Title

SEEKR: Simulation Enabled Estimation of Kinetic Rates, A Computational Tool to Estimate Molecular Kinetics and Its Application to Trypsin–Benzamidine Binding

Journal

The Journal of Physical Chemistry B, 121(15)

ISSN

1520-6106

Authors

Votapka, Lane W
Jagger, Benjamin R
Heyneman, Alexandra L
et al.

Publication Date

2017-04-20

DOI

10.1021/acs.jpcb.6b09388

Peer reviewed

HHS Public Access

Author manuscript

J Phys Chem B. Author manuscript; available in PMC 2018 April 20.

Published in final edited form as:

J Phys Chem B. 2017 April 20; 121(15): 3597–3606. doi:10.1021/acs.jpcb.6b09388.

SEEKR: Simulation Enabled Estimation of Kinetic Rates, A Computational Tool to Estimate Molecular Kinetics and its Application to Trypsin-Benzamidine Binding

Lane W. Votapka[†,‡,§], **Benjamin R. Jagger**[‡,§], **Alexandra L. Heyneman**[‡], and **Rommie E. Amaro**[‡,*]

[†]Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY 11794

[‡]University of California San Diego, 9500 Gilman Dr., La Jolla, CA 92093

Abstract

We present the Simulation Enabled Estimation of Kinetic Rates (SEEKR) package, a suite of open-source scripts and tools designed to enable researchers to perform multi-scale computation of the kinetics of molecular binding, unbinding, and transport using a combination of molecular dynamics, Brownian dynamics, and milestoning theory. To demonstrate its utility, we compute the k_{on}, k_{off}, and ΔG_{bind} for the protein trypsin with its noncovalent binder, benzamidine, and examine the kinetics and other results generated in the context of the new software, and compare our findings to previous studies performed on the same system. We compute a k_{on} estimate of $2.1\pm0.3\bullet10^{7}$ $M^{-1}s^{-1}$, a k_{off} estimate of 83 ± 14 s^{-1}, and a ΔG_{bind} of -7.4 ± 0.2 kcal•mol^{-1}, all of which compare closely to the experimentally measured values of $2.9\bullet10^{7}$ $M^{-1}s^{-1}$, 600 ± 300 s^{-1}, and −6.7 kcal•mol^{-1}, respectively.

Graphical abstract

ramaro@ucsd.edu, Phone: 1 (858) 534-9629.

[§]Author Contributions

LWV and BRJ contributed equally to this work. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Introduction

Elucidating the kinetics and thermodynamics of binding and unbinding processes between a biomolecule and a substrate remains an important challenge in the field of molecular biophysics. Countless processes within the cell involve the association of a biomolecule with a metabolite, signaling molecule, toxin, drug, or another biomolecule.[1–2] Many of these interactions have important kinetic considerations: for instance, the speed of reactions or the residence time of an intermolecular encounter.[2–3]

Significant effort has been expended to accurately estimate the thermodynamics of binding using a variety of methods, particularly in the field of drug discovery, where the identification of a tight binder is an integral step towards obtaining a potential drug molecule that would accomplish a desired medical result.[4–8] While the thermodynamics of binding, encapsulated in the quantity of the free energy ΔG_{bind} of receptor-ligand complex formation, is an important factor in the binding process, a comprehensive understanding of the binding process requires consideration of binding kinetics and reaction rates.

Many theoretical approaches and simulation methods have been used to estimate both the thermodynamics and kinetics of binding. For instance, specialized machinery and long molecular dynamics (MD) simulations can be used in a ‘brute force’ approach, although it is relatively costly compared to other methods.[9–12] Markov models[13–19] can also be used to investigate the kinetics of binding,[20–22] as can milestoning.[7] Additional clever methodologies can be used to speed the computation using MD.[6, 8, 23–25] Brownian dynamics (BD) can also be used to approach the problem of binding kinetics,[5, 26–29] as can Smoluchowski equation solvers.[30]

Our past work[31–33] has focused on using a multi-scale combination of MD and BD, unified through the theoretical framework of milestoning. In our previous study, we presented a hybrid MD/BD/milestoning methodology to conduct our investigations into the kinetics of binding between superoxide dismutase and its natural substrate, the superoxide anion, and between troponin C and its natural substrate, the calcium ion.[31] Here, we make available a software package, SEEKR, that implements this method with significant improvements in automation, usability, and analysis. We demonstrate the utility of SEEKR by applying it to estimate the k_{on}, the k_{off}, and ΔG_{bind} between the serine protease trypsin and its ligand, benzamidine. In addition to the SEEKR software to perform milestoning calculations on any receptor-ligand system, we also make available a user guide, tutorial, and workflow to allow users to repeat our simulations and analysis for the trypsin-benzamidine system, and compute kinetics and thermodynamics for additional receptor-ligand systems.

Theory

The rationale and methodology behind our usage of milestoning to estimate kinetics using both MD and BD has been described recently in detail[31] for multiple applications. Our implementation to the trypsin-benzamidine receptor-ligand system in this study was adapted with few changes, the majority consisting of improvements in software efficiency.

In the case of bimolecular association, the kinetics of binding and unbinding can be represented respectively by two quantities, k_{on} and k_{off}, which are frequently depicted according to the following equation:

$$A+B \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} AB \quad \text{Eq. 1}$$

Which is shorthand for specifying that the values k_{on} and k_{off} function as parameters within the following differential equations:

$$\frac{d[AB]}{dt}=k_{on}[A][B]-k_{off}[AB] \quad \text{Eq. 2}$$

$$\frac{d[A]}{dt}=k_{off}[AB]-k_{on}[A][B] \quad \text{Eq. 3}$$

$$\frac{d[B]}{dt}=k_{off}[AB]-k_{on}[A][B] \quad \text{Eq. 4}$$

Where *[A]*, *[B]*, and *[AB]* represent the concentrations of chemical species *A*, *B*, and their complex *AB*. The k_{on} and k_{off} relate to the dissociation constant K_D, and by extension, a free energy of association ΔG_{bind}.[6]

$$\frac{k_{off}}{k_{on}}=K_D=K_{\ominus}e^{\Delta G_{bind}/RT} \quad \text{Eq. 5}$$

Where R is the gas constant, T is temperature, and $K_{\ominus}$ is a factor equal to one, in units of concentration.

The theory of milestoning has been formulated to compute kinetic and thermodynamic details of a process if the states of that process are represented as carefully chosen surfaces in phase space. These surfaces are known as “milestones”.[34] In this study, the milestones are represented as concentric spherical shells (figure 1) that encapsulate the binding site of the receptor. These spherical milestones are used for the computation of k_{on}, k_{off}, and ΔG_{bind}. Milestoning theory allows us to approach the problem of kinetics by utilizing a multi-scale strategy. We use highly-detailed, but computationally expensive MD simulations to observe transitions between milestones closer to the binding site so that molecular flexibility will be a component of the transitions between milestones. We then use BD for the larger and more widely-spaced milestones far from the binding site, where fast sampling of long trajectories is required and rigid body dynamics and implicit solvent are adequate[21, 28, 35] to model

transition times and probabilities. In this way, we take advantage of fully flexible MD where molecular flexibility is required, and also take advantage of the computation efficiency of BD where molecular flexibility is less important. Milestoning is the theory that combines the MD and BD components, by allowing statistics to be obtained in each regime independently, and then unifying the statistics through a rigorous theory that is agnostic to the method that was used to obtain them. Since the statistics of each milestone are obtained independently from the others, and since milestoning theory is a robust framework that can utilize information obtained by either Brownian or Newtonian dynamics,[36] we can choose whichever simulation method is most appropriate and convenient for that milestone.

By sampling transition statistics and times between the milestones using numerous short simulations, one can construct a transition kernel **K** that represents the transition probabilities and an incubation time vector $\langle \boldsymbol{t} \rangle$ that represents the average times of a system traversing the milestones.[37–38]

The transition kernel **K** is a square matrix whose elements are constructed according to the following formula:

$$K_{ij} = n_{i \to j} \Big/ \sum_k n_{i \to k} \quad \text{Eq. 6}$$

Where $n_{i \to j}$ is the number of trajectories that begin at a given milestone *i* and end at an adjacent milestone *j*. And the incubation time vector $\langle \boldsymbol{t} \rangle$ has elements that are constructed according to the following formula:

$$\langle t \rangle_i = \sum_l t_l \Big/ \sum_k n_{i \to k} \quad \text{Eq. 7}$$

Where t_l is the time of the *l*th successful forward trajectory starting at milestone *i*, and $n_{i \to k}$, as before, is the number of trajectories beginning at milestone *i* and ending at milestone *k*. Therefore, $\langle t \rangle_i$ represents the average time spent by the system after crossing *i* and before crossing any other milestone.

In order to compute a free energy profile along the milestones, we must first obtain the stationary flux vector $\mathbf{q_{stat}}$ along the milestones by computing the principle eigenvector of **K**.

$$\mathbf{K} \cdot \mathbf{q_{stat}} = \mathbf{q_{stat}} \quad \text{Eq. 8}$$

Then $\mathbf{q_{stat}}$ must be multiplied elementwise by $\langle \boldsymbol{t} \rangle$ to find the stationary probability vector $\mathbf{p_{stat}}$.

$$p_{stat,i}=q_{stat,i}\cdot\langle t\rangle_i \quad \text{Eq. 9}$$

Finally, $p_{stat,i}$ relates to the relative free energy ΔG_i at milestone i according to the following:

$$\Delta G_i=-RT\ln\left(p_{stat,i}/p_{stat,ref}\right) \quad \text{Eq. 10}$$

Where the index of $p_{stat,ref}$ is any reference state, such as the lowest energy, bound state. The value of $p_{stat,ref}$ is found by applying Eq. 9 to the chosen reference state.

To compute the k_{on}, we utilize the formula that is also used in BD theory[26]:

$$k_{on}=k(b)\beta \quad \text{Eq. 11}$$

Where $k(b)$ is computed using the following formula:

$$k(b)=\left[\int_b^\infty \frac{e^{W(r)/k_BT}}{4\pi r^2 D(r)}dr\right]^{-1} \quad \text{Eq. 12}$$

The value $k(b)$ represents the rate constant at which the ligand particles are crossing the b-surface, $W(r)$ and $D(r)$ are the potential of mean force and diffusion coefficient, respectively, that the ligand experiences at a distance r from the center of the receptor beyond the b-surface.[26] $D(r)$ is computed by generating a Rotne-Prager diffusion tensor to approximate the hydrodynamics of a two body interaction in a viscous medium.[39] The value $k(b)$ is computed automatically in BrownDye.

To find β, which represents the proportion of ligands crossing the b-surface that continue on to bind to the receptor, a starting probability vector $\mathbf{q_0}$ must be obtained in BD simulations by running a large number of conventional BD simulations where ligand molecules are started on a b-surface surrounding a receptor molecule. As the simulations run, and the proportion of trajectories that touch the outermost milestone(s) that encompasses a binding site on the biomolecule, rather than escaping to an infinite distance, are counted. In this case, $\mathbf{q_0}$ becomes:

$$\mathbf{q_0}=[0,\ldots,0,q_{0,i},0,\ldots,0,q_{0,j},0,\ldots,0,q_{0,\infty},0,\ldots,0]^T \quad \text{Eq. 13}$$

Where i and j are the indices of one or more of these outermost site-encompassing milestones, $q_{0,i}$, $q_{0,j}$, are the probabilities that a BD trajectory started on the b-surface descend and touch these milestones, and $q_{0,\infty}$ is the probability that a trajectory diffuses away to an infinite distance. All the entries in $\mathbf{q_0}$ must be normalized such that their sum

equals a value of one. An “infinity” state in both vector $\mathbf{q_0}$ and in matrix **K**, represents the condition in which the ligand has escaped to an infinite distance from the receptor.

Next the transition matrix **K** must be modified to a new matrix $\hat{\mathbf{K}}$ such that the milestones representing the bound and “infinity” states are sink states. That is, they all must have a probability of one that they transition only to themselves, and a zero probability to transition to anything else.

$$\hat{K}_{ii}=1;\text{ if } i \text{ is a bound state, or the “infinity”state} \quad \text{Eq. 14a}$$

$$\hat{K}_{ij}=0;\text{if also } i \neq j \quad \text{Eq. 14b}$$

Once $\hat{\mathbf{K}}$ and $\mathbf{q_0}$ are properly defined, we compute the static flux vector[21] $\mathbf{q_\infty}$.

$$\mathbf{q}_\infty=\lim_{a\to\infty}\hat{\mathbf{K}}^a\cdot\mathbf{q}_0 \quad \text{Eq. 15}$$

Finally, we obtain β:

$$\beta=\sum_i q_{\infty,i} \quad \text{Eq. 16}$$

Where i is the index of one of the bound states.

To compute the k_{off}, we must return to the initial definition of matrix **K** as specified in Eq. 6. But it must be modified by introducing a “draining” state i by changing **K** into a draining matrix $\tilde{\mathbf{K}}$ according to the following:

$$\tilde{K}_{ij}=0,\forall j \quad \text{Eq. 17}$$

That is, once we have decided that i is the draining state, we set that entire column of the matrix $\tilde{\mathbf{K}}$ to zeros, while all other columns are kept the same as they were in **K**. In the SEEKR implementation, the outermost non-infinite milestone is considered to be the draining state. Then, we compute a mean first passage time (MFPT) τ:

$$\tau=\mathbf{p_0}\left(\mathbf{I}-\tilde{\mathbf{K}}^{\boldsymbol{T}}\right)^{-\mathbf{1}}\langle\mathbf{t}\rangle \quad \text{Eq. 18}$$

Where $\mathbf{p_0}$ is a starting distribution of probabilities along each milestone, and $\tilde{\mathbf{K}}^{\boldsymbol{T}}$ is the transpose of matrix $\tilde{\mathbf{K}}$. We set $p_{0,i}$ to be 1 if i was a bound state, and set $p_{0,i}$ to be equal to 0

otherwise. The MFTP τ is equivalent to a residence time of the ligand within the binding site, and can be related to the k_{off} according to the following relation:

$$k_{off} = \frac{1}{\tau} \quad \text{Eq. 19}$$

Materials and Methods

Description of the SEEKR package

SEEKR is a collection of scripts and files designed to automate the preparation and analysis of ligand-receptor kinetic calculations that use a multi-scale MD/BD/milestoning framework.

SEEKR does not run the simulations themselves, but instead relies on the well-established NAMD[40] and BrownDye[41] programs. In this case, SEEKR is more of a specialist interface or tool that automates the cumbersome process of preparing, running, and analyzing a particular type of multi-scale milestoning calculation so that researchers will be able to run them more easily than if the process were done manually.

SEEKR programs are classified into three general categories:

1. Preparation: These scripts and modules accept input from the user in order to construct all the necessary files needed by both NAMD and BrownDye to run their respective simulations. The files are organized into a file tree whose branches represent the various independent milestones, which simulation method is being used (MD or BD), and the various stages of the calculations. When run, the user will have all the required files arranged and poised for simulation and milestoning calculations.
2. Running: Other scripts aid the user in running the MD and BD simulations locally and on supercomputers. For instance, SEEKR contains a script to prepare the submission of the computationally-intensive MD simulation jobs to a SLURM supercomputer queue, and when the allotted time runs out, the script prepares all the necessary resubmission files for one, some, or all of the milestones with a single command. Other scripts use previous BD trajectory output to prepare and run ensembles of BD simulations from first hitting point distributions (FHPD).
3. Analysis: When all the simulations are complete, the user can run an analysis script that descends into the file tree, gathering all the simulation output. It then combines this information to construct the milestoning model, and performs all the milestoning and error calculations, providing the user with kinetic and thermodynamic information, including k_{on}, k_{off}, and the free energy profile. It also has the option to perform convergence analysis on these values. Additional analysis scripts can be utilized to generate a single file containing the ligand equilibrium distribution or FHPD of each milestone for easy visualization.

The Python scripts have been tested using Python 2.7 and can be safely run in any version of Python 2 at version 2.7 or later. The remainder of the scripts are written in TCL, particularly those interfacing with NAMD, which has a TCL-based interface. SEEKR also uses the Numpy, Scipy, and MDAnalysis python libraries. The Adaptive Poisson-Boltzmann Solver (APBS)[42] is used to generate the electrostatic potential maps for input to BrownDye, and the AmberTools program LEaP[43] is also used to prepare structures for MD simulation.

Trypsin structure preparation and SEEKR creation of milestoning structures

Atomic coordinates of the trypsin-benzamidine system were obtained from the high resolution crystal structure Protein Databank (PDB) ID: 3PTB.[44] Hydrogens were added using Molprobity with ring flips allowed.[45–46] The system was then further prepared using LEaP with the Amber forcefield, ff14SB.[47] Disulfide bonds were added manually. The appropriate protonation states of ASP, GLU, and HIS residues at a pH of 7.7 were determined using PROPKA.[48–49] This pH was selected to align with the experimental conditions of Guillian and Thusius[50]. The structure was then solvated in a truncated octahedron of TIP4Pew[51–52] waters and eight Cl^- ions were added to neutralize the overall charge. The benzamidine ligand was parameterized using Antechamber with the GAFF force field.[52–53] The total size of the system was approximately 23,000 atoms. To allow for relaxation from the crystallographic starting structure, the benzamidine ligand was removed and a 20 ns simulation of the apo structure was performed at a constant temperature of 298 K using the Langevin thermostat and a constant pressure of 1 atm using the Langevin piston with a damping coefficient of 5 ps^{-1}. A representative structure from this simulation was then used as the SEEKR input structure to generate all the necessary inputs for the MD simulations to be run using NAMD, and the BD simulations using Browndye.

The benzamidine bound-state coordinates were defined from the center of mass of the alpha carbons of residues 190, 191, 192, 195, 213, 215, 216, 219, 220, 224, 228 of PDB: 3PTB because these residues form the binding pocket in the bound-state crystal structure by manual inspection. Spherical milestones were defined with radii of 1, 1.5, 2, 3, 4, 6, 8, 10, 12, 14 Å, with the origin being the bound state coordinates defined above. This spacing of the milestones was chosen to facilitate the simulation of transitions between milestones while still ensuring the Markov assumptions required by formal milestoning theory. Ten copies of the apo structure were generated, each with the benzamidine ligand inserted on one of the ten spherical milestones (figure 2A). Water molecules that clashed with the ligand structure were removed. The first nine milestones correspond to the MD simulation regime, with the innermost milestone (1 Å) representing the bound state, as the center of mass of the bound benzamidine ligand falls well within the 1 Å sphere that defines this milestone (figure 2B). Furthermore, in a ~170 ns unrestrained MD simulation with the ligand in the bound pose, the 1 Å sphere contained the center of mass of the ligand over 71% of the simulation. The tenth and outermost milestone (14 Å) corresponds to the BD simulation regime. The distribution along any milestone where BD was started was constructed by first running conventional BD simulations and obtaining the distribution of hitting points along that milestone.

The *b*-surface is a relatively large spherical shell that encloses the entire receptor molecule, with a radius of sufficient size that the entire surface sits well out into the bulk solvent where forces between the ligand and receptor would be largely unaffected by molecular orientation, and are therefore centrosymmetric.

MD simulations

A modified version of NAMD 2.11 was used for all MD calculations. The numerous MD inputs, including input files, integrator parameters, boundary conditions, temperature and pressure controls, etc. are either defined by the user or set by SEEKR to default values. Relevant settings and procedures implemented for each milestone in the MD regime are described here.

For each milestone system generated by SEEKR as described above, the solvent molecules were allowed to relax around the newly placed benzamidine ligand by minimizing for 5000 steps with both the ligand and receptor restrained. The solvent was then further relaxed through a series of 2 ps heating simulations, where the temperature was increased from 298 K to 350 K and then cooled back to 298 K in 10 K increments, keeping the atoms of the ligand and receptor restrained. Following this relaxation of the solvent, an equilibrium distribution of the ligand on the milestone surface was obtained from 1 μs of constant volume simulation at a temperature of 298 K where a harmonic spring force of 90 $kcal \bullet mol^{-1} \bullet Å^{-2}$ was imposed to restrain the ligand at the appropriate radius from the binding site center for each milestone to generate an equilibrium distribution (figure 3A). This is also known as the umbrella sampling stage. From this equilibrium distribution, a FHPD (figure 3B) was obtained by selecting 4700 position and velocity configurations from times 60 ns – 1 μs of the equilibrium trajectory and allowing them to propagate backwards in time by reversing their velocities at constant energy and volume (reverse stage). Any trajectories that struck another milestone before re-crossing the milestone from which they originated were counted as part of the FHPD. All members of the FHPD were then brought back to their original positions and velocities and subsequently allowed to propagate forward in time at constant energy and volume (forward stage). When a simulation crossed its starting milestone again, it was then monitored for transitions to adjacent milestones and the incubation time for these transitions was also recorded. Once a trajectory crossed an adjacent milestone, the simulation was terminated. Any trajectories in this forward stage that crossed adjacent milestones before re-crossing their starting milestone were rejected. The 1, 1.5, 2, 4, and 10 Å milestones produced results with significantly fewer transitions than the other milestones. Therefore, to improve the robustness of our statistics, we performed additional reverse and forward simulations where 10 more trajectories were initiated at random Maxwell-Boltzmann velocities from each equilibrium distribution point, in addition to the one described above (a total of 470,000 reversals for each of these milestones), increasing the number of transitions observed. For each milestone, successful forward stage statistics were inserted into the transition kernel $\mathbf{K}$ and incubation time vector $\langle t \rangle$.

BD simulations

All BD calculations were conducted with BrownDye, a software package specializing in the rigid-body diffusion of two biological molecules in an implicit solvent.[41] The electric

potential map used as input for the BD simulation was calculated with the APBS version 1.4. All BD inputs, as well as the necessary APBS inputs for creation of the electrostatics map, are user defined in the SEEKR input file or generated as SEEKR default values.

In an attempt to recreate the ionic conditions used in the experiment,[50] a nonlinear APBS calculation was run at 298 K, with a solvent dielectric of 78 and a solute dielectric of 2, with the following ions: Ca^{2+} at a concentration of 0.02 mM with a charge of +2.0 *e* and a radius of 1.14 Å, Cl^{-} at a concentration of 0.10 mM with a charge of −1.0 *e* and a radius of 1.67 Å, and tris at a concentration of 0.06 mM with a charge of +1.0 *e* and a radius of 4.0 Å.[54] At the specified concentrations, these ions generate a Debye length of 8 Å, which is used as input to BrownDye. Both the *b*-surface BD simulations and BD trajectories starting from a milestone ran with a solvent dielectric 78 and a solute dielectric of 2, at 298 K. We ran three additional sets of BD simulations at different ionic concentrations to examine the effect of ionic strength in the BD simulations on the k_{on}. Therefore, three additional simulations were run: one with an ion concentration of zero, another with half of the ion concentrations of the experimental procedure, and another with double the ion concentration of the experimental procedure. Although an electrolyte solution technically has a Debye length equal to infinity, we approximated the Debye length with a value of 99 Å in the BrownDye program.

For each k_{on} calculation, we performed 10^6 BD simulations initiated at random points distributed on the *b*-surface, which were used to construct the vector $\mathbf{q_0}$ in Eq. 13. Once these simulations completed, the trajectories that successfully reached the outermost milestone were used as that milestone's FHPD. From that FHPD, an additional 10^6 BD trajectories were run until reaching the second-outermost milestone or escaping to the *q*-surface. These statistics were also included in the transition kernel **K** and incubation time vector $\langle \boldsymbol{t} \rangle$.

Milestoning calculations

Using the statistics obtained from all the milestones in both the MD and BD regimes, the SEEKR software was used to construct the milestoning model and compute the k_{on}, k_{off}, ΔG_{bind}, and other quantities of interest. Additional scripts used to generate some of the figures and data are also included in the SEEKR package. Error estimates were computed according to our previously defined procedure[31].

The vast majority of the procedure outlined in the Materials and Methods section is automated within the SEEKR software package.

Results

Using the MD/BD/milestoning methodology through the SEEKR interface yielded a k_{on} of $2.1\pm0.3\bullet10^7$ $M^{-1}s^{-1}$ for the trypsin-benzamidine system. This value deviates from the experimentally measured k_{on} for the same system at $2.9\bullet10^7$ $M^{-1}s^{-1}$ by a factor of ~1.5 (no experimental error margins were reported). We also estimate a k_{off} of 83 ± 14 s^{-1}, which is within an order of magnitude of the experimentally determined value of 600 ± 300 s^{-1} though our value is slower than expected. similar phenomenon is observed in other computational k_{off} estimations of this system. An examination of the effect of ionic concentration on the

k_{on} convergence of the rate constants as a function of the length of umbrella sampling performed is provided in the SI. Using Eq. 5, we obtain a ΔG_{bind} estimate of −7.3±0.2 kcal•mol^{-1} from a K_d of 4.3±1.2•10^{-6} M compared to the experimental ΔG_{bind} of −6.71±0.05 kcal•mol^{-1}, computed at 298 K using Eq. 5 and an experimental K_d of 1.2±0.1•10^{-5} M.[50]

In addition, we obtained a relative free energy at each of the milestones along the binding pathway using the vector $\mathbf{p}_{\mathbf{stat}}$ in combination with Eq. 10. This free energy profile is displayed in figure 4.

Aside from the predicted thermodynamic and kinetic quantities, we used the trajectories generated during the SEEKR run to make other observations about the system during the binding and unbinding process.

By removing the benzamidine molecule and the solvent, we used POVME2[55] to provide pocket volume measurement and characterization during the course of the MD runs. The same origin and radius of the inclusion region that defined the binding pocket were used for all umbrella sampling trajectories. The pocket itself remains relatively rigid when the benzamidine is deep in the binding site during the umbrella sampling stage, however, more variation in volume was observed when the benzamidine was constrained to a milestone nearer to the entrance of the opening of the binding site (figure 5).

Closer analysis of the umbrella sampling trajectories for the 6, 10, and 12 Å milestones in conjunction with the POVME data indicates sampling of multiple conformations of the trypsin S1 binding pocket (figure 6A, 6B, and 6C). The binding pocket conformation is primarily dependent on the motion of two loops; the loop containing TRP215 and the loop containing ASP189, a critical residue for benzamidine recognition. The opening and closing of the S1 pocket is greatly influenced by the orientation of TRP215 When oriented downward as in figure 6A, the S1 pocket is open. This is the conformation observed in the crystal structure 3PTB with benzamidine bound. When TRP215 rotates upwards as in figure 6B, the binding pocket is closed, and pocket volume significantly decreases. The dramatic change in pocket volume for the 10 Å milestone also occurs when TRP215 moves to close the S1 binding site.

We also observe the formation of an S1* pocket, that results from the motion of these two loops (figure 6C). This pocket provides an alternate binding pathway, in which benzamidine can approach ASP189 from a different orientation. These observations are in agreement with the study of Plattner and Noé[22] where these results were observed through several hundred independent MD trajectories totaling over 100 μs of aggregate simulation time.

We also observed significant positional and rotational sampling by the benzamidine along most of the milestones during the umbrella sampling stages. This information can provide an idea for the likelihood of pathways that benzamidine follows on its route to binding. Figure 3A shows the equilibrium distribution along each of the milestones, and figure 3B shows the FHPD for each of the milestones. Figure 7 shows the angle between a vector pointing along the amidine group and a vector pointing out from the opening of the binding site as a function of time during the equilibrium simulations. Several flips are observed in all but the

lowest milestones, where benzamidine rotation was restricted because these milestones are located deep within the binding pocket. The 10 Å also experiences a decrease in rotational sampling because benzamidine is interacting extensively with TRP215 and thus adopts an orientation that favors stacking of the aromatic rings.

The crystal structure of the trypsin/benzamidine complex shows the amidine group pointing downward toward the binding site (figure 2B). This structural feature is confirmed by our own simulations, and a relatively narrow arrangement of ligand orientations are observed along the lowest milestone.

The entire calculation cost approximately 1.4 million CPU hours on the Stampede supercomputer and local machines, with a total MD cost of approximately 19 μs of simulation.

Discussion

Compared to the experimental k_{on}, our estimated k_{on} is slower by about a factor of 1.3, but falls well within an order of magnitude. We attempted to closely recreate the experimental ionic conditions within our simulations, which has a pronounced effect on the k_{on} (details in the SI). Our k_{on} of 2.2±0.3•10^7 $M^{-1}s^{-1}$ is much closer to the experimental value of 2.9•10^7 $M^{-1}s^{-1}$ than the k_{on}s obtained by Buch et. al.[20] (15±2•10^7 $M^{-1}s^{-1}$) and comparable to what was obtained by Plattner et. al.[22] (6.4±1.6•10^7 $M^{-1}s^{-1}$), although ours was obtained with significantly less computational resources, smaller by an order of magnitude. Our result is also very close to what was obtained by Tiwary et. al.[25] (1±1•10^7 $M^{-1}s^{-1}$). Our estimated k_{off} of 83±14 s^{-1} is within an order of magnitude of the experimental k_{off}, far closer than the value obtained by Buch et. al. (9.5±3.3•$10^4 s^{-1}$), and comparable to the values obtained by Plattner et. al. (131±109 s^{-1}), Teo et. al.[24] (260±240 s^{-1}), and Tiwary et. al (9.1±2.5 s^{-1}). To our knowledge, this is the first successful estimate of k_{off} using a hybrid MD/BD/milestoning model.

An advantage of our approach is that both k_{off} and k_{on} can be determined from the same calculation. We can use our calculated k_{off} and k_{on} values in Eq. 5 to obtain an entirely computationally-determined dissociation constant K_D of 3.8±0.8•10^{-6} M, and by extension a free energy of binding ΔG_{bind} estimate of −7.4±0.2 kcal•mol^{-1}. This is in good agreement with the experimental K_D of 1.2•10^{-5} M, which when put through eq. 5 at a temperature of 298 K, yields a free energy of −6.7 kcal•mol^{-1}.

The accurate determination of kinetics using milestoning requires the proper generation of equilibrium and FHPD distributions. It is important to ensure adequate sampling in the generation of equilibrium distributions. Figure 3A shows the equilibrium distribution of benzamidine center-of-mass along the 1 Å to 12 Å milestones in the MD regime. The benzamidine appears to have explored all solvent-accessible regions along the milestones. Along with positional sampling, the observed diversity of benzamidine orientation in figure 7 indicates that the ligand orientational degree of freedom is well-sampled in all but the lowest milestones. In addition to the ligand, it is important that receptor conformations that may affect ligand binding are also well sampled. By using POVME2, we observed

conformational states that have been observed in other studies such as the S1* pocket (figure 6).[22] We do not however observe any complete binding events via the S1* pocket, presumably as a result of our simplified spherical milestoning model. This may provide some explanation as to why our calculated rates are somewhat slower than experiment, as we do not capture this alternate pathway. However, we may reasonably assume that we are capturing most of the effects of slower receptor conformational changes and subsequently, that our kinetics predictions are reasonable.

While, of course, verification of SEEKR as a computational kinetics and thermodynamics estimator will need to be performed on additional systems, this similarity between experimental and theoretical free energies and rate constants in our accessible and highly parallel framework is encouraging.

Conclusions

In this work, we use our multi-scale MD/BD/milestoning methodology to examine ligand-protein binding events with a larger, more complex, and more drug-like ligand than in our previous work. Furthermore, we present the first successful k_{off} calculation to within one order of magnitude of experiment using this approach. Using the obtained values of k_{on} and k_{off}, and entirely computational estimate of K_D and ΔG_{bind} in good agreement with experiment were obtained. These results are further evidence that the MD/BD/milestoning methodology can be successfully applied to the investigation of binding and unbinding kinetics in receptor-ligand systems. We also present the SEEKR software package, which automates much of the preparation, submission, and analysis of these types of calculations. We have made SEEKR freely available and open-source on Github, and hope that it will be used and improved by the community to run predictive multi-scale milestoning calculations. SEEKR downloads, tutorials, and the user guide may be found at http://amarolab.ucsd.edu/seekr.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Jim Philips, Wen Ma, Jamie Schiffer, Gary Huber, Robert Malmstrom, Rob Swift, J. Andrew McCammon, and Carlos Simmerling for their assistance to the SEEKR project. We dedicate this work to the memory of Klaus Schulten, a pioneer who inspired so many.

LWV acknowledges support from the National Science Foundation Graduate Research Fellowship Program (DGE-1144086). BRJ acknowledges support from the NIH Molecular Biophysics Training Program (T32-GM008326). REA acknowledges the NIH Directors New Innovator Award DP2 OD007237, the National Biomedical Computation Resource (NBCR) NIH P41 GM103426, and supercomputing resources provided by XSEDE (NSF TG-CHE060073).

REA is a co-founder of Actavalon, Inc.

References

1. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, Milo R. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. Biochemistry-Us. 2011; 50(21):4402–4410.
2. Copeland RA, Pompliano DL, Meek TD. Drug-Target Residence Time and Its Implications for Lead Optimization (Vol 5, Pg 730, 2006). Nat Rev Drug Discov. 2007; 6(3):249–249.
3. Copeland RA, Pompliano DL, Meek TD. Opinion - Drug-Target Residence Time and Its Implications for Lead Optimization. Nat Rev Drug Discov. 2006; 5(9):730–739. [PubMed: 16888652]
4. Jorgensen WL. Foundations of Biomolecular Modeling. Cell. 2013; 155(6):1199–1202. [PubMed: 24315087]
5. Held M, Noe F. Calculating Kinetics and Pathways of Protein-Ligand Association. Eur J Cell Biol. 2012; 91(4):357–364. [PubMed: 22018914]
6. Swegat W, Schlitter J, Kruger P, Wollmer A. Md Simulation of Protein-Ligand Interaction: Formation and Dissociation of an Insulin-Phenol Complex. Biophys J. 2003; 84(3):1493–1506. [PubMed: 12609856]
7. Yu TQ, Lapelosa M, Vanden-Eijnden E, Abrams CF. Full Kinetics of Co Entry, Internal Diffusion, and Exit in Myoglobin from Transition-Path Theory Simulations. J Am Chem Soc. 2015; 137(8): 3041–3050. [PubMed: 25664858]
8. Cavalli A, Spitaleri A, Saladino G, Gervasio FL. Investigating Drug-Target Association and Dissociation Mechanisms Using Metadynamics-Based Algorithms. Accounts Chem Res. 2015; 48(2):277–285.
9. Dror RO, Pan AC, Arlow DH, Borhani DW, Maragakis P, Shan YB, Xu HF, Shaw DE. Pathway and Mechanism of Drug Binding to G-Protein-Coupled Receptors. P Natl Acad Sci USA. 2011; 108(32):13118–13123.
10. Shan YB, Eastwood MP, Zhang XW, Kim ET, Arkhipov A, Dror RO, Jumper J, Kuriyan J, Shaw DE. Oncogenic Mutations Counteract Intrinsic Disorder in the Egfr Kinase and Promote Receptor Dimerization. Cell. 2012; 149(4):860–870. [PubMed: 22579287]
11. Shan YB, Kim ET, Eastwood MP, Dror RO, Seeliger MA, Shaw DE. How Does a Drug Molecule Find Its Target Binding Site? J Am Chem Soc. 2011; 133(24):9181–9183. [PubMed: 21545110]
12. Pan AC, Borhani DW, Dror RO, Shaw DE. Molecular Determinants of Drug-Receptor Binding Kinetics. Drug Discov Today. 2013; 18(13–14):667–673. [PubMed: 23454741]
13. Chodera JD, Noe F. Probability Distributions of Molecular Observables Computed from Markov Models. Ii. Uncertainties in Observables and Their Time-Evolution. J Chem Phys. 2010; 133(10)
14. Noe F. Probability Distributions of Molecular Observables Computed from Markov Models. J Chem Phys. 2008; 128(24)
15. Prinz JH, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schutte C, Noe F. Markov Models of Molecular Kinetics: Generation and Validation. J Chem Phys. 2011; 134(17)
16. Sarich M, Noe F, Schutte C. On the Approximation Quality of Markov State Models. Multiscale Model Sim. 2010; 8(4):1154–1177.
17. Pande VS, Beauchamp K, Bowman GR. Everything You Wanted to Know About Markov State Models but Were Afraid to Ask. Methods. 2010; 52(1):99–105. [PubMed: 20570730]
18. Lane TJ, Bowman GR, Beauchamp K, Voelz VA, Pande VS. Markov State Model Reveals Folding and Functional Dynamics in Ultra-Long Md Trajectories. J Am Chem Soc. 2011; 133(45):18413–18419. [PubMed: 21988563]
19. Schutte C, Noe F, Lu J, Sarich M, Vanden-Eijnden E. Markov State Models Based on Milestoning. J Chem Phys. 2011; 134(20):204105. [PubMed: 21639422]
20. Buch I, Giorgino T, De Fabritiis G. Complete Reconstruction of an Enzyme-Inhibitor Binding Process by Molecular Dynamics Simulations. P Natl Acad Sci USA. 2011; 108(25):10184–10189.
21. Luty BA, Elamrani S, Mccammon JA. Simulation of the Bimolecular Reaction between Superoxide and Superoxide-Dismutase - Synthesis of the Encounter and Reaction Steps. J Am Chem Soc. 1993; 115(25):11874–11877.

22. Plattner N, Noe F. Protein Conformational Plasticity and Complex Ligand-Binding Kinetics Explored by Atomistic Simulations and Markov Models. Nat Commun. 2015:6.
23. Pan AC, Sezer D, Roux B. Finding Transition Pathways Using the String Method with Swarms of Trajectories. J Phys Chem B. 2008; 112(11):3432–3440. [PubMed: 18290641]
24. Teo I, Mayne CG, Schulten K, Lelievre T. Adaptive Multilevel Splitting Method for Molecular Dynamics Calculation of Benzamidine-Trypsin Dissociation Time. J Chem Theory Comput. 2016; 12(6):2983–2989. [PubMed: 27159059]
25. Tiwary P, Limongelli V, Salvalaglio M, Parrinello M. Kinetics of Protein-Ligand Unbinding: Predicting Pathways, Rates, and Rate-Limiting Steps. P Natl Acad Sci USA. 2015; 112(5):E386–E391.
26. Northrup SH, Allison SA, Mccammon JA. Brownian Dynamics Simulation of Diffusion-Influenced Bimolecular Reactions. J Chem Phys. 1984; 80(4):1517–1526.
27. Zhou HX. On the Calculation of Diffusive Reaction-Rates Using Brownian Dynamics Simulations. J Chem Phys. 1990; 92(5):3092–3095.
28. Mccammon JA, Northrup SH, Allison SA. Diffusional Dynamics of Ligand Receptor Association. J Phys Chem-Us. 1986; 90(17):3901–3905.
29. Lindert S, Kekenes-Huskey PM, McCammon JA. Long-Timescale Molecular Dynamics Simulations Elucidate the Dynamics and Kinetics of Exposure of the Hydrophobic Patch in Troponin C. Biophys J. 2012; 103(8):1784–1789. [PubMed: 23083722]
30. Cheng YH, Suen JK, Zhang DQ, Bond SD, Zhang YJ, Song YH, Baker NA, Bajaj CL, Holst MJ, McCammon JA. Finite Element Analysis of the Time-Dependent Smoluchowski Equation for Acetylcholinesterase Reaction Rate Calculations. Biophys J. 2007; 92(10):3397–3406. [PubMed: 17307827]
31. Votapka LW, Amaro RE. Multiscale Estimation of Binding Kinetics Using Brownian Dynamics, Molecular Dynamics and Milestoning. Plos Comput Biol. 2015; 11(10)
32. Boras BW, Hirakis S, Votapka L, Malmstrom RD, Amaro RE, McCulloch AD. Bridging Scales through Multiscale Modeling: A Case Study on Protein Kinase A. Front Physiol. 2015:6.
33. Votapka LW. Numerical and Computational Solutions for Biochemical Kinetics, Druggability, and Simulation. 2016
34. Kirmizialtin S, Elber R. Revisiting and Computing Reaction Coordinates with Directional Milestoning. J Phys Chem A. 2011; 115(23):6137–48. [PubMed: 21500798]
35. Ermak DL, Mccammon JA. Brownian Dynamics with Hydrodynamic Interactions. J Chem Phys. 1978; 69(4):1352–1360.
36. Faradjian AK, Elber R. Computing Time Scales from Reaction Coordinates by Milestoning. J Chem Phys. 2004; 120(23):10880–9. [PubMed: 15268118]
37. Vanden-Eijnden E, Venturoli M, Ciccotti G, Elber R. On the Assumptions Underlying Milestoning. J Chem Phys. 2008; 129(17)
38. Votapka LW, Lee CT, Amaro RE. Two Relations to Estimate Membrane Permeability Using Milestoning. J Phys Chem B. 2016; 120(33):8606–8616. [PubMed: 27154639]
39. Skolnick J. Perspective: On the Importance of Hydrodynamic Interactions in the Subcellular Dynamics of Macromolecules. J Chem Phys. 2016; 145(10)
40. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable Molecular Dynamics with Namd. J Comput Chem. 2005; 26(16):1781–1802. [PubMed: 16222654]
41. Huber GA, McCammon JA. Browndye: A Software Package for Brownian Dynamics. Comput Phys Commun. 2010; 181(11):1896–1905. [PubMed: 21132109]
42. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. P Natl Acad Sci USA. 2001; 98(18):10037–10041.
43. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, Debolt S, Ferguson D, Seibel G, Kollman P. Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. Comput Phys Commun. 1995; 91(1–3):1–41.

44. Marquart M, Walter J, Deisenhofer J, Bode W, Huber R. The Geometry of the Reactive Site and of the Peptide Groups in Trypsin, Trypsinogen and Its Complexes with Inhibitors. Acta Crystallogr B. 1983; 39(Aug):480–490.
45. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, Snoeyink J, Richardson JS, Richardson DC. Molprobity: All-Atom Contacts and Structure Validation for Proteins and Nucleic Acids. Nucleic Acids Res. 2007; 35:W375–W383. [PubMed: 17452350]
46. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. Molprobity: All-Atom Structure Validation for Macromolecular Crystallography. Acta Crystallogr D. 2010; 66:12–21. [PubMed: 20057044]
47. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. Ff14sb: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99sb. J Chem Theory Comput. 2015; 11(8):3696–3713. [PubMed: 26574453]
48. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. Pdb2pqr: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. Nucleic Acids Res. 2004; 32:W665–W667. [PubMed: 15215472]
49. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA. Pdb2pqr: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations. Nucleic Acids Res. 2007; 35:W522–W525. [PubMed: 17488841]
50. Guillain FTD. The Use of Proflavin as an Indicator in Temperature-Jump Studies of the Binding of a Competitive Inhibitor to Trypsin. J Am Chem Soc. 1970; 92(18):5534–5536. [PubMed: 5449454]
51. Horn HW, Swope WC, Pitera JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: Tip4p-Ew. J Chem Phys. 2004; 120(20):9665–9678. [PubMed: 15267980]
52. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and Testing of a General Amber Force Field. J Comput Chem. 2004; 25(9):1157–1174. [PubMed: 15116359]
53. Wang JM, Wang W, Kollman PA, Case DA. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. J Mol Graph Model. 2006; 25(2):247–260. [PubMed: 16458552]
54. Schindler P, Robinson RA, Bates RG. Solubility of Tris(Hydroxymethyl)Aminomethane in Water-Methanol Solvent Mixtures and Medium Effects in Dissociation of Protonated Base. J Res Nbs a Phys Ch. 1968; A 72(2):141-+.
55. Durrant JD, Votapka L, Sorensen J, Amaro RE. Povme 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. J Chem Theory Comput. 2014; 10(11):5047–5056. [PubMed: 25400521]

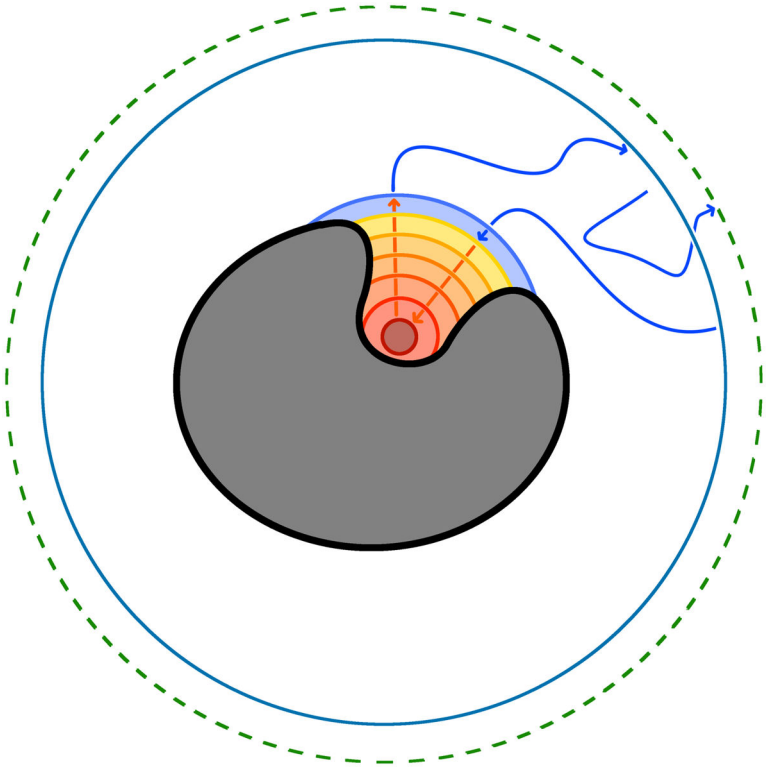

Figure 1.

A cartoon schematic of trypsin (grey shape) with the concentric spherical milestones (orange and blue circular curves) surrounding the binding site. Also, the *b*- and *q*-surfaces are represented as the outer blue and dashed green curves, respectively, that sit away from the molecule. Blue arrows represent BD trajectories, and orange arrows represent MD trajectories. Any surface with a blue arrow coming from or going to it represents the starting or ending surface for BD trajectories, respectively. Similarly, a surface with an orange arrow coming from or going to it represents the starting or ending surface for MD simulations, respectively.

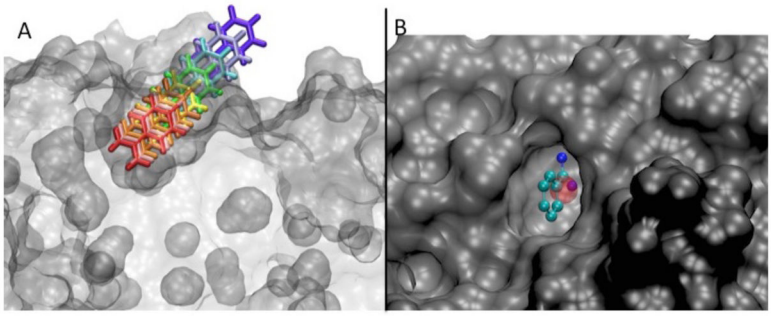

Figure 2.

Panel A: before beginning the simulations, benzamidine has been placed along each of the milestones in gradually increasing distances from the center of the binding site on trypsin. Panel B: The center-of-mass of the benzamidine molecule in the trypsin 3PTB crystal structure lies within the lowest 1Å milestone (red sphere), which we define as the bound state

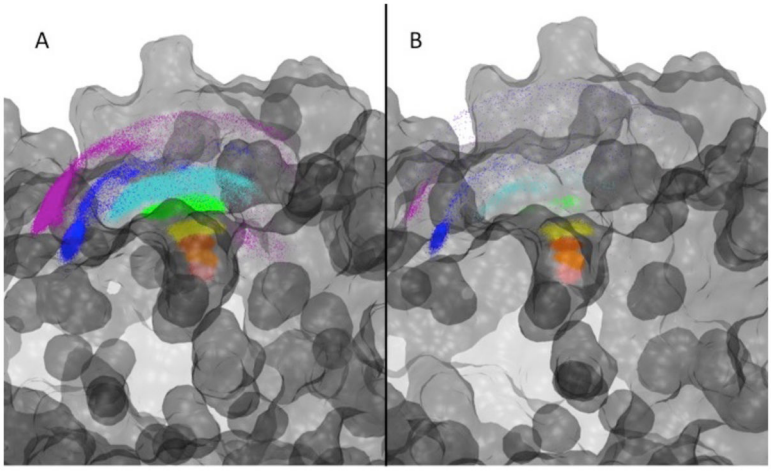

Figure 3.

Panel A: The equilibrium distribution of the center of mass of benzamidine generated along all of the milestones from 2 Å (red) to 12 Å (green) at the end of the umbrella sampling. No umbrella sampling is performed for the BD stages, so there are no points representing the 14 Å milestone. Panel B: The FHPD of benzamidine centers of mass generated from the equilibrium distribution that succeeded in the reverse stage. The milestones between 1 Å (red) and 12 Å (green) were generated during the MD simulations. In addition, the blue distribution at 14 Å represents the FHPD obtained from the BD simulation. This FHPD is used to start forward stage trajectories for generating milestoning statistics.

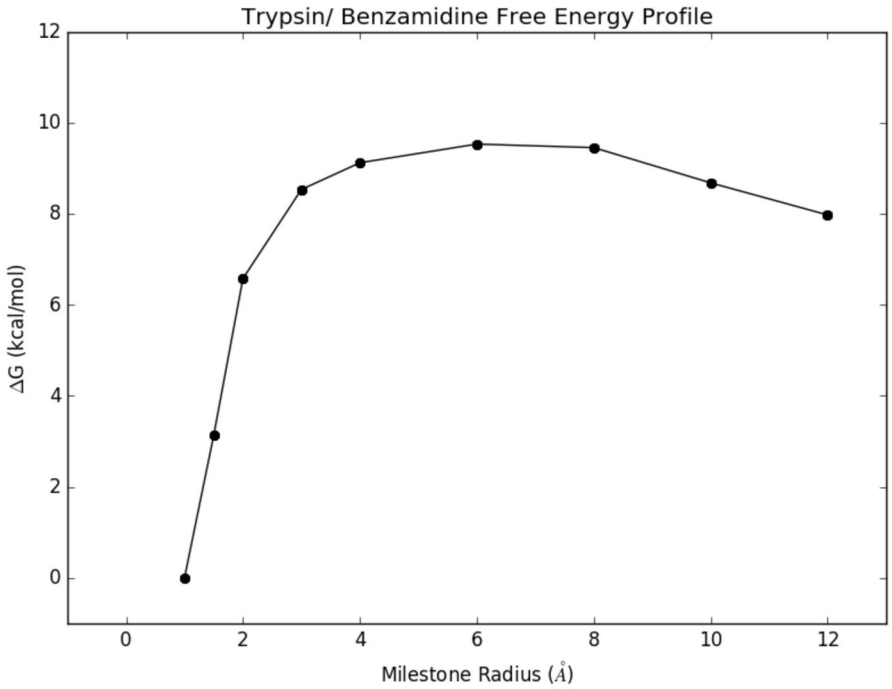

Figure 4.

The free energy profile of benzamidine along each of the milestones leading to the binding site. The free energy barrier peaks around the milestone located at 6 Å.

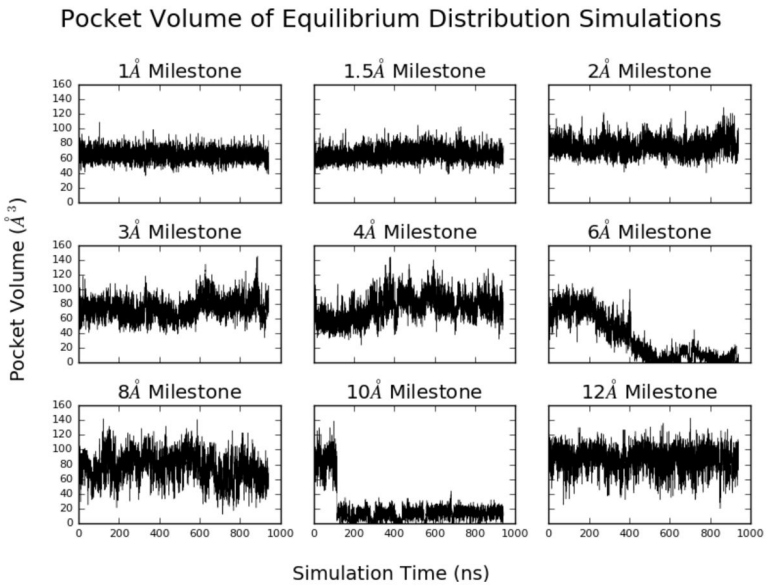

Figure 5.

The volume of the S1 binding site with benzamidine restrained to the milestones as computed using the POVME2 program. Stabilization of the binding site pocket volume is observed as the ligand moves closer to the binding site.

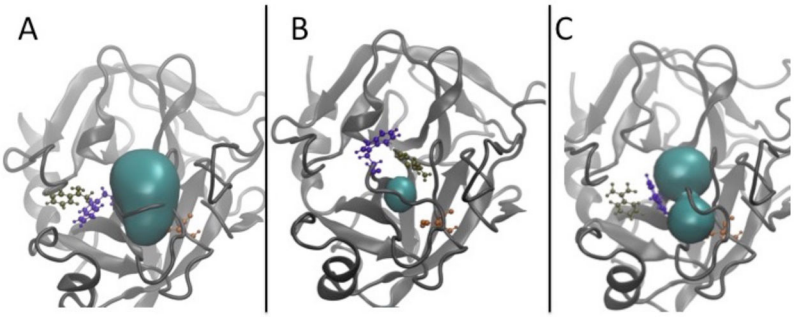

Figure 6.

Dynamics of the apo trypsin S1 binding pocket umbrella sampling simulations. Pocket conformations are significantly influenced by the motions of the loop containing TRP215 (violet) and the loop containing ASP189 (orange), which is important for benzamidine recognition. Benzamidine is shown in tan. POVME calculated volumes are shown in cyan. **A)** The open S1 pocket, where TRP215 is pointed in a downward orientation. **B)** Closed conformation of the S1 pocket as a result of TRP215 rotating to an upward pointing conformation. **C)** Formation of the S1* pocket where benzamidine can approach via an alternate pathway and interact with ASP189 from a different angle.

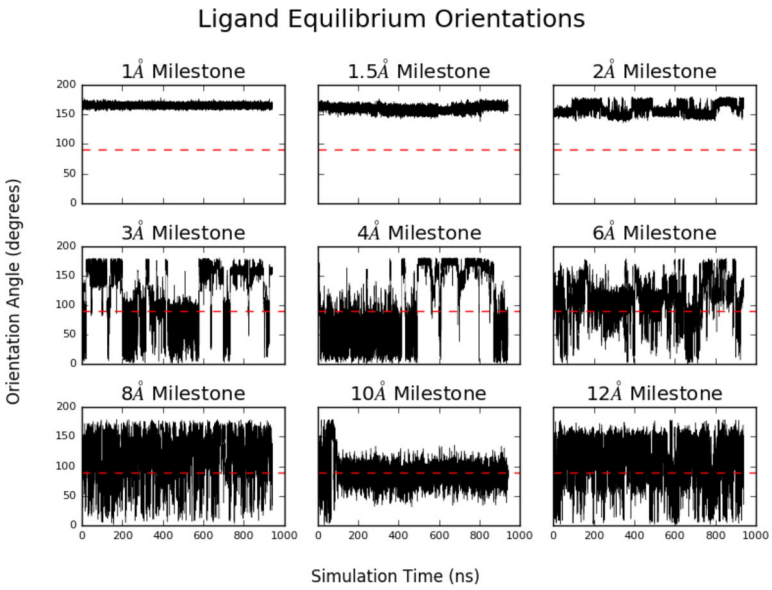

Figure 7.

The angle of benzamidine along the center-of-mass/amidine axis compared to a vector pointing outward from the binding site. An angle larger than 90° represents a conformation where the amidine group is pointing toward the binding site. Several flips were observed in all milestones above 2 Å, implying that the orientation of the ligand is well sampled along all of the milestones except for those deepest in the binding pocket, where the orientation found in the crystal structure is preferred, and the amidine group is pointing down into the site.