

# Accelerating the Calculation of Protein-Ligand Binding Free Energy and Residence Times using Dynamically Optimized Collective Variables

Z. Faidon Brotzakis,<sup>†,‡</sup> Vittorio Limongelli,<sup>\*,¶,¶</sup> and Michele Parrinello<sup>\*,†,‡,||</sup>

<sup>†</sup>*Department of Chemistry and Applied Bioscience, ETH Zürich, c/o USI Campus, Via Giuseppe Buffi 13 Lugano, CH-6900, Lugano, Ticino, Switzerland*

<sup>‡</sup>*Institute of Computational Science, Università della Svizzera Italiana (USI), Via Giuseppe Buffi 13 Lugano, CH-6900, Lugano, Ticino, Switzerland.*

<sup>¶</sup>*Università della Svizzera Italiana (USI), Faculty of Biomedical Sciences, Institute of Computational Science - Center for Computational Medicine in Cardiology, Via G. Buffi 13, CH-6900 Lugano, Switzerland*

<sup>§</sup>*Department of Pharmacy, University of Naples "Federico II", via D. Montesano 49, I-80131 Naples, Italy*

<sup>||</sup> *Instituto Italiano di Tecnologia, Via Morego 30, 16163 Genova, Italy*

E-mail: vittorio.limongelli@usi.ch; parrinello@phys.chem.ethz.ch

## Abstract

Elucidation of the ligand/protein binding interaction is of paramount relevance in pharmacology to increase the success rate of drug design. To this end a number of computational methods have been proposed, however all of them suffer from limitations since the ligand binding/unbinding transitions to the molecular target involve many slow degrees of freedom that hamper a full characterization of the binding process.

Being able to express this transition in simple and general slow degrees of freedom, would give a distinctive advantage, since it would require minimal knowledge of the system under study, while in turn it would elucidate its physics and accelerate the convergence speed of enhanced sampling methods relying on collective variables. In this study we pursue this goal by combining for the first time Variation Approach to Conformational dynamics with Funnel-Metadynamics. In so doing, we predict for the benzamidine/trypsin system the ligand binding mode, and we accurately compute the absolute protein-ligand binding free energy and unbinding rate at unprecedented low computational cost. Finally, our simulation protocol reveals the energetics and structural details of the ligand binding mechanism and shows that water and binding pocket solvation/desolvation are the dominant slow degrees of freedom.

## 1 Introduction

Elucidating the way a drug interacts with its molecular target is of paramount relevance for drug design. The development of reliable models of ligand-protein interaction has led to the discovery of new compounds.<sup>1,2</sup> In this context, a plethora of methods aimed at identifying the ligand binding mode and estimating the protein-ligand binding free energy has been reported, see for instance recent reviews at Refs.<sup>1,2</sup> From a technical point of view, we can group the available methodologies into three categories: i) molecular docking;<sup>3-5</sup> ii) end-point methods such as linear interaction energy,<sup>6</sup> Poisson-Boltzmann surface area,<sup>7</sup> Generalized Born Surface Area;<sup>8</sup> and iii) free energy pathway methods such as Free Energy Perturbation,<sup>9</sup> Thermodynamic Integration,<sup>10</sup> Transition Path Sampling,<sup>11</sup> Umbrella Sampling,<sup>12</sup> Steered-MD,<sup>13</sup> and Funnel-Metadynamics (FM).<sup>14</sup>

Although docking methods are very fast in screening and ranking ligand bound poses, they suffer from inefficient sampling and scoring function inaccuracy. One of the major limitations of docking is the lack of solvation and protein flexibility dynamics. In order to obtain accurate binding free-energy, one might use either end-point methods or free energy-

pathway methods. Despite the robustness of these methods, most of them require a priori knowledge of the ligand binding mode and assume that the ligand studied have similar properties (e.g., Free Energy Perturbation), thus hampering general applicability.

At variance with the other methods, FM does not need any preliminary information on the ligand binding mode. Once the approximate location of the binding pocket is defined, the user is able to capture the whole ligand binding mechanism from the fully solvated state to its final binding mode. The calculation of a well-characterized binding free-energy surface (BFES) allows identifying the lowest energy state (i.e., ligand binding mode) and other energetically relevant minima, which might represent important alternative binding poses. Furthermore, the absolute protein-ligand binding free energy is accurately computed from the free-energy difference between the ligand bound and unbound free energy minima. FM has been successfully used in a wide number of studies to investigate ligand binding to both proteins and DNA.<sup>14-16</sup>

Like in standard metadynamics, in FM a history dependent bias is added to the system Hamiltonian as a function of a set of Collective Variables (CVs), thereby enhancing the fluctuations along those CVs that mediate the rare transitions. The choice of CVs is critical. Bad CVs lead to hysteresis and poorly converged calculations.<sup>17,18</sup> Unfortunately, the simulation time required to converge the free energy increases exponentially with the number of CVs, thus limiting the choice of the system's degrees of freedom that can be included. This scenario poses first the need to discover the slow degrees of freedom to bias, and then develop CVs able to describe these modes. Here, we have pursued this goal by combining two enhanced sampling techniques: the Variation Approach to Conformational dynamics Metadynamics (VAC-MetaD)<sup>19</sup> with the Funnel-Metadynamics. The first, allows capturing the details of the intermolecular interactions during binding while the second ensures an exhaustive exploration of the bound and unbound state and a faster convergence.

The benzamidine/trypsin system has been widely studied with different computational techniques.<sup>14,20-23</sup> Its solvent exposed binding site makes it an ideal system for testing new

ligand-protein binding methods. Therefore we choose benzamidine-trypsin complex as case study<sup>14,20,23</sup> using a basis set that includes general descriptors of the protein-ligand binding process. The combination of the two methods allows describing the binding mechanism of benzamidine to trypsin. We demonstrate the ability of FM using VAC-MetaD optimized CVs to reproduce the known ligand binding mode and to provide an accurate estimation of the absolute protein-ligand binding free energy, reducing the convergence time. Furthermore, this simulation provides a high resolution free-energy surface that highlights the key role played by the water molecules in the different energy minima visited by the ligand during its binding path. We also show that using these optimized CVs with infrequent deposition of the bias as reported in,<sup>23</sup> leads to an accurate estimate of the ligand unbinding rate  $k_{off}$ . The flexibility of the method permits to include in one CV a number of molecular descriptors of the protein-ligand binding process such as the number of protein-ligand contacts, ligand solvation, pocket solvation and ligand orientation. The analysis of the results indicates which degrees of freedom play a major role during ligand binding. In our case, we have found that the slowest degrees of freedom are the desolvation of the binding site and of the ligand rather than the formation of direct protein-ligand contacts.

## 2 Methods

### 2.1 Metadynamics

Metadynamics (MetaD) is an established enhanced sampling method used to study rare events.<sup>17,24,25</sup> MetaD accelerates the sampling by adding a history-dependent bias in the form of Gaussian kernels on selected system's degrees of freedom, also known as collective variables (CVs).

In well tempered Metadynamics<sup>26</sup> this scope is achieved by periodically adding a bias that is updated according to the iterative procedure

$$V_n(\mathbf{s}) = V_{n-1}(\mathbf{s}) + G(\mathbf{s}, \mathbf{s}_n) \exp\left(-\frac{1}{\gamma-1} V_{n-1}(\mathbf{s}_n)\right) \quad (1)$$

where  $V_n(\mathbf{s})$  is the total bias deposited at iteration  $n$  and is obtained by adding at the previous bias  $V_{n-1}$  a contribution that results from the product between a Gaussian kernel  $G(\mathbf{s}, \mathbf{s}_n)$  and a multiplicative factor  $\exp\left(-\frac{1}{\gamma-1} V_{n-1}(\mathbf{s}_n)\right)$  that makes the height of the added Gaussian diminished with time. The bias factor  $\gamma > 1$  determines the rate with which the added bias decreases and regulates the amplitude of the  $\mathbf{s}$  fluctuations. Thanks to the MetaD potential, the system explores the FES escaping the free-energy minima and crossing large energy barriers in a reasonable computational time.

An additional advantage of well tempered MetaD is that after a transient period the simulation reaches a quasi-stationary limit in which one can estimate the expectation value of any observable  $\langle O(\mathbf{R}) \rangle$ ,<sup>27</sup> as an average over the simulation time according to:

$$\langle O(\mathbf{R}) \rangle = \lim_{T \rightarrow \infty} \frac{\int_0^T O(\mathbf{R}_t) e^{\beta(V(\mathbf{s}(\mathbf{R}_t)) - c(t))} dt}{\int_0^T e^{\beta(V(\mathbf{s}(\mathbf{R}_t)) - c(t))} dt} \quad (2)$$

Here,  $\mathbf{R}_t$  are atomic positions at time  $t$ ,  $\beta$  is the inverse of the temperature multiplied by the Boltzmann constant  $k_B$  and the time dependent offset  $c(t)$  is defined as:

$$c(t) = -\frac{1}{\beta} \log \frac{\int e^{-\beta(F(\mathbf{s}) + V(\mathbf{s}, t))} d\mathbf{s}}{\int e^{-\beta(F(\mathbf{s}))} d\mathbf{s}} \quad (3)$$

where  $F(\mathbf{s})$  is the free energy as a function of the CV. Following this procedure, one can reweight to Boltzmann averages of any variable even if they are not included in the biased CV.

## 2.2 Funnel-Metadynamics

The binding process between a ligand and its molecular target is typically represented as chemical reaction with the equilibrium constant  $K_b$  (Fig. 1A). The latter is the ratio between

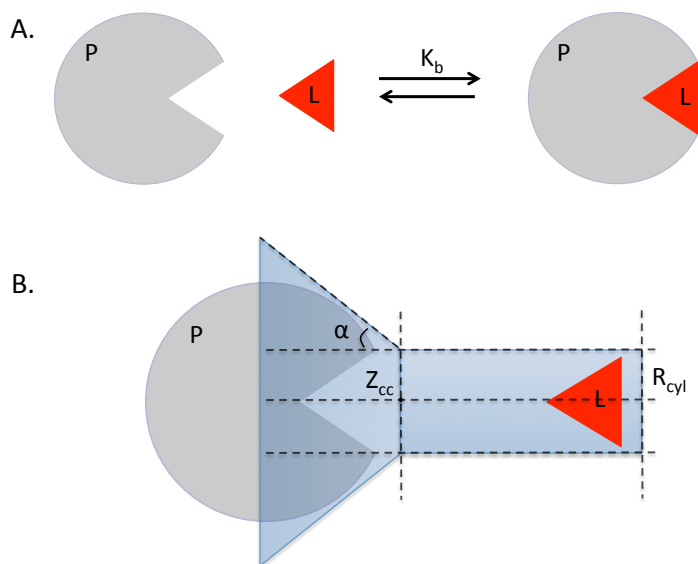


Figure 1: A) A cartoon representation of a ligand (L) binding to a protein (P) . B) The funnel restraining potential illustrated in blue, along with its parameters. More specifically  $Z$  is the exit binding path axis,  $Z_{cc}$  is the distance along the axis  $Z$ , where the cone restraining potential turns into a cylindrical shape, and  $R_{cyl}$  is the radius of the cylinder, and  $\alpha$  is the angle defining the cones amplitude.

the concentration of the system in the bound state and that in the unbound state. In thermodynamics,  $K_b$  is related to the absolute protein-ligand binding free energy  $G_b^0$ , which is the free-energy difference between the unbound and the bound state, through the formula:

$$\Delta G_b^0 = -\frac{1}{\beta} \ln (C^0 K_b) \quad (4)$$

where  $C^0=1/1660 \text{ \AA}^{-3}$  is the standard concentration. However, calculating the free energy along a binding transition is difficult. In particular, in the case of protein/ligand binding an additional complication arises from the fact that once the ligand leaves the protein, it has to explore a large conformational space. A remedy to this problem has been suggested in Refs.<sup>14,28,29</sup> with the imposition of a restraint potential in the binding site. In this study, we will use the Funnel-Metadynamics method, where the restraining potential has in the shape of a funnel as illustrated in Fig. 1B. The design of the funnel is such that the restraint does not act on the ligand when in the binding site while confines the ligand to a cylindrical

region once outside. One can also look at this procedure as having imposed an entropic restraint. The net effect of the cylindrical restraint can be accounted for, leading to the following formula:

$$K_b = \pi R_{cyl}^2 \int_{site} dz \exp(-\beta(W(z) - W(ref))) \quad (5)$$

where  $W(z)$  and its value in the unbound state  $W(ref)$  can be derived from the Potential of Mean Force (PMF) along the funnel axis  $Z$ , while  $R_{cyl}$  is the radius of the cylindrical part of the funnel potential and  $\pi R_{cyl}^2$  accounts for the restraining potential in the unbound state (see Ref.<sup>14</sup> for details). Finally, one can calculate the absolute protein-ligand binding free energy  $\Delta G_b^0$  combining Eq. 4 and Eq. 5. Due to the funnel restraint, several forth-and-back events between the bound and the unbound state are observed during the simulation leading to the free-energy convergence of the calculation in a shorter computational time.<sup>14,15,30-33</sup>

### 2.3 Variational Approach to Conformational dynamics Metadynamics

As already discussed, in MetaD calculations the number of CVs should be kept as small as possible, since the cost of reconstructing the free energy grows exponentially with the number of CVs used. This limitation is particularly severe in protein systems where a slow process often involves several degrees of freedom. Recently Mc Carty and Parrinello developed a Variational Approach to Conformational dynamics in Metadynamics (VAC-MetaD)<sup>19</sup> that allows combining several degrees of freedom into a number of CVs that are a combination of many descriptors. So far this has led to a better understanding of chemical reactions and folding of small peptides.<sup>34,35</sup>

Here we briefly review the steps needed to apply this procedure. First one identifies a set of  $N$  descriptors  $d_k(\mathbf{R})$  that one thinks are important for the phenomenon under investigation. Then one searches for the best CV  $s_i(\mathbf{R})$  that can be expressed as a weighted linear

combination of  $d_k(\mathbf{R})$ , with corresponding coefficients  $b_{ik}$ :

$$\mathbf{s}_i(\mathbf{R}) = \sum_k^N b_{ik} d_k(\mathbf{R}) \quad (6)$$

This is obtained by a variant of Time-lagged Independent Conformational Analysis (TICA) developed in Noe's group.<sup>36,37</sup> In this setting, one computes the time-lagged covariance matrix  $C(\tau)$  defined as:

$$C_{mn} = \langle r_m(0)r_n(\tau) \rangle \quad (7)$$

where  $r_k(\tau) = d_k(\tau) - \langle d_k(\tau) \rangle$ . We then determine the  $b_{ik}$  coefficients by solving the generalized eigenvalue equation:

$$\overline{\mathbf{C}}(\tau)\mathbf{b}_i = \overline{\mathbf{C}}(0)\lambda_i\mathbf{b}_i \quad (8)$$

where  $\lambda_i$  is the  $i$ th eigenvalues,  $\overline{\mathbf{C}}(\tau)$  is the time lagged covariance matrix,  $\overline{\mathbf{C}}(0)$  is the covariance matrix at time zero. In the variant of TICA proposed by Noe's group,<sup>36,37</sup> they assume a long trajectory which spontaneously transits between metastable states. Mc Carty and Parrinello<sup>19</sup> noticed that a similar procedure can be applied from a MetaD trajectory, provided that the MetaD timescale is properly rescaled, as follows:

$$d\tau = e^{\beta(V(\mathbf{s}(\mathbf{R}_t)) - c(t))} dt \quad (9)$$

For the purpose of studying a ligand unbinding transition, the eigenvectors of the eigenvalue equation obtained from a MetaD trajectory, correspond to the slowest decaying modes, and can be used as CVs in the production MetaD simulation.<sup>19,38</sup>



## 2.4 Ligand unbinding rate $k_{off}$ from Infrequent Metadynamics

Inspired by Voter<sup>39</sup> and Grubmüller,<sup>40</sup> in a recent study Tiwary and Parrinello<sup>41</sup> introduced the Infrequent Metadynamics (InMetaD) framework illustrating how MetaD could be used to calculate transition rates of activated processes. The assumptions of this method are that i) there is no bias deposited in the transition state ensemble (TSE) region and ii) that the biased CVs are able to distinguish between relevant metastable states. This goal can be achieved by using an infrequent bias deposition which reduces the probability to add bias to the short transition time where the system is in the TSE region.<sup>41</sup> Metadynamics accelerates the dynamics by a factor  $\alpha$ , given by a running average obtained through MetaD as:

$$\alpha = \langle e^{\beta V(\mathbf{s}, t_M)} \rangle_V, \quad (10)$$

where  $t_M$  is the Metadynamics escape time,  $\mathbf{s}$  is the biased collective variable at time  $t_M$ ,  $\beta$  is the inverse of  $k_B T$  and  $V(\mathbf{s}, t_M)$  is the bias deposited at time  $t_M$ . The real time  $t$  can be related to  $t_M$  via Eq. 11.

$$t = t_M \alpha \quad (11)$$

A rigorous quantitative way to assess the method's assumption is to perform a Kolmogorov-Smirnov statistical test.<sup>42,43</sup> This effectively measures if the  $M$  observed escape times follow a Poisson distribution, as expected in a rare event scenario.

## 2.5 Protocol

In this section we present a schematic description of our protocol. The protein and water are modeled using the Amber-14SB and TIP3P classical forcefields, while the ligand using the Amber GAFF library.<sup>44,45</sup> At variance with our previous studies,<sup>14,23</sup> here we use a new benzamidine force-field parametrization based on quantum mechanics calculations. These calculations show that in the ground state of the isolated benzamidine the amidine group has a non planar torsion angle with respect to the benzene ring. Similar results have been

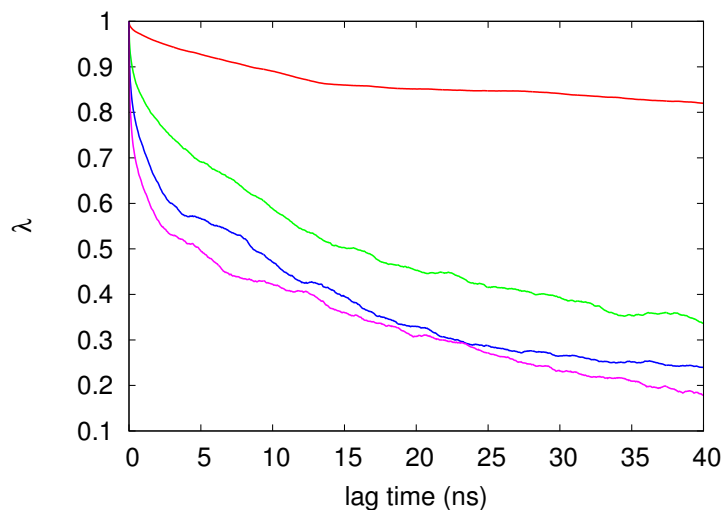


Figure 2: a) Eigenvalues decay as a function of lag time and b) coefficients of the basis set for the two eigenvectors associated with the two first slowest decaying eigenvalues (red and green respectively).

found in a previous study by Pophristic et al.<sup>46</sup> and through X-ray diffraction by Liebschner et al.<sup>47</sup> All of our simulations start from a solvated and equilibrated configuration of the benzamidine-trypsin X-ray crystal structure (PDB ID: 2oxs). Initially, we define a simple CV, namely the distance between the ligand C7 carbon and the D189 C $\gamma$  of the pocket. Then, we run a 500 ns FM simulation which enhances the fluctuations of the binding transition. Using this FM trajectory we perform the VAC-MetaD optimization procedure implemented in the PLUMED software<sup>48</sup> to generate the CVs biased in a 280 ns VAC-FM production run. Both simulations are performed at room temperature and atmospheric pressure. For more information about the system preparation we refer the interested reader to the SI. For the sake of clarity, we describe below the VAC-MetaD procedure in the following steps.

Made wise by the many studies showing that direct ligand/protein contacts, as well as the solvation and the orientation of the ligand are involved in the stability of the intermediate states,<sup>14,18,23,49</sup> we define a basis set of descriptors for ligand/protein binding ( $N_{pr/lig}, W_{pock}, \alpha\beta(\theta), W_{lig}$ ). In particular,  $N_{pr/lig}$  corresponds to the protein-ligand contacts,  $W_{pock}$ , corresponds to the number of water molecules in the pocket,  $\alpha\beta(\theta)$  is a transformation of the torsion angle describing the rotation of the ligand relative to the protein binding site,

and  $W_{lig}$  is the number of contacts between water molecules and the ligand. For more details see Table S1. After defining the basis set we perform a VAC-MetaD optimization on the first 120 ns of the preliminary FM simulation to identify the eigenvectors and eigenvalues associated with the slowest modes of the system.

We select the first two eigenvectors associated with the slowest decaying eigenvalues (see Fig. 2a,b), and use them as CVs. Since reaching convergence using one CV proves difficult (see Fig. S1) it does seem natural then to include the second eigenvector as CV, given the fact that its associated eigenvalue time decay, while faster than of the first one's is slower than the others. Furthermore, the eigenvalues of the first two eigenvectors decay more slowly than of the other (see Fig. 2a). The two optimal CVs obtained from the VAC-MetaD analysis described above are,  $\mathbf{s1} = -0.38 N_{pr/lig} + 0.78 W_{pock} - 0.19 \alpha \beta(\theta) + 0.45 W_{lig}$  and  $\mathbf{s2} = 0.01 N_{pr/lig} - 0.85 W_{pock} - 0.30 \alpha \beta(\theta) + 0.44 W_{lig}$ . The most striking feature of this two CVs is the dominant role played by water both in pocket hydration/dehydration and in ligand solvation/desolvation.

### 3 Results

In the following paragraphs we address the binding mechanism of benzamidine/trypsin complex, the convergence speed obtained using FM with VAC-MetaD optimized CV and finally the unbinding rate.

#### 3.1 Free energy

First, we perform a production run of 280 ns FM simulation, biasing two CVs  $\mathbf{s1}$  and  $\mathbf{s2}$  obtained earlier. Using these two CVs we computed the corresponding free energy surface (see Fig. 3a). Three stable bound states are found, A, P and B. The first corresponds to the crystal structure, the second to a presolvated state, the third to a rotated presolvated still bound state. A detailed description of the stable bound states follows.

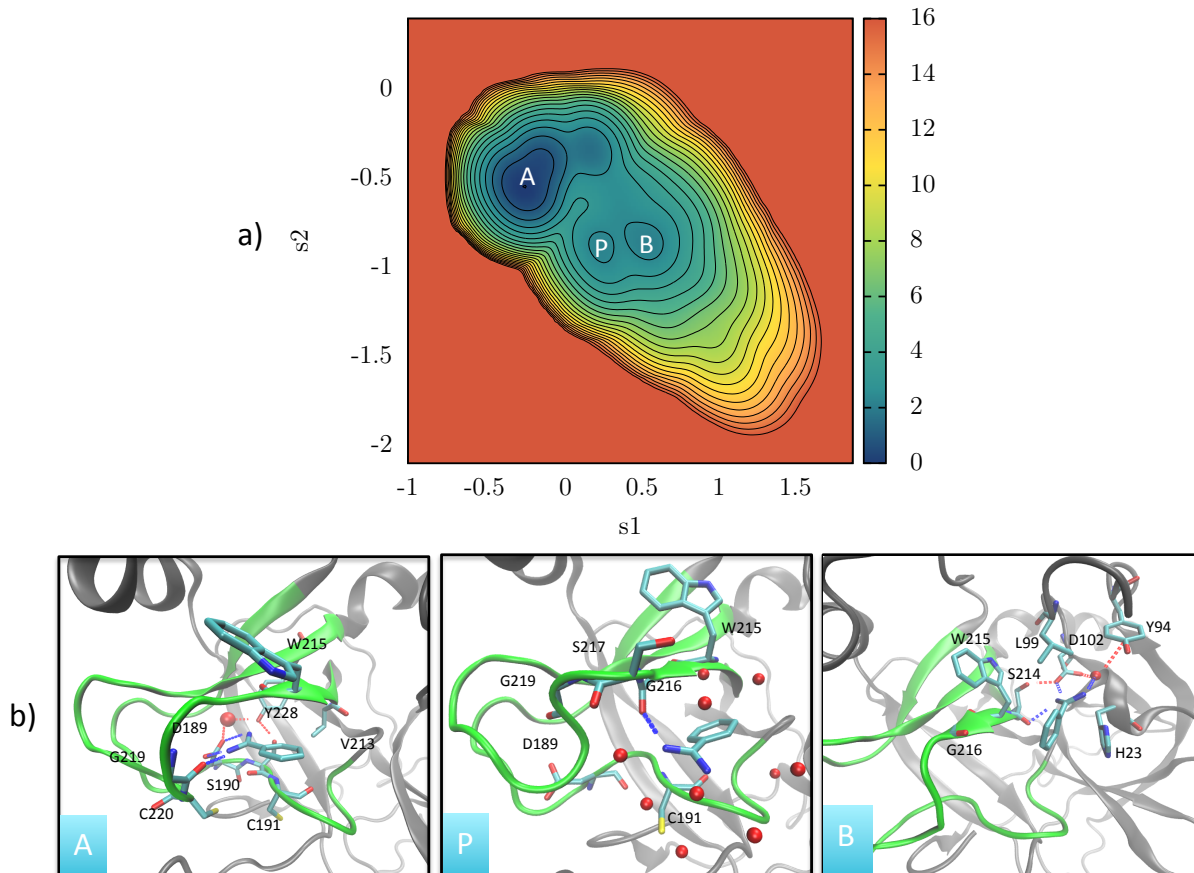


Figure 3: a) Free energy surface as a function of the two biased eigenvectors  $s_1$  and  $s_2$ . where in white the stable states and b) a detailed description of the bound poses.

**Minimum A.** Minimum A corresponds to the crystal structure. In this minimum the ligand's diamino group forms direct hydrogen bonds with the hydrogen bond accepting groups of the negatively charged D189, polar S190 and G219. Moreover, this minimum is stabilized by a hydrogen bond network formed by a structural water, bridging the interactions of Y228, D189 and S190. Finally, the ligand's aromatic ring, forms hydrophobic contacts with V213, C191 and W215. These interactions are in agreement with previous studies.<sup>14,20,23,50,51</sup>

**Minimum P.** This minimum corresponds to a presolvated state. In P, the ligand is rotated compared to state A, with its diamino group exposed to solution. The phenyl ring of the ligand is sandwiched between the  $C\alpha$  of W215 and C191. It forms direct hydrogen bond with G216 while at the vicinity of the triad G216, S217, G219. Minimum P has been also discussed in Ref.,<sup>23</sup> where is also called P, and it is the same as state P2 of Ref.<sup>50</sup>

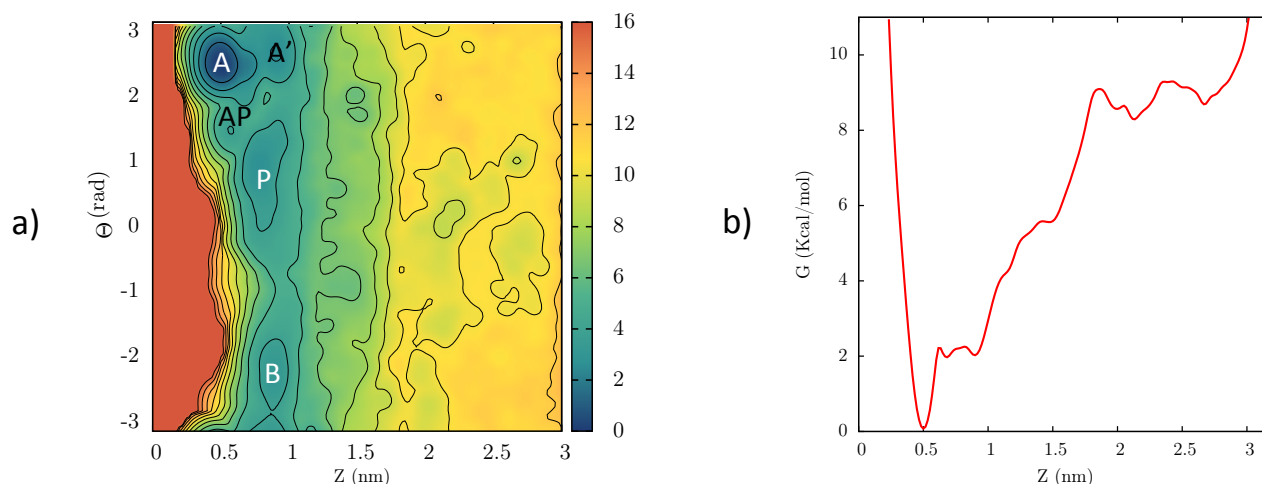


Figure 4: BFES as a function of the position of a) the ligand's center of mass along the  $Z$  axis of the funnel and the torsion  $\Theta$ , where in white the stable bound states, in black the unstable binding poses. b) The  $Z$  axis obtained from the VAC-MetaD optimized CVs FM simulation.

**Minimum B.** This minimum corresponds to a rotated presolvated bound state. In B, the ligand is rotated compared to minimum A and sits just outside the pocket. The ligand's diamino group forms direct H-bond with the negatively charged D102, and S214. The ligand is sandwiched between W215 and H23 forming  $\pi$ -stacking interactions.<sup>52</sup> Finally, a water forms a hydrogen bond network between the diamino group of the ligand, Y94 and D102. This minimum, is very similar to state P12 and/or TS2 in ref.<sup>50</sup> The combination of these interactions is allowed only by a rotated torsion, and can be captured by the new parametrization of the ligand.

**Binding Free Energy Surface (BFES).** The representation of the BFES shown in Fig. 3 does not clearly describe the position of these states relatively to the unbound state region. Thus we represent the free energy landscape as a function of the same two variables used in our previous study.<sup>14</sup> One,  $Z$ , gives the ligand's center of mass position projected along the funnel axis. The other,  $\Theta$ , is the orientation of the ligand relative to the binding site (see Table S1). In this BFES representation ( Fig. 4a), A, B and P can also be clearly identified. Here, it is possible to distinguish two additional local minima A and AP in which most of the ligand/protein interactions are mediated by water molecules.

These minima represent transient and preliminary binding poses that easily evolve to the energetically most stable states A and P within 20 ns of unbiased MD calculations. The interested reader can find more information about these minima in the SI.

**The role of water.** It is known that explicit protein/water/ligand H-bond bridging interactions can increase the stability of the bound pose by several kcal/mol.<sup>18</sup> In the benzamidine system it is known<sup>14</sup> and verified here that state A is stabilized by a hydrogen bond network formed by a structural water bridging the interactions of Y228, D189 and S190. Given such a positively charged pocket,<sup>53</sup> it makes sense that upon breakage of some protein-ligand interactions of minimum A, they will be replaced by water mediated H-bonds between the ligand and the protein, thus forming nearby minima A' and AP (see Fig. S3). Here, the stability of state P reflects this balance of solvation and direct protein-ligand forces. Importantly, at all stages of binding, water is involved in the process, be it from a ligand desolvation point of view, or from a structural water inside the pocket point of view.

### 3.2 Fast prediction of the binding free energy

From the projection of the BFES on the  $Z$  axis of the funnel (see Fig. 4b) we can compute the absolute protein-ligand binding free energy  $\Delta G_b^0$  ( $-6.7 \pm 0.4$  Kcal/mol), whose value after the correction in Eq. 5 gives the value in embarrassing agreement with the experimentally reported value of 6.4 to 7.3 kcal/mol.<sup>54,55</sup> It must be also underlined that the VAC-MetaD optimized CVs simulation converges smoothly to this value, while the FM with suboptimal CVs after 500 ns still exhibits relative large fluctuation (see Fig. S2). Regarding the technical aspects of this calculation we refer the interested reader to the SI.

Finally, a remarkable finding is that the the main features of the BFES as the local minima of the bound state, are already captured from the first 60 ns of the VAC-MetaD optimized CVs FM simulation (see Fig. S8d), posing a great improvement compared to the FM with suboptimal CVs where one would have to simulate at least 240 ns (see Fig. S8b) to obtain a  $\Delta G_b^0$  value in the right ballpark.

### 3.3 Unbinding rate

Using 20 simulations initiated in state A, we perform Metadynamics with infrequent deposition of the bias (see SI) and stop the simulation once the trajectories reach the unbound state ( $Z > 2$  nm). The associated  $k_{off}$  found is  $4176 \pm 324$  s<sup>-1</sup>, which given the uncertainties intrinsic to these type of calculations, can be considered as a good agreement with the experiment  $k_{off}^{exp}$  of  $600 \pm 300$ .<sup>56</sup> Note that the p-value of the KS test performed was 0.77, indicating the quality of the optimized CVs (see Fig. S9). The different result compared to Ref.<sup>23</sup> can be explained by the new ligands torsion potential used in this study, which allows a much easier rotation of the ligands diamino group. In so doing, the benzamidine can disentangle itself from the interactions formed at the lowest minimum A and reach the solvated state more rapidly.

## 4 Discussion

In this study we combined for the first time Variation Approach to Conformational dynamics with Funnel-Metadynamics, in order to accurately predict the binding energy, binding mechanism, and unbinding rate of the benzamidine-trypsin complex. In this way we find ligand and pocket solvation/desolvation to be the slow degrees of freedom by looking at the slowly decaying eigenvectors of the VAC-MetaD, which when biased significantly accelerate the convergence of the bound state and the overall BFES, compared to FM with suboptimal CVs. The estimated absolute binding free energy of FM using optimized CVs is in excellent agreement with experiments. In addition it gives a better resolved BFES -including more resolved water rich local minima A' and AP- compared to the FM simulation of suboptimal CVs. Moreover, when VAC-MetaD optimized CVs are used with the Infrequent Metadynamics scheme, it predicts the unbinding rate in good agreement with the experiment, thus providing a relatively cheap, rigorous and accurate calculation with respect to already existing methods.<sup>56</sup>

## 5 Acknowledgements

Z. F. B. and M. P. acknowledge VARMET European Union Grant ERC–2014–ADG–670227 for funding. V. L. acknowledges the support from the Swiss National Science Foundation (Project N. 200021\_163281) and the COST action CA15135 (Multi-target paradigm for innovative ligand identification in the drug discovery process MuTaLig). All the authors acknowledge the Swiss National Supercomputing Centre (CSS) for providing the computational resources.

## References

- (1) Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* **2004**, *303*, 1813–1819.
- (2) Gilson, M. K.; Zhou, H.-x. Calculation of Protein-Ligand Binding. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.
- (3) Goodsell, D. S.; Olson, A. J. Automated Docking of Substrates to Proteins by Simulated Annealing. *Proteins Struct. Funct. Bioinforma.* **1990**, *8*, 195–202.
- (4) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated Docking with Grid-Based Energy Evaluation. *J. Comput. Chem.* **1992**, *13*, 505–524.
- (5) Abagyan, R.; Totrov, M.; Kuznetsov, D. ICM A New Method for Protein Modeling and Design: Applications to Docking and Structure Prediction from the Distorted Native Conformation. *J. Comput. Chem.* **1994**, *15*, 488–506.
- (6) Åqvist, J.; Medina, C.; Samuelsson, J.-E. A new Method for Predicting Binding Affinity in Computer-Aided Drug Design. *Protein Eng.* **1994**, *7*, 385–391.
- (7) Gouda, H.; Kuntz, I. D.; Case, D. A.; Kollman, P. A. Free Energy Calculations for



- Theophylline Binding to an RNA Aptamer: Comparison of MM-PBSA and Thermodynamic Integration Methods. *Biopolymers* **2003**, *68*, 16–34.
- (8) Im, W.; Lee, M. S.; Brooks, C. L. Generalized Born Model with a Simple Smoothing Function. *J. Comput. Chem.* **2003**, *24*, 1691–1702.
- (9) Zwanzig, R. High-Temperature Equation of State by a Perturbation Method. I. Non-polar Gases. *J. Chem. Phys.* **1954**, *1420*.
- (10) Straatsma, T. P.; McCammon, J. A. Multiconfiguration Thermodynamic Integration. *J. Chem. Phys.* **1991**, *95*, 1175–1188.
- (11) Brotzakis, Z. F.; Bolhuis, P. G. A One-Way Shooting Algorithm for Transition Path Sampling of Asymmetric Barriers. *J. Chem. Phys.* **2016**, *145*, 164112.
- (12) Izrailev, S.; Stepaniants, S.; Isralewitz, B.; Kosztin, D.; Lu, H.; Molnar, F.; Wriggers, W.; Schulten, K. Steered Molecular Dynamics. *Comput. Mol. Dyn. Challenges, Methods, Ideas*. Berlin, Heidelberg, 1999; pp 39–65.
- (13) Torrie, G. M Valleau, J. Nonphysical Sampling Distribution in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187.
- (14) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel Metadynamics as Accurate Binding Free-Energy Method. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 6358–6363.
- (15) Moraca, F.; Amato, J.; Ortuso, F.; Artese, A.; Pagano, B.; Novellino, E.; Alcaro, S.; Parrinello, M.; Limongelli, V. Ligand Binding to Telomeric G-quadruplex DNA Investigated by Funnel-Metadynamics Simulations. *Proc. Natl. Acad. Sci.* **2017**, *114*, E2136–E2145.
- (16) Cavalli, A.; Spitaleri, A.; Saladino, G.; Gervasio, F. L. Investigating Drug Target Association and Dissociation Mechanisms Using Metadynamics-Based Algorithms. *Acc. Chem. Res.* **2015**, *48*, 277.

- (17) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 826–843.
- (18) Limongelli, V.; Marinelli, L.; Cosconati, S.; La Motta, C.; Sartini, S.; Mugnaini, L.; Da Settimo, F.; Novellino, E.; Parrinello, M. Sampling Protein Motion and Solvent Effect during Ligand Binding. *Proc. Natl. Acad. Sci.* **2012**, *109*, 1467–1472.
- (19) Mccarty, J.; Parrinello, M.; Mccarty, J.; Parrinello, M. A Variational Conformational Dynamics Approach to the Selection of Collective Variables in Metadynamics. *J. Chem. Phys.* **2017**, *147*, 204109–1.
- (20) Buch, I.; Giorgino, T.; Fabritiis, G. D. Complete Reconstruction of an EnzymeInhibitor Binding Process by Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10184–10189.
- (21) Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible Docking in Solution Using Metadynamics. *J. Am. Chem. Soc.* **2005**, *127*, 2600–2607.
- (22) Doudou, S.; Burton, N. A.; Henchman, R. H. Standard Free Energy of Binding from a One-Dimensional Potential of Mean Force. *J. Chem. Theory Comput.* **2009**, *5*, 909–918.
- (23) Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M. Kinetics of ProteinLigand Unbinding: Predicting Pathways, Rates, and Rate-Limiting Steps. *Proc. Natl. Acad. Sci.* **2015**, *112*, 201424461.
- (24) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–6.
- (25) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics : A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 1–4.
- (26) Bonomi, M.; Parrinello, M. Enhanced Sampling in the Well-Tempered Ensemble. *Phys. Rev. Lett.* **2010**, *104*, 1–4.

- (27) Tiwary, P.; Dama, J. F.; Parrinello, M. A Perturbative Solution to Metadynamics Ordinary Differential Equation. *J. Chem. Phys.* **2015**, *143*, 1–4.
- (28) Allen, T. W.; Andersen, O. S.; Roux, B. Energetics of Ion Conduction through the Gramicidin Channel. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 117–122.
- (29) Deng, Y.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113*, 2234–2246.
- (30) Glc, B. C. S. F.; Bruno, A.; Scrima, M.; Novellino, E.; Errico, G. D.; Ursi, A. M. D.; Limongelli, V. The Glycan Role in the Glycopeptide Immunogenicity Revealed by Atomistic Simulations and Spectroscopic Experiments on the Multiple Sclerosis. *Sci. Rep.* **2015**, *114*, 1–10.
- (31) Troussicot, L.; Guillièrè, F.; Limongelli, V.; Walker, O.; Lancelin, J. M. Funnel-Metadynamics and Solution NMR to Estimate Protein-Ligand Affinities. *J. Am. Chem. Soc.* **2015**, *137*, 1273–1281.
- (32) Comitani, F.; Limongelli, V.; Molteni, C. The Free Energy Landscape of GABA Binding to a Pentameric Ligand-Gated ion Channel and its Disruption by Mutations. *J. Chem. Theory Comput.* **2016**, *12*, 3398–3406.
- (33) Yuan, X.; Raniolo, S.; Limongelli, V.; Xu, Y. The Molecular Mechanism Underlying Ligand Binding to the Membrane-Embedded Site of a G-Protein-Coupled Receptor. *J. Chem. Theory Comput.* **2018**, *14*, 2761–2770.
- (34) Piccini, G.; Polino, D.; Parrinello, M. Identifying Slow Molecular Motions in Complex Chemical Reactions. *J. Phys. Chem. Lett.* **2017**, *8*, 4197–4200.
- (35) Yang, Y. I.; Parrinello, M. Refining Collective Coordinates and Improving Free Energy Representation in Variational Abstract Keywords I Introduction. *J. Chem. Theory Comput.* **2018**, *14*, 2889–2894.

- (36) Pérez-hernández, G.; Paul, F.; Giorgino, T.; Fabritiis, G. D.; Noé, F. Identification of Slow Molecular Order Parameters for Markov Model Construction. **2013**, *015102*.
- (37) Noé, F.; Nüske, F. A Variational Approach to Modeling Slow Processes In Stochastic Dynamic Systems. *Multiscale Model. Simul.* **2013**, *11*, 635–655.
- (38) Sultan, M. M.; Pande, V. S. TICA-Metadynamics: Accelerating Metadynamics by Using Kinetically Selected Collective Variables. *J. Chem. Theory Comput. Theory Comput.* **2017**, *13*, 2440–2447.
- (39) Voter, A. F. A Method for Accelerating the Molecular Dynamics Simulation of Infrequent Events. *J. Chem. Phys.* **1997**, *106*, 4665–4677.
- (40) Grubmuller, H. Structural Transitions. *Phys. Rev. E* **1995**, *52*, 2893–2906.
- (41) Tiwary, P.; Parrinello, M. From Metadynamics to Dynamics. *Phys. Rev. Lett.* **2013**, *111*, 230602.
- (42) Salvalaglio, M.; Tiwary, P.; Parrinello, M. Assessing the Reliability of the Dynamics Reconstructed from Metadynamics. *J. Chem. Theory Comput.* **2014**, *10*, 1420–1425.
- (43) Casasnovas, R.; Limongelli, V.; Tiwary, P.; Carloni, P.; Parrinello, M. Unbinding Kinetics of a p38 MAP Kinase Type II Inhibitor from Metadynamics Simulations. *J. Am. Chem. Soc.* **2017**, *139*, 4780–4788.
- (44) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- (45) Jorgensen, W. L.; Madura, J. D. Solvation and Conformation of Methanol in Water. *J. Am. Chem. Soc.* **1983**, *105*, 1407–1413.

- (46) Pophristic, V.; Vemparala, S.; Ivanov, I.; Liu, Z.; Klein, M. L.; DeGrado, W. F. Controlling the Shape and Flexibility of Arylamides: a Combined Ab Initio, Ab Initio Molecular Dynamics, and Classical Molecular Dynamics Study. *J. Phys. Chem. B* **2006**, *110*, 3517–26.
- (47) Liebschner, D.; Dauter, M.; Brzuszkiewicz, A.; Dauter, Z. On the Reproducibility of Protein Crystal Structures: Five Atomic Resolution Structures of Trypsin. *Acta Cryst.* **2013**, *69*, 1447–1462.
- (48) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.
- (49) Grazioso, G.; Limongelli, V.; Branduardi, D.; Novellino, E.; De Micheli, C.; Cavalli, A.; Parrinello, M. Investigating the Mechanism of Substrate Uptake and Release in the Glutamate Transporter Homologue Glt Ph through Metadynamics Simulations. *J. Am. Chem. Soc. Chem. Soc.* **2012**, *134*, 453–463.
- (50) Soderhjelm, P.; Tribello, G. A.; Parrinello, M. Locating Binding Poses in Protein-Ligand Systems using Reconnaissance Metadynamics. *Proc. Natl. Acad. Sci.* **2012**, *109*, 5170–5175.
- (51) Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible Docking in Solution using Metadynamics. *J. Am. Chem. Soc.* **2005**, *127*, 2600–2607.
- (52) Chelli, R.; Gervasio, F. L.; Procacci, P.; Schettino, V. Stacking and T-Shape Competition in Aromatic-Aromatic Amino Acid Interactions. *J. Am. Chem. Soc.* **2002**, *124*, 6133–6143.
- (53) Essex, J. W.; Severance, D. L.; Tirado-Rives, J.; Jorgensen, W. L. Monte Carlo Simulations for Proteins: Binding Affinities for Trypsin-Benzamidine Complexes via Free-Energy Perturbations. *J. Phys. Chem. B* **1997**, *101*, 9663–9669.

- (54) Katz, B. A.; Elrod, K.; Luong, C.; Rice, M. J.; Mackman, R. L.; Sprengeler, P. A.; Spencer, J.; Hataye, J.; Janc, J.; Link, J.; Litvak, J.; Rai, R.; Rice, K.; Sideris, S.; Verner, E.; Young, W. A Novel Serine Protease Inhibition Motif Involving a Multi-centered Short Hydrogen Bonding Network at the Active Site. *J. Mol. Biol.* **2001**, *307*, 1451.
- (55) Talhout, R.; Engberts, J. B. F. N. Thermodynamic Analysis of Binding of p-Substituted Benzamidines to Trypsin. *Eur. J. Biochem.* **2001**, *1560*, 1554–1560.
- (56) Bruce, N. J.; Ganotra, G. K.; Kokh, D. B.; Sadiq, S. K.; Wade, R. C. New Approaches for Computing Ligand-Receptor Binding Kinetics. *Curr. Opin. Struct. Biol.* **2017**, *49*, 1–10.