Fast dynamic docking guided by adaptive electrostatic bias: the MD-Binding approach

Andrea Spitaleri,[a, ‡] *Sergio Decherchi,*[a,b, ‡] *Andrea Cavalli,*[c,d,*] *Walter Rocchia*[a,*]

[a]CONCEPT Lab, Istituto Italiano di Tecnologia, via Morego, 30, I-16163 Genoa, Italy

[b]BiKi Technologies srl, Via XX Settembre 33/10, 16121 Genoa, Italy

[c]CompuNet, Istituto Italiano di Tecnologia, Via Morego 30, 16163 Genova, Italy

[d]Department of Pharmacy and Biotechnology, University of Bologna, Via Belmeloro 6, I-40126 Bologna, Italy

Keywords: Molecular dynamics, enhanced sampling, molecular recognition, drug discovery, binding path.

ABSTRACT

Engineering chemical entities to modify how pharmaceutical targets function, as it is done in drug design, requires a good understanding of molecular recognition and binding. In this context, the limitations of statically describing bimolecular recognition, as done in docking/scoring, call for insightful and efficient dynamical investigations. On the experimental side, the

characterization of dynamical binding processes is still in its infancy. Thus, computer simulations, particularly molecular dynamics (MD), are compelled to play a prominent role, allowing a deeper comprehension of the binding process and its causes and thus a more informed compound selection, making more significant the computational contribution to drug discovery[1]. Unfortunately, MD-based approaches cannot yet describe complex events without incurring prohibitive time and computational costs. Here, we present a new method for fully and dynamically simulating drug-target-complex formations, tested against a real world and pharmaceutically relevant benchmark set. The method, based on an adaptive, electrostatics-inspired bias, envisions a campaign of trivially parallel short MD simulations, and a strategy to identify a near native binding pose from the sampled configurations. At an affordable computational cost, this method provided predictions of good accuracy also when the starting protein conformation was different from that of the crystal complex, a known hurdle for traditional molecular docking[2]. Moreover, along the observed binding routes, it identified some key features also found by much more computationally expensive plain-MD simulations. Overall, this methodology represents a significant progress in the description of binding phenomena.

1. INTRODUCTION

Understanding the mechanisms underlying molecular recognition and binding is pivotal for several fields (e.g. chemistry, biophysics, drug discovery)[3-5]. The scientific community is historically used to judge the entire binding process by just looking at the "end of the movie", namely the structure of the complex. This is mainly due to the availability of experimental crystallographic structures, which describe the complex with atomic detail. The computational counterpart of this is molecular docking, which was conceived many years ago and is now a well

established technology, the reference in the field.[6] However, affordable computational requirements and practical applicability, rather than inherent predictive power and accuracy, have made molecular docking the choice of election[7]. Since its inception, this approach underwent several significant but still incremental improvements.[8] We believe that the time is now ripe for a scientific and technological shift in mindset, where the dynamic character of binding is considered and analyzed in addition to the single crystallographic structure[9]. The knowledge of binding paths can point to key molecular events, such as gate openings, encounter complexes, and other intermediates, which can in turn suggest modifications to hit compounds[10]. Unfortunately, there are not yet established experimental techniques for dynamically and structurally describing such processes at the atomistic level. In contrast, this is a momentous time for computer-aided predictions of events at the molecular scale. Computational capabilities have grown tremendously in recent years, and there has been continuous development of dedicated techniques in this area[11]. *In silico* approaches, particularly molecular dynamics (MD), are thus the most promising resource for addressing this need[12, 13]. Indeed, long *plain* MD simulations have already described the spontaneous binding of small organic molecules to biological targets of pharmacological interest, demonstrating both the adequacy of the present force fields and also the feasibility, at least in some cases, of observing rare events[10, 14, 15]. However, the sampling provided by this approach rarely meets the requirements for statistical significance, which is desirable for quantitatively estimating thermodynamic observables. Moreover, it requires a huge amount of computational resources, far more than the average time allocated to a real-world drug discovery project.

One solution is to combine MD with so-called *enhanced sampling* approaches, whereby shorter but biased simulations energetically and mechanistically characterize binding events[16-20].

Indeed, in recent years, the impact of MD in drug design has increased enormously due to several enhanced sampling techniques exploited to investigate the ligand binding process[12]. In this field, the most broadly applied techniques are FEP [21], umbrella sampling [22], steered-MD [23], and, markedly, metadynamics, which has been used several times, in its original or funnel version [24], to study protein-ligand binding, including for a variety of targets of pharmacological interest[25-28]. More recently, Moro and coworkers addressed the issue of fast MD-based docking simulations for the human adenosine receptor[29]. Interestingly, several biased approaches perform free-energy estimations as a function of the progress coordinate along a guess binding path, when available[30, 31]. Overall, if *enhanced sampling* approaches are to be effective, they usually need a quite detailed a priori knowledge of the most relevant degrees of freedom, or collective variables (CVs), for the process at hand. CVs that prove successful for one system cannot easily be extended to different systems, limiting their applicability. Consequently, the success of these methods can be user- and system-dependent.

Here, we propose a new method, dubbed MD-binding, which systematically exploits MD to predict binding complexes and identify plausible binding paths. The aim of the method is enabling the scientist studying binding to make, at a reasonable computational cost, interpretations and predictions in a more informed way, that includes the observation of the "full movie" of the process together with the intrinsic benefit of fully flexible docking. To do this, it takes inspiration from the well-established notion that *electrostatic interactions* play a fundamental role in molecular recognition and binding, being pivotal at both long and short ranges[32]. Indeed, in nature, the long-range electrostatic behavior allows the interacting partners to spend sufficient time in relative proximity, so that they assume the most suitable conformation for binding. When approaching the final bound conformation, the short-range components begin

providing high specificity to the forming complex. This power has been exploited by some computational approaches applied to systems where electrostatics is known to play an important role[33, 34]. With respect to those approaches, the dramatic difference of the presented one lies in the assumption that electrostatics-shaped field lines can effectively guide binding also in processes that are not naturally driven by electrostatics. In line with this assumption, we add a contrived attractive multi-centered external bias that acts between the heavy atoms of the binding partners. In protein-ligand binding cases, the bias acts between a subset of the residues of the binding site and the ligand.

The other key aspect of the method is the *adaptivity* of the bias. In this respect, the intensity of the external force is always kept at a fraction of the intensity of the actual physical force felt by the ligand. Moreover, the biasing force gradually switches off as the process moves forward so that, after the conjectured passing of the transition state has occurred, it slowly recovers the behavior of classical unbiased MD. In summary, the added forces are everywhere tangent to the electrostatic field lines generated by the fictitious charges on ligand and binding site. Their intensity is regulated by the adaptivity rules. The method requires the prior identification of binding site residues. This information is usually available before any structure-based study.

The protocol envisions campaigns of several short (20 ns) simulation runs. Here, we demonstrate its performance on a diverse and challenging set of protein-ligand complexes of pharmacological interest, encompassing enzymes, kinases, and GPCRs. Some of the systems have been chosen due to the availability of long plain MD trajectories describing the binding, to be used as a reference. We also tested our approach against a protein-peptide system. In campaigns of about 20 runs each, the method attained at least once near-native binding poses. We also compared some mechanistic details of our paths with those obtained by plain MD. This

comparison pointed to some common features, indicating a new and powerful way to exploit molecular simulations in real-case drug design settings at a reasonable computational cost.

2. METHODS

The protocol consists of a number of steps:

- Characterization of the binding pocket. This is done via the NanoShaper tool[35], which identifies the atoms facing the lumen of the pocket, we call it the "entrance" of the pocket. Several entrances can be detected.
- Initial ligand positioning. The ligand is positioned with a random orientation at a predetermined distance, measured in terms of the thickness of the solvation shell around the ligand, from the entrance of the binding pocket.
- Identification of attracting atoms: the *attractive* (i.e. the atoms of a subset S of the site and those of the ligand L) and the *switch-off* (S') residues are chosen, as detailed in subsection 2.3.
- Run of the simulation campaign. A number of independent simulation runs are launched in parallel. In this work, campaigns of 20 runs per entrance gate, each of them during 20ns, resulted in at least one near native binding pose.
- Post-processing of the resulting trajectories. The unbiased final parts of the trajectories are taken and clustered with the k-medoids approach presented in Decherchi et al.[31], imposing k=20. If the bias never vanishes during a simulation run, then that run is discarded.
- Ranking of the poses. A Scaled MD approach, as described in subsection 2.5, is used to identify the best candidate binding pose and, consequently, a physically sound binding path.

2.1. Characterization of the binding pocket and initial ligand positioning. NanoShaper is used to characterize the binding site in terms of volume, surface area and constituting residues. Large and small probe radii of 3.0 and 1.4 Å, respectively, were used. At the same time,

NanoShaper samples the part of the pocket surface that is accessible from the solvent and defines a set of points that lie at the entrance of the pocket. These dots, and the corresponding outward normal vectors, are clustered in *N* "gate" sets. The centroid of each cluster (i.e. the representative of the corresponding gate) identifies a possible access of the ligand to the pocket. Consistently, the ligand is positioned away from each gate along the normal, with random orientation. Indeed, no information is provided about the "correct" orientation or the final contacts between the receptor and the ligand.

If there is ligand-protein overlap, or if the hydration shell between the ligand and protein is thinner than 10.0 Å, the ligand is translated farther away from the protein.

2.2. The electrostatic-shaped bias. The MD-binding method envisions an additive external force that is summed to the regular potential energy of the system in a molecular dynamics simulation. The bias consists in a moving umbrella (as per classical steered MD) where, however, the force constant $K(t)$ is adaptively modified along time. The functional form of the bias hence is:

$$\frac{1}{2} C * K(t) * (CV(t) - CV_0(t))^2$$

The pair of attracting atom sets (S as site, L as ligand) must thus be defined. The collective variable $CV(t)$, is the electrostatic energy of the biasing system, and can be expressed as the sum of terms having the following form:

$$CV(t) = \sum_{i \in L, s \in S}^{n} \frac{Q_i Q_s}{r_{i,s}(t)} \exp(-r_{i,s}(t)/\lambda)$$

where, r are the interatomic distances, exp(-r/λ) is a decay function aimed at avoiding unnecessary long ranged forces, and λ is the parameter that rules the spatial range of the decay (heuristically set to 10 Å).

Fictitious charges are therefore positioned on every attracting atom. In the present implementation: $Q_l = -Q_s \ \forall l \in L, s \in S$. The absolute value of the charges is not relevant since the overall intensity is ruled by the C parameter.

The total force field lines result from the vectorial composition of regular and biasing forces. "By design", the total force field lines identify a path leading to the binding site, which the ligand is invited to follow. This path automatically and dynamically updates as the binding site changes shape along its natural conformational evolution. Importantly, spreading the attraction across all the atoms of the S and L sets increases the likelihood of synergistic motions and reduces the likelihood of conformations that are incompatible with the binding due to steric hindrance.

2.3. Choice of the attracting and switch-off atoms. The set L contains all of the ligand's heavy atoms. The set S is chosen as a subset of the overall list of atoms constituting the binding site, which is created with NanoShaper. The choice of the subset is based mostly on geometric factors. Namely, the heavy atoms of the innermost part of the site are made to be attracting (for the presently studied systems, see Table S1). The choice of the attracting residues is an important aspect of the method, and affects its performance. Results of three cases of MD-Binding approach where the residues at the bottom of the binding pocket have been removed from the S set are shown in Table S7. The choice of how many atoms should be included in S depends on several factors, including ligand size, and the size and shape of the binding site. The switch-off atoms, defined as set S', should ideally provide an estimate of whether the transition state for the binding process has been overcome. When the S' set is approached by the ligand, the intensity of the biasing force irreversibly and gradually decreases.

2.4. Dynamics and adaptivity of the bias. The application of the bias envisages a steering schedule on the collective variable target value $CV_0(t)$, enforcing the reduction of the energy of the contrived electrostatic system. The form of the steer is piecewise linear, and consists of a first part, where the decrease in the target value of the energy is of limited amount to allow reorientation and a minimal conformational rearrangement of the ligand (see Figure S3 in the Supplementary Information). In the following steps of the plan, the decrease becomes more pronounced, increasing the thrust applied to the ligand. The actual target value of the bias, especially in the final part of the steering, is however of limited importance due to the modulation induced by the adaptivity of the protocol, as hereby explained.

The protocol is adaptive in four different respects.

First, the biasing force is always kept below a user-defined threshold. More specifically, the modulus of the biasing force is rescaled at every 0.2 ps of simulation, so that it equals a predefined fraction (10%, of the "real" resulting force acting on the ligand), which originates from the rest of the system. This is done to avoid an overly strong biasing, which could distort the structure of the protein, increasing the probability of following high-energy binding paths.

Second, the bias is further reduced based on the distance between the ligand L and a subset S' of S. S' is a switch-off set comprising atoms that are expected to interact with the ligand (either geometrically or chemically guessed, or a priori known). By monitoring the decrease of the above-mentioned distance, this option seeks to identify the final steps of the process, where the transition state for the binding is already overcome. This makes the bias in this phase particularly unobtrusive, so that the simulation becomes closer to unbiased MD.

The switch-off is obtained via a scaling pre-factor γ, which multiplies the biasing force (this is a component of *K(t)*). This factor is calculated via a switching function as follows:

$$\gamma = \frac{1}{1 + \exp(-ss * (dist - th))}$$

where ss and th are two suitable parameters, regulating the extent and the steepness of the switching. The value of *dist* equals:

$$dist = min_{y \in S'} min_{x \in L} d(x, y),$$

where d(x,y) is the pairwise distance between the x and y atoms.

A third flavor of adaptivity is introduced into the MD-binding method by adding a "memory effect". Once the *K(t)* has begun to decrease due to the approach of the S' set, it can no longer increase even if the ligand bounces back. This reduces the likelihood that the ligand reaches the bottom of the pocket, but increases the probability that, if it does, it does this by following physically sound trajectories, which is the highest priority.

Lastly, the dynamical variation of the spring constant is time-averaged over a circular buffer window of 2 ps to ensure that its time variation is consistently smooth.

2.5. Post-processing of the resulting trajectories and ranking of the putative binding conformations. Once the simulation campaign is finished, a pruning process is carried out. This process discards those replicas where either the unbiased MD is not recovered, i.e. the final value of *K(t)* is greater than 10^{-9} (the ligand did not approach the switch-off set S'), or the ligand bounces back from the binding site during the unbiased MD (final distance between ligand and the S' set greater than 7 Å). Based on our present experience, roughly half of the simulations can be discarded in this way. Table S2 shows the number of remaining replicas for each case study. The frames present in the unbiased chunks are aligned on the set of attracting residues and then clustered based on the heavy atoms of the ligand using the previously published k-medoids clustering method[31]. Again, based on our experience, 20 clusters characterize the sampled conformations with sufficient detail. The ranking process is based on the evaluation of the

stability of each medoid. Therefore, each medoid conformation is iteratively minimized with the steepest descent and conjugate gradient methods. Its stability is then estimated along short unrestrained Scaled Molecular Dynamics (SMD)[36, 37] runs in the NVT ensemble for 5 ns and with scaling factor "s" equal to 0.6. In SMD, the potential energy of the entire system is scaled by a factor s, consequently reducing energetic barriers, including that of ligand unbinding from a given pose. The rationale underlying this approach is that the best pose should be that exhibiting the largest stability despite the applied scaling. SMD, used in a more complex protocol, proved successful in ranking congeneric ligands based on residence time in reasonable amount of computational time[38] and already proved a viable tool to assess pose stability[31]. Table S3 reports the average RMSD of the SMD simulations and the scoring values from Autodock, Vina, and Drugscore for each of the 20 medoids along with the RMSD with respect to the X-ray. For the RMSD calculation, we first aligned the heavy atoms of both backbone and side chains of the residues constituting the protein binding site, as defined by NanoShaper, between each putative binding pose and the X-Ray structure of the complex. The RMSD was then calculated on the heavy atoms of the ligand. In case of RAD51-BRCA2 we used a proper protein-peptide scoring function, namely HADDOCK score, to rank the 20 medoids[39]. The fitting has been carried out on the RAD51 backbone and the RMSD calculation on the BRCA2 backbone, residues from 1523 to 1534, which define the peptide interface contact region (iRMSD). iRMSD, or "interface RMSD", which is the RMSD calculated only on the backbone of the residues that are at the binding interface, is a commonly accepted criterion used for protein-protein docking evaluation[40].

The figures of merit used to assess the performance of SMD and other scoring functions in selecting the near-native binding poses are: i) the coincidence between the best ranked and the

nearest to native medoid (Min to Min); ii) the ability to position the nearest to native medoid within 1 Å (in the case of SMD) or within 2kcal/mol (in the case of scoring functions) of the best ranked pose; and iii) the ability to position the nearest to native medoid within 1 σ (standard deviation) of the best ranked pose. Standard deviation is calculated from the set of RMSDs with respect to initial configuration in the case of SMD and from the set of scores in the remaining cases.

2.6. Computational molecular dynamics setup. The algorithm setup and the analyses, excluding the ligand geometric optimization, were performed using BiKi Life Sciences suite 1.3 of BiKi Technologies s.r.l. Each compound geometry was optimized via a quantum mechanical approach: electron density calculations were performed at the HF/6-31G* level of theory using the NWChem program[41]. Partial charges were derived using the RESP method in Antechamber[42], while a GAFF[43] parameterization was used to achieve the complete topological description of each ligand.

Once the ligand was positioned through the "Residue Placement" tool in BiKi, the system was solvated in an orthorhombic box using TIP3P water model[44]. Adding a suitable number of counter-ions neutralized the overall system. Amber99SB-ildn force field was used[45]. Then, the whole system underwent energy minimization. Four different consecutive equilibration steps were performed: 1) 100 ps in NVT ensemble at 100K with both the protein backbone and ligand restrained (1000 kJ/mol nm^2), 2) 100 ps in NVT ensemble at 200K with both the protein backbone and the ligand restrained, 3) 100 ps in NVT ensemble at 300K with the protein free and the ligand restrained, and 4) 1000 ps in NPT ensemble at 300K with the protein free and the ligand restrained. Electrostatics was treated with the cutoff method for short-range interactions

and with the Particle Mesh Ewald method for long-range interactions (rlist = 1.1 nm, cutoff distance = 1.1 nm, VdW distance = 1.1 nm, PME order = 4). The constant temperature conditions were provided using the velocity rescale thermostat[46], which is a modification of the Berendsen's coupling algorithm[47]. The coordinate output from the last simulation was then used as input to produce the biased molecular dynamics. Finally, 20 replica production runs, 20 ns long in NVT ensemble at 300K, were performed for each complex using C=0.1 (the fraction of the felt force, here 10%), th=0.4 (in nm, representing the switch-off residue distance at which the bias is turned off), a smoothing window size of 1000 samples, and a maximal K(t) of 0.001 (maximal steering constant).

3. RESULTS

3.1. Sampling putative binding poses. Here, we analyze the method's ability to sample the correct binding pose as a result of a fully flexible and explicitly solvated dynamics that starts from the apo conformation of the target protein, available for 5 out of 8 targets. Results are shown in Table 1. Our figure of merit is the RMSD with respect to the crystal structure of a chosen medoid obtained by clustering the final unbiased part of the binding simulations (see Methods for details). In only two of the analyzed systems the RMSDs of the nearest to native medoid are larger than 2Å, a commonly accepted threshold in docking benchmarks. Also in these cases, namely AChE- Galantamine and GPCR β2- Alprenolol with RMSDs of 2.1Å (crystal resolution 2.4Å) and 2.5Å (3.16Å), respectively, however, the main molecular interactions are reproduced. Remarkably, the simulations sampled configurations very close to the crystal ones in all cases (see Table 1). The same approach was also attempted on a protein-peptide system, namely the RAD51-BRCA2 complex, observing similarly good results. Here, a restraint on the

local secondary structure of the peptide was applied, to reduce the conformational space to be explored, as detailed in the SI (Table S4).

Table 1. Sampling of binding poses for a diverse set of targets and ligands

#	Target (initial struct.)	Ligand	Holo PDB	Best Medoid RMSD (Å)	Min Obs RMSD (Å)	Crystal Resolution (Å)	RMSD Target initial struct. vs. holo (Å)
1	CDK2 (4EK3)	Staurosporine	4ERW	1.5	1.3	2.0	3.5
		Hymenialdisine	1DM2	1.0	0.5	2.1	4.7
2	CK1D (3UYS)	PF670462	3UZP	0.6	0.3	1.94	1.5
3	AChE (4EY4)	Donepezil	4EY7	2.0	1.4	2.35	1.3
		Galantamine	4EY6	2.1	1.4	2.4	0.9
4	GPCR β2	ICI 118,551	3NY8	1.0	0.9	2.84	0.9
		Alprenolol	3NYA	2.5	1.6	3.16	0.0
5*	GPCR A2A	T4E	3UZC	1.6	1.2	3.34	0.0
6	SRC (1Y57)	PP1	1QCF	0.9	0.3	2.0	0.7
7*	PNP (1M73)	DADMe Imm-H	1RSZ	1.9	1.0	2.2	3.1
8	RAD51	BRCA2	1N0W	1.9	0.9	1.7	0.0

* for these targets 3 entrances to the site were identified and 3 campaigns were performed

Column 4: PDB code of the holo structure of the target. Column 5: the RMSD wrt. crystal of the nearest to crystal medoid obtained by clustering (see the Methods). Column 6: minimum RMSD observed throughout the simulation. Column 7: resolution of the crystal of the complex. Column 8: RMSD between the initial conformation of the binding site and that of the complex.

Although dynamic docking is different in many respects from classic molecular docking, conventional docking simulations were performed on targets #1 to #7 using Autodock Vina[48] and rDock[49]. In Autodock Vina all the attracting residues and the dihedral angles of the ligand have been left flexible while in rDock only terminal OH and NH3 groups on the protein in vicinity of the binding site were allowed to be fully flexible. Further details can be found in Tables S5 and S6 and are briefly summarized here in Figure 1. With these settings, no pose with RMSD< 2Å was identified by Autodock Vina among the first 20 ones, while at least one pose with RMSD< 2Å was observed with rDock for 6 out of 10 complexes (see Figure 1a). Then, we considered separately the performance of rDock on the systems where the RMSD between the input conformation of the target and that of the holo complex, reported in column 8 of Table 1, is larger or smaller than 1Å. The results reported in Figure 1b, rDock1 and rDock2 series, show a remarkable sensitivity to this aspect, in contrast to what occurs with MD-Binding. Finally, we also tested the sensitivity of the performance of rDock with respect to the input conformation of the ligand, by feeding the program with the same ligand structures obtained after the minimization and equilibration steps, which precede the execution of MD-Binding. Results, reported in Figure1b, (rDock3 series) show a marginal deterioration of the performance. To highlight the importance for molecular docking tools of the complementarity between the conformations of the target and the ligand and of the amplitude of the conformational space that needs to be explored, we performed a rigid self-docking with Autodock Vina on target #5, where we have only the holo conformation, obtaining very good results (see Table S8). Overall, these results are in agreement with what it is known about the performance of molecular docking when non-cognate structures are used[7, 50].

As a further test, we performed two Steered MD campaigns of 20 replicas each, using the distance between the centers of mass of the ligand and the same residues composing the S set in the MD-binding method. This was done for the CDK2-STU and the CDK2-HYM complexes. In no replica the final RMSD wrt. crystal goes below 4.5Å. Related details concerning the methods and results are in the Supplementary Information (see Table S9).

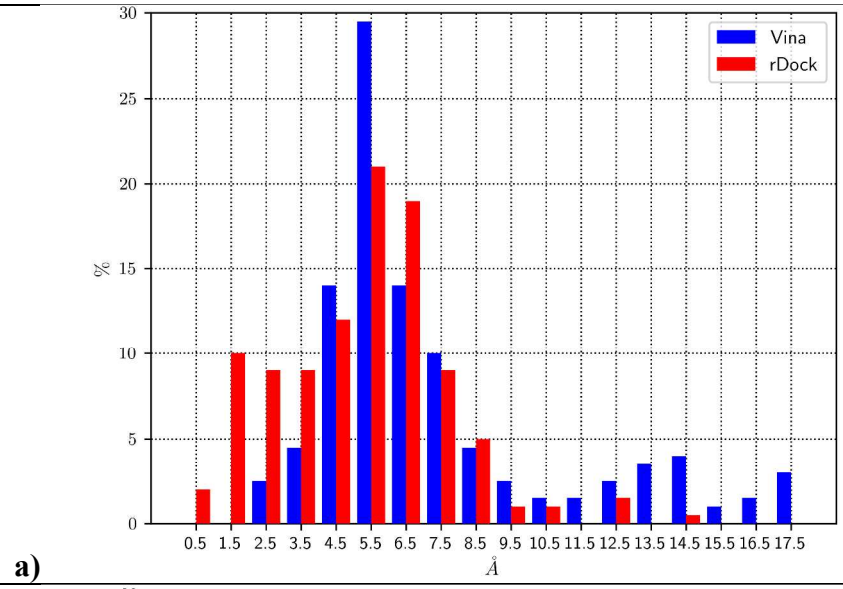

Figure 1. Performance of Molecular Docking tools. a) Statistics of RMSD wrt. crystal of the

first 20 poses provided by Autodock Vina (in blue) and rDock (in red) when they are inputted with crystal ligand structures. **b**) Performance of rDock on the complexes presenting a RMSD in protein binding site between inputted and crystal structure larger than 1Å (rDock1 series), and lower than 1Å (rDock2 series). rDock3 series shows the performance of rDock when the inputted ligand structures are subjected to energy minimization and equilibration (the same preprocessing done in the MD Binding procedure).

3.2. Identification of near-native binding poses. Here, we analyze the method's ability to identify near-native binding poses from among the configurations explored during the simulation campaign. This ability is essential for the global viability of a dynamical docking approach. It is achieved by first exploiting the adaptivity of the bias. As detailed in the Methods, by construction, the bias gradually switches off as the algorithm judges that the ligand is approaching the bound state, so that the last part of the simulation becomes unbiased. In some runs, this condition is never met and they are therefore immediately discarded (see, for example, Figure S1). Then, the last unbiased pieces of the remaining trajectories are collected and clustered via a k-medoid approach, imposing 20 final clusters. Finally, the corresponding medoids are ranked based on their stability, assessed via RMSD deviation, during a short run of unrestrained scaled MD (SMD)[37] (see the Methods). As shown in Figure 2, this approach provided equal or better performance relative to scoring functions traditionally adopted in molecular docking, namely Autodock4[51], Vina[48] and DSX DrugScore[52]. In this respect, the most relevant result is the ability of SMD to map the best rank to the best pose, which makes the overall approach particularly efficient. Interestingly, the different ranking approaches were challenged to recognize a very bad conformation: a head-to-tail inversion in the case of the drug

donepezil binding to the enzyme acetylcholinesterase. Unlike the tested scoring functions, which could not reject it so clearly, SMD identified this unlikely binding pose and ranked it among the lowest positions. (see Supplementary Information for complete data). Another interesting outcome is that the main interactions between the binding partners were recapitulated. Moreover, the simulated pose correctly reports the water-mediated interaction that can be observed in the crystal structure of one of the examined cases, namely the in the CK1D case (see Figure S2).

Interestingly, the discriminating ability of SMD seems to extend somehow also to peptide ligands.

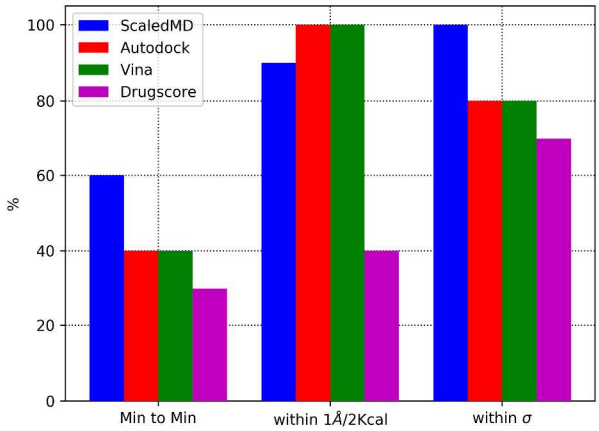

Figure 2. Ranking of putative bound poses. Performance of SMD and widely adopted docking scoring functions in selecting the near-native binding poses. Three figures of merit were used: the ability to map the best rank to the best configuration (Min to Min), the ability to position the best configuration within 1 Å or 2kcal/mol of the best ranked pose, and the ability to position the actual best configuration within 1 σ (standard deviation) of the best ranked pose. Target #8 was not included in this comparison, since the considered scoring functions have been tuned for small chemical compounds rather than for peptides.

To assess this aspect, we compared the results it provided on the RAD51-BRCA2 complex with those of a scoring function specifically designed for the protein-protein binding, namely that of HADDOCK[53]. As shown in Table S3, the difference of SMD-based score, that is the iRMSD averaged over 5 ns of SMD, between the nearest to native pose and the best scored one was of 0.7 Å, corresponding to 0.16 σ (standard deviation of the SMD-scores of that campaign). The HADDOCK score difference between nearest native and best scoring pose was instead of slightly more than one standard deviation.

3.3. Similarity to near-native binding paths. To the best of our knowledge, no experimental study to date has provided a full atomistic characterization of an entire binding process. However, in principle, long plain-MD approaches can model a realistic binding process, provided that sufficient computational resources are available. Ligand-binding to three of our targets (namely #4, #6 and #7) has already been studied via plain MD, and these simulated binding events are the best available comparison reference for our biased trajectories. We therefore focus on the formation of three complexes, namely β2 adrenergic receptor and alprenolol, SRC kinase and PP1 inhibitor, and purine nucleoside phosphorylase and DADMe-Immucillin-H.

3.4. β_2-adrenergic binding alprenolol. The G-protein-coupled receptor β_2-adrenergic (β_2-AR) is an important target for hypertension and several heart diseases. DE Shaw and co-workers have used extensive plain MD to study how this receptor interacts and binds with the inverse agonist alprenolol[10]. We analyzed our best binding route with two binding trajectories obtained via plain MD and kindly provided by DE Shaw Research. The comparison showed interesting shared

features. The recognition is very similar to that described in Dror et al., involving the same contacts in the same temporal order. As shown in Figure 3, although the alprenolol molecule reached the entrance from different positions and with different orientations, the binding paths followed the same two-steps molecular mechanism. First, alprenolol reorients so that its hydrophobic group binds to the extracellular vestibule surface (Figure 3, upper row). Then, it penetrates into the binding pocket through the gate closing-opening mechanism between Tyr308 and Phe193 to form the salt bridges between the ammonium moiety and the carboxylate of Asp113 (Figure 3, lower row).

3.5. PP1 binding SRC kinase. Another complex that was previously studied by extensive plain MD is PP1 inhibiting SRC kinase, a relevant biological target for cancer therapy. The PP1 binding mode is characterized by a well-defined hydrogen bond network between the receptor, the ligand, and the surrounding water molecules. These interactions are only reproduced in one out of 20 medoids, which the SMD-based approach correctly identified. Characterization of the SRC-PP1 binding path was done by means of two different campaigns (20 runs each) with the ligand starting in different positions. The two binding paths leading to near-native binding, one per campaign, were compared with each other and with one long plain-MD trajectory kindly provided by DE Shaw Research[15].

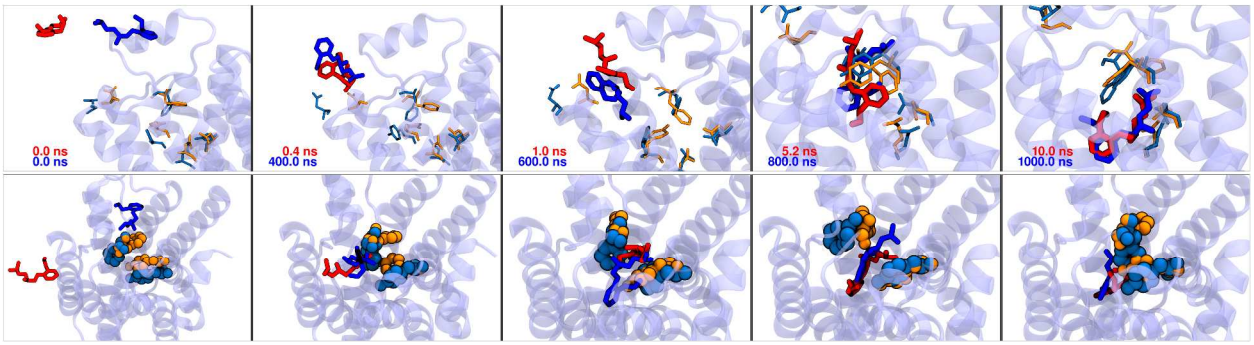

Figure 3. Binding of alprenolol to the β2-adrenergic receptor. Time sequence of the alprenolol binding process studied by MD-Binding (ligand in red and relevant side chains in orange) and plain MD[10] (ligand in blue and relevant side chains in ice blue). Simulated times are reported in red for MD-Binding and in blue for plain MD. In both simulations alprenolol approaches the extracellular vestibule with the hydrophobic ring first and then it rotates to form the salt bridges (upper row). Moreover, in both simulations, Tyr308 and Phe193 separate and let the ligand's hydrophobic group pass through them. After the binding occurs, the gate closes (lower row).

Interestingly, despite the different starting positions and the much larger sterically available space, in all cases the PP1 ligand accessed the site passing near a relatively small hydrophobic patch, as summarized in Figure 4. In the accelerated simulation, after spending a few ns in this region, the ligand enters the binding site with the heteroaromatic moiety, i.e. pyrazole and pyrimidinylamine. Only later does the ligand rearrange, positioning the toluene ring at the bottom of the binding site, recapitulating the main interactions between the heteroaromatic ring and the Met341 and Glu339 residues.

3.6. DADMe binding purine nucleoside phosphorylase. We compared the successful binding trajectory obtained with our dedicated bias to the set of binding trajectories obtained

with our previous work[31]. Due to the peculiar structure of the binding site, which presents 3 sizeable entrances, we conducted one campaign for each entrance for a total of 60 runs. The best obtained medoid had an RMSD of 1.9 Å and the corresponding route followed the so-called "upper" path, recapitulating the main interactions observed in the crystal. Similarly to what was observed in the unbiased simulations, the ligand entered the binding site only when the α-helix facing the site left enough space and the site became exposed. The first observed interactions were between the ligand dihydroxypyrrolidine and the phosphate. Upon binding, DADMe Imm-H recapitulated the bi-dentate interaction with Asn243, and the bond with the PO_4^{2-} group, as observed in the plain simulation.

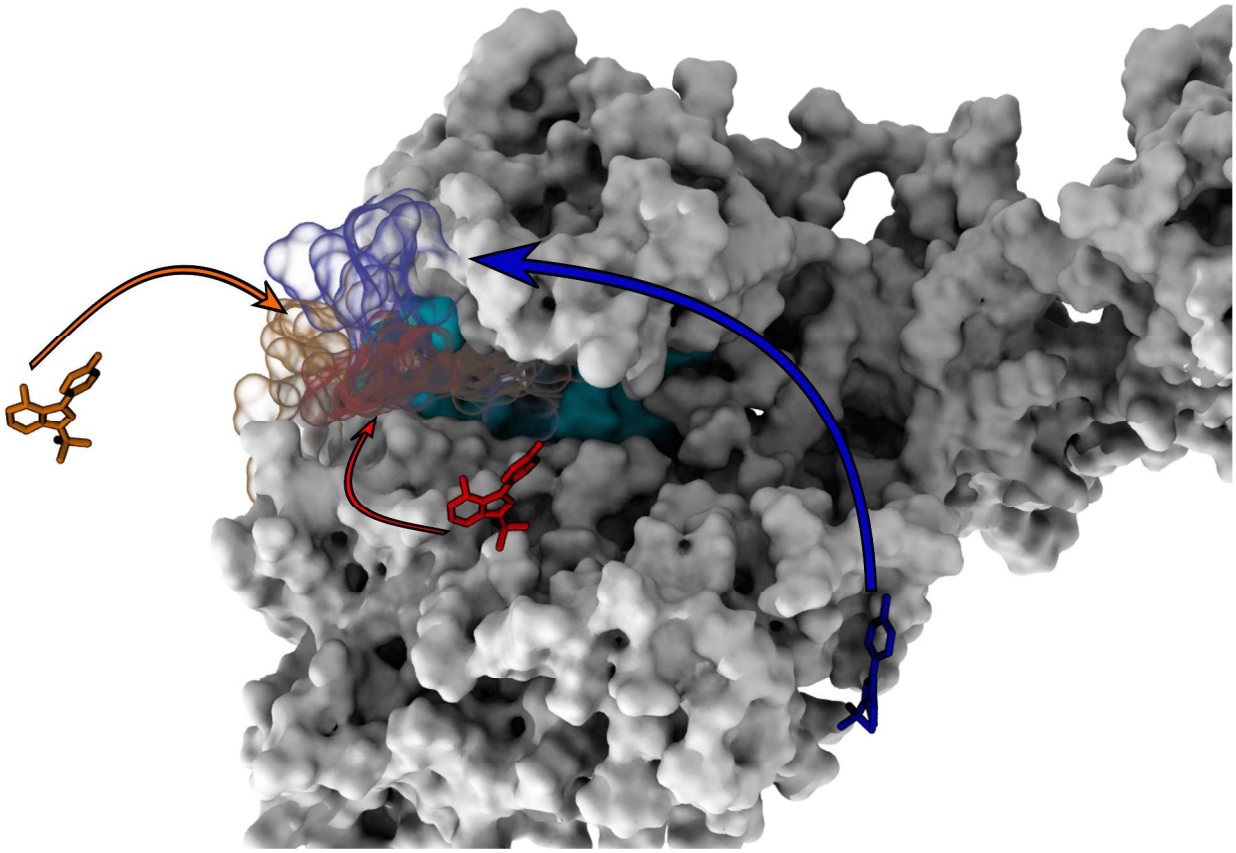

Figure 4. Binding of PP1 to SRC kinase. Binding paths observed in our biased runs (in orange and red) and in the plain MD described in Shan et al. (in blue)[15]. Despite initial starting positions and orientations are different, and that the sterically accessible region (in cyan) is quite large, all ligands go through the same narrower region (indicated by the convergence of the three differently colored clouds representing the three ligands' paths). Subsequent analysis showed that the preferred paths pass near the only hydrophobic patch found in the site entrance.

4. DISCUSSION and CONCLUSIONS

The method presented here addresses the main hurdles faced by the computer simulation of protein-ligand binding processes. Protein-ligand binding poses can sometimes be reproduced with a high level of accuracy by plain-MD simulations in the microsecond-to-millisecond timescale, or by shorter, but more numerous, simulations. However, these approaches require a great deal of time and resources, which is rarely compatible with routine drug discovery projects, where, ideally, many compounds would be computationally investigated prior to medicinal chemistry experiments[4]. The results presented herein indicate that a partially guided approach can provide insightful information in an aggregate time which is, on average, 2-3 orders of magnitude shorter than that required by plain MD even in cases where a non-cognate site conformation is chosen to start the simulation, therefore requiring a substantial rearrangement of binding site residues. Moreover, this approach carries the additional advantage of being divisible into trivially parallel short simulations.

To attain its aim, the method exploits electrostatic attraction to induce recognition and binding. Rather than localizing the attraction on centers of mass, as often reported in the literature, it spreads it over a large number of atoms, facilitating concerted movements. Moreover, it favors

the exploration of low-energy, physically sound paths via a suitably adaptive behavior. At the end of the simulations, near-native binding poses were identified by clustering and SMD-based conformation ranking procedures. Choosing the attracting residues on the protein side is a critical part of the method and examples of the deterioration of the performance induced by a wrong choice are reported in the SI. Our criterion here consists in a down-selection, chiefly based on geometric criteria, of the residues composing the binding site. Remarkably, the same choice of residues provided equally good results in type-1 inhibitors of three different kinases, namely CDK2, CK1D, and SRC kinase. A useful feature of the method is the possibility to include a priori knowledge about the binding with no difficulty, as it was done in the β2-adrenergic case, where the interaction of ligands with Asp113 is widely accepted. In the performed validation tests, we targeted the known orthosteric binding sites. However, the approach is expected to work also in case of targeting allosteric or cryptic sites.

Interestingly, the availability of the holo conformation of the protein, which lacks in the most challenging, and also interesting, cases, does not seem to be an essential requisite to achieve a good performance of the MD-Binding method. This is one of the main differences with conventional molecular docking, where the differences between self- and cross-docking results may be striking.

To challenge the extent of applicability of the method, we applied it to a relevant protein-peptide case, obtaining very promising results, although further studies and some adaptation are required to assess whether the novel methodology is fully suitable for protein-peptide and, in the long term, protein-protein systems.

A promising feature of our approach might be the ability to identify prominent mechanistic features of the binding process. There is increasing interest in the scientific community in

understanding the path that characterizes the binding and thus identifying the relevant intermediates that could explain non-obvious kinetic behaviors, such as traps and allosteric sites, as a new dimension to be exploited in drug design. In this respect, the binding trajectories generated by our method were compared to the available reference paths (generated via plain MD) and showed interesting mechanistic commonalities.

A further aspect to consider is that a reasonable initial guess binding path can be exploited to feed traditional path-based enhanced sampling techniques to obtain free-energy profile estimates along binding and, finally, chemical affinity[54, 55]. Interested readers can apply the techniques described in this paper by obtaining a trial version of the BiKi LifeSciences software suite by requesting it on the site http://www.bikitech.com.

The present methodology seeks to fulfill two divergent requirements. These are a) accuracy, which can, in principle, be obtained by extensive MD simulations, and b) speed, which is particularly necessary when many compounds must be investigated to drive a lead discovery campaign. Here, using electrostatics as the bias to accelerate MD simulations, we obtained accurate binding poses and routes much faster than via plain MD. The obtained results corroborate our initial hypothesis that electrostatics is suitable for building a good attractive basin even for binding processes that naturally are not driven by electrostatics. The presented approach paves the way for a feasible and efficient dynamic docking simulation, which delivers the "full movie" rather than just the "final scene". Dynamic docking is a quite different concept from conventional molecular docking,[56] providing a wider wealth of information at a larger computational cost. It can also be used as a second line alternative, in cases where the dynamic phenomena accompanying binding prevent static docking from predicting the correct conformation of the complex. We believe this will remarkably broaden the scope and the impact

of MD in every discipline concerning protein-ligand interaction and opens up new avenues for developing innovative drug discovery methods.

AUTHOR INFORMATION

Corresponding Authors

* To whom correspondence should be addressed. E-mail: walter.rocchia@iit.it, andrea.cavalli@iit.it

CONFLICT OF INTEREST

The Authors are shares owners of BiKi Technologies s.r.l., a company that commercializes software solutions for medicinal chemistry including the presented method.

ACKNOWLEDGMENTS

We thank D.E. Shaw Research for providing unbiased binding trajectories for two systems, BiKi Technologies s.r.l., and PRACE for awarding us access to resource on FERMI based in Italy at CINECA. We acknowledge AIRC MFAG n. 11899 for funding. We thank Dr. Simone Sciabola and Dr. Marco De Vivo for useful discussions.

ASSOCIATED CONTENT

Supporting Information.

Supplementary results with accompanied figures and tables referred to in the main text (.pdf). Supplementary Movies (in .avi format) showing the simulated binding of Alprenolol to the β2 Adrenergic GPCR, PP1 inhibitor to Src kinase and DADMe Imm-H to Purine Nucleoside

Phosphorylase. The corresponding structures (.gro) and trajectories (.xtc) format are also provided.

This information is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

(1) Carlson, H. A., Protein flexibility and drug design: how to hit a moving target. *Curr Opin Chem Biol* **2002,** *6* (4), 447-452.
(2) Lexa, K. W.; Carlson, H. A., Protein flexibility in docking and surface mapping. *Q Rev Biophys* **2012,** *45* (3), 301-343.
(3) Durrant, J. D.; McCammon, J. A., Molecular dynamics simulations and drug discovery. *BMC Biol* **2011,** *9*, 71.
(4) Kuhn, B.; Guba, W.; Hert, J.; Banner, D.; Bissantz, C.; Ceccarelli, S.; Haap, W.; Korner, M.; Kuglstatter, A.; Lerner, C.; Mattei, P.; Neidhart, W.; Pinard, E.; Rudolph, M. G.; Schulz-Gasch, T.; Woltering, T.; Stahl, M., A Real-World Perspective on Molecular Design. *J Med Chem* **2016,** *59* (9), 4087-4102.
(5) Jorgensen, W. L., The many roles of computation in drug discovery. *Science* **2004,** *303* (5665), 1813-1818.
(6) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E., A geometric approach to macromolecule-ligand interactions. *J Mol Biol* **1982,** *161* (2), 269-288.
(7) Chen, Y. C., Beware of docking! *Trends Pharmacol Sci* **2015,** *36* (2), 78-95.
(8) Ferreira, L. G.; Dos Santos, R. N.; Oliva, G.; Andricopulo, A. D., Molecular docking and structure-based drug design strategies. *Molecules* **2015,** *20* (7), 13384-13421.
(9) Nussinov, R.; Udgaonkar, J. B., Editorial overview: Folding and binding: Dynamic conformational heterogeneity is pivotal to cell life. *Curr Opin Struct Biol* **2016,** *36*, iv-vi.
(10) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E., Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc Natl Acad Sci U S A* **2011,** *108* (32), 13118-13123.
(11) Ruiz-Carmona, S.; Schmidtke, P.; Luque, F. J.; Baker, L.; Matassova, N.; Davis, B.; Roughley, S.; Murray, J.; Hubbard, R.; Barril, X., Dynamic undocking and the quasi-bound state as tools for drug discovery. *Nat Chem* **2017,** *9* (3), 201-206.
(12) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A., Role of Molecular Dynamics and Related Methods in Drug Discovery. *J Med Chem* **2016,** *59* (9), 4035-4061.
(13) Ganesan, A.; Coote, M. L.; Barakat, K., Molecular dynamics-driven drug discovery: leaping forward with confidence. *Drug Discov Today* **2017,** *22* (2), 249-269.
(14) Buch, I.; Giorgino, T.; De Fabritiis, G., Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc Natl Acad Sci U S A* **2011,** *108* (25), 10184-10189.
(15) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E., How does a drug molecule find its target binding site? *J Am Chem Soc* **2011,** *133* (24), 9181-9183.

(16) Colizzi, F.; Perozzo, R.; Scapozza, L.; Recanatini, M.; Cavalli, A., Single-molecule pulling simulations can discern active from inactive enzyme inhibitors. *J Am Chem Soc* **2010,** *132* (21), 7361-7371.
(17) Patel, J. S.; Berteotti, A.; Ronsisvalle, S.; Rocchia, W.; Cavalli, A., Steered molecular dynamics simulations for studying protein-ligand interaction in cyclin-dependent kinase 5. *J Chem Inf Model* **2014,** *54* (2), 470-480.
(18) Lin, Y. L.; Roux, B., Computational analysis of the binding specificity of Gleevec to Abl, c-Kit, Lck, and c-Src tyrosine kinases. *J Am Chem Soc* **2013,** *135* (39), 14741-14753.
(19) Woo, H. J.; Roux, B., Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc Natl Acad Sci U S A* **2005,** *102* (19), 6825-6830.
(20) Wang, J.; Deng, Y.; Roux, B., Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys J* **2006,** *91* (8), 2798-2814.
(21) Jorgensen, W. L.; Thomas, L. L., Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria. *J Chem Theory Comput* **2008,** *4* (6), 869-876.
(22) Torrie, G. M.; Valleau, J. P., Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977,** *23* (2), 187-199.
(23) Grubmuller, H.; Heymann, B.; Tavan, P., Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* **1996,** *271* (5251), 997-999.
(24) Limongelli, V.; Bonomi, M.; Parrinello, M., Funnel metadynamics as accurate binding free-energy method. *Proc Natl Acad Sci U S A* **2013,** *110* (16), 6358-6363.
(25) Besker, N.; Gervasio, F. L., Using metadynamics and path collective variables to study ligand binding and induced conformational transitions. *Methods Mol Biol* **2012,** *819*, 501-513.
(26) Branduardi, D.; Gervasio, F. L.; Cavalli, A.; Recanatini, M.; Parrinello, M., The role of the peripheral anionic site and cation-pi interactions in the ligand penetration of the human AChE gorge. *J Am Chem Soc* **2005,** *127* (25), 9147-9155.
(27) 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* (2), 277-285.
(28) Gervasio, F. L.; Laio, A.; Parrinello, M., Flexible docking in solution using metadynamics. *J Am Chem Soc* **2005,** *127* (8), 2600-2607.
(29) Cuzzolin, A.; Sturlese, M.; Deganutti, G.; Salmaso, V.; Sabbadin, D.; Ciancetta, A.; Moro, S., Deciphering the Complexity of Ligand-Protein Recognition Pathways Using Supervised Molecular Dynamics (SuMD) Simulations. *J Chem Inf Model* **2016,** *56* (4), 687-705.
(30) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G., String method in collective variables: minimum free energy paths and isocommittor surfaces. *J Chem Phys* **2006,** *125* (2), 24106.
(31) Decherchi, S.; Berteotti, A.; Bottegoni, G.; Rocchia, W.; Cavalli, A., The ligand binding mechanism to purine nucleoside phosphorylase elucidated via molecular dynamics and machine learning. *Nat Commun* **2015,** *6*, 6155.
(32) Sheinerman, F. B.; Norel, R.; Honig, B., Electrostatic aspects of protein-protein interactions. *Curr Opin Struct Biol* **2000,** *10* (2), 153-159.
(33) Samsonov, S. A.; Gehrcke, J. P.; Pisabarro, M. T., Flexibility and explicit solvent in molecular-dynamics-based docking of protein-glycosaminoglycan systems. *J Chem Inf Model* **2014,** *54* (2), 582-592.

(34) Do, T. N.; Carloni, P.; Varani, G.; Bussi, G., RNA/Peptide Binding Driven by Electrostatics-Insight from Bidirectional Pulling Simulations. *J Chem Theory Comput* **2013,** *9* (3), 1720-1730.
(35) Decherchi, S.; Rocchia, W., A general and robust ray-casting-based algorithm for triangulating surfaces at the nanoscale. *PLoS One* **2013,** *8* (4), e59744.
(36) Tsujishita, H.; Moriguchi, I.; Hirono, S., Potential-Scaled Molecular-Dynamics and Potential Annealing - Effective Conformational Search Techniques for Biomolecules. *Journal of Physical Chemistry* **1993,** *97* (17), 4416-4420.
(37) Sinko, W.; Miao, Y.; de Oliveira, C. A.; McCammon, J. A., Population based reweighting of scaled molecular dynamics. *J Phys Chem B* **2013,** *117* (42), 12759-12768.
(38) Mollica, L.; Decherchi, S.; Zia, S. R.; Gaspari, R.; Cavalli, A.; Rocchia, W., Kinetics of protein-ligand unbinding via smoothed potential molecular dynamics simulations. *Sci Rep* **2015,** *5*, 11539.
(39) van Zundert, G. C.; Rodrigues, J. P.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S.; van Dijk, M.; de Vries, S. J.; Bonvin, A. M., The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol* **2016,** *428* (4), 720-725.
(40) Mendez, R.; Leplae, R.; De Maria, L.; Wodak, S. J., Assessment of blind predictions of protein-protein interactions: current status of docking methods. *Proteins* **2003,** *52* (1), 51-67.
(41) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W., NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Computer Physics Communications* **2010,** *181* (9), 1477-1489.
(42) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* **2006,** *25* (2), 247-260.
(43) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J Comput Chem* **2004,** *25* (9), 1157-1174.
(44) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983,** *79* (2), 926-935.
(45) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** *78* (8), 1950-1958.
(46) Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *J Chem Phys* **2007,** *126* (1), 014101.
(47) Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984,** *81* (8), 3684-3690.
(48) Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **2010,** *31* (2), 455-461.
(49) Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Garmendia-Doval, A. B.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R. E.; Morley, S. D., rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput Biol* **2014,** *10* (4), e1003571.

(50) Cozzini, P.; Kellogg, G. E.; Spyrakis, F.; Abraham, D. J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L. A.; Morris, G. M.; Orozco, M.; Pertinhez, T. A.; Rizzi, M.; Sotriffer, C. A., Target flexibility: an emerging consideration in drug discovery and design. *J Med Chem* **2008,** *51* (20), 6237-6255.
(51) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **2009,** *30* (16), 2785-2791.
(52) Neudert, G.; Klebe, G., DSX: a knowledge-based scoring function for the assessment of protein-ligand complexes. *J Chem Inf Model* **2011,** *51* (10), 2731-2745.
(53) de Vries, S. J.; van Dijk, M.; Bonvin, A. M., The HADDOCK web server for data-driven biomolecular docking. *Nat Protoc* **2010,** *5* (5), 883-897.
(54) Branduardi, D.; Faraldo-Gomez, J. D., String method for calculation of minimum free-energy paths in Cartesian space in freely-tumbling systems. *J Chem Theory Comput* **2013,** *9* (9), 4140-4154.
(55) Yu, T. Q.; Lapelosa, M.; Vanden-Eijnden, E.; Abrams, C. F., Full kinetics of CO entry, internal diffusion, and exit in myoglobin from transition-path theory simulations. *J Am Chem Soc* **2015,** *137* (8), 3041-3050.
(56) De Vivo, M.; Cavalli, A., Recent advances in dynamic docking for drug discovery. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2017,** *7* (6), e1320-n/a.

Table of Contents