Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations

Ignasi Buch [*], **Toni Giorgino** [*], **and Gianni De Fabritiis** [*]

[*]Computational Biochemistry and Biophysics Laboratory (GRIB-IMIM), Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB), C/ Doctor Aiguader 88, 08003 Barcelona, Spain

Submitted to Proceedings of the National Academy of Sciences of the United States of America

The understanding of protein-ligand binding is of critical importance for biomedical research, yet the process itself has been very difficult to study because of its intrinsically dynamic character. Here, we have been able to quantitatively reconstruct the complete binding process of the enzyme-inhibitor complex trypsin-benzamidine by performing 495 molecular dynamics simulations of free ligand binding of 100 ns each, 187 of which produced binding events with an RMSD less than 2 Å compared to the crystal structure. The binding paths obtained are able to capture the kinetic pathway of the inhibitor diffusing from solvent (S0) to the bound (S4) state passing through two metastable intermediate states S2 and S3. Rather than directly entering the binding pocket the inhibitor appears to roll on the surface of the protein in its transition between S3 and the final binding pocket, while the transition between S2 and the bound pose requires rediffusion to S3. An estimation of the standard free energy of binding gives $\Delta G^\circ = -5.2 \pm 0.4$ kcal/mol (cf. the experimental value -6.2 kcal/mol), and a two-states kinetic model $k_{on} = (1.5 \pm 0.2) \times 10^8$ $\mathbf{M}^{-1}\mathbf{s}^{-1}$ and $k_{off} = (9.5 \pm 3.3) \times 10^4$ $\mathbf{s}^{-1}$ for unbound to bound transitions. The ability to reconstruct by simple diffusion the binding pathway of an enzyme-inhibitor binding process demonstrates the predictive power of unconventional high-throughput molecular simulations. Moreover, the methodology is directly applicable to other molecular systems, and thus of general interest in biomedical and pharmaceutical research.

Markov state model | free ligand binding | distributed computing | binding affinity | transition state kinetics

Abbreviations: MD, Molecular dynamics; MSM, Markov state model; PMF, Potential of mean force; MFPT, Mean first passage time; GPU, Graphics processing unit

Introduction

Understanding protein-ligand binding processes is undoubtedly of critical importance in structure-based drug design and much effort is being invested in experimental and computational methods to resolve binding. The focus has generally resided on predicting the lowest energy binding pose of a ligand, (1, 2) but resolving the kinetic mechanisms and structure activity relationships of the ligand has increasingly been recognized to provide additional mechanisms for elucidating therapeutically safe and differentiated responses (3, 4). The kinetics of binding depend on the characteristic transition states of a system. Hence, characterizing the binding pathway is crucial to understanding how to control and re-engineer the process of binding.

Commonly-used methods to experimentally determine kinetic data for biomolecular interactions are available (5), but fast time-scale resolution of a binding mechanism with atomic resolution remains difficult due to the intrinsic dynamic and volatile nature of the process of binding. From a computational perspective, the difficulty lies in accurately measuring binding affinities and kinetic parameters, but it has become easier to try to predict binding free energies on a limited number of targets and to qualitatively interpret binding mechanisms using molecular dynamics. Although it still requires substantial computational resources, the use of special MD engines running on graphical processing units (GPUs) have greatly reduced its cost (6, 7). In this work we introduce the first, to our knowledge, complete reconstruction of the binding process for an enzyme-inhibitor complex by free diffusion molecular dynamics simulations. Not only we reproduce with atomic resolution the crystallographic mode of binding, but also provide the kinetically and energetically meaningful transition states of the process.

Free ligand binding has been used in the past to describe computational experiments in which, typically, a ligand is placed at a certain distance from the target protein and first by diffusion and then by specific interactions, binds to one or more sites in the protein. These works can be classified into two groups: those mainly trying to predict binding sites and modes for one or more ligands, and those adding some degree of mechanistic information about the process of binding. Nonetheless, previous attempts to perform free ligand binding to proteins could not recover more than a few binding events due to the computational cost, therefore providing some qualitative information on the process but lacking a quantitative validation of the results with experimental data. Proper validation is necessary to make sure that the results provide the correct strategy for understanding biological function.

Wu et al. (8) reported possible binding modes of thioflavin-T (ThT) to β-rich peptide self-assemblies complementing previous experimental work in which ThT binding sites could not be determined (9). Other predictions of single or multiple binding sites were carried out in nicotinic acetylcholine receptors where Brannigan et al. (10) proposed multiple binding sites for anesthetic isoflurane, and the more recent description of different binding modes of agonists to β-adrenergic receptors also with all-atom molecular dynamics simulations (11). In Ref. (12), we could recover the experimental binding site for sodium ions obtaining several hundred binding events on D2-dopaminergic receptors. On the more mechanistic description, there has been the work on the fast recognition of proline-rich peptides by SH3 domains (13) as well as the pH-dependent mechanism of NO transport by nitrophorins (14). Finally, results on initial conformational changes upon binding are reported for Glycerol 3-phosphate transporter (GlpT), which mediates the import of glycerol 3-phosphate using a phosphate gradient (15). Despite such progress having been made, none of the aforementioned studies have provided a complete re-

Reserved for Publication Footnotes

construction of a ligand binding process in terms of pathway and quantitative information of the energetics and kinetics.

An ideal method for resolving the binding process would provide not only the binding affinity and kinetics of the reaction, but also atomic resolution information on its pathway. Binding sites, transition states and metastable states are potentially useful to broaden the probability of success in the context of drug design. Here, we present a kinetic model for the binding process of serine protease β-trypsin inhibitor benzamidine obtained from extensive high-throughput all-atom molecular dynamics (MD) simulations of free ligand binding using the ACEMD (6) software on the GPUGRID distributed computing network (7). An aggregate of 50 μs of trajectory data have been used to construct a Markov State Model (MSM) (16) of the binding process of benzamidine to trypsin. Previous computational studies on trypsin date back to the 80s with work by Warshel et al. (17) on binding free energies computed from potentials of mean force for trypsin and subtilisin. More recent efforts on the trypsin-benzamidine system in particular, include studies of the conformational variability of the bound complex (18, 19) and benzamidinium-derivatives binding affinity calculations via end-point methods (20, 21) and via potential of mean force free energy biased methods such as umbrella sampling (22) and metadynamics (23). In this work, we report 187 events of binding from bulk to bound based on free diffusion simulations, a quantitative analysis of the process of binding, and the identification of the rate-limiting step in the pathway of binding, which consists of a set of transient hydrogen-bond interactions on the surface of trypsin close to the binding pocket.

Results and Discussion

The results of this work are based on the analysis of 495 trajectories of free diffusion of benzamidine around trypsin each of 100 ns of length (see Figure 1 for a schematic representation of the system). Among these, 187 trajectories (37%) successfully recovered the bound pose in the binding pocket with an RMSD compared to the crystal structure of less than 2 Å (Supporting Video S1 shows a typical binding event). As shown in Figure 2a, benzamidine placed at 35 Å from its binding pocket explores the entire simulation box (as monitored by its C7 carbon). Interestingly, several clusters of benzamidine on the surface of trypsin can be observed, which indicates that a rather complex mechanism of binding takes place instead of a of simple pathway directly from the bulk. The RMSD of each trajectory in Figure 2b shows the time-evolution for a sample (one every ten) of the trajectories. Some trajectories reach the bound crystallographic pose just after 10-15 ns of simulation while some reach the binding pocket only after 90 ns. Finally, the majority of the trajectories do not enter the binding site within 100 ns, as should be expected in such a short time frame. Nevertheless, there is more than enough data to carry out a detailed quantitative analysis of the binding pathway. In fact, being able to obtain 187 bound trajectories at less than 2 Å RMSD by high-throughput molecular dynamics is a compelling result.

In order to formalize the observable conformations of the system and its kinetic relationships we built several Markov state models of the process (16): (a) a two dimensional projection in the xz reaction plane, (b) a simple 5 metastable-state model, and finally (c) a full three-dimensional projection to explicitly compute kinetic and energetic information.

Identification of metastable states. Starting from the two dimensional model of the process, we chose the projection of the coordinates of benzamidine C7-atom on the xz-plane which is the plane which better capture the binding process (see Supporting Figure S1 for other projections). A two-dimensional 2500-states MSM was constructed with a lag time of 50 ns (see Supporting Figure S2). The clustering of the coordinates was performed on a matrix of 50×50 bins of 1.44 Å^2 each. From the first eigenvector of the transition probability matrix we obtained the potential of mean force (PMF) of the binding process (Figure 3a). Several metastable binding sites are identifiable by this energetic map. These are respectively named: S0, to represent the bulk, i.e. all those conformations with z greater than 30 Å; S1, where a 1.0 kcal/mol minima indicates the first interactions of benzamidine with trypsin; S2, the isolated 2.5 kcal/mol minima in the lower-left corner of the PMF plane; S3, the second deepest minima of 3.0 kcal/mol at the top-left corner of the PMF plane; and finally S4, the bound conformation with relative free energy of 6.0 kcal/mol, centered at the origin. The metastable states are structurally characterized by visual inspection to highlight the underlying molecular interactions, as shown in Figure 3b. In state S1, several mild hydrophobic interactions between benzamidine and Y59, Y94 and V90 are identified as representing the first recognition contacts of the complex in our experiments. In S2, benzamidine is stabilized by a stronger π-π stacking interaction with Y151 and Y39 sidechains. In S3, several hydrogen bonds are present between the amidine moiety of benzamidine and Q175, T98, L99 and T177 main chains in a rather electronegative region of trypsin. Finally S4, shows the bound conformation of benzamidine and its strong interaction with D189. (24) Additionally, the crystal waters reported with the original structure in the binding pocket were reproduced by the binding simulations (a video of water molecules in the binding pocket is available in Supporting Video S2).

Characterization of binding pathway. The two-dimensional PMF is further coarse-grained into 5 states using a Voronoi tessellation, with the centroids selected to represent the different metastable states of the system. This simple model of the process provides a direct interpretation of the metastable binding sites through analysis of the structure of the eigenvectors of the 5-state MSM (available in Supporting Figure S3), which describe the characteristic modes between the states as well as their implied timescales. Considering the slowest modes, we see transitions from site S0 to S1 (Figure 4a) and collectively from sites S0/S1 to sites S3/S4 (Figure 4b) on an implied timescale of between 6-10 ns, corresponding to the diffusion of the ligand from bulk to the first structural contact with the protein. With an implied timescale of $\approx$ 20 ns, the model shows the transitions between sites S2 and S3 (Figure 4c). Therefore, the ligand is more likely to diffuse back and bind to the S3 site rather directly transit from S2 to the binding pocket. Site S2 is a secondary binding pocket but not directly involved in the binding pathway. Finally, the rate-limiting step of the process is the transition to the bound site S4 (Figure 4d) which is seen to happen at an implied timescale of around 60 ns in the approximate 5-state model, and preferentially coming from S3 (the complete probability transition matrix is available in Supporting Table S1). This is consistent with the energetics in Figure 3a, since the energetic barrier between the two states is only 1.5 kcal/mol high, lower than that to any other metastable state.

To further understand the rate-limiting transition S3–S4 we inspected the trajectories that underwent binding to define a consensus pathway. As shown in Figure 5a benzamidine appears to "roll" over the surface of trypsin in its rate-limiting transition where three transition states are identified. TS1, TS2 and TS3 are found to be three path-defining interactions along which most of the binding trajectories transit. In higher

detail, Figure 5b displays TS1 as a single hydrogen bond of benzamidine with N89 main chain. TS2 is a stronger hydrogen bond interaction with S210 and H57, the latter belonging to trypsin's catalytic triad with S195 and N102. The last transition state identified prior to binding is TS3 corresponding to a hydrogen bond formed with the S214 side chain. Not all binding trajectories follow sequentially the transition through TS1, TS2 and TS3. Nonetheless, the vast majority bind passing through TS2 and TS3 or directly to TS3 before burying into the pocket.

Estimation of energetics and kinetics. In order to provide a more precise quantitative interpretation of binding, we compute both the binding affinity and the kinetics of the overall process of binding. For this, a higher resolution three-dimensional MSM model was built with 9700 states after clustering the coordinates into $18 \times 18 \times 30$ bins with a volume of 36.33 Å^3 each. Using the new model it was possible to isolate with higher precision the binding site by making a single binary distinction of the conformations into *bound* and *unbound*. After a sensitivity analysis in the three-dimensional PMF (Supporting Figure S4) we defined the bound conformation as those states with $G_{PMF} \leq 3$ kcal/mol and the unbound as $G_{PMF} > 3$ kcal/mol.

The standard free energy of binding (ΔG°), is calculated using the following expression (derivation available in Supporting Text S1):

$$\Delta G^\circ = -\Delta W_{3\mathrm{D}} - k_B T \log\left(\frac{V_\mathrm{b}}{V^\circ}\right)$$

where $\Delta W_{3\mathrm{D}}$ is the depth of the PMF, k_B is the Boltzmann constant, T is the temperature, V_b is the bound volume calculated as the integral of the PMF, and V° is the standard-state volume. The three-dimensional PMF (shown in Supporting Figure S5) yields a standard free-energy of binding of $\Delta G^\circ = -5.2 \pm 0.4$ kcal/mol, where the error is calculated by block averaging over 5 sets of data of 10 μs each. Our calculated free energy differs by 1 kcal/mol from the experimentally determined value of -6.2 kcal/mol (25). Previous computational attempts to determine trypsin-benzamidine standard free energy of binding reported the ample range of -5.5 to -9.0 kcal/mol via pathway-based free energy methods (22), while Ref. (23) recovered a PMF depth similar to ours, although obtained in a different setting and without reporting the standard free energy of binding.

In order to compare the reconstructed binding pathway with experimental kinetics data, we mapped the binding process onto a simple two-state model for the reaction of an inhibitor (I) with an enzyme (E), i.e.

$$[E] + [I] \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} [EI]$$

where, assuming first order kinetics, k_{on} and k_{off} directly relate to the mean times of binding and unbinding, respectively (26,27). In a binary view of the states of the system as having a bound and unbound conformations, computing the mean first passage time (MFPT) for the *on* and *off* reactions would directly give us the constant rates for the reactions as

$$\mathrm{MFPT} = \frac{1}{k}$$

where k_{on}, measured in units of $\mathrm{M}^{-1}\mathrm{s}^{-1}$, is inversely dependent on the ligand concentration, $C^{comp} = 0.015$ M, according to the relation $k_{on}^{-1} = \mathrm{MFPT}_{on} \cdot C^{comp}$. The calculated mean times for the process are $\mathrm{MFPT}_{on} = 444 \pm 59$ ns and $\mathrm{MFPT}_{off} = (1.17 \pm 0.43) \times 10^4$ ns. The derived kinetic parameters for the binding process therefore are $k_{on} = (1.5 \pm 0.2) \times 10^8\ \mathrm{M}^{-1}\mathrm{s}^{-1}$ and $k_{off} = (9.5 \pm 3.3) \times 10^4\ \mathrm{s}^{-1}$. Both results deviate from the experimental reference values (28) by at least an order of magnitude, with a higher error on k_{off} ($k_{on}^{exp} = 0.29 \times 10^8\ \mathrm{M}^{-1}\mathrm{s}^{-1}$ and $k_{off}^{exp} = 0.06 \times 10^4\ \mathrm{s}^{-1}$). A larger uncertainty is coupled to the precision of k_{off} since a larger extrapolation is done by inferring long-timed events from various short-time transitions. The complete unbinding rate is recovered by the MSM from the probability of transitions moving up the binding well in the PMF. So, while it is not necessary to produce a full unbinding event, the error in the estimation of the transitions may be higher in the *off* direction rather than *on* (29). For higher accuracy in the computation of the kinetic parameters, a better sampling of the slowest mode of the MSM might improve the estimations. Methodologies that increase the unbiased sampling on the states involved in the slowest modes could be used to improve the efficiency of sampling and possibly yield more accurate estimation of quantitative parameters (30, 31). Finally, it could also be that these results are simply within the limitations of the force-field employed, as our results are in agreements with previous estimations (32).

Conclusions

This study proposes an approach to reconstruct a complete enzyme-inhibitor binding process from free ligand binding MD simulations. The quantitative reconstruction is achieved through the construction of Markov state models of the process. In 187 of the trajectories (37% of total) benzamidine was found to bind with RMSD less than 2 Å from the crystallographic pose. Investigation of the energetics for the interaction revealed two additional metastable states of the complex, S2 and S3, in addition to the bound state S4 (Figure 3a). Structural characterization of these states revealed persistent interactions with key residues. Interestingly, transitions to the bound state S4 tend to come from the metastable binding site S3, and not from the bulk. Further investigation of the transition revealed how benzamidine "rolls" over the surface of trypsin. Mechanistically, this is explained by the existence of three representative transition states (TS1, TS2 and TS3) along this transition corresponding to hydrogen-bond interactions between amidine moiety of benzamidine and trypsin residues. In particular, TS2 and TS3 characterized as interactions with catalytic triad residue H57 and S210 for TS2 and S214 for TS3. We propose that putative mutations to the transition-state residues may alter the binding kinetics of benzamidine. Further examples and validation of this approach need to be carried out comparing the binding of several ligands to several proteins (33). We plan to validate and improve this methodology in the future to extend its applicability, also taking advantage of improved computational capabilities. In a broader sense, the present approach should be useful in studying the binding pathways of complexes of similar size. According to our results, 20% of the data generated in the current study could be sufficient to reconstruct similar protein-ligand binding processes. Surely, not all cases would be amenable to this methodology – larger, flexible ligands and proteins would present more complex binding processes requiring higher computational effort. We don't think that this approach is that relevant for high throughput screening for several reasons. First, the amount of calculations required for a single compound is not scalable to thousands or millions of ligands, and second because the degree of detail provided by the methodology is probably not needed during the screening phase. Instead, we envisage a successful application in the lead optimization phase on possible tens of lead compounds

to completely reconstruct and understand their binding pathways before moving forward in the pipeline. The method, although computationally expensive, is certainly feasible if performed on just few ligands as indicated. Here, we have made use of our volunteer GPU project GPUGRID.net (7) based on ACEMD (6). Extrapolating from the current study, it would be possible to reconstruct the binding pathways of 5 to 10 ligands per week in-house on a moderately sized GPU cluster of 32-64 GPUs. This level of complexity should be already within the reach of many interested in using the methodology, while future hardware, software and improved protocols could further reduce the computational costs.

Materials and Methods

System preparation. The input model was based on the crystallographic structure of the inhibitor benzamidine binding to the enzyme bovine trypsin (PDB: 3PTB, Figure 1a). (24) The protein was modeled using the AMBER 99SB force field (34) and the ligand with the GAFF force field (35) as described by Doudou et al. (22). The complex was rotated to align the vector defined by the C7 atom of benzamidine and the Cγ atom of ASP 189 to the z-axis of the box. The ligand was placed in the bulk at $z = 35$ Å from the binding pocket, on the z axis (see Figure 1b). The complex was then solvated with 10604 TIP3P water molecules (36) and neutralized with the addition of 9 Cl^- ions. The setup resulted in a system of 35060 atoms in a simulation box of initial dimensions $71 \times 64 \times 82$ Å^3.

The system was minimized and equilibrated under NPT conditions at 1 atm and 298 K using a timestep of 2 fs, non-bonded cutoff of 9 Å, rigid bonds and PME for long range electrostatic with a grid of $72 \times 64 \times 88$. During minimization and equilibration, the heavy protein atoms were restrained by a 10 kcal mol^{-1}Å^{-2} spring constant. Two rounds of velocity re-initialization and 2 ps simulation were performed under NVT conditions. The magnitude of the restraining spring constant was then reduced to 1 kcal mol^{-1}Å^{-2} during 10 ps of NVT before the barostat was switched on at 1 atm for a further 10 ps of NPT simulation. A final 40 ps of NPT simulation was conducted with a restraint constant of 0.05 kcal mol^{-1}Å^{-2}. Finally, the volume was allowed to relax for 10 ns under NPT conditions reaching a final box size of $69.7 \times 63.2 \times 80.2$ Å^3. During this run, Cα atoms of the complex were restrained with a 1 kcal mol^{-1}Å^{-2} harmonic potential to prevent the system reorienting. The equilibration run was performed using NAMD2.6 (37) on a standard CPU cluster. All production runs using ACEMD (6) on the distributed computing project GPUGRID (7).

Free-binding production simulations. 500 different conformations were generated by sampling from a pre-production run of 50 ns. The ligand was kept at a minimum distance of 20 Å from the protein and allowed to diffuse in the xy plane only. All production runs were carried in the NVT ensemble. A longer timestep of 4 fs was used thanks to the use of the hydrogen mass repartitioning scheme (38) implemented in ACEMD. Temperature, non-bonded forces cutoff and PME grid size were kept the same as in the equilibration phase. In order to keep the protein orientation fixed and to avoid the benzamidine diffusing away, two restraint schemes were applied. First, an harmonic restraint of $k = 1$ kcal mol^{-1}Å^{-2} was applied on Cα atoms of the protein further than 9 Å from the binding pocket. Second, a flat-bottomed harmonic restraint was applied to the C7 atom of benzamidine with box size $\mathbf{b} = (30, 30, 80)$ centered around the binding pocket and restraint potential of 0.1 kcal mol^{-1}Å^{-2}. 500 production runs were submitted to GPUGRID. Upon data retrieval and analysis, 495 trajectories had reached the target of 100 ns of simulated time. The completed 495 trajectories ($\sim 50\ \mu s$ total aggregate data) constitute the total data used for trypsin-benzamidine binding analysis.

Markov State Models. Markov State Models (MSM) are useful to describe the dynamics of a system as a sequence of transitions between coarse-grained states (16). We obtained a dynamic description of the binding process, defining the states according to the ligand-protein geometry. Two- and three-dimensional reaction coordinates were defined from the Cartesian components of the position of the C7 atom in benzamidine, assuming its bound position as the origin of the axes. The values of the coordinates were discretized in a grid of states by rounding them to the closest multiple of a chosen bin size. The complet MSM model is then built and analyzed using the methods highlighted in Supporting Text S1 (16, 26, 39–41).

ACKNOWLEDGMENTS. We thank Dr. R. H. Henchman for his help revising the PMF derivation and also Matthew J. Harvey, Dr. S. Kashif Sadiq and all the other colleagues that carefully revised the manuscript. IB acknowledges support from the Fundació La Marató de TV3. TG acknowledges support from the Beatriu de Pinós programme of the Generalitat de Catalunya. GDF acknowledges support from the Ramón y Cajal scheme and support by the Spanish Ministry of Science and Innovation (Ref. FIS2008-01040). We finally thank all the volunteers of GPUGRID who donate GPU computing time to the project.

1. Bleicher KH, Bohm H, Muller K, Alanine AI (2003) Hit and lead generation: beyond high-throughput screening. Nat Rev Drug Discov 2:369–378.
2. Jorgensen WL (2004) The many roles of computation in drug discovery. Science 303:1813–1818.
3. Swinney DC (2004) Biochemical mechanisms of drug action: what does it take for success? Nat Rev Drug Discov 3:801–808.
4. Swinney DC (2009) The role of binding kinetics in therapeutically useful drug action. Curr Opin Drug Discov Devel 12:31–39.
5. Rich RL, Myszka DG (2004) Why you should be using more SPR biosensor technology. Drug Discov Today 1:301–308.
6. Harvey MJ, Giupponi G, De Fabritiis G (2009) ACEMD: accelerating biomolecular dynamics in the microsecond time scale. J Chem Theory Comput 5:1632–1639.
7. Buch I, Harvey MJ, Giorgino T, Anderson DP, De Fabritiis G (2010) High-throughput all-atom molecular dynamics simulations using distributed computing. J Chem Inf Model 50:397–403.
8. Wu C, Biancalana M, Koide S, Shea J (2009) Binding modes of Thioflavin-T to the Single-Layer β-Sheet of the peptide Self-Assembly mimics. J Mol Biol 394:627–633.
9. Biancalana M, Makabe K, Koide A, Koide S (2009) Molecular mechanism of Thioflavin-T binding to the surface of β-Rich peptide Self-Assemblies. J Mol Biol 385:1052–1063.
10. Brannigan G, LeBard DN, Hénin J, Eckenhoff RG, Klein ML (2010) Multiple binding sites for the general anesthetic isoflurane identified in the nicotinic acetylcholine receptor transmembrane domain. Proc Natl Acad Sci USA 107:14122.
11. Vanni S, Neri M, Tavernelli I, Rothlisberger U (2011) Predicting novel binding modes of agonists to β adrenergic receptors using All-Atom molecular dynamics simulations. PLoS Comput Biol 7:e1001053.
12. Selent J, Sanz F, Pastor M, Fabritiis GD (2010) Induced effects of sodium ions on dopaminergic G-Protein coupled receptors. PLoS Comput Biol 6:e1000884.
13. Ahmad M, Gu W, Helms V (2008) Mechanism of fast peptide recognition by SH3 domains. Angew Chem Int Ed 47:7626–7630.
14. Martí MA, Lebrero MCG, Roitberg AE, Estrin DA (2008) Bond or cage effect: how nitrophorins transport and release nitric oxide. J Am Chem Soc 130:1611–1618.
15. Enkavi G, Tajkhorshid E (2010) Simulation of spontaneous substrate binding revealing the binding pathway and mechanism and initial conformational response of GlpT. Biochemistry 49:1105–1114.
16. Noé F, Fischer S (2008) Transition networks for modeling the kinetics of conformational change in macromolecules. Curr Opin Struct Biol 18:154–162.
17. Warshel A, Sussman F, Hwang J (1988) Evaluation of catalytic free energies in genetically modified proteins. J Mol Biol 201:139–159.
18. Li X, He X, Wang B, Merz K (2009) Conformational variability of Benzamidinium-Based inhibitors. J Am Chem Soc 131:7742–7754.
19. Guvench O, Price DJ, Brooks CL (2005) Receptor rigidity and ligand mobility in trypsin–ligand complexes. Proteins 58:407–417.
20. Essex JW, Severance DL, Tirado-Rives J, Jorgensen WL (1997) Monte carlo simulations for proteins: Binding affinities for Trypsin-Benzamidine complexes via Free-Energy perturbations. J Phys Chem B 101:9663–9669.
21. Resat H, Marrone TJ, McCammon J (1997) Enzyme-Inhibitor association thermodynamics: Explicit and continuum solvent studies. Biophys J 72:522–532.
22. Doudou S, Burton NA, Henchman RH (2009) Standard free energy of binding from a One-Dimensional potential of mean force. J Chem Theory Comput 5:909–918.
23. Gervasio FL, Laio A, Parrinello M (2005) Flexible docking in solution using metadynamics. J Am Chem Soc 127:2600–2607.
24. Marquart M, Walter J, Deisenhofer J, Bode W, Huber R (1983) The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. Acta Crystallogr B 39:480–490.
25. Mares-Guia M, Shaw E (1965) Studies on the active center of trypsin. J Biol Chem 240:1579–1585.
26. Singhal N, Snow CD, Pande VS (2004) Using path sampling to build better markovian state models: Predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. J Chem Phys 121:415.
27. Huang D, Caflisch A (2011) The free energy landscape of small molecule unbinding. PLoS Comput Biol 7:e1002002.
28. Guillain F, Thusius D (1970) Use of proflavine as an indicator in temperature-jump studies of the binding of a competitive inhibitor to trypsin. J Am Chem Soc 92:5534–5536.

29. Prinz J, et al. (2011) Markov models of molecular kinetics: Generation and validation. J Chem Phys 134 In press.
30. Bowman GR, Ensign DL, Pande VS (2010) Enhanced modeling via network theory: Adaptive sampling of markov state models. J Chem Theory Comput 6:787–794.
31. Voelz VA, Bowman GR, Beauchamp K, Pande VS (2010) Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1-39). J Am Chem Soc 132:1526–1528.
32. Deng Y, Roux B (2009) Computations of standard binding free energies with molecular dynamics simulations. J Phys Chem B 113:2234–2246.
33. Singh N, Warshel A (2010) Absolute binding free energy calculations: On the accuracy of computational scoring of protein–ligand interactions. Proteins 78:1705–1723.
34. Hornak V, et al. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins 65:712–725.
35. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25:1157–1174.
36. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935.
37. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781–1802.
38. Feenstra KA, Hess B, Berendsen HJC (1999) Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. J Comput Chem 20:786–798.
39. Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR (2009) Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. Proc Natl Acad Sci USA 106:19011–19016.
40. Noé F, Horenko I, Schütte C, Smith JC (2007) Hierarchical analysis of conformational dynamics in biomolecules: transition networks of metastable states. J Chem Phys 126:155102.
41. Bowman GR, Beauchamp KA, Boxer G, Pande VS (2009) Progress and challenges in the automated construction of markov state models for full protein systems. J Chem Phys 131:124101.

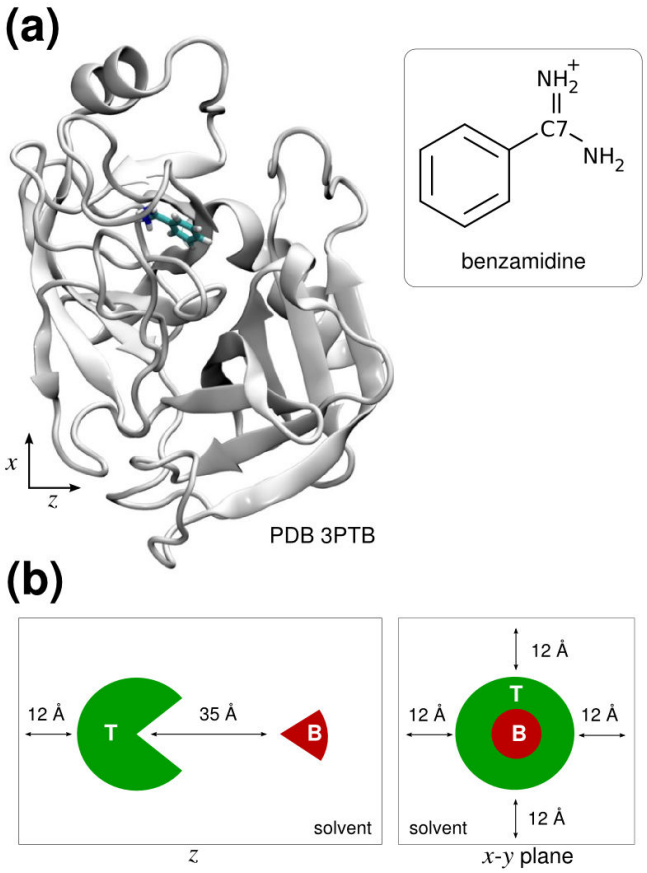

Figure 1. (a) Crystallographic bound structure of benzamidine and trypsin (PDB 3PTB). Parameters for benzamidine were derived from the GAFF forcefield, and RESP partial charges fitted with Gaussian. The C7 atom was used as a reference to define the reaction coordinate. (b) System setup. Trypsin (T) and benzamidine (B) are separated by a distance of 35 Å along z first, and then solvated with a TIP3P water box extending 12 Å from any atom in all directions.

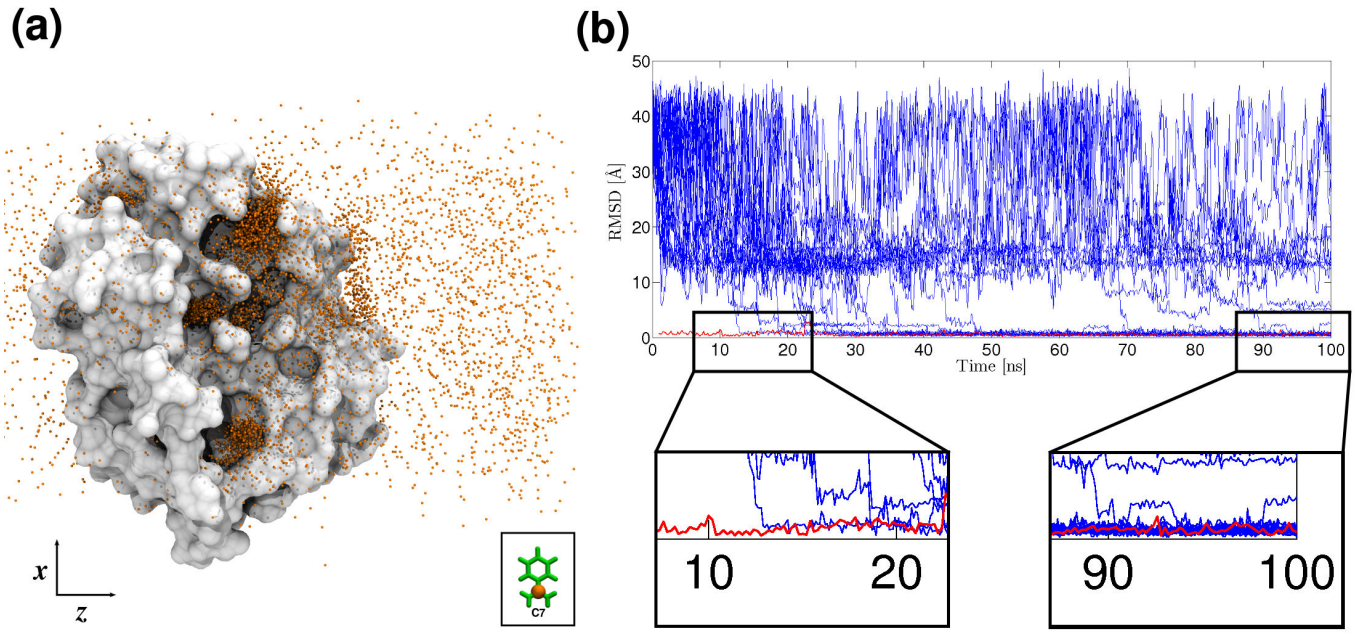

Figure 2. Raw data and binding events. (a) Visualization showing snapshots of the benzamidine C7 atom position during multiple trajectories. Some clusterings of benzamidine can already be seen on trypsin's surface. The kinetic relationship between these clusters is analyzed with the construction of a Markov State Model. (b) Time evolution of RMSD for benzamidine C3, C6 and C7 atoms. The plot shows only a fraction (5%) of the total runs for simplicity. Binding trajectories are considered as those 2 Å RMSD from the crystallographic pose, also calculated for reference (red line). While some trajectories are seen to bind as soon as 10 ns after the start of the simulation, others do it beyond the 90 ns or did not bind.

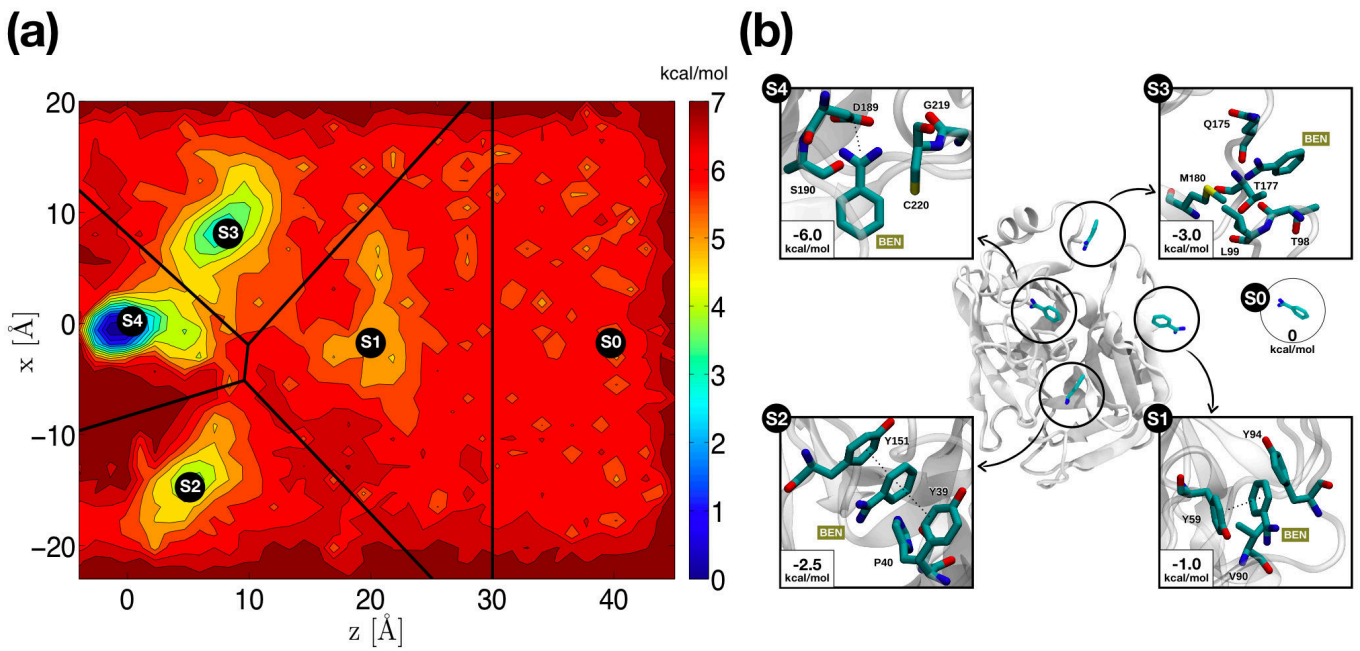

Figure 3. Identification of metastable states. (a) Potential of mean force (PMF) in the xz plane. 5 different metastable states can be identified from the different free energy minima (S0 to S4). The relative free energy between the unbound state S0 and the bound state S4 is -6 kcal/mol. Most probable transition to the bound state S4, may be from S3 from the fact that the barrier between the two states is of just 1.5 kcal/mol. (b) Structural characterization of metastable states. In states S1 and S2, benzamidine is stabilized by π-π stacking interactions with Y151 and Y39 sidechains. In S3, an hydrogen bond may be formed between NH_2 groups of benzamidine (only heavy atoms shown for clarity) and Q175 sidechain, or by a cation-π interaction between Q175 sidechain again, and benzamidine's benzene ring.

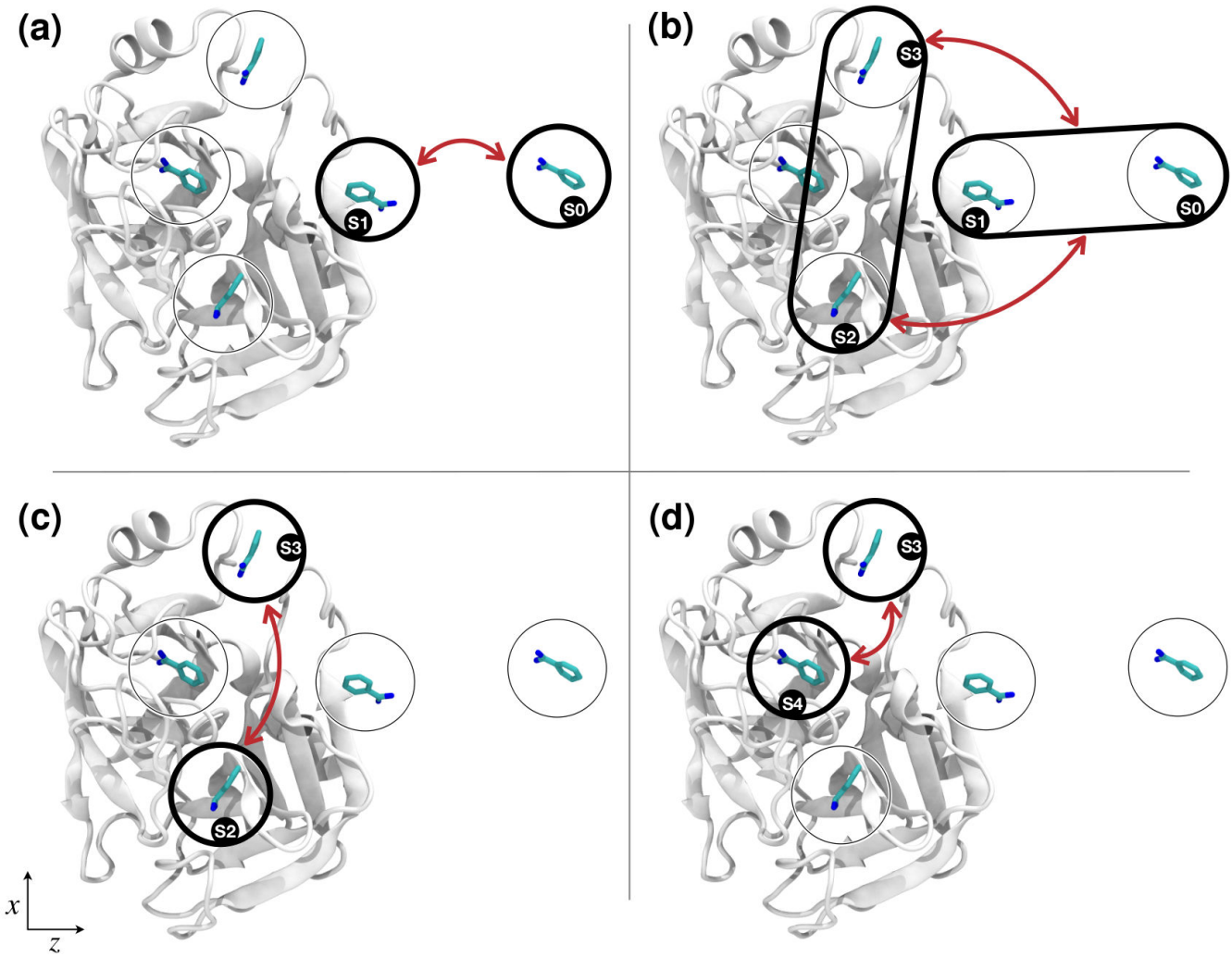

Figure 4. Characteristic transition modes for the metastable-states model. Each mode corresponds to a left eigenvector of the transition matrix in the 5-state coarse-grained model. Considering the slowest modes, on an 'implied timescale' (τ^*) between 6 and 10 ns we have transitions from state S0 to S1 (a) and collectively from states S0/S1 to states S3/S4 (b). With $\tau^* \approx 20$ ns, the model shows the transitions between states S2 and S3 (c). Transitions from lower energy states take longer times on average. The rate-limiting step of the process is the transition to the bound state S4, which is seen to happen at a $\tau^* \approx 60$ ns and preferentially from state S3 (d).

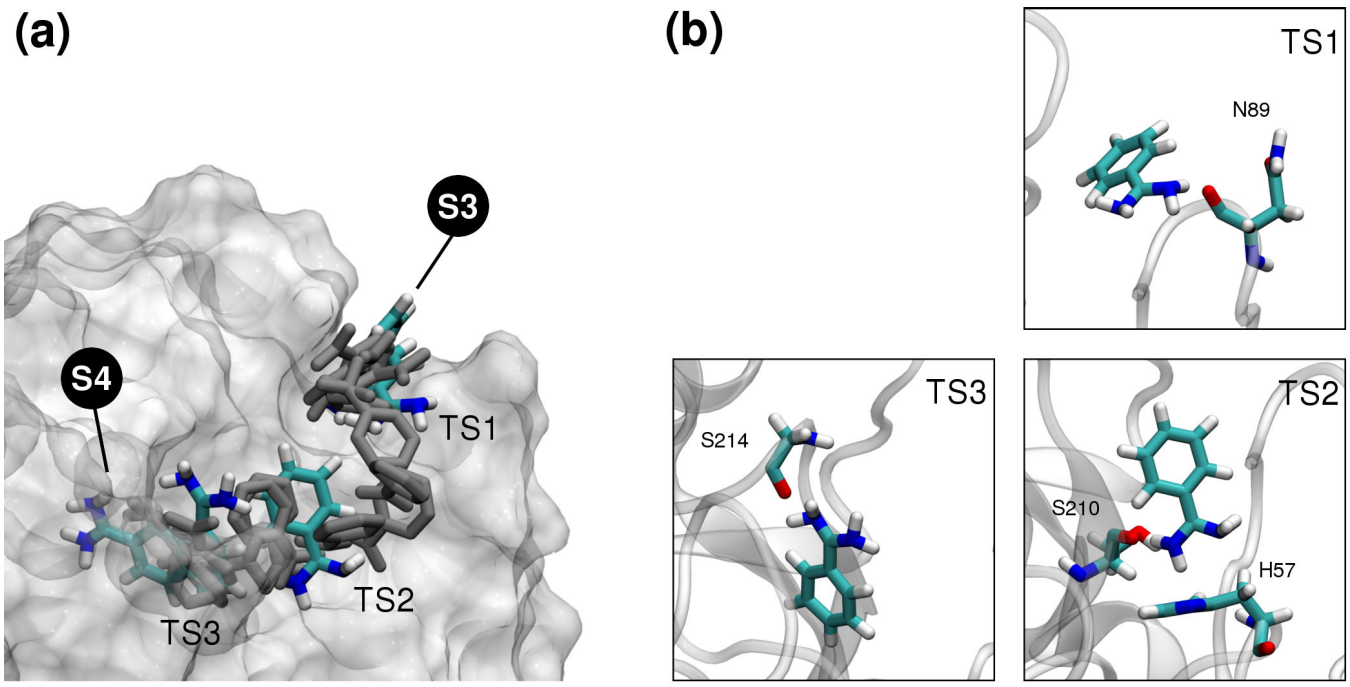

Figure 5. **Rate limiting transition**. (a) Transition between states S3-S4 often occurs as pictured. Benzamidine "rolls" over trypsin's surface by a set of successive transient interactions (transition states) TS1, TS2 and TS3 that canalizes the penetration of the inhibitor to its binding pocket. (b) Transition states are defined by the hydrogen-bond interactions formed between benzamidine and N89 main chain (TS1), with H57 main chain and S210 side chain (TS2), and finally with S214 main chain (TS3). Far from all binding trajectories following exactly this path; some trajectories bind passing only from TS2 and TS3, or directly to TS3, prior to burying into the binding pocket at the known bound conformation. As a side note, H57 belongs to the trypsin catalytic triad, responsible of the peptide bond cleavage activity of serine proteases.