data (chosen from the interval where k_{off} appears to be independent of both lag times). A transition matrix for the unbiased ensemble was computed according to Eq. **16**. k_{off} was computed as the reciprocal of the mean-first-passage time from the bound macrostates (US coordinate < 14.5Å) to the unbound state.

Bootstrapping. Errors bars for the different estimates were obtained from a bootstrap. Every sample of the bootstrap was generated by first partitioning the trajectories by ensemble, and then independently for every partition drawing whole trajectories and finally merging the trajectories.

Appendix

Solution of the TRAM Problem. Ignoring constants, the constrained optimization problem of the TRAM log likelihood [**9**] can be written as follows:

$$\begin{aligned} \min_{\{\mu(x)\},\{\mathbf{P}^k\}} \quad & -\sum_{k,i,j} c_{ij}^k \ln p_{ij}^k - \sum_{k,i} N_i^k f_i^k - \sum_{x\in X} \ln\mu(x) \\ s.t. \quad & \mathrm{e}^{-f_i^k} p_{ij}^k = \mathrm{e}^{-f_j^k} p_{ji}^k \qquad \text{for all } i,\ j,\ k \\ & \sum_j p_{ij}^k = 1, \qquad \text{for all } i,\ k \end{aligned} \qquad [20]$$

with

$$f_i^k = -\ln \sum_{x\in X_i} \mu(x)\mathrm{e}^{-b^k(x)}. \qquad [21]$$

We omit the normalization constraint [**17**], because the normalization of $\mu(x)$ does not affect the optimality of the solution of Eq. **20**, and we can thus normalize $\mu(x)$ a posteriori.

Using the Lagrange duality lemma of discrete TRAM problem (56), it can be shown that Eq. **20** is equivalent to the following unconstrained min-max optimization problem:

$$\min_{\{\mu(x)\}}\max_{\{\mathbf{v}^k\}} L_{\text{dual}} = \sum_{k,i,j} c_{ij}^k \ln\left(\mathrm{e}^{-f_i^k} v_j^k + \mathrm{e}^{-f_j^k} v_i^k\right) - \sum_{k,i}\left(N_i^k - \sum_j c_{ji}^k\right) f_i^k - \sum_{i,k} v_i^k - \sum_{x\in X} \ln \mu(x), \quad [22]$$

where $\mathbf{v}^k = [v_i^k]$ are Lagrange multipliers. Equivalence means that the optimal solution of Eq. **20** can be obtained from that of Eq. **22** by using Eq. **16**.

We now consider solving Eq. **22**. Because L_{dual} is a concave function of $\{\mathbf{v}^k\}$ and a convex function of $\{\ln\mu(x)\}$, the optimal solution of Eq. **22** can be characterized as a saddle point with $\partial L_{\text{dual}}/\partial v_i^k = 0$ and $\partial L_{\text{dual}}/\partial \mu(x) = 0$ for all i, k, and x (see section 10.3.4 in ref. 78). Because

$$\frac{\partial L_{\text{dual}}}{\partial v_i^k} = \sum_j \frac{c_{ij}^k + c_{ji}^k}{\exp\left(f_j^k - f_i^k\right) v_j^k + v_i^k} - 1, \quad [23]$$

$$\frac{\partial L_{\text{dual}}}{\partial \mu(x)} = \sum_k R_{i(x)}^k \mathrm{e}^{f_{i(x)}^k - b^k(x)} - \mu(x)^{-1}, \quad \text{for } x \in S_i \quad [24]$$

where R_i^k is defined in Eq. **15**, we can conclude that the optimal solution of Eq. **22** should satisfy Eq. **13**, and Eq. **17** holds. Substituting Eq. **17** into Eq. **21**, we can get the optimality condition [**14**].

Asymptotic Correctness of TRAM. We use $\overline{b}$ to denote the exact value of an unknown variable b without any statistical error, and denote by $c_i^k = \sum_j c_{ij}^k$ the sum of row i in count matrix $\mathbf{C}^k = [c_{ij}^k]$, and by N_i the number of samples in S_i (c_i^k is different from N_i^k for finite statistics).

Now we show that the TRAM estimates of local partition functions, transition matrices, and reference distribution converge to the correct ones under the condition that the size of simulation data (either length or number of simulation trajectories) tends to infinity and the ratio N_i^k/N_i tends to a constant w_i^k for any i,k under the assumption that the local equilibrium within each configuration S_i is achieved in simulations. In this limit, the transition counts become the following:

$$c_{ij}^k = c_i^k \overline{p}_{ij}^k. \quad [25]$$

Substituting Eq. **25** into Eqs. **16**, **15**, and **13**, and replacing f_i^k, v_i^k with $\overline{f}_i^k, c_i^k$, we have $p_{ij}^k = \overline{p}_{ij}^k$, $R_i^k = N_i^k$, and

$$\sum_j \frac{c_{ij}^k + c_{ji}^k}{\exp\left[\overline{f}_j^k - \overline{f}_i^k\right] c_j^k + c_i^k} = \sum_j \overline{p}_{ij}^k = 1. \quad [26]$$

Define the average equilibrium distribution within S_i over multiple ensembles as $\overline{\mu}_i'(x) = \sum_k w_i^k \overline{\mu}_i^k(x)$. According to the law of large numbers, we can get

$$\frac{1}{N_i}\sum_{x\in X\cap S_i} a(x) = \int a(x)\overline{\mu}_i'(x)dx \quad [27]$$

in the statistical limit for an arbitrary function $a(x)$. According to the above equation, we have the following:

$$\begin{aligned}
&\sum_{x\in X\cap S_i} \frac{\exp\left[\overline{f}_i^k - b^k(x)\right]}{\sum_l N_i^l \exp\left[\overline{f}_i^l - b^l(x)\right]} \\
&= \frac{1}{N_i}\sum_{x\in X\cap S_i} \frac{\exp\left[\overline{f}_i^k - b^k(x)\right]}{\sum_l w_i^l \exp\left[\overline{f}_i^l - b^l(x)\right]} \\
&= \int \frac{\exp\left[\overline{f}_i^k - b^k(x)\right]}{\sum_l w_i^l \exp\left[\overline{f}_i^l - b^l(x)\right]} \left(\sum_l w_i^l \overline{\mu}_i^l(x)\right) dx \\
&= \int_{S_i} \exp\left[\overline{f}_i^k - b^k(x)\right] \overline{\mu}(x) dx = 1.
\end{aligned} \quad [28]$$

From the above, we can conclude that in the statistical limit, the TRAM iterative algorithm converges to $v_i^k = c_i^k$ and $f_i^k = \overline{f}_i^k$, and the estimates of p_{ij}^k given by Eq. **16** are also equal to $\overline{p}_{ij}^k$ in the limit. Moreover, the corresponding estimated reference distribution is as follows:

$$\mu(x) = \frac{1}{\sum_k N_{i(x)}^k \exp\left[\overline{f}_{i(x)}^k - b^k(x)\right]}, \quad \text{for } x \in S_i, \quad [29]$$

and it satisfies that

$$\begin{aligned}
\mathbb{E}_\mu[a(x)] &= \sum_i \sum_{x\in X\cap S_i} \frac{a(x)}{\sum_l N_i^l \exp\left[\overline{f}_i^l - b^l(x)\right]} \\
&= \sum_i \frac{1}{N_i} \sum_{x\in X\cap S_i} \frac{a(x)}{\sum_l w_i^l \exp\left[\overline{f}_i^l - b^l(x)\right]} \\
&= \sum_i \int_{S_i} a(x)\ \overline{\mu}(x)dx = \mathbb{E}_{\overline{\mu}}[a(x)]
\end{aligned} \quad [30]$$

for any function $a(x)$ of the system configuration. So the discrete distribution $\mu(x)$ given by the TRAM algorithm is also a consistent estimate of the reference distribution $\overline{\mu}(x)$.

Proofs That TRAM Is a Generalization of Discrete TRAM, WHAM, MSMs, and MBAR.

MBAR/binless WHAM. Suppose that all simulations are in global equilibrium and there is only one configuration state S_1 for the whole configuration space, i.e., $S_1 = \Omega$. Then we can rewrite the TRAM equations [**13**] and [**14**] by dropping all of the subscripts as $v^k = c^k$ and

$$\sum_{x\in X} \frac{\exp\left[f^k - b^k(x)\right]}{\sum_l N^l \exp\left[f^l - b^l(x)\right]} = 1. \quad [31]$$

Eq. **31** is exactly the MBAR estimation equation for free energies f^k (49, 51, 66, 79).

Discrete (histogram-based) TRAM. Discrete (histogram-based) TRAM (56) can be expressed in the TRAM nomenclature by using bias potentials $b^k(x)$ that are step functions with

$$\mathrm{e}^{-b^k(x)} \equiv \gamma_i^k, \quad \text{for } x \in S_i. \quad [32]$$

Then, $\mu(x)$ in Eq. **17** takes a constant value $\mu_i = (\sum_k R_i^k \exp[f_i^k]\gamma_i^k)^{-1}$ on S_i, yielding the following estimate of the stationary probability of S_i in the unbiased ensemble:

$$\pi_i = N_i \mu_i. \quad [33]$$

Substituting Eqs. **32**, **33**, and **17** into the TRAM equation [**14**], we can obtain $\exp[-f_i^k]=\gamma_i^k\pi_i$ and rewrite Eq. **33** as follows:

$$\frac{N_i}{\pi_i}=\mu_i^{-1}=\sum_{k,j}\frac{\left(c_{ij}^k+c_{ji}^k\right)v_j^k\gamma_i^k}{\gamma_i^k\pi_i v_j^k+\gamma_j^k\pi_j v_i^k}+\frac{N_i}{\pi_i}-\frac{\sum_{k,j}c_{ji}^k}{\pi_i}$$
$$\Rightarrow\frac{\sum_{k,j}c_{ji}^k}{\pi_i}=\sum_{k,j}\frac{\left(c_{ij}^k+c_{ji}^k\right)v_j^k\gamma_i^k}{\gamma_i^k\pi_i v_j^k+\gamma_j^k\pi_j v_i^k}. \quad [34]$$

Eqs. **13** and **34** are identical to the self-consistent equations of discrete TRAM (56) with bias factors γ_i^k, which means that discrete TRAM is a special case of TRAM and applies if the bias energies can be discretized without error.

WHAM. From the discrete TRAM equations [**13**] and [**34**], we can further derive the WHAM equations under the assumption that the global equilibrium is achieved with $p_{ij}^k=\pi_j^k\propto\gamma_j^k\pi_j$ and $c_{ij}^k=\pi_i^k\sum_{j'}c_{ij'}^k$ (56).

MSMs. If simulations are only performed at one ensemble, the self-consistent equations for reversible maximum-likelihood estimation of MSMs are a special case of discrete TRAM (56).

Interpretation of the Effective Counts, R_i^k. Supposing that we apply MBAR only to the samples in a given configuration state S_i, the estimates of the local free energies $\{f_i^k\}$ are given by the following:

$$\sum_{x\in X_i}\frac{\exp\left(f_i^k-b^k(x)\right)}{\sum_l N_i^l\exp\left(f_i^l-b^l(x)\right)}=1. \quad [35]$$

This equation has the same form as the TRAM Eq. **14** except that R_i^k is replaced by N_i^k. Thus, we can interpret R_i^k as counts. By using Eqs. **15** and **16**, we find $R_i^k=N_i^k-\sum_j c_{ji}^k+\sum_j v_j^k p_{ji}^k$. $N_i^k-\sum_j c_{ji}^k$ is the number of visits to S_i in the initial frames of all trajectories. $\sum_j v_j^k p_{ji}^k$ can be interpreted as the corrected number of incoming transitions to S_i: First, $\sum_j v_j^k p_{ji}^k$ converges to $\sum_j c_{ji}^k$ in the limit of infinite statistics. Second, the term $N_i^k-\sum_j c_{ji}^k$ in R_i^k accounts for the first visit to S_i (which cannot be computed from the MSM alone). What remains to be included into R_i^k is the effective number of visits to S_i after the first state transition has happened. A suitable candidate would be the number of incoming transitions to S_i. What distinguishes $\sum_j v_j^k p_{ji}^k$ from $\sum_j c_{ji}^k$ is that the transition matrix is used in the computation of the former. Moreover, although $\sum_j c_{ji}^k$ and $\sum_j c_{ij}^k$ in principle are two independent variables, the quantities $\sum_j v_j^k p_{ji}^k$ and v_i^k (which can be interpreted as the corrected number of outgoing transitions) are linked by the equation $\sum_j v_j^k p_{ji}^k+v_i^k=\sum_j c_{ji}^k+\sum_j c_{ij}^k$ (which can be derived from Eq. **16**). So both $\sum_j v_j^k p_{ji}^k$ and v_i^k are counts that are corrected by the Markov model, which itself fulfills detailed balance.

Implementation Notes. In applications of this paper, we initialize the TRAM iteration with $v_i^k:=1$ and $f_i^k:=1$ as the convergence of TRAM does not seem to depend on the choice of initial point. We terminate the TRAM algorithm when the maximum change in normalized free energies $\max_{i,k}|f_i^k-f_i^{k,\text{new}}|<\text{tol}$ with tol being a small number (e.g., 10^{-10}). Considering that the TRAM equations are invariant with respect to a global shift $f_i^k\to\alpha+f_i^k$, we perform the normalization after every iteration such that $\sum_i\exp[-f_i^k]=1$ for the first ensemble $k=1$ to avoid an uncontrolled drift of the f_i^k.

The bias factors $\exp[-b^k(x)]$ can easily exceed the maximum range of double-precision floating point numbers, so we perform most calculations in log-space to avoid the numerical overflow or underflow. For all summations of the form $\log\sum_i\exp[a_i]$, we use the log-sum-exp formula $\log\sum_i\exp[a_i]=\hat{a}+\log\sum_i\exp[a_i-\hat{a}]$, where $\hat{a}=\max_i(a_i)$.

In addition, according to our experience, the convergence of the TRAM algorithm can be significantly sped up by adding an extra update step to each iteration that shifts local free energies f_i^k by δ_i as follows:

$$f_i^{k,\text{ new}}=f_i^k+\delta_i, \quad [36]$$

with

$$\delta_i=\ln\sum_{k,j}\frac{\left(c_{ij}^k+c_{ji}^k\right)v_j^k}{v_j^k+\exp\left[f_i^k-f_j^k\right]v_i^k}-\ln\sum_{k,j}c_{ji}^k. \quad [37]$$

Note that we can obtain from Eqs. **15** and **14** that

$$\begin{aligned}&\sum_{k,j}\frac{\left(c_{ij}^k+c_{ji}^k\right)v_j^k}{v_j^k+\exp\left[f_i^k-f_j^k\right]v_i^k}-\sum_{k,j}c_{ji}^k\\&=\sum_k R_i^k\cdot\sum_{x\in X\cap S_i}\frac{\exp\left[f_i^k-b^k(x)\right]}{\sum_l R_i^l\exp\left[f_i^l-b^l(x)\right]}-\sum_k N_i^k\\&=\sum_{x\in X\cap S_i}1-\sum_k N_i^k=0.\end{aligned} \quad [38]$$

Hence $\delta_i=0$ is a necessary condition for the TRAM equations, and the update step [**36**] does not influence the optimality of the limit of the algorithm.

ACKNOWLEDGMENTS. We are grateful to A. S .J. S. Mey for sharing the Alanine dipeptide simulations and to P. G. Bolhuis, J. D. Chodera, C. Clementi, G. De Fabritiis, G. Hummer, A. Laio, J.-H. Prinz, E. Rosta, B. Roux, B. Trendelkamp-Schroer, and G. A. Voth for enlightening discussions.

1. Jensen MO, et al. (2012) Mechanism of voltage gating in potassium channels. *Science* 336(6078):229–233.
2. Zhu F, Hummer G (2010) Pore opening and closing of a pentameric ligand-gated ion channel. *Proc Natl Acad Sci USA* 107(46):19814–19819.
3. Bernèche S, Roux B (2001) Energetics of ion conduction through the K^+ channel. *Nature* 414(6859):73–77.
4. Köpfer DA, et al. (2014) Ion permeation in K^+ channels occurs by direct Coulomb knock-on. *Science* 346(6207):352–355.
5. Kohlhoff KJ, et al. (2014) Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nat Chem* 6(1):15–21.
6. Dror RO, et al. (2011) Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc Natl Acad Sci USA* 108(32):13118–13123.
7. Nygaard R, et al. (2013) The dynamic process of β_2-adrenergic receptor activation. *Cell* 152(3):532–542.
8. Blood PD, Voth GA (2006) Direct observation of Bin/amphiphysin/Rvs (BAR) domain-induced membrane curvature by means of molecular dynamics simulations. *Proc Natl Acad Sci USA* 103(41):15068–15072.
9. Arkhipov A, Yin Y, Schulten K (2009) Membrane-bending mechanism of amphiphysin N-BAR domains. *Biophys J* 97(10):2727–2735.
10. Reubold TF, et al. (2015) Crystal structure of the dynamin tetramer. *Nature* 525(7569): 404–408.
11. Voelz VA, Bowman GR, Beauchamp K, Pande VS (2010) Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1-39). *J Am Chem Soc* 132(5): 1526–1528.
12. Silva DA, et al. (2014) Millisecond dynamics of RNA polymerase II translocation at atomic resolution. *Proc Natl Acad Sci USA* 111(21):7665–7670.
13. Plattner N, Noé F (2015) Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat Commun* 6:7653.
14. Sadiq SK, Noé F, De Fabritiis G (2012) Kinetic characterization of the critical step in HIV-1 protease maturation. *Proc Natl Acad Sci USA* 109(50):20449–20454.
15. Silva DA, Bowman GR, Sosa-Peinado A, Huang X (2011) A role for both conformational selection and induced fit in ligand binding by the LAO protein. *PLoS Comput Biol* 7(5):e1002054.

16. Shaw DE, et al. (2010) Atomic-level characterization of the structural dynamics of proteins. *Science* 330(6002):341–346.
17. Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC (2013) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J Chem Theory Comput* 9(9):3878–3888.
18. Harvey MJ, Giupponi G, Fabritiis GD (2009) ACEMD: Accelerating biomolecular dynamics in the microsecond time scale. *J Chem Theory Comput* 5(6):1632–1639.
19. Eastman P, et al. (2013) Openmm 4: A reusable, extensible, hardware independent library for high performance molecular simulation. *J Chem Theory Comput* 9(1): 461–469.
20. Shirts M, Pande VS (2000) Screen savers of the world unite! *Science* 290(5498): 1903–1904.
21. Buch I, Harvey MJ, Giorgino T, Anderson DP, De Fabritiis G (2010) High-throughput all-atom molecular dynamics simulations using distributed computing. *J Chem Inf Model* 50(3):397–403.
22. Schütte C, Fischer A, Huisinga W, Deuflhard P (1999) A direct approach to conformational dynamics based on hybrid Monte Carlo. *J Comput Phys* 151(1):146–168.
23. Swope WC, Pitera JW, Suits F (2004) Describing protein folding kinetics by molecular dynamics simulations: 1. Theory. *J Phys Chem B* 108:6571–6581.
24. Singhal N, Pande VS (2005) Error analysis and efficient sampling in Markovian state models for molecular dynamics. *J Chem Phys* 123(20):204909.
25. Sriraman S, Kevrekidis IG, Hummer G (2005) Coarse master equation from Bayesian analysis of replica molecular dynamics simulations. *J Phys Chem B* 109(14):6479–6484.
26. Noé F, Horenko I, Schütte C, Smith JC (2007) Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states. *J Chem Phys* 126(15):155102.
27. Chodera JD, Singhal N, Pande VS, Dill KA, Swope WC (2007) Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J Chem Phys* 126(15):155101.
28. Prinz JH, et al. (2011) Markov models of molecular kinetics: Generation and validation. *J Chem Phys* 134(17):174105.
29. Pérez-Hernández G, Paul F, Giorgino T, De Fabritiis G, Noé F (2013) Identification of slow molecular order parameters for Markov model construction. *J Chem Phys* 139(1): 015102.
30. Schwantes CR, Pande VS (2013) Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. *J Chem Theory Comput* 9(4):2000–2009.
31. Noé F, Clementi C (2015) Kinetic distance and kinetic maps from molecular dynamics simulation. *J Chem Theory Comput* 11(10):5002–5011.
32. Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR (2009) Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc Natl Acad Sci USA* 106(45):19011–19016.
33. Bowman GR, Voelz VA, Pande VS (2011) Atomistic folding simulations of the five-helix bundle protein λ(6–85). *J Am Chem Soc* 133(4):664–667.
34. Buch I, Giorgino T, De Fabritiis G (2011) Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc Natl Acad Sci USA* 108(25):10184–10189.
35. Tummino PJ, Copeland RA (2008) Residence time of receptor-ligand complexes and its effect on biological function. *Biochemistry* 47(20):5481–5492.
36. Torrie GM, Valleau JP (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J Comput Phys* 23:187–199.
37. Souaille M, Roux B (2001) Extension to the weighted histogram analysis method: Combining umbrella sampling with free energy calculations. *Comput Phys Commun* 135:40–57.
38. Marinari E, Parisi G (1992) Simulated tempering: A new Monte Carlo scheme. *Europhys Lett* 19:451–458.
39. Hansmann UH (1997) Parallel tempering algorithm for conformational studies of biological molecules. *Chem Phys Lett* 281:140–150.
40. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314:141–151.
41. Laio A, Parrinello M (2002) Escaping free-energy minima. *Proc Natl Acad Sci USA* 99(20):12562–12566.
42. Grubmüller H (1995) Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 52(3):2893–2906.
43. Hénin J, Chipot C (2004) Overcoming free energy barriers using unconstrained molecular dynamics simulations. *J Chem Phys* 121(7):2904–2914.
44. Hamelberg D, Mongan J, McCammon JA (2004) Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J Chem Phys* 120(24): 11919–11929.
45. Gumbart JC, Roux B, Chipot C (2013) Efficient determination of protein-protein standard binding free energies from first principles. *J Chem Theory Comput* 9(8): 3789–3798.
46. Tiwary P, Limongelli V, Salvalaglio M, Parrinello M (2015) Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proc Natl Acad Sci USA* 112(5):E386–E391.
47. Ferrenberg AM, Swendsen RH (1989) Optimized Monte Carlo data analysis. *Phys Rev Lett* 63(12):1195–1198.
48. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* 13:1011–1021.
49. Shirts MR, Chodera JD (2008) Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys* 129(12):124105.
50. Kong A, McCullagh P, Meng XL, Nicolae D, Tan Z (2003) A theory of statistical models for Monte Carlo integration. *J R Stat Soc Series B Stat Methodol* 65:585–604.
51. Bartels C (2000) Analyzing biased Monte Carlo and molecular dynamics simulations. *Chem Phys Lett* 331:446–454.
52. Rosta E, Woodcock HL, Brooks BR, Hummer G (2009) Artificial reaction coordinate "tunneling" in free-energy calculations: The catalytic reaction of RNase H. *J Comput Chem* 30(11):1634–1641.
53. Trendelkamp-Schroer B, Noé F (2016) Efficient estimation of rare-event kinetics. *Phys Rev X* 6:011009.
54. Wu H, Noé F (2014) Optimal estimation of free energies and stationary densities from multiple biased simulations. *Multiscale Model Simul* 12:25–54.
55. Mey ASJS, Wu H, Noé F (2014) xTRAM: Estimating equilibrium expectations from time-correlated simulation data at multiple thermodynamic states. *Phys Rev X* 4:041018.
56. Wu H, Mey ASJS, Rosta E, Noé F (2014) Statistically optimal analysis of state-discretized trajectory data from multiple thermodynamic states. *J Chem Phys* 141(21): 214106.
57. Rosta E, Hummer G (2015) Free energies from dynamic weighted histogram analysis using unbiased Markov state model. *J Chem Theory Comput* 11(1):276–285.
58. Minh DDL, Chodera JD (2009) Optimal estimators and asymptotic variances for non-equilibrium path-ensemble averages. *J Chem Phys* 131(13):134110.
59. Chodera JD, et al. (2011) Dynamical reweighting: Improved estimates of dynamical properties from simulations at multiple temperatures. *J Chem Phys* 134(24):244107.
60. Prinz JH, et al. (2011) Optimal use of data in parallel tempering simulations for the construction of discrete-state Markov models of biomolecular dynamics. *J Chem Phys* 134(24):244108.
61. Tiwary P, Parrinello M (2013) From metadynamics to dynamics. *Phys Rev Lett* 111(23): 230602.
62. Bowman GR, Beauchamp KA, Boxer G, Pande VS (2009) Progress and challenges in the automated construction of Markov state models for full protein systems. *J Chem Phys* 131(12):124101.
63. Trendelkamp-Schroer B, Wu H, Paul F, Noé F (2015) Estimation and uncertainty of reversible Markov models. *J Chem Phys* 143(17):174101.
64. Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE* 77:257–286.
65. Zhu F, Hummer G (2012) Convergence and error estimation in free energy calculations using the weighted histogram analysis method. *J Comput Chem* 33(4):453–465.
66. Tan Z, Gallicchio E, Lapelosa M, Levy RM (2012) Theory of binless multi-state free energy estimation with applications to protein-ligand binding. *J Chem Phys* 136(14): 144102.
67. Zhang BW, Xia J, Tan Z, Levy RM (2015) A stochastic solution to the unbinned WHAM equations. *J Phys Chem Lett* 6(19):3834–3840.
68. Tan Z (November 17, 2015) Optimally adjusted mixture sampling and locally weighted histogram analysis. *J Comput Graph Stat*, 10.1080/10618600.2015.1113975.
69. Gallicchio E, Andrec M, Felts AK, Levy RM (2005) Temperature weighted histogram analysis method, replica exchange, and transition paths. *J Phys Chem B* 109(14): 6722–6731.
70. Kass RE, Carlin BP, Gelman A, Neal RM (1998) Markov chain Monte Carlo in practice: A roundtable discussion. *Am Stat* 52:93–100.
71. Guillain F, Thusius D (1970) The use of proflavin as an indicator in temperature-jump studies of the binding of a competitive inhibitor to trypsin. *J Am Chem Soc* 92(18): 5534–5536.
72. Wojtas-Niziurski W, Meng Y, Roux B, Bernèche S (2013) Self-learning adaptive umbrella sampling method for the determination of free energy landscapes in multiple dimensions. *J Chem Theory Comput* 9(4):1885–1895.
73. Chipot C, Pohorille A (2007) *Free Energy Calculations* (Springer, Berlin).
74. Frenkel D, Smit B (2001) *Understanding Molecular Simulation: From Algorithms to Applications*. (Academic, Orlando, FL).
75. Scherer MK, et al. (2015) PyEMMA 2: A software package for estimation, validation and analysis of markov models. *J Chem Theory Comput* 11(11):5525–5542.
76. Geyer CJ (1992) Practical Markov chain Monte Carlo. *Stat Sci* 7:473–483.
77. Röblitz S, Weber M (2013) Fuzzy spectral clustering by PCCA+: Application to Markov state models and data classification. *Adv Data Anal Classif* 7:147–179.
78. Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge Univ Press, Cambridge, UK).
79. Vardi Y (1985) Empirical distributions in selection bias models. *Ann Stat* 13:178–203.