HHS Public Access

Author manuscript

J Chem Theory Comput. Author manuscript; available in PMC 2017 December 11.

Published in final edited form as:

J Chem Theory Comput. 2016 June 14; 12(6): 2983–2989. doi:10.1021/acs.jctc.6b00277.

Adaptive multilevel splitting method for molecular dynamics calculation of benzamidine-trypsin dissociation time

Ivan Teo[†,¶], **Christopher G. Mayne**[†], **Klaus Schulten**[†,¶], and **Tony Lelièvre**[‡]

[†]Beckman Institute, 405 N. Mathews Ave, Urbana, IL, USA

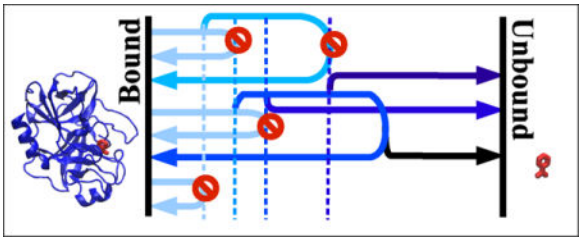
[‡]CERMICS, École des Ponts ParisTech, 6-8 avenue Blaise Pascal, 77455 Marne La Vallée, France

[¶]Department of Physics, University of Illinois at Urbana-Champaign

Abstract

Adaptive Multilevel Summation (AMS) is a rare event sampling method that requires minimal parameter tuning and that allows unbiased sampling of transition pathways of a given rare event. Previous simulation studies have verified the efficiency and accuracy of AMS in the calculation of transition times for simple systems in both Monte Carlo and molecular dynamics (MD) simulations. Now, AMS is applied for the first time to a MD simulation of protein-ligand dissociation, representing a leap in complexity from the previous test cases. Of interest is the dissociation rate, which is typically too low to be accessible to conventional MD. The present study joins other recent efforts to develop advanced sampling techniques in MD to calculate dissociation rates, which are gaining importance in the pharmaceutical field as indicators of drug efficacy. The system investigated here, benzamidine bound to trypsin, is an example common to many of these efforts. The AMS estimate of the dissociation rate was found to be $(2.6 \pm 2.4) \times 10^2$ s^{-1}, which compares well with the experimental value.

TOC image



Correspondence to: Tony Lelièvre.

Supporting Information Available

Supporting information for this article includes a description of how the MD simulations performed in this study were set up, an explanation of how AMS was implemented within the MD simulation software NAMD, and a description of how the average loop time and associated uncertainty were extracted from the simulation data. This material is available free of charge via the Internet at http://pubs.acs.org/.

1 Introduction

In simulations of physical processes, a powerful and expedient means of studying the behavior of the system is to project the trajectory of the system on a small number of reaction coordinates or order parameters, and sample along the reaction coordinates. The physical process in question is often a rare event with respect to the time scale of the simulation, characterized by high free energy barriers along the reaction coordinate(s), particularly in the case of molecular dynamics (MD). Rare event sampling techniques enable sampling of regions in the reaction coordinate that are normally difficult to access.

One class of rare event sampling techniques involves the setting up of branching points along the reaction coordinate, at which trajectories are initialized. At each branching point, the sampling method builds upon the reference prior probability of the branching point and focuses on sampling the posterior probabilities of trajectories emanating from the branching point to the next branching point, thus avoiding the difficulty of sampling a very small overall unconditional probability of the event. Members of this class of techniques include importance sampling,[1,2] transition interface sampling,[3,4] forward flux sampling,[5,6] and multilevel splitting.[10] Adaptive Multilevel Splitting (AMS)[11,12] is a variant of multilevel splitting, designed to minimize the need for prior knowledge, such as good choices of reference probabilities in importance sampling or branching locations and frequencies in transition interface sampling, forward flux sampling and multilevel splitting, by adaptively determining the branching points during the simulation. In contrast to most other rare event sampling methods, AMS does not require prior definition of branching point locations and frequencies, thus enabling easy implemention to a potentially high degree of automation even for processes that are highly complex and/or for which little information apart from the initial and final states is available.

The present study serves as a proof-of-principle of the applicability of AMS to MD simulations, in the context of measuring drug-target dissociation rates by using the benzamidine-trypsin complex as an example system. The drug binding rate has traditionally been a quantity of interest in the field of drug discovery, fueling numerous *in silico* studies of binding kinetics.[7–9,16,18,20] However, dissociation rates are increasingly considered an equally important indicator of drug efficacy[13,14] and efforts are being made to develop computational means of determining these rates.[15–18] The complexity of unbinding processes underlies the difficulty of obtaining dissociation rates. In the case of benzamidine-trypsin, previous computational studies[16–18] have identified multiple dissociation pathways and utilized Markov State Models[19,20] to characterize the entire dissociation process and obtain estimates of the overall dissociation rate. In the present study, AMS is applied along a simple reaction coordinate to estimate the dissociation rate. In contrast to the other computational studies, prior determination of pathways and specific metastable states was not required in the AMS calculation, but it is noted that such knowledge may be helpful in obtaining better convergence of results. It should also be noted that the retrospective reconstruction of pathways and metastabilities is possible through the reactive pathways obtained in the AMS algorithm,[12] but has not been undertaken in the present study.

This article is organized as follows. In the Methods section, a concise description of the basic AMS algorithm and theory is given, followed by issues of practicality that arise in adapting AMS to MD simulations, and finally the simulation of the benzamidine-trypsin system. The next section, Results and Discussion, begins with details on a prior equilibrium MD simulation for the purpose of exploration and parameter selection, followed by a discussion of the AMS simulation results, and ends with an error analysis. The error analysis addresses an auxiliary aim of the present study, which is to identify some common pitfalls in AMS and suggest solutions to mitigate them.

2 Methods

2.1 Adaptive Multilevel Splitting - Basic Algorithm

The output of AMS is the committor probability, defined as the probability that a system, after leaving a given initial state, reaches the given final state before returning to the initial state. Following the formulation by Cérou and Guyader,[11] let $\{Z_t\}$ be a Markov process along some continuous reaction coordinate z, with $Z_0 = z_0$. Let an event be defined by $Z_t = z_{\max}$ for some $z_{\max} > z_0$. Define also the committor probability P_C that a given realization of $\{Z_t\}$ exceeds $z_{\max}$ before returning to z_0 for $t > 0$, given that $Z_{t>0} \geq z_0$. The event is rare if P_C is small, namely, $P_C < 10^{-9}$.

The AMS algorithm begins with the initialization of N replica trajectory segments $\{Z_t^n\}$, $n = 1, \ldots, N$. Simulate the replicas until all of them have returned to z_0 (1a). Any of these replicas may also exceed $z_{\max}$, at which point the replica is stopped, but the probability is presumably negligible for such an event to occur within N replicas. Obtain the farthest point along z attained by each replica,

$$S_n^1 = \sup\left(Z_t^n\right), \quad (1)$$

and identify the minimum of these points,

$$q_1 = \min_n \left(S_n^1\right). \quad (2)$$

Note that at the k^{th} iteration, the proportion of surviving replicas, $1 - 1/N$, provides an estimate $\hat{p}_k$ of the conditional probability that a process starting at z_0 attains a supremum $S > q_k$, given that its supremum is greater than q_{k-1}, i.e. $P(S > q_k | S > q_{k-1})$. In the first iteration, $\hat{p}_1 = 1 - 1/N$ estimates simply the probability $P(S > q_1)$. The probability that a process starting at z_0 exceeds $z_{\max}$ before returning to z_0 is by definition the committor probability P_C, given by

$$P_C = P\left(S > z_{\max}\right) = P\left(S > q_1\right) \prod_{k=2}^{M-1} P\left(S > q_k \middle| S > q_{k-1}\right), \quad (3)$$

where M is the number of iterations taken by the AMS algorithm to reach completion.

It can be shown that the product $\hat{p}$ of estimators $\hat{p}_k$ of $P(S > q_k/S > q_{k-1})$ is itself an estimator of the committor probability P_C,[11] achieving equality in the $N \to \infty$ limit, where

$$\hat{p} = \prod_k \hat{p}_k = \left(1 - \frac{1}{N}\right)^M. \quad (4)$$

The algorithm is here presented for continuous time diffusion process. Slight modifications have to be made for discrete time processes where it is possible for more than one replica to reach the same supremum level.[23]

In an idealized setting (namely when the chosen reaction coordinate is the committor function associated with the two sets $A = \{z\,;\, z < z_0\}$ and $B = \{z\,;\, z > z_{max}\}$), it can be shown that the symptotic variance as $N \to \infty$ is:[21,22]

$$\mathrm{Var}\,(\hat{p}) = \frac{-P_C^2 \log P_C}{N} \quad (5)$$

which can be estimated in practice using $\frac{-\hat{p}^2 \log \hat{p}}{N}$, the square root of which provides an estimate of the uncertainty in $\hat{p}$. This estimate should be used with care since it is an asymptotic result and since, in practice, the reaction coordinate is not the committor function. However, it can be used to get a lower bound on the variance. We will discuss in the next paragraph another way to get a safer estimate of the variance.

In the interest of improving efficiency via parallelism, a few variations can be made to the original algorithm described above. Regardless of the choice of N, the mean of $\hat{p}$ is P_C,[23] so that instead of a single AMS simulation with large N, several smaller simulations can be run in parallel to obtain $\hat{p}$ to a similar degree of accuracy through simple averages. Additionally, obtaining multiple estimates of $\hat{p}$ allows for a safe estimation of the variance, by treating $\hat{p}$ itself as a random variable, without having to rely on the need for a large number of replicas through 5. Parallelism may also be incorporated into the algorithm itself, by re-initializing the $(k/N)^{\text{th}}$ quantile at each iteration, killing and restarting $k > 1$ replicas.[11,25] Nevertheless, it should be noted although the number of iterations can be reduced by using a larger quantile, the variance on the estimator of the probability P_C will also be larger, and it has been suggested that killing only one replica is the best compromise.[24]

2.2 Practical Implementation

The AMS simulations described in the present study run on two different time steps, namely the conventional MD time step and the interval between reaction coordinate measurements, which we term the AMS time step. The MD platform used in the present study, NAMD[26] (see Section S2 of Supporting Information for details on implementation), does not perform on-the-fly evaluation and comparisons of reaction coordinate values or replica kill-and-

restart operations without incurring a large computational overhead. Instead, reaction coordinate evaluation, as well as inter-replica communication and decisions, are performed at regular time intervals larger than the MD time step in the interest of computational efficiency. This modification does not affect the reliability of the AMS algorithm, which is therefore applied to the subsampled process considered at multiples of the AMS time step. Indeed, as shown in,[23] the AMS algorithm is unbiased for a discrete-in-time process.

It is also important to note that for physical systems, the initial condition is typically a collection of initial states occupying a continuum on the reaction coordinate, rather than a single value. In a scheme adapted from Cérou *et al*,[12] instead of starting the simulation at $z = z_0$, the replicas are initialized and allowed to reach quasi-equilibrium within the defined subspace A of initial states (hereafter called the initial metastable state), assumed to be characterized by the condition $z < z_0$. A value, $z_{\min} > z_0$, is chosen as discussed below, and the replicas are then evolved in time until every replica has reached $z_{\min}$, so as to obtain a representative distribution of trajectories up to the $z = z_{\min}$ hyperplane in configuration space. Thence, the AMS algorithm described above can proceed as shown in 2b. The final state B is similarly defined to be the subspace characterized by $z > z_{\max}$, but this definition does not affect the present implementation of the algorithm.

The reason for using initial conditions with reaction coordinate value greater than zmin>z0 is to avoid the situation where the replicas have a very small probability to reach B before A, leading to a small estimated committor probability that is typically difficult to estimate accurately. Thus, $z_{\min}$ should not be chosen too close to z_0. Suitable choices of z_0 and $z_{\min}$ can be determined heuristically from a quasi-equilibrium distribution of the system, such that z_0 and $z_{\min}$ are far enough from each other that the average trajectory from $z_{\min}$ to z_0 is resolvable given the AMS time step. Note that $z_{\min}$ is also bounded above by the requirement that direct simulation can adequately sample trajectory times to and from z_0 to $z_{\min}$ within reasonable computational time.

Another practical consideration is that the AMS algorithm, as described above, allows no parallelization after the initial preparation of replicas. Consider that while one replica is running, if the process of that replica surpasses the lowest supremum of the non-running replicas, then it is certain that the replica corresponding to that lowest supremum will be killed in the next iteration. Instead of waiting for the running replica to finish its course, the replica to be killed can be restarted immediately, thus achieving two parallel runs although conceptually, each iteration still corresponds to the killing of one replica. This parallelization process is easily generalized to the case of multiple currently running replicas. However, it should be noted that the maximum possible number of concurrently running replicas at any time depends on the stochastic occurrence of the current iteration exceeding the current lowest supremum. For example, during the AMS simulation in the present study, the number of running replicas at any one time ranged from 1 to 20. The fluctuating number of replicas is not easily implemented in traditional computing clusters (the present implementation relies on manual monitoring and initialization of replica runs) and would likely require sophisticated distributed computing resources for complete automation. An easier-to-implement and more extensive parallelism may be introduced, although not in the present

study, by running multiple independent AMS simulations and making use of the unbiasedness of the estimator $\hat{p}$, as discussed previously.

2.3 Calculation of mean first passage time and determination of AMS parameters

The committor probability obtained from AMS provides a means of estimating the mean first passage time from the initial state *A* to the final state *B*. The inverse of the committor probability $1/\hat{p}$ is an estimate of the expected number of non-reactive *A*-to-*A* trajectory segments that the system undergoes before a reactive *A*-to-*B* trajectory is observed. To make precise the notion of trajectory loops, define t_1 to be the time taken for a trajectory starting at z_0 to reach z_{min} and t_2 to be the time taken for a non-reactive trajectory starting at z_{min} to reach z_0. The expectation of the time taken for one loop is then $\overline{T}=(t_1+t_2)$. Additionally, define t_3 to be the time taken by a reactive trajectory, that is one starting at z_{min} and reaching z_{max} without first returning to z_{min}. The expected time for a reactive path is then (t_1+t_3).

The total time spent by the system in non-reactive trajectory loops is obtained by multiplying the number of trajectory loops by the average time per segment, $\overline{T}$. The time taken for the single reactive trajectory segment at the end then added to the time spent in the the non-reactive loops to produce the AMS estimate of the mean first passage time

$$\hat{\tau}=\frac{\overline{T}}{\hat{p}}+(t_1+t_3)\,. \tag{6}$$

An equilibrium simulation, separate from the AMS simulation, is performed to estimate $\overline{T}$. The distribution of loop times is obtained from a projection of the trajectory on the reaction coordinate *z*, thus providing an estimate of the average loop time and the associated uncertainty. Conveniently, the trajectory also provides the quasi-equilibrium distribution within the initial metastable state, from which the initial AMS replica states can be drawn and suitable values for the parameters z_0 and z_{min} can be chosen. z_{max} is chosen by other means, depending on the process being studied. In the present study, a steered MD pulling simulation was employed to obtain a suitable value.

The average time of a reactive trajectory segment, (t_1+t_3) can be obtained from a reconstruction of reactive paths by piecing together the successive trajectory segments traversed by each replica. The resulting collection of paths represent an unbiased distribution of reactive paths, from which the average reactive path time can be obtained. At present, the AMS implementation on NAMD does not retain trajectory segments, hence such a reconstruction cannot be undertaken. However, it is expected that the mean reactive path time is small compared to the time spent in unreactive loops. This assumption is retrospectively confirmed by the final AMS estimate of the mean first passage time being in at least the millisecond time scale, as compared to the sum of AMS trajectory times, which is in microseconds.

3 Results and discussion

3.1 Application to Benzamidine-Trypsin

Trypsin is a protease found in many vertebrate species. The complex of trypsin and competitive inhibitor benzamidine is a well studied exemplar of molecular binding and unbinding kinetics.[16–19,27,28] In particular, the rate of dissociation of the bound complex has been measured both experimentally[29] and computationally.[16–18] The benzamidine-trypsin complex was set up and pre-equilibrated for MD simulation as described in Section S1 of the Supporting Information.

For the AMS simulation, the reaction coordinate z was defined as the center-of-mass distance between Cαs of residues proximal to the binding site (D171, S172, C173, Q174, G175, D176, S177, V191, S192, W193, G194, G196, C197, A198, G204, V205) and benzamidine. The initial (bound) state is characterized by $z < z_0 = 1.6$ Å. z_{min} and z_{max} were chosen to be 1.7 Å and 15 Å respectively, as described in the following section. 3 provides visual examples of benzamidine conformations corresponding to the defined AMS levels, with the structure from a randomly picked equilibration frame displayed for reference.

3.2 Determination of AMS Parameters

A 130-ns equilibrium MD simulation of the benzamidine-trypsin complex was performed, with trajectory frames recorded at 0.1-ps time intervals. The projection of the trajectory on z and the normalized distributions of z values are shown in the inset and main figure of 4a respectively. The last ~ 30-ns portion of the trajectory (red) reveal the presence of at least one metastable state distinct from the one that the system was initialized in (green). Trajectory frames taken from each portion of the trajectory (4b) show that they correspond to proximal but distinct regions of benzamidine occupancy within the binding site.

Parameters for the AMS simulation were chosen as follows. z_0 was set at 1.6 Å, so that the initial state includes the apex of the distribution of the initial bound state, but not too far out so that loop times within the state are kept small and easily sampled. z_{min} was set at 1.7 Å, close to the value of z_0, but such that the AMS simulation was able to proceed without spending a great amount of time making loops that quickly return to z_0 before making progress along the z coordinate. Initial configurations for the AMS replicas were obtained by randomly drawing frames from the equilibrium simulation that satisfied $z < z_0$.

Ideally, the initial AMS ensemble should contain exclusively configurations in the initial metastable state. The overlap of other, less well-sampled, metastable states in the initial ensemble introduces some error in the estimation of the average loop time as well as the calculation of $\hat{p}$, since the statistical weight of the other states cannot be determined exactly by simulations of the present study. However, we will argue in a later section that the impact of other metastable states on the final result is small.

To obtain a suitable value for z_{max}, a constant velocity steered MD simulation[30] was performed. The average pulling force profile, shown in 5 provides a qualitative survey of the potential of mean force along the reaction coordinate. The force drops to zero at around $z =$

13 Å, thus providing a range at which benzamidine can be considered to have completely dissociated. The endpoint of the simulation, at z_{max} = 15 Å, is thus justified.

3.3 Determination of average loop times

The equilibrium simulation was also utilized to estimate the average time taken for the trajectory to loop from z_{min} to z_0 and back to z_{min}. Loops occurring in the two distinct portions of the equilibrium trajectory are considered separately. In the first portion, corresponding to the initial metastable state, loop times have the distribution shown in 6, across 2.8×10^4 loops. The estimated mean loop time $\hat{T}_{\mathrm{lo}}$ was calculated from the range defined by the lower and higher standard deviations about the median (described in Section S3 of Supporting Information, to account for the asymmetry of the distribution), giving $\hat{T}_{\mathrm{lo}} = (10 \pm 9)$ ps.

Similarly, the average loop time for the trajectory portion corresponding to other metastable states has a range of $\hat{T}_{\mathrm{hi}} = (250 \pm 249)$ ps, from a sample of size 804. An analysis of the relative statistical weights in the following section reveals that the overall average loop time is much closer to $\hat{T}_{\mathrm{lo}}$ than to $\hat{T}_{\mathrm{hi}}$, as reflected in 7.

3.4 AMS Results

Shown in 7 is a histogram of z-coordinates of branching points during the AMS simulation. A total of 20,376 branching points were required for the 1000 replicas to run to completion. The total simulation time over all replicas was about 2.1 μs. ?? yield the committor probability estimate $\hat{p} = (5.2 \pm 0.8) \times 10^{-10}$, where the uncertainty is just the root of the variance estimate given by 5.

Of all the branching points, 12,579 fell within the range 0 Å to 2.7 Å, roughly corresponding to the initial metastable state. 4 can again be applied to obtain the committor probability of the particle exceeding the heuristic boundary z = 2.7 Å of the initial states. This committor probability was found to be $\hat{p}' = 3.4 \times 10^{-6}$. In other words, any loop, on condition that it begins at z_{min}, has only a small probability $\hat{p}'$ of leaving the initial metastable state. Assuming that loops venturing outside the initial metastable state have an average loop time of $\hat{T}_{\mathrm{hi}}$, the overall average loop time is hence estimated as

$$\overline{T} = \hat{p}'\hat{T}_{\mathrm{hi}} + (1 - \hat{p}')\,\hat{T}_{\mathrm{lo}} = (10 \pm 9)\ \mathrm{ps}, \quad (7)$$

which is practically equal to the average loop time within the initial metastable state.

The dissociation time estimate is then obtained by applying $\hat{p}$ and $\overline{T}$ to 6 while assuming that the contribution from $(t_1 + t_3)$ is negligible. The dissociation time estimate thus obtained is $\hat{\tau} = 0.023 \pm 0.021$ s. The corresponding estimated dissociation rates, given by the reciprocal of the estimated dissociation time, is k_{off} = (260 ± 240) s^{-1}, within the same order of magnitude as the experimentally measured rate of 600 ± 300 s^{-1}.[29] The overall simulation time taken, summed over all 1000 replicas, was 2.1 μs (2.3 μs after including direct MD and

steered MD simulations), which is four orders of magnitude shorter than the estimated dissociation time of one event.

The dissociation rate estimates obtained in other computational studies of benzamidine-trypsin[16–18] differ from the experimental measurement and the present study by between one to two orders of magnitude. However, it should be noted that these studies treated the dissociation process more comprehensively, incorporating multiple distinct bound states that were not considered in the present study. The additional bound states could not have been sampled in the initial equilibration within the AMS initial state, which was shorter than the average transition times between the crystallographic state and the other bound states, reported to be on the order of microseconds.[16–18] In the same spirit, it should also be noted that the difference in results between the present and reference studies may be due in part to the use of the CHARMM36[31,32] force field in the present study, in contrast to the AMBER force fields employed in the other studies.

4 Conclusion

The results of the present study demonstrate the potential of AMS in the estimation of dissociation rates of protein ligand complexes, producing a k_{off} that compares very well with the experimental value. However, significant sources of error exist in the methodology and it cannot be ruled out that the excellent agreement of k_{off} with the experimental value is due to a serendipitous cancellation of errors. One such source of error was the omission of alternative initial bound states, as discussed in Results and Discussion. A related problem is the overlap of metastable states proximal to the initial bound state, which although detected during the AMS run, were not extensively sampled. Although it was argued that the impact of the overlap was small, measures to prevent such overlaps should nevertheless be undertaken in future AMS studies of complex systems. A number of remedies may be employed for this purpose.

One solution is to impose a more restrictive definition of the initial metastable state (i.e. lower z_0 and z_{min} levels), thereby excluding all others. Overlaps may also come about as artifacts of the choice of reaction coordinate. Cleanly separating two metastable states in sampling would then be a matter of selecting the right reaction coordinate. For this purpose, visual inspection of the equilibrium simulation is recommended as a good practice for AMS methodology. In cases where the metastable bound states can be characterized, detailed information about the system can be obtained by applying AMS piecewise to determine transition rates between the various states of the system in a Markov State Model, in similar fashion to the other studies referenced.[16–18] One may also consider the alternative approach of obtaining better sampling of the initial distribution of states. The major problem with having a mixture of metastable states within the initial distribution is that transitions between these states are typically too slow for a long equilibrium simulation to sample adequately. Enhanced sampling methods can be applied to obtain an accurate initial distribution of states instead of attempting to isolate one state. Such an approach is especially relevant in cases where significant overlap is too difficult to avoid.

It should also be noted that unaccounted-for sources of error, inherent in MD-based methods, were not incorporated into the uncertainty of the dissociation rate estimate. In particular, the use of non-polarizable force fields may significantly affect the interaction strength between ligand and substrate. Tiwary *et al*[17] suggest that the use of a non-polarizable force field results in an overestimate of the dissociation time. Polarizable force fields such as the Drude model[33,34] may be used to obtain more accurate descriptions of dissociation dynamics where solvation and other polarization-dependent effects can be foreseen to play a major role.

Nonetheless, the favorable result reported in the present study merits further developmental effort to address the present inadequacies and improve the technique; experimental evidence suggests that significant variation in drug efficacy occur with within-order-of-magnitude differences in residence times. For example, a study of a set of inhibitor compounds of the FabI enoyl reductase in mice infected with *Francisella tularensis* showed an approximate 1.2% increase in survival rate for each 1-min increase in residence time in the range of 20 to 140 mins.[35] In a study of A_{2A} receptor agonists, an almost twofold increase in efficacy was found for each increase in order of magnitude of residence time over the 1- to 100-minute range.[36] In order to further test the applicability of AMS in this respect, future efforts are required to address the challenges of initial state definition and loop time sampling encountered in the present study, as well as to validate AMS in other molecular systems. Various technical issues also need to be resolved, such as improving time resolution by enabling more frequent reaction coordinate queries, efficient communication between replicas, and increased parallelism of the algorithm.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The work of I. Teo, C. G. Mayne, and K. Schulten is supported by NIH Grant 9P41GM104601. The work of T. Lelièvre is supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement number 614492. The authors benefited from the scientific environment of the Laboratoire International Associé between the Centre National de la Recherche Scientifique and the University of Illinois at Urbana-Champaign.

References

1. Glynn PW, Iglehart DL. Management Sci. 1989; 35:1367–1392.
2. Perilla JR, Beckstein O, Denning EJ, Woolf TB. J Comput Chem. 2011; 32:196–209. [PubMed: 21132840]
3. van Erp TS, Moroni D, Bolhuis PG. J Chem Phys. 2003; 118:7762–7774.
4. van Erp TS, Bolhuis PG. J Comp Phys. 2005; 205:157–181.
5. Allen RJ, Warren PB, ten Wolde PR. Phys Rev Lett. 2005; 94:018104. [PubMed: 15698138]
6. Allen RJ, Valeriani C, ten Wolde PR. J Phys-Condens Mat. 2009; 21:463102.
7. Votapka, L., Amaro, RE. Multiscale Estimation of Binding Kinetics Using Brownian Dynamics, Molecular Dynamics and Milestoning. PLoS Comput Biol. 2015. In press. http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004381 (accessed 05/01/2016)
8. Shan Y, Kim E, Eastwood MP, Dror RO, Seeliger MA, Shaw DE. J Am Chem Soc. 2012; 133:9181–9183.

9. Hurst DP, Grossfield A, Lynch DL, Feller S, Romo TD, Gawrisch K, Pitman MC, Reggio PH. J Biol Chem. 2010; 285:17954–17964. [PubMed: 20220143]
10. Kahn H, Harris TE. National Bureau of Standards Appl Math Series 01. 1951; 12:27–30.
11. Cérou F, Guyader A. Stoch Anal Appl. 2007; 25:417–443.
12. Cérou F, Guyader A, Leliévre T, Pommier D. J Chem Phys. 2011; 134:054108. [PubMed: 21303093]
13. Lu H, Tonge P. J Curr Opin Chem Biol. 2010; 14:467–474.
14. Copeland RA. Nat Rev Drug Discov. 2015
15. Huang D, Caflisch A. PLoS Comput Biol. 2011; 7:e1002002. [PubMed: 21390201]
16. Buch I, Giorgino T, De Fabriitis G. Proc Natl Acad Sci USA. 2011; 108:10184–10189. [PubMed: 21646537]
17. Tiwary P, Limongelli V, Salvalaglio M, Parrinello MP. Natl Acad Sci USA. 2014; 112:E386–E391.
18. Plattner N, Noé F. Nat Commun. 2015; 6:7653. [PubMed: 26134632]
19. Noé F, Fischer S. Curr Opin Struct Biol. 2008; 18:154–162. [PubMed: 18378442]
20. Prinz J-H, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schütte C, Noé F. J Chem Phys. 2011; 134:174105. [PubMed: 21548671]
21. Guyader A, Hengartner N, Matzner-Løber E. Appl Math Optim. 2011; 64:171–196.
22. Cérou F, Del Moral P, Furon T, Guyader A. Stat Comput. 2012; 22:795–808.
23. Bréhier, CE., Gazeau, M., Goudenège, L., Lelièvre, T., Rousset, M. Unbiasedness of some generalized Adaptive Multilevel Splitting algorithms. 2015. arXiv:1505.02674 [math.PR]. arXiv.org ePrint archive. http://arxiv.org/abs/1505.02674 (accessed 01/20/2016). To appear in Ann. Appl. Probab
24. Bréhier L, Lelièvre T, Rousset M. ESAIM Proc Surv. 2015; 19:361–394.
25. Aristoff D, Lelièvre T, Mayne CG, Teo I. ESAIM Proc Surv. 2015; 48:215–225. [PubMed: 26005670]
26. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. J Comput Chem. 2005; 26:1781–1802. [PubMed: 16222654]
27. Essex JW, Severance DL, Tirado-Rives J, Jorgensen WL. J Phys Chem B. 1997; 101:9663–9669.
28. Resat H, Marrone TJ, McCammon JA. Biophys J. 1997; 72:522–532. [PubMed: 9017183]
29. Guillain F, Thusius D. J Am Chem Soc. 1970; 92:5534–5536. [PubMed: 5449454]
30. Lu H, Isralewitz B, Krammer A, Vogel V, Schulten K. Biophys J. 1998; 75:662–671. [PubMed: 9675168]
31. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. J Comput Chem. 1983; 4:187–217.
32. Huang J Jr, A DM. J Comput Chem. 2013; 34:2135–2145. [PubMed: 23832629]
33. Jiang W, Hardy DJ, Phillips JC Jr, A DM, Schulten K, Roux B. J Phys Chem Lett. 2011; 2:87–92. [PubMed: 21572567]
34. Lopes PEM, Huang J, Shim J, Luo Y, Li H, Roux B Jr, A DM. J Chem Theory Comput. 2013; 9:5430–5449. [PubMed: 24459460]
35. Lu H, England K, am Ende C, Truglio JJ, Luckner S, Reddy BG, Marlenee NL, Knudson SE, Knudson DL, Bowen RA, Kisker C, Slayden RA, Tonge PJ. ACS Chem Biol. 2009; 4:221–231. [PubMed: 19206187]
36. Guo D, Mulder-Krieger T, IJzerman AP, Heitman LH. Br J Pharmacol. 2012; 166:1846–1859. [PubMed: 22324512]

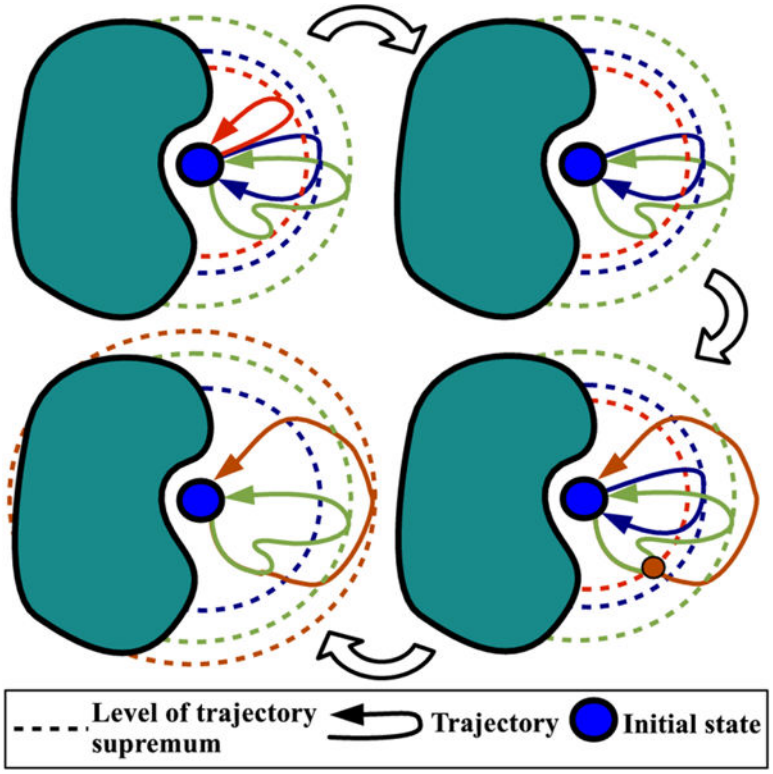


Figure 1.

Schematic of basic AMS algorithm for a dissociation process of a ligand in an initially bound state from a binding site (blue-green). In this case, $N = 3$ replicas are used and the reaction coordinate is defined as the radius about the initial state. In **(a)**, an initial trajectory segment is generated for each replica. The suprema of the segments are compared, and the replica with the segment of lowest supremum (red) is killed, as shown in **(b)**. Subsequently in **(c)**, a surviving replica is randomly picked (green in this case); its trajectory segment up to the supremum of the killed replica is cloned into the killed replica and simulation is restarted until the trajectory returns to the initial state. Once again, the replica which has the least progress along the reaction coordinate (blue) is identified and killed, as shown in **(d)**. The process is repeated until all replicas have surpassed z_{max} (not shown).

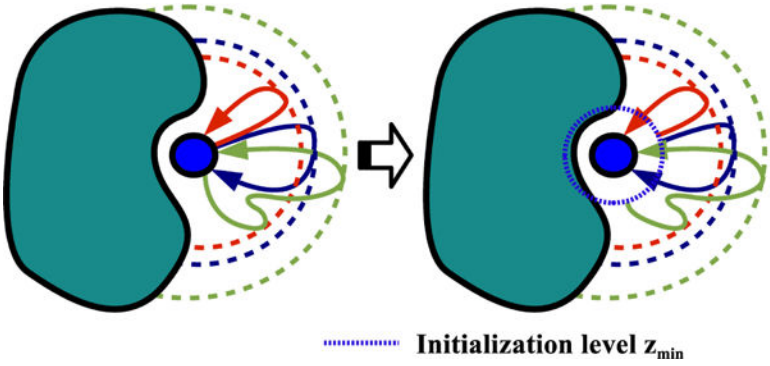


Figure 2.

To avoid being stuck in small trajectory loop segments near the starting point, the first step of the original algorithm, shown in **(a)**, is altered to incorporate a starting point z_{min}, shown in **(b)**, a small distance away from z_0. Replica trajectory segments now begin at z_{min}, but still terminate at z_0.

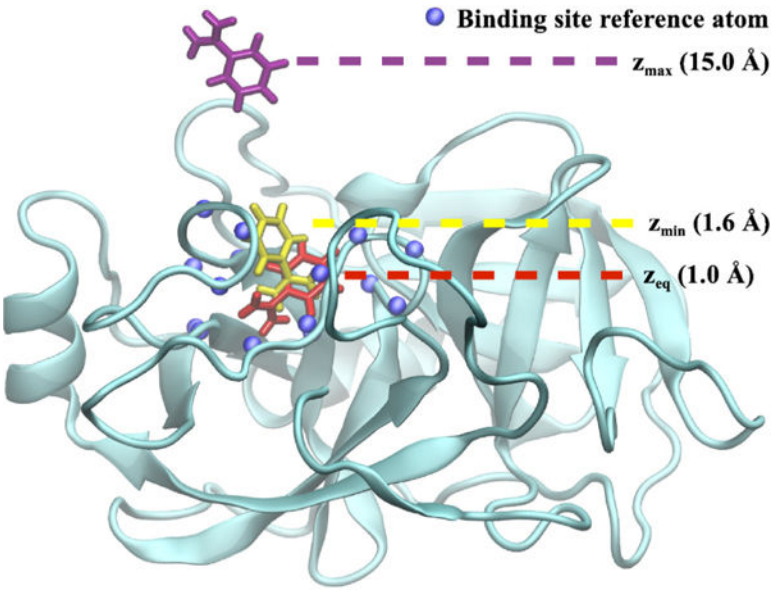


Figure 3.

Trypsin (light blue) and benzamidine in a randomly picked equilibration frame at reaction coordinate z_{eq} (red), AMS initial (yellow) bound states, and AMS final unbound state (purple). The reaction coordinate z is defined as the center-of-mass distance between non-hydrogen benzamidine atoms and the Cα atoms of 16 residues near the binding site (blue spheres). AMS initial and final state structures were extracted from trajectory frames during AMS simulation.

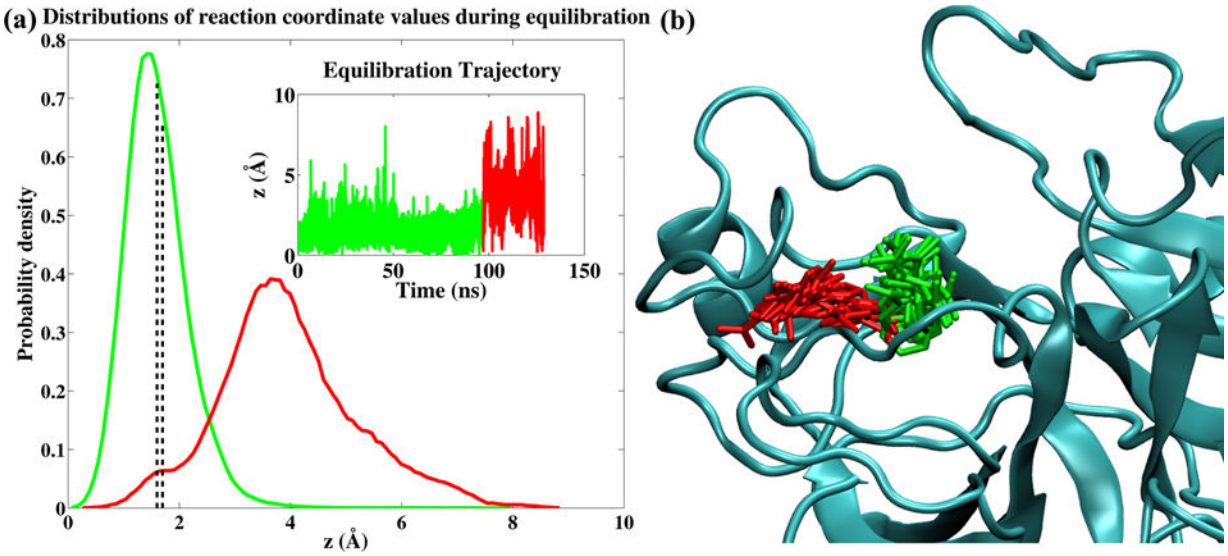


Figure 4.

(a) Normalized sample distributions of z values within initial (green) and other (red) metastable portions of the trajectory during equilibrium simulation. The initial conditions of the AMS simulation are defined by z_0 and z_{min} levels, indicated by dashed lines. The inset shows the time evolution of the z value. **(b)** Simulation trajectory frames of benzamidine corresponding to initial (green) and other (red) metastable state portions, each 20 ns long, within the binding site on trypsin (grey).

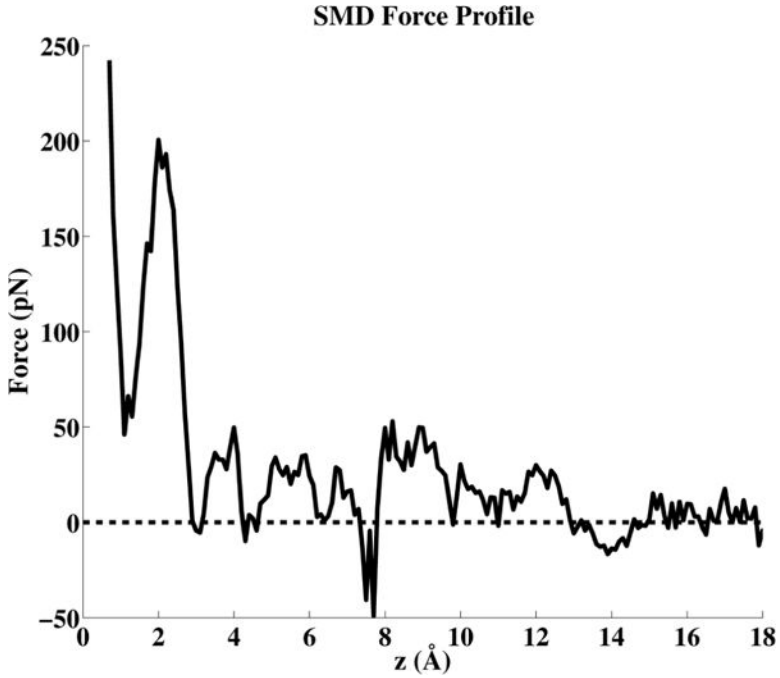


Figure 5.

Benzamidine is gradually pulled away from trypsin (details in Section S1 of the Supporting Information). The force profile height reflects the amount of resistance against the pulling force, which drops to near zero when benzamidine is far enough to escape the influence of trypsin. While being a crude measurement of the potential of mean force, this calculation adequately serves as a quick and simple means of locating the unbound state along the reaction coordinate. Note the correspondence of the force peak around $z = 2$ Å to the potential of mean force barrier of the initial metastable state.

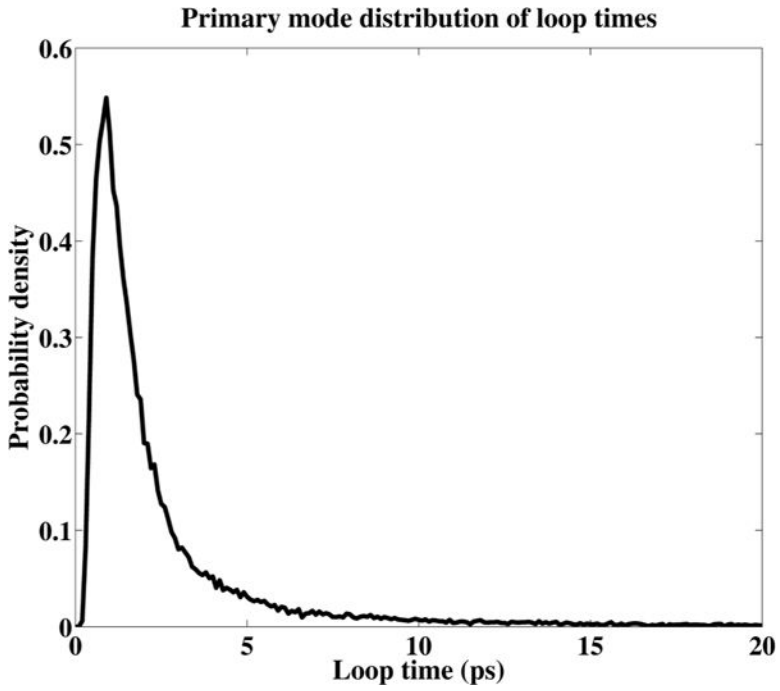


Figure 6.

Loop times are defined as the time taken for the system, starting from z_{min}, to reach z_0 and return to z_{min} without reaching z_{max} first. The distribution of loop times shown above was obtained from the portion of the equilibrium trajectory corresponding to the initial metastable state. The average loop time is extracted from the distribution, as described in Section S3 of the Supporting Information, as a necessary step in calculating the dissociation rate.

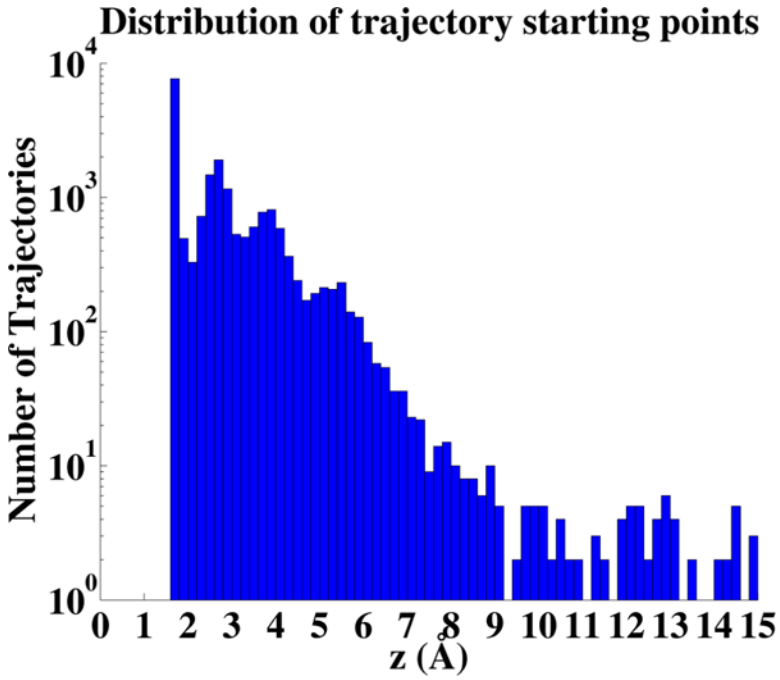


Figure 7.

AMS trajectories are histogrammed by branching point *z* coordinates. Displayed with a logarithmic scale in the *y* axis, the histogram clearly shows a large concentration of loops about the initial metastable state.