Funnel metadynamics as accurate binding free-energy method

Vittorio Limongelli[a,1], Massimiliano Bonomi[b], and Michele Parrinello[c,d,1]

[a]Department of Pharmacy, University of Naples Federico II, I-80131 Naples, Italy; [b]Department of Bioengineering and Therapeutic Sciences, and California Institute of Quantitative Biosciences, University of California, San Francisco, CA 94158; [c]Department of Chemistry and Applied Biosciences, Eidgenössiche Technische Hochschule (ETH), 8006 Zürich, Switzerland; and [d]Facoltà di Informatica, Istituto di Scienze Computazionali, Università della Svizzera Italiana, CH-6900 Lugano, Switzerland

Contributed by Michele Parrinello, March 7, 2013 (sent for review December 20, 2012)

A detailed description of the events ruling ligand/protein interaction and an accurate estimation of the drug affinity to its target is of great help in speeding drug discovery strategies. We have developed a metadynamics-based approach, named funnel metadynamics, that allows the ligand to enhance the sampling of the target binding sites and its solvated states. This method leads to an efficient characterization of the binding free-energy surface and an accurate calculation of the absolute protein–ligand binding free energy. We illustrate our protocol in two systems, benzamidine/trypsin and SC-558/cyclooxygenase 2. In both cases, the X-ray conformation has been found as the lowest free-energy pose, and the computed protein–ligand binding free energy in good agreement with experiments. Furthermore, funnel metadynamics unveils important information about the binding process, such as the presence of alternative binding modes and the role of waters. The results achieved at an affordable computational cost make funnel metadynamics a valuable method for drug discovery and for dealing with a variety of problems in chemistry, physics, and material science.

enhanced sampling | protein/ligand binding | ligand docking

Studying the molecular interactions between a drug and its target helps in understanding the target functional mechanism and offers the possibility for exogenous control of its physiological activity. In recent years, a vast experimental and computational effort has revealed in ever-more-precise detail the ligand/target recognition mechanism (1, 2). In this context, an accurate estimation of the ligand-binding affinity is in great demand because it would facilitate many steps of the drug discovery pipeline, such as structure-based drug design and lead optimization; this is not, however, a simple task. In fact, an accurate estimation of the binding affinity or, equivalently, the absolute protein–ligand binding free energy, requires an accurate description of the ligand/protein interactions, their flexibility, and the solvation process. Many methods have been proposed to tackle this problem. For instance, docking protocols are widely used to generate and rank candidate poses based on empirical scoring functions, either physically or statistically based (3–5). These techniques have been proven to be highly efficient in screening a large number of compounds in a short time (6); this, however, at the price of limited accuracy in estimating affinities (7).

Alternatively, a variety of methods to describe ligand/protein interactions in a more accurate way at higher computational cost have been proposed. These techniques can be grouped in two categories: (*i*) endpoint and (*ii*) pathway methods. The former group is composed of those techniques that sample ligand and protein in unbound and bound states and compute the protein–ligand binding free energy by taking the difference between the absolute free energy of these two states. Examples include microscopic linear response approximation (8), linear interaction energy (9, 10), protein dipoles Langevin dipoles (11), as well as molecular mechanics Poisson–Boltzmann surface area, and generalized Born surface area (12).

At variance with endpoint methods, in pathway methods, the ligand is gradually separated from the protein. The binding free energy is then obtained by summing different contributions coming from a discretized path that connects the initial and final state. This class includes methods in which the ligand/protein interactions are gradually switched off, such as thermodynamic integration (13), free-energy perturbation (14, 15), double-decoupling method (16), and double-annihilation method (17). Techniques such as steered molecular dynamics (SMD) (18) and umbrella sampling (19), where the ligand and the protein are physically separated from each other, also belong to this group. While in SMD, the ligand is dragged out from the protein using a moving restraining potential, in umbrella sampling, the path from the bound to the unbound state is divided in a finite number of windows, which are independently sampled.

Though these methods have been successfully used to compute the ligand binding free energy in many cases (20–22), the requirement of knowing in advance the binding mode hampers a more general applicability. The intensity of the efforts in developing these methods reflects both the great potential of these calculations and their difficulties. In particular, the difficulties arise mainly from the fact that the ligand/protein binding process is a rare event, difficult to sample with standard techniques such as molecular dynamics (MD). Even the most ambitious efforts in this direction, though revealing precious details of the binding process (23, 24), have not been able to determine accurately the binding energy. To achieve this result, the use of enhanced sampling methods is mandatory.

Among the emerging techniques, metadynamics (25) has proven to be very useful in studying long-timescale processes (26, 27), particularly in complex ligand/protein binding cases (28–30). Metadynamics works by adding an external history-dependent potential that acts on few degrees of freedom, named collective variables (CVs). In such a way, the sampling is accelerated, and the free-energy surface (FES) of the process can be calculated from the added potential. Unfortunately, only a qualitative estimation of the protein–ligand binding free energy could be obtained for the binding processes studied so far (28, 31). In fact, once the ligand leaves the binding pocket, it has difficulty finding its way back, and starts exploring all of the possible solvated states. These conformations represent a vast part of the configuration space that cannot be sampled thoroughly in a limited computation time. Therefore, once out, the ligand does not again find the binding site, and multiple binding/unbinding events, which are the key to an accurate determination of the binding free energy in metadynamics, cannot be observed.

Here, we present a metadynamics-based approach, named funnel metadynamics (FM), which overcomes all these limitations and allows an accurate estimation of the absolute

Author contributions: V.L., M.B., and M.P. designed research; V.L. performed research; V.L., M.B., and M.P. contributed new reagents/analytic tools; V.L., M.B., and M.P. analyzed data; and V.L., M.B., and M.P. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. E-mail: vittoriolimongelli@gmail.com or parrinello@phys.chem.ethz.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1303186110/-/DCSupplemental.

protein–ligand binding free energy. In particular, in FM, a funnel-shaped restraint potential is applied to the system, reducing the space to explore in the unbound state. In such a way, the sampling of ligand-bound and -unbound states is highly enhanced, thus leading to an accurate estimation of the binding FES within a reasonable simulation time. We have used FM to study the binding process in two different systems: benzamidine/trypsin and SC-558/cyclooxygenase 2 (COX-2). The latter system represents a particularly challenging case study with the coexistence of protein motions, solvent effect, and complex binding pathways. In both cases, the X-ray pose turned out to have the lowest free-energy value, and the computed absolute protein–ligand binding free energy was in good agreement with experiments. Furthermore, from FM simulations, important information on the binding process has been disclosed, such as the role of waters in the benzamidine/trypsin case and the presence of an alternative ligand binding mode in COX-2.

Using FM, no a-priori information about the ligand binding mode is required, and the exploration capability of the original method remains unaltered. Thus, the exploration of buried binding sites is possible, while taking into account slow protein motion and solvent effects. The present protocol represents a most valuable method to study ligand/protein interaction, and its relatively low computational cost renders its use appealing even in industrial applications where speed is valued.

Theoretical Background

Funnel Idea. Sampling ligand/protein binding with all-atoms MD simulations in explicit solvent is extremely attractive because it can provide molecular information at high resolution. Unfortunately, these processes usually have a long timescale and therefore cannot be sampled at reasonable computational costs. Nevertheless, the advance in computer power and the advent of graphics processor units have allowed the study of ligand/protein binding processes using plain MD simulations, however still at great computational cost and using dedicated hardware (23, 24). In fact, tens of microseconds of MD simulations were necessary to collect enough statistics to describe the ligand binding process.

An alternative method is to use enhanced sampling techniques to access long-timescale events within a reasonable computational time. One of these techniques is umbrella sampling, which has largely been used to study ligand/protein interactions and to compute the absolute protein–ligand binding free energy (20, 21). However, this technique fails in exploring thoroughly the fully solvated state of the ligand. This limitation has been remedied by using a cylindrical restraint potential to reduce the sampling space (32, 33). The effect of the external potential can be rigorously taken into account (34), and the binding constant K_b in the presence of the restraint is given by

$$K_b = e^{\beta \Delta G_{site}} \int_{site} dz \; e^{-\beta\left[W(z)-W_{ref}\right]} S_u. \qquad [1]$$

Here, ΔG_{site} is the change in the free energy of the bound state caused by the presence of the restraint, $\beta = (k_B\, T)^{-1}$, k_B is the Boltzmann constant, T the temperature of the system, and S_u is the cross-section of the cylinder. The potential $W(z)$ and its value in the unbound state, W_{ref}, can be derived from the potential of mean force (PMF). Eq. **1** provides an unbiased estimator of K_b, independent of the choice of the restraint potential. If the radius of the cylinder is chosen to be much larger than the lateral fluctuations of the ligand in the binding site, the restraint potential is not felt by the ligand in the binding site, thus $\Delta G_{site} = 0$. Unfortunately, the choice of the radius is not simple. In fact, though a small radius value limits the exploration in the unbound state, it reduces also the exploration of the binding site. However, choosing a large radius is not advantageous because it increases the space to be sampled in the unbound state and hence the computational time. As a result, the PMF calculation might be affected by this choice and, consequently, the estimation of the protein–ligand binding free energy (21).

To overcome this limitation, we have developed a funnel-shaped restraint potential that can be applied to the target protein. This potential is a combination of a cone restraint, which includes the binding site, and a cylindric part, which is directed toward the solvent (Fig. 1). Using the funnel potential during the simulation, the system does not feel any repulsive bias when the ligand explores regions inside the funnel area. As the ligand reaches the edge of the funnel, a repulsive bias is applied to the system, disfavoring it from visiting regions outside the funnel. As can be seen in Fig. 1, if the shape of the funnel is properly chosen, the sampling at the binding site is not affected by the external bias, whereas in the bulk water the repulsive potential reduces the space to be explored to a cylindric region; this favors the observation of multiple binding/unbinding events leading to a faster convergence of the results.

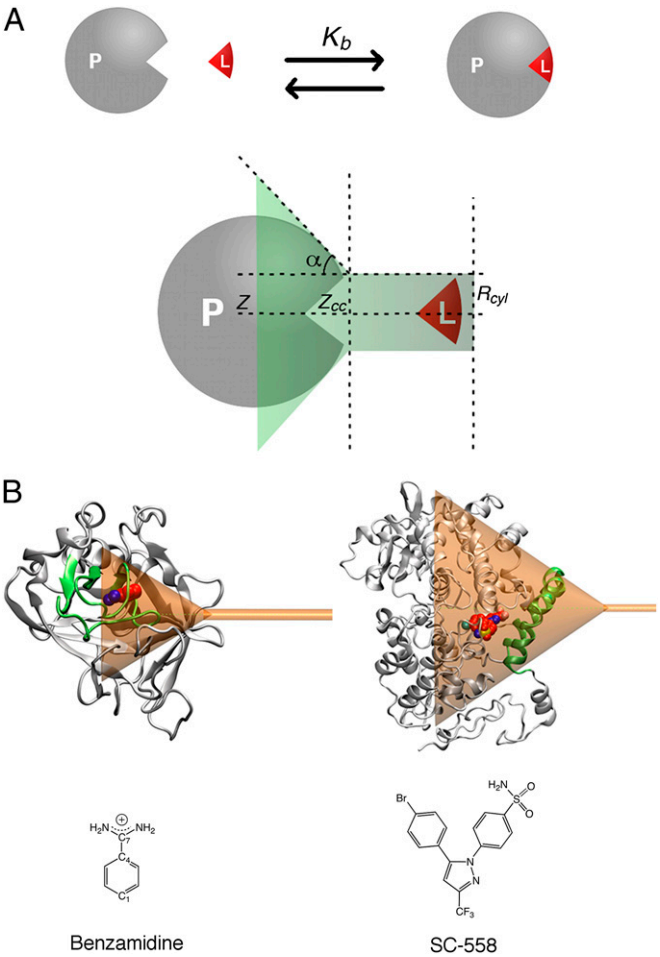


Fig. 1. (*A*) Schematic representation of the ligand/protein binding process and the funnel restraint potential used in FM calculations. The shape of the funnel can be customized on the target by setting a few parameters. In particular, given z, the axis defining the exit-binding path of the ligand, z_{cc} is the distance where the restraint potential switches from a cone shape into a cylinder. The α-angle defines the amplitude of the cone, and R_{cyl} is the radius of the cylindrical section. (*B*) The funnel restraint potential applied to trypsin (*Upper Left*) and COX-2 (*Upper Right*) enzymes with the ligands considered in the study, benzamidine (*Lower Left*) and SC-558 (*Lower Right*). In the trypsin case, α is 0.55 rad and z_{cc} is 18 Å (Table S1). In the COX case, α is 0.6 rad and z_{cc} is 44 Å (Table S2). In both cases, R_{cyl} is set to 1 Å.

Furthermore, if the funnel restraint potential is used in free-energy calculations with techniques such as metadynamics, the free-energy difference between bound and unbound states depends exclusively on the free-energy value at the two states, independently from the path that connects one state to the other; this is advantageous with respect to other methods, such as umbrella sampling. In fact, in umbrella sampling, where the PMF is calculated using the different contributions coming from the simulation windows that connect the bound state to the unbound one, all of the simulation windows in the bound states must not feel the restraint potential to provide an easy estimate of the PMF ($\Delta G_{site} = 0$). If this condition is not fulfilled, the PMF calculation and, consequently, the binding free-energy estimate, is more complex ($\Delta G_{site} \neq 0$) and depends on the chosen path.

Absolute Protein–Ligand Binding Free Energy. In free-energy calculations, the absolute protein–ligand binding free energy $\Delta G_b{}^0$ is typically computed using the following formula:

$$\Delta G_b^0 = -\frac{1}{\beta} ln\left(C^0 \; K_b\right), \qquad [2]$$

where K_b is the equilibrium binding constant, and $C^0 = 1/1{,}660$ Å^{-3} is the standard concentration. Using metadynamics calculations with the funnel restraint potential and using Eq. **1** with some rearrangements, as reported in *Methods*, Eq. **2** becomes

$$\Delta G_b^0 = \Delta G - \frac{1}{\beta} ln\left(\pi R_{cyl}^2 C^0\right), \qquad [3]$$

where, ΔG is the free-energy difference between bound and unbound states, and πR^2_{cyl} is the surface of the cylinder used as restraint potential. With β and C^0 being constant, the absolute protein–ligand binding free energy is equal to ΔG minus the analytical correction in Eq. **3** (see *Methods* for details). It is important to stress that the funnel restraint potential ensures a number of recrossing events between the different states visited by the system during the simulation, leading to a quantitatively well-characterized free-energy profile and a converged estimation of ΔG.

Results

We have used the funnel restraint potential in combination with well-tempered metadynamics (35), hereafter named FM, to study two ligand/protein binding cases and compute the absolute protein–ligand binding free energies. The first case is the trypsin/benzamidine complex, mainly used as a reference model, and the second is COX-2 in complex with the potent inhibitor SC-558. The latter, which has been previously studied by us (28), represents a challenging ligand/protein binding case.

Benzamidine/Trypsin System. The benzamidine/trypsin system has been studied using several different computational approaches (21, 23, 31, 36, 37). The fact that the binding pocket in trypsin is almost solvent-exposed makes this system a good test model for new docking methods. The funnel restraint potential has been chosen in such a way that the cone section includes the whole binding site. This condition can be fulfilled by properly setting two parameters, the angle α and the distance z_{cc}, to ensure that during the ligand exploration of the binding site, z values $< z_{cc}$, the sampling is not affected by the external bias (Fig. 1; *Methods*). When the ligand is completely in bulk water, $z > z_{cc}$, a cylindric restraint potential is applied to the system and the free energy of the ligand-unbound state can be computed in a similar way as in Allen et al. (32).

Bound State. The whole sampling took ∼0.5 μs of FM simulations. Looking at the FES computed as a function of the projection on z axis, where z is the axis of the funnel restraint potential and a torsion CV (*Methods*; Table S1), two main basins can be detected (Fig. 2). The deepest one, basin A in Fig. 2, corresponds to the benzamidine in its crystal pose, where a number of strong interactions with the enzyme are formed (Fig. 2; *SI Text*). It is interesting to note that in this pose, a water molecule is present in the binding pocket and forms an H-bond network with the side chains of Tyr228, Asp189, and Ser190. A water molecule is present in a similar position in many X-ray structures (e.g., PDB ID code 3atl) (38), thus suggesting a structural role for this molecule in the trypsin binding site (Fig. S1). The second energy minimum, basin B, is ∼1 kcal/mol (1 kcal = 4.18 kJ) higher than basin A. Here, the ligand is slightly rotated in the binding site. Overall, however, the interactions established in pose A are conserved (Fig. 2). In particular, the diamino group of benzamidine engages a direct H-bond with Ser190, and a water bridge interaction is established with the Asp189 side chain. Furthermore, the aromatic ring of the ligand is involved in an interaction with the sulfur atoms forming the disulfide bridge between Cys191 and Cys220. It is interesting to note that in basin B the diamino group of benzamidine engages H-bond interactions with the carbonyl oxygens of Val227, Val213, and Ser214 via two water molecules. Water molecules located at similar positions in the binding site can be found in the X-ray structures of trypsin in the apo and ligated form (PDB ID codes 1s0q and 3atl, respectively; Fig. S1).

Our results show an important role played by waters during ligand binding, and a similar functional role has been reported also in other studies (29, 39). Therefore, the use of atomistic simulations with explicit solvent is mandatory to take into account the solvation effect and have an accurate estimation of ligand/protein interactions.

The stability of the basin B pose has been further assessed through an over 100-ns-long unbiased MD simulation. During the whole simulation, the ligand binding mode is stable conserving all of the interactions described above (*SI Text*; Fig. S2). The depth of basin B and the good stability of the ligand/protein interactions lead us to consider this pose the first binding event of benzamidine in the active site before reaching its final position in basin A. Alternatively, the basin B pose can be considered the first unbinding event of the ligand from the catalytic site of the enzyme. It is important to stress that both basin A and B are within the cone region, and their exploration is not affected by the funnel restraint potential (Fig. S3).

Finally, we note that in a recent study by Söderhjelm et al. (37), a binding mode highly similar to pose B was found, and that in our simulations the ligand often occupies, along its way to the binding site, positions close to the states described in this study.

A movie showing the benzamidine binding/unbinding to trypsin under the action of FM is provided (Movie S1).

Unbound State. In the unbound state, the ligand has no contact with the protein and can assume a large number of conformations, which are represented by states at z values higher than z_{cc}. As shown in Fig. 2, in the solvated state the free energy is completely flat. Although this is to be expected, because when the ligand is fully solvated and outside the interaction range of the protein its free energy should be position- and orientation-independent, the flatness of the FES gives a measure of the convergence of our calculations.

Protein–Ligand Binding Free Energy. To have a quantitatively well-characterized free-energy profile, a number of recrossing events between the different states visited by the system should be observed (26). As shown in Fig. S4, during the simulation, the system visits several times the bound ($z < 7$ Å) and the unbound states ($z > 30$ Å). At the end of the simulation, ΔG is equal to −12.3 kcal/mol. Considering the analytical correction of 3.8 kcal/mol, calculated as reported in Eq. **3**, the final binding free energy of benzamidine to trypsin is −8.5 ± 0.7 kcal/mol (Fig. S3). This

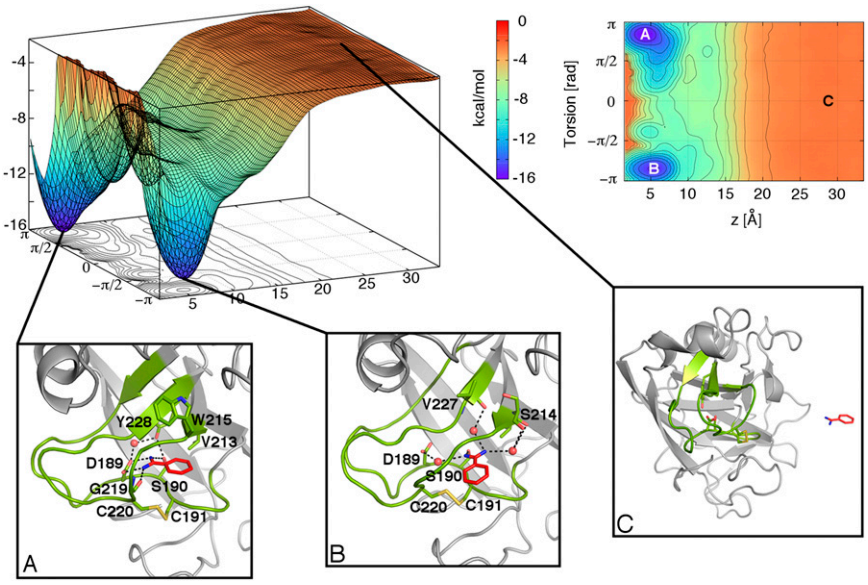


Fig. 2. The FES of the benzamidine/trypsin binding is computed using a reweighting algorithm (50) as a function of the projection on the *z* axis of the ligand center of mass and a torsion CV, where *z* is the axis of the funnel (Table S1). Isosurfaces are shown every 1 kcal/mol. (*Insets A–C*) Conformations representing the bound poses, basin A (*Inset A*) and basin B (*Inset B*). *C* shows one of the isoenergetic conformations representing the unbound state. In the unbound state, $z > 20$ Å, the ligand can assume a large number of orientations, which are represented by states with similar energy values, as shown in the FES.

value falls in the range of the previous calculations of trypsin/benzamidine binding free energy (−5.5 to −9.0 kcal/mol) (21) and is in line with the experimental measurements (−6.4 to −7.3 kcal/mol) (40, 41).

To provide a picture of the convergence of the binding free-energy estimation, the free-energy difference between bound and unbound states has been computed as function of the simulation time (Fig. S3). Fig. S3 clearly shows that after 400 ns, the free energy is converged. In fact, though the system continues to go from bound to the unbound states (Fig. S4), the estimation of the free-energy difference between these two states does not change significantly.

SC-558/COX-2 System. Encouraged by the results obtained on the benzamidine/trypsin system, we decided to use FM in a more complex case study: the binding of SC-558 to COX-2. This study is particularly challenging because the binding site is buried, and substantial motion of the protein scaffold is needed to allow the ligand in. Furthermore, the ligand needs to dispose of its solvation water to access to the binding site. This system has already been studied by us, and the existence of a second binding mode different from the crystallographic one has been revealed (28). In this work, we could not give an accurate estimation of the binding energy due to the difficulty of exploring the solvated state. However, the binding cavity and the free-energy difference between the two poses were properly and fully described.

We used a CV setting similar to that of Limongelli et al. (28). In particular, the projection on the *z* axis of the funnel of the ligand center of mass and a torsion have been chosen as CVs to describe the ligand motion with respect to the protein. Furthermore, a path CV has been used to take into account the motion of the helices that form the gate to access the active site (*Methods*; Tables S2 and S3). As in the trypsin/benzamidine case, the funnel parameters (z_{cc} and α) were chosen so as not to alter the potential seen by the ligand inside the binding site (Fig. 1; *Methods*).

The FES was converged after ∼350 ns, and for the part inside, the binding cavity was very similar to that described in ref. 28. In fact, as in ref. 28, two binding poses were found, basin A, which corresponds to the X-ray pose, and basin B, that was first described in ref. 28 and corresponds to an alternative binding mode of SC-558 in COX-2 (Fig. 3; *SI Text*). As can be seen in Fig. 3, these two basins are within the cone region, and their exploration is not affected by the funnel restraint potential. Contrary to ref. 28, we can now thoroughly sample the unbound state, which is reached at $z > 44$ Å from the binding site. Here, the ligand is outside the interaction range of the protein and can assume a large number of orientations. As seen in the trypsin case, in the unbound state the FES is flat, corresponding to different states of the system with similar free-energy values.

The free-energy difference between bound and unbound states, ΔG, of SC-558 is equal to −14.9 kcal/mol, which, considering the analytical correction of 3.8 kcal/mol, leads to an estimate of the absolute binding free energy of -11.1 ± 1.5 kcal/mol (Fig. S5). This value is in good agreement with the experimental one of ∼−12 kcal/mol, derived from the ligand IC_{50} value (42) via the Cheng–Prusoff relation $\Delta G_b^{\,0} = -K_b\, T\, ln\, IC_{50}/(1 + (S/K_m))$ (43) using the substrate concentration (S) and the concentration of substrate at which enzyme activity is at half maximal (K_m), reported in ref. 44.

These results show that FM preserves the exploration capability of metadynamics in the binding cavity. At the same time, sampling of bound and unbound states is considerably enhanced (Fig. S6), and an accurate estimation of the absolute protein–ligand binding free energy can finally be computed. One might note that COXs are monotopic membrane proteins, and that in our model we do not consider the membrane. This represents an approximation that can be however tolerated, because COXs binding site is hydrated with the waters playing a key role in the cyclooxygenase site (45). Furthermore, most COX inhibitors (e.g., aspirin, ibuprofen, flurbiprofen) are not lipophilic enough to cross the membrane, suggesting that they can follow an extracellular binding pathway to reach the catalytic site. Therefore, this approximation does not affect the estimate of the ligand/protein binding free energy, and the agreement with the experimental value supports the reliability of our model. In this difficult case, in which the binding site is buried and the use of a larger number of CVs is necessary, FM has proven to be very powerful in studying the binding process and accurately computing the protein–ligand binding free energy.

Discussion

Here, we have developed a metadynamics-based approach, named FM, which allows us to compute the absolute protein–ligand binding free energy, leaving unaltered the exploration capability of the method. Thanks to a funnel-shaped restraint potential, the sampling of bound and unbound states is highly enhanced, leading to a quantitatively well-characterized binding free-energy surface within a reasonable simulation time. We have used FM

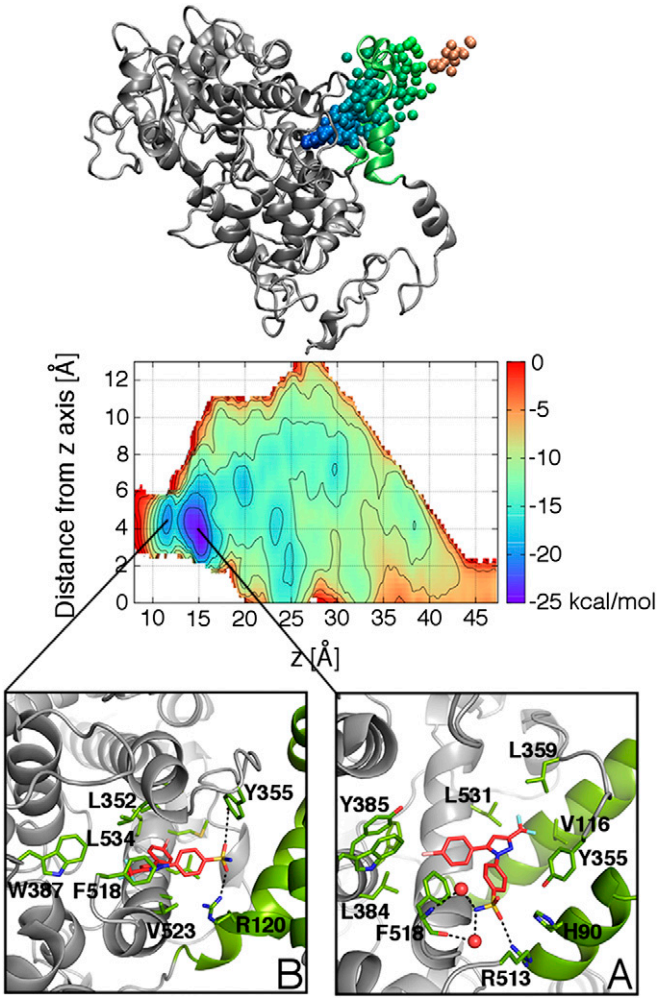


Fig. 3. The FES of the SC-558/COX-2 binding is computed using a reweighting algorithm (50) as a function of the projection on the *z* axis and distance from *z* axis of the center of mass of the ligand, where *z* is the axis of the funnel (Table S2). Isosurfaces are shown every 3 kcal/mol. (*Insets A and B*) Binding modes corresponding to the two deepest energy basins, basin A (X-ray) and basin B (alternative pose). (*Upper*) Ligand position relative to the enzyme during FM simulations. The spheres represent the ligand center of mass and are colored according to their corresponding free-energy values. For clarity, only selected frames are shown.

to sample the ligand binding process in two different systems: benzamidine/trypsin and SC-558/COX-2. In both cases, the X-ray pose was the lowest-energy pose, and the estimated binding free energy was in good agreement with experiments. In both systems, the whole simulation of the binding process took less than 0.5 μs, which is definitely less demanding than other MD-based approaches that take tens of microseconds of simulations (23, 24). Furthermore, FM allows us to compute the absolute protein–ligand binding free energy by overcoming some of the difficulties related to other methods. For instance, with respect to umbrella sampling, FM allows the use of a larger number of collective variables, as in the COX case, still achieving convergence of the results at an affordable computational cost; this allows us to describe the slow degrees of freedom of the system, which is often necessary in complex ligand/protein interactions. Furthermore, using FM, information about the ligand binding mode is not required in advance, although the approximate location of the binding pocket in the target structure should be known. However, if this information is not available, FM can be combined with other methods, such as virtual screening protocols, or more advanced techniques (37), that are more efficient in finding candidate binding sites. Furthermore, FM can also be combined with methods, such as free-energy perturbation, which are able to estimate the ligand binding affinity for a series of analogs (46); and whereas the former helps in finding the correct ligand binding mode with an accurate energy estimation of the binding event, the latter can be used to assess changes in the ligand/protein binding affinity with respect to different ligand substitutions.

We note that the application of FM can go beyond ligand/protein binding studies. In fact, it can be exploited in many other research fields where two-body interactions are important. For instance, FM simulations can be used to study the interactions between atoms in crystal growth (47). Furthermore, the funnel-shaped restraint potential can be used in combination with other enhanced sampling methods, such as bias exchange, parallel tempering, or replica exchange (26), to achieve convergence in even shorter computational time. Finally, one might customize on the studied system the shape of the restraint potential, changing the cone part and leaving unaltered the cylindrical region of the potential for the unbound state.

The results achieved here at a reasonable computational cost make FM a valuable method to tackle a wide variety of problems in chemistry, physics, and material science.

Methods

FM Theory. The absolute protein–ligand binding free energy ΔG_b^0 can be defined in terms of the equilibrium binding constant K_b, as seen in Eq. **2**. Following the approach of Allen et al. (32), we can write K_b in terms of the PMF $W(\boldsymbol{r})$ as

$$K_b = \frac{\int d\boldsymbol{r}\ H_{site}(\boldsymbol{r})\ e^{-\beta W(\boldsymbol{r})}}{e^{-\beta W(\boldsymbol{r}')}}, \qquad [4]$$

where $\boldsymbol{r}'$ is a reference ligand position in the bulk water, and $H_{site}(\boldsymbol{r})$ is equal to 1 inside the binding site and 0 elsewhere. We want to express K_b in terms of the 1D PMF $W(z)$, which is obtained by integrating over the allowed lateral displacement in the *xy* space. We assume that the binding site is contained in the range $z_{min} \le z \le z_{max}$. The allowed displacement is determined by a hybrid truncated conical and cylindrical restraint $H_R(x,y,z)$ with axis aligned to the *z* axis, as follows:

$$H_R(x,y,z) = \begin{cases} H_{cone}(x,y,z); & z \le z_{cc} \\ H_{cyl}(x,y); & z > z_{cc} \end{cases} \qquad [5]$$

with the condition $z_{cc} > z_{max}$. The truncated conical restraint is defined as

$$H_{cone}(x,y,z) = \begin{cases} 1; & \sqrt{x^2+y^2} \le R(z) \\ 0; & \sqrt{x^2+y^2} > R(z) \end{cases}, \qquad [6]$$

with $R(z) = R_{cyl} + \tan(\alpha)\,(z_{cc} - z)$, and α modulating the aperture of the cone. The cylindrical restraint is defined as

$$H_{cyl}(x,y) = \begin{cases} 1; & \sqrt{x^2+y^2} \le R_{cyl} \\ 0; & \sqrt{x^2+y^2} > R_{cyl} \end{cases}. \qquad [7]$$

The 1D PMF $W(z)$ is given by

$$e^{-\beta W(z)} = C \int dx\ dy\ H_R(x,y,z)\, e^{-\beta W(x,y,z)}. \qquad [8]$$

We define the 1D PMF to be zero in the bulk water at $z = z_{bulk}$, with $z_{bulk} > z_{cc}$. In this region, $H_R(x,y,z) = H_{cyl}(x,y)$, and the constant C in Eq. **8** can be written as

$$C = \frac{1}{\int dx\ dy\ H_{cyl}(x,y) e^{-\beta W(x,y,z_{bulk})}}. \qquad [9]$$

Because in the bulk water, $W(x,y,z)$ is independent from *x* and *y*, we can write

$$W(x,y,z_{bulk}) = W(0,0,z_{bulk}), \qquad [10]$$

and the constant C can be written as

$$C = \frac{e^{+\beta W(0,0,z_{bulk})}}{\pi R_{cyl}^2}. \quad [11]$$

By using the expression above, we can then rewrite Eq. **8** as

$$e^{-\beta W(z)} = \frac{1}{\pi R_{cyl}^2} \int dx\ dy\ H_R(x,y,z)\ e^{-\beta(W(x,y,z)-W(0,0,z_{bulk}))}. \quad [12]$$

Let us assume that in the binding site, the (truncated conical) restraint does not act on the ligand, i.e., the free-energy cost introduced by the restraint is zero when the ligand is in the binding site. We can write

$$H_{site}(\boldsymbol{r}) = h_{site}(z)\, H_R(x,y,z), \quad [13]$$

where

$$h_{site}(z) = \begin{cases} 1; & z_{min} \le z \le z_{max} \\ 0; & z < z_{min} \\ 0; & z > z_{max} \end{cases}. \quad [14]$$

If we substitute Eqs. **13** and **12** in Eq. **4**, we obtain an expression for the binding constant in terms of the PMF $W(z)$:

$$\begin{aligned} K_b &= \int dx\ dy\ dz\ H_{site}(x,y,z)\ e^{-\beta(W(x,y,z)-W(0,0,z_{bulk}))} \\ &= \int dz\ h_{site}(z) \int dx\ dy\ H_R(x,y,z)\ e^{-\beta(W(x,y,z)-W(0,0,z_{bulk}))} \\ &= \pi R_{cyl}^2 \int_{z_{max}} dz\ h_{site}(z)\ e^{-\beta W(z)} \\ &= \pi R_{cyl}^2 \int_{z_{min}} dz\ e^{-\beta W(z)} \end{aligned}. \quad [15]$$

FM simulations have been carried out with NAMD code (48) using the PLUMED plugin (49). Details on FM simulations are reported in *SI Text*.

ACKNOWLEDGMENTS. V.L. thanks Prof. Benoit Roux for fruitful discussions. Funding for this study was provided by European Union Grant ERC-2009-AdG-247075 and Swiss National Supercomputing Center Project s233.

1. Jorgensen WL (2004) The many roles of computation in drug discovery. *Science* 303(5665):1813–1818.
2. Gilson MK, Zhou HX (2007) Calculation of protein–ligand binding affinities. *Annu Rev Biophys Biomol Struct* 36:21–42.
3. Goodsell DS, Olson AJ (1990) Automated docking of substrates to proteins by simulated annealing. *Proteins* 8(3):195–202.
4. Meng EC, Shoichet BK, Kuntz ID (1992) Automated docking with grid-based energy evaluation. *J Comput Chem* 13(4):505–524.
5. Abagyan RA, Totrov MM, Kuznetsov DA (1994) ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J Comput Chem* 15(5):488–506.
6. Schlessinger A, et al. (2011) Structure-based discovery of prescription drugs that interact with the norepinephrine transporter, NET. *Proc Natl Acad Sci USA* 108(38): 15810–15815.
7. Leach AR, Shoichet BK, Peishoff CE (2006) Prediction of protein–ligand interactions. Docking and scoring: Successes and gaps. *J Med Chem* 49(20):5851–5855.
8. Lee FS, Chu ZT, Bolger MB, Warshel A (1992) Calculations of antibody-antigen interactions: Microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603. *Protein Eng* 5(3):215–228.
9. Aqvist J, Medina C, Samuelsson JE (1994) A new method for predicting binding affinity in computer-aided drug design. *Protein Eng* 7(3):385–391.
10. Jones-Hertzog DK, Jorgensen WL (1997) Binding affinities for sulfonamide inhibitors with human thrombin using Monte Carlo simulations with a linear response method. *J Med Chem* 40(10):1539–1549.
11. Madura JD, Nakajima Y, Hamilton RM, Wierzbicki A, Warshel A (1996) Calculations of the electrostatic free energy contributions to the binding free energy of sulfonamides to carbonic anhydrase. *Struct Chem* 7(2):131–137.
12. Srinivasan J, Cheatham TE III, Cieplak P, Kollman PA, Case DA (1998) Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J Am Chem Soc* 120(37):9401–9409.
13. Straatsma TP, McCammon JA (1991) Multiconfiguration thermodynamic integration. *J Chem Phys* 95(2):1175–1188.
14. Zwanzig RW (1954) High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J Chem Phys* 22(8):1420–1426.
15. Jorgensen WL, Ravimohan C (1985) Monte Carlo simulation of differences in free energies of hydration. *J Chem Phys* 83(6):3050–3054.
16. Gilson MK, Given JA, Bush BL, McCammon JA (1997) The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys J* 72(3): 1047–1069.
17. Jorgensen WL, Buckner JK, Boudon S, Tirado-Rives J (1988) Efficient computation of absolute free energies of binding by computer simulations. Application to methane dimer in water. *J Chem Phys* 89(6):3742–3746.
18. Izrailev S, et al. (1999) Steered molecular dynamics. *Computational Molecular Dynamics: Challenges, Methods, Ideas* (Springer, Berlin), pp 39–64.
19. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* 13(8):1011–1021.
20. Woo HJ, Roux B (2005) Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc Natl Acad Sci USA* 102(19):6825–6830.
21. 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(4):909–918.
22. Gumbart JC, Roux B, Chipot C (2013) Standard binding free energies from computer simulations: What is the best strategy? *J Chem Theory Comput* 9(1):794–802.
23. Buch I, Giorgino T, De Fabritiis G (2011) Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc Natl Acad Sci USA* 108(25):10184–10189.
24. Shan Y, et al. (2011) How does a drug molecule find its target binding site? *J Am Chem Soc* 133(24):9181–9183.
25. Laio A, Parrinello M (2002) Escaping free-energy minima. *Proc Natl Acad Sci USA* 99(20):12562–12566.
26. Barducci A, Bonomi M, Parrinello M (2011) Metadynamics. *WIREs Comput Mol Sci* 1:826–843.
27. Limongelli V, et al. (2013) The G-triplex DNA. *Angew Chem Int Ed Engl* 52(8): 2269–2273.
28. Limongelli V, et al. (2010) Molecular basis of cyclooxygenase enzymes (COXs) selective inhibition. *Proc Natl Acad Sci USA* 107(12):5411–5416.
29. Limongelli V, et al. (2012) Sampling protein motion and solvent effect during ligand binding. *Proc Natl Acad Sci USA* 109(5):1467–1472.
30. Grazioso G, et al. (2012) Investigating the mechanism of substrate uptake and release in the glutamate transporter homologue $Glt_{(Ph)}$ through metadynamics simulations. *J Am Chem Soc* 134(1):453–463.
31. Gervasio FL, Laio A, Parrinello M (2005) Flexible docking in solution using metadynamics. *J Am Chem Soc* 127(8):2600–2607.
32. Allen TW, Andersen OS, Roux B (2004) Energetics of ion conduction through the gramicidin channel. *Proc Natl Acad Sci USA* 101(1):117–122.
33. Deng Y, Roux B (2009) Computations of standard binding free energies with molecular dynamics simulations. *J Phys Chem B* 113(8):2234–2246.
34. Roux B, Andersen OS, Allen TW (2008) Comment on "Free energy simulations of single and double ion occupancy in gramicidin A" [J. Chem. Phys. 126, 105103 (2007)]. *J Chem Phys* 128(22):227101, , author reply 227102.
35. Barducci A, Bussi G, Parrinello M (2008) Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys Rev Lett* 100(2):020603.
36. 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(46):9663–9669.
37. Söderhjelm P, Tribello GA, Parrinello M (2012) Locating binding poses in protein–ligand systems using reconnaissance metadynamics. *Proc Natl Acad Sci USA* 109(14): 5170–5175.
38. Yamane J, et al. (2011) In-crystal affinity ranking of fragment hit compounds reveals a relationship with their inhibitory activities. *J Appl Cryst* 44:798–804.
39. Dror RO, et al. (2011) Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc Natl Acad Sci USA* 108(32):13118–13123.
40. Talhout R, Engberts JB (2001) Thermodynamic analysis of binding of p-substituted benzamidines to trypsin. *Eur J Biochem* 268(6):1554–1560.
41. Katz BA, et al. (2001) A novel serine protease inhibition motif involving a multicentered short hydrogen bonding network at the active site. *J Mol Biol* 307(5): 1451–1486.
42. Kurumbail RG, et al. (1996) Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents. *Nature* 384(6610):644–648.
43. Cheng Y, Prusoff WH (1973) Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction. *Biochem Pharmacol* 22(23):3099–3108.
44. Gierse JK, et al. (1995) Expression and selective inhibition of the constitutive and inducible forms of human cyclo-oxygenase. *Biochem J* 305(Pt 2):479–484.
45. Selinsky BS, Gupta K, Sharkey CT, Loll PJ (2001) Structural analysis of NSAID binding by prostaglandin H2 synthase: Time-dependent and time-independent inhibitors elicit identical enzyme conformations. *Biochemistry* 40(17):5172–5180.
46. Price MLP, Jorgensen WL (2000) Analysis of binding affinities for celecoxib analogues with COX-1 and COX-2 from combined docking and Monte Carlo simulations and insight into the COX-2/COX-1 selectivity. *J Am Chem Soc* 122(39):9455–9466.
47. Salvalaglio M, Vetter T, Giberti F, Mazzotti M, Parrinello M (2012) Uncovering molecular details of urea crystal growth in the presence of additives. *J Am Chem Soc* 134(41):17221–17233.
48. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781–1802.
49. Bonomi M, et al. (2009) PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput Phys Commun* 180(10):1961–1972.
50. Bonomi M, Barducci A, Parrinello M (2009) Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *J Comput Chem* 30(11):1615–1621.

Supporting Information

Limongelli et al. 10.1073/pnas.1303186110

SI Text

Description of Energy-Basin Poses in Benzamidine/Trypsin Complex. Looking at the free-energy surface (Fig. 2) computed as a function of the projection on the z axis and torsion collective variable (CV) using a reweighting algorithm, two main energy minima can be detected. The deepest one, basin A in Fig. 2, shows the trypsin in its crystal-binding conformation. At this site, the ligand is bound to the binding pocket, forming a number of strong interactions with the enzyme. In particular, the diamine group of trypsin engages a salt bridge with Asp189, and H-bond interactions with Ser190 and Gly219, while the aromatic ring of the ligand is involved in hydrophobic contacts with the Val213 side chain and the Cα atoms of Cys191 and Trp215. It is interesting to note that a water molecule is present in the binding pocket, favoring the formation of an H-bond network between the side chains of Tyr228, Asp189, and Ser190. A water molecule at similar position is present in many X-ray crystals, suggesting a structural role for this molecule in the formation of the ligand binding site. The second energy minimum, basin B, is ∼1 kcal/mol higher than basin A. At variance with the basin A pose, here the ligand is slightly rotated in the binding site. However, the diamino group is able to interact with the same residues described at basin A, although some of these interactions are mediated by water molecules. In particular, benzamidine engages directly with Ser190 through an H-bond, whereas its interaction with the Asp189 side chain is mediated by a water molecule. The aromatic ring of trypsin is instead involved in an interaction with the sulfur atoms forming the disulfide bridge between Cys191 and Cys220. At this basin the diamino group is involved in H-bond interactions with the carbonyl oxygens of Val227, Val213, and Ser214 via two water molecules. Waters at similar positions are found in the apo and ligated form of the enzyme (PDB ID codes 1s0q and 3atl, respectively), suggesting a functional role of the solvent during the ligand binding. The stability of the basin B pose has been further assessed through an extensive molecular dynamics simulation, over 100-ns long. During the whole simulation, the ligand binding mode is stable with all of the above described interactions well conserved (Fig. S2). This result, together with the low free-energy value of the basin, suggests an important role for this pose during the ligand binding and unbinding, as discussed in the main text.

Description of Energy-Basin Poses in SC-558/Cyclooxygenase 2 Complex. Looking at the free-energy surface (FES) shown in Fig. 3, two deepest-energy minima can be found, basin A and basin B. The lowest-energy pose represents SC-558 in its X-ray conformation (Fig. 3). Here, the bromophenyl ring is placed in a hydrophobic cavity surrounded by Phe381, Leu384, Tyr385, Trp387, and Phe518, whereas on the other branch, the trifluoromethyl moiety resides in a pocket surrounded by Met113, Val116, Tyr355, Leu359, and Leu531. Finally, the phenylsulphonamide moiety engages H-bond interactions with Arg513 and with the backbone of Phe518 via two water molecules. All these favorable interactions render this pose highly stable with a very low-energy value.

The second energy minimum, basin B, corresponds to the alternative binding mode of SC-558 in cyclooxygenase 2 (COX-2), first discovered and largely discussed by us in the original paper (1). At this site, the trifluoromethylpyrazole occupies the cavity formed by Leu352, Trp387, Phe518, Val523, Gly526, and Ala527, and the bromophenyl moiety is placed in the hydrophobic pocked defined by Ile345, Val349, Leu531, Leu534, and Met535. Finally, the sulphonamide group engages H-bond interactions with Arg120 and Tyr355 (Fig. 3).

SI Methods

The starting conformation for the benzamidine/trypsin and SC-558/COX-2 complexes was taken from the Protein Data Bank (PDB ID codes 2oxs and 1cx2, respectively). All simulations were carried out with the AMBER99SB-ILDN force field (2–4) for the protein, and the TIP3P water model (5) for the explicit solvent. The Amber charges were applied to protein and water atoms, and the restrained electrostatic potential charges were used for the ligands. The systems were simulated using periodic boundary conditions. Before well-tempered metadynamics simulations, the benzamidine/trypsin and SC-558/COX-2 complexes were equilibrated using 5-ns-long molecular dynamics simulations in the isothermal–isobaric ensemble NPT at 1 atm and 300 K. The PLUMED plugin (6) was used to carry out metadynamics calculations in the isothermal–isochoric ensemble NVT with the NAMD code (7).

In benzamidine/trypsin complex, the bias was added on a distance CV (Table S1). A Gaussian width of 0.05 Å was used, and a Gaussian deposition rate of 0.5 kcal·mol·ps (1 kcal = 4.18 kJ) was initially used and gradually decreased on the basis of the adaptive bias with a ΔT of 3,300 K.

In SC-558/COX-2 complex, the bias was added on the projection on the z axis of the ligand center of mass, a torsion, and a contact map CV (Tables S2 and S3) using a Gaussian width of 0.32 Å, 0.43 rad, and 0.12, respectively. The λ value for the contact map CV was set to 8.51 (1). A Gaussian deposition rate of 0.5 kcal·mol·ps was initially used and gradually decreased on the basis of the adaptive bias with a ΔT of 2,700 K.

1. Limongelli V, et al. (2010) Molecular basis of cyclooxygenase enzymes (COXs) selective inhibition. *Proc Natl Acad Sci USA* 107(12):5411–5416.
2. Cornell WD, et al. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117(19):5179–5197.
3. Hornak V, et al. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65(3):712–725.
4. Lindorff-Larsen K, et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78(8):1950–1958.
5. Jorgensen WL, Madura JD (1983) Solvation and conformation of methanol in water. *J Am Chem Soc* 105(6):1407–1413.
6. Bonomi M, et al. (2009) PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput Phys Commun* 180(10):1961–1972.
7. Phillips J-C, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781–1802.

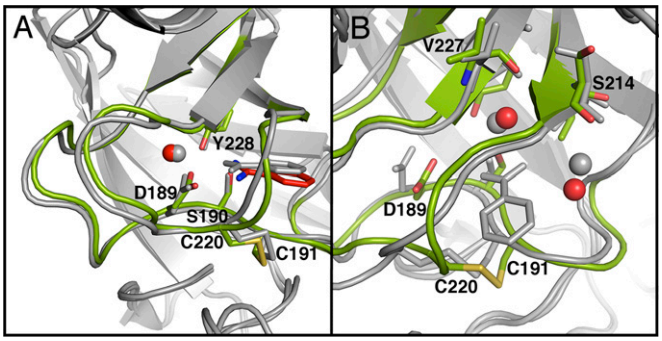


Fig. S1. (*A*) Overlay of the benzamidine/trypsin conformation corresponding to basin A found in the funnel metadynamics (FM) simulations (gray) with that of the X-ray structure of trypsin in the ligated form (PDB ID code 3atl; color). (*B*) Overlay of the benzamidine/trypsin conformation corresponding to basin B found in the FM simulations (gray) with that of the X-ray structure of trypsin in the apo form (PDB ID code 1s0q; color). The superimpositions show that water molecules at similar positions in the binding site are found in the FM simulations and in the X-ray structures. This finding suggests the important role played by waters during ligand binding, and the need to use explicit solvent simulations to take into account solvation effect and obtain an accurate estimation of ligand/protein interactions.

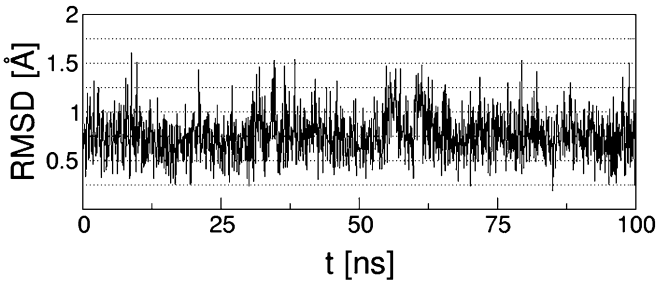


Fig. S2. Plot of the rmsd of the heavy atoms of benzamidine during over 100 ns of molecular dynamics simulation with the ligand in the binding conformation corresponding to basin B. The very low average rmsd of 0.75 Å reflects the good stability of this pose with all of the ligand/protein interactions conserved.

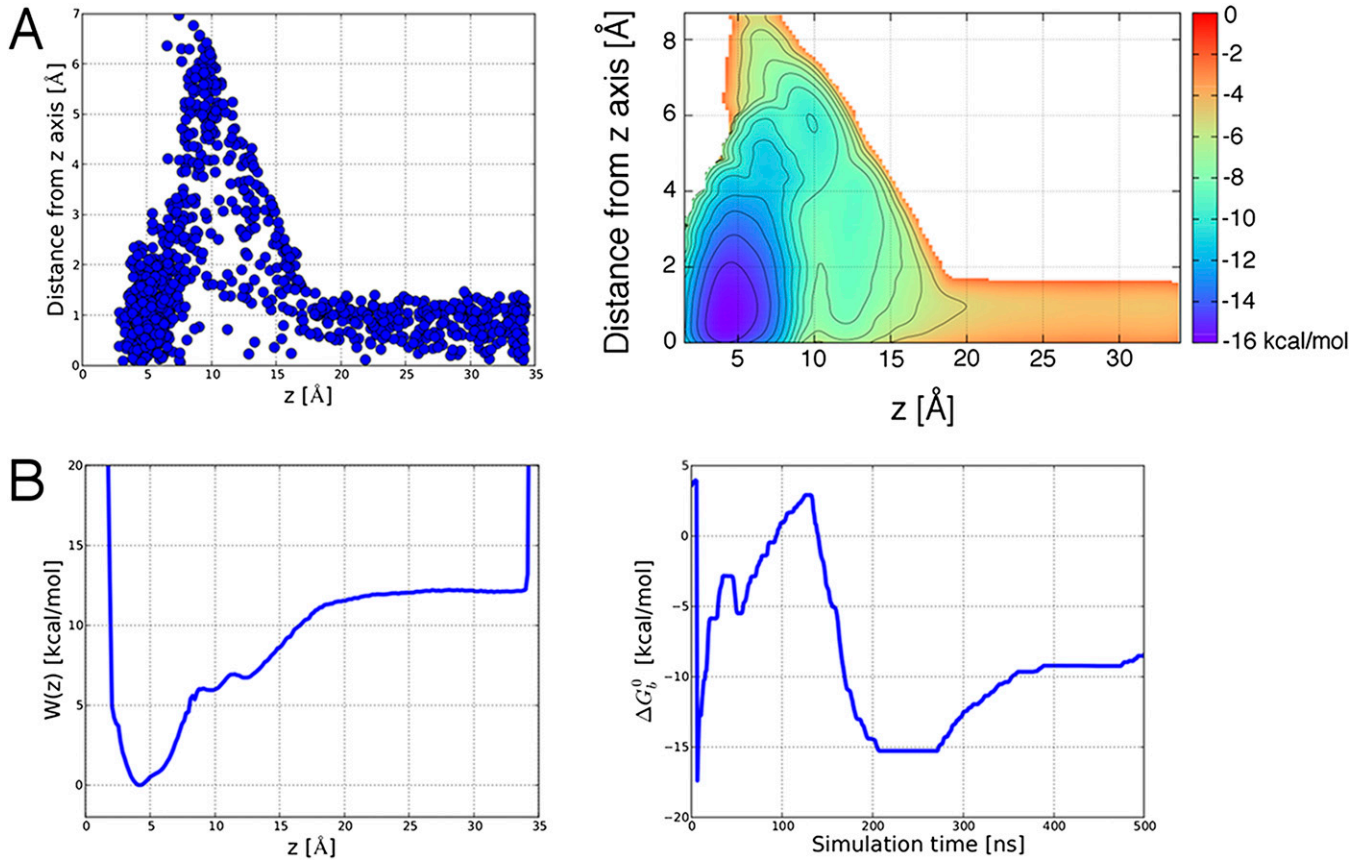


Fig. S3. (*A*) (*Left*) Scatter plot of the projection on the *z* axis vs. distance from the *z* axis of the ligand center of mass, collecting all of the configurations sampled during FM simulations in the benzamidine/trypsin system. The *z* is the axis of the funnel (Table S1). In the binding site region ($3.0 \leq z \leq 7.0$ Å), the conical restraint is not acting on the ligand. (*Right*) Representation of the FES computed as a function of the projection on the *z* axis and distance from *z*-axis CVs using a reweighting algorithm. Isosurfaces are shown every 1 kcal/mol. These CVs are not able to distinguish the two binding modes, basin A and B, which are merged in one deep-energy basin. However, the FES shows that this basin is almost at the center of the conical section of the funnel where the simulation is not affected by the funnel restraint potential. (*B*) (*Left*) Potential of mean force $W(z)$ obtained by reweighting the metadynamics simulations. The binding site region is defined at $3.0 \leq z \leq 7.0$ Å, and the fully solvated at $z_{bulk} = 32.0$ Å. (*Right*) Convergence of the absolute protein–ligand binding free-energy calculation. ΔG_b^0 is calculated using $z_{min} = 3.0$ Å, $z_{max} = 7.0$ Å, and $z_{bulk} = 32.0$ Å, at different times along the simulation to assess the convergence. Considering $R_{cyl} = 1$ Å, and applying the analytical correction (*Methods*), the estimate of ΔG_b^0 is -8.5 ± 0.7 kcal/mol, in line with previous calculations and experiments. The uncertainty is calculated as the SD from the asymptotic value of the absolute protein–ligand binding free energy obtained from the last part of the simulation.

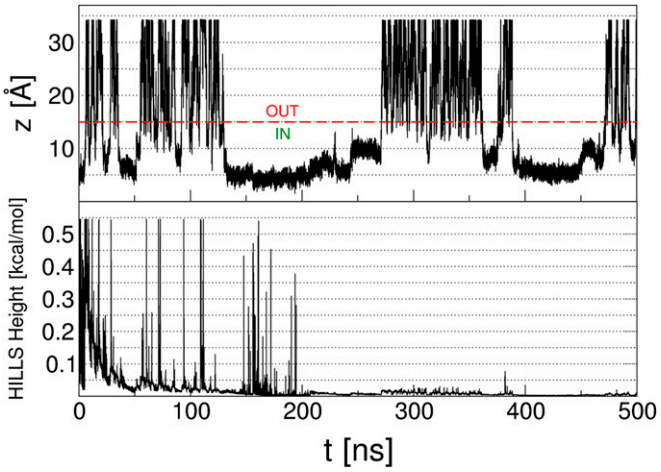


Fig. S4. (*Upper*) Plot of the projection on the *z* axis CV, where *z* is the axis of the funnel restraint potential, along the FM simulations in the benzamidine/trypsin system. (*Lower*) Plot of the Gaussian height added to the system along the FM simulations. Thanks to the funnel restraint potential, several recrossing events between bound (IN) and unbound (OUT) states are possible, even when the added Gaussians have become very small. These events lead to the convergence in the estimation of the protein–ligand binding free energy (Fig. S3).

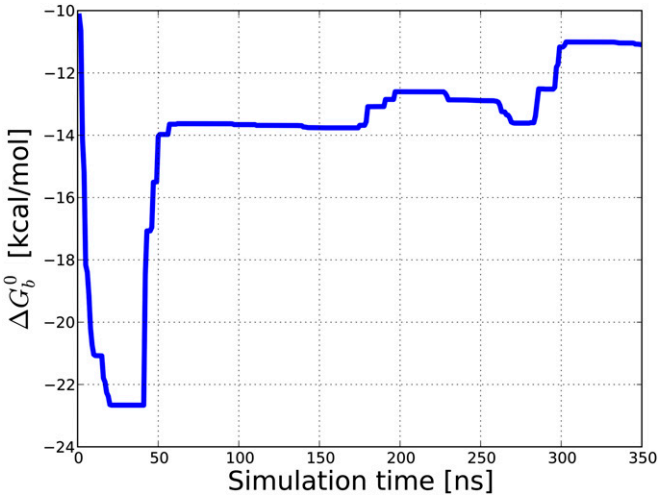


Fig. S5. Convergence of the absolute SC-558/COX-2 binding free energy. The $\Delta G_b{}^0$ is calculated every 50 ps along the simulation to assess the convergence. The bound state is defined at $8.0 \leq z \leq 27.0$ Å ($z_{min} = 8.0$ Å, $z_{max} = 27.0$ Å), and the unbound one at $z_{bulk} = 46.0$ Å. Considering $R_{cyl} = 1$ Å and applying the analytical correction (*Methods*), the estimate of $\Delta G_b{}^0$ is -11.1 ± 1.5 kcal/mol, in good agreement with the experimental value of ~ -12 kcal/mol. The uncertainty is calculated as the SD from the asymptotic value of the absolute protein–ligand binding free energy obtained from the last part of the simulation.

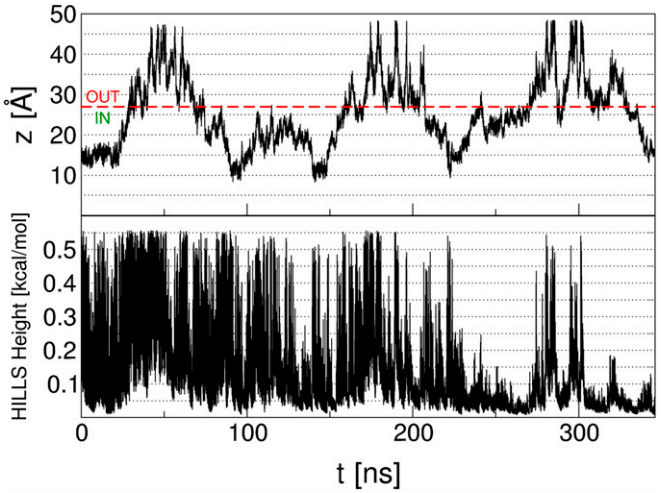


Fig. S6. (*Upper*) Plot of the projection on the *z* axis CV, where *z* is the axis of the funnel restraint potential along the FM simulations in the SC-558/COX-2 system. (*Lower*) Plot of the Gaussian height added to the system along the FM simulations. As seen in the benzamidine/trypsin case, FM favors the recrossing events between bound (IN) and unbound (OUT) states. These events lead to the convergence in the estimation of the protein–ligand binding free energy within a reasonable computational time (Fig. S5).

Table S1. List of atoms and parameters that define the different CVs used in the FM simulations in the benzamidine/trypsin system

CV type	Parameters	Protein (P), ligand (L)
Funnel*	z_{cc} = 18 Å	
	α = 0.55 rad	
	R_{cyl} = 1.0 Å	
	Atom	
Distance†	Cδ of Asp189	P
Distance†	C7 (C endowing the diamine group)	L
Torsion	Gly226Cα	P
Torsion	Gly216Cα	P
Torsion	C1	L
Torsion	C4	L

Note that the residue labels are taken from the Protein Data Bank using the trypsin structure with PDB ID code 2oxs.
*The *z* axis of the funnel is defined in the space *x,y,z* by two points, one corresponding approximately to the center of mass of Leu185C, Pro225Cα, and Pro225N, and the other to the center of mass of Cys220Cα, Lys224C, and Gly226N; the first one also defines the origin of the *z* axis.
†Distance CV is defined as the distance between the atoms.

Table S2. List of atoms and parameters that define the different CVs used in the FM simulations in the SC-558/COX-2 system

CV type	Parameters	Protein (P), ligand (L)
Funnel*	z_{cc} = 44 Å	
	α = 0.6 rad	
	R_{cyl} = 1.0 Å	
	Atom	
Torsion	Ala527Cα	P
Torsion	Val523Cα	P
Torsion	C endowing CF3	L
Torsion	C endowing S	L

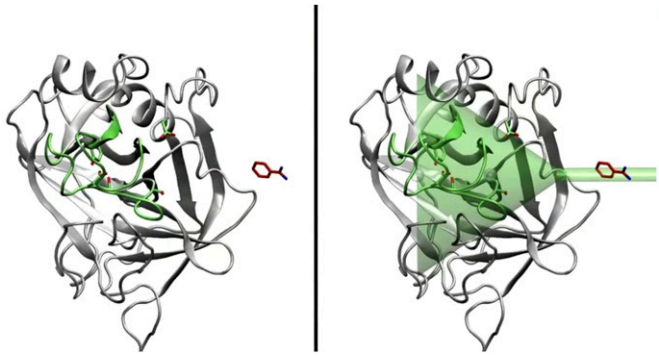
*The *z* axis of the funnel is defined in the space *x,y,z* by two points, one corresponding to the center of mass of Phe201Cα, Tyr385Cα, and Pro389Cα, and the other to the center of mass of His351Cα, Leu352Cα, and Ala527Cα; the first one also defines the origin of the *z* axis.

Table S3. List of atoms whose contacts are included in the contact map CV to sample the motion of the ligand exit door in COX-2 (1)

Contact no.	Atom *i*	Atom *j*	R_0	*p*	*q*
1	Val88Cα	Ile112Cα	12	6	10
2	Val88Cα	Val116Cα	12	6	10
3	Val88Cα	Arg120Cα	12	6	10
4	Ile92Cα	Ile112Cα	12	6	10
5	Ile92Cα	Val116Cα	12	6	10
6	Ile92Cα	Arg120Cα	12	6	10
7	Ile112Cα	Tyr115Cζ	6	6	10
8	Ser119Cα	Tyr115Cζ	10	6	10
9	Glu524Cδ	Arg120Cζ	6	4	10

R_0, *p*, and *q* are parameters of the switching function that defines the contact between the atom *i* and *j* (please see ref. 1 for details).

1. Limongelli V, et al. (2010) Molecular basis of cyclooxygenase enzymes (COXs) selective inhibition. *Proc Natl Acad Sci USA* 107(12):5411–5416.



Movie S1. Binding/unbinding events of benzamidine to trypsin under the action of FM simulations. Using a funnel restraint potential (*Right*), several recrossing events are possible, leading to a well-converged free-energy surface and an accurate estimation of the absolute protein–ligand binding free energy.

Movie S1