Holo-like and Druggable Protein Conformations from Enhanced Sampling of Binding Pocket Volume and Shape

Andrea Basciu,[†] Giuliano Malloci,[†] Fabio Pietrucci,[‡] Alexandre M. J. J. Bonvin,[§] and Attilio V. Vargiu*[†,§]

[†]Dipartimento di Fisica, Università di Cagliari, Cittadella Universitaria, I-09042 Monserrato (CA), Italy

[‡]Sorbonne Université, Muséum National d'Histoire Naturelle, UMR CNRS 7590, IRD, Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, IMPMC, F-75005 Paris, France

[§]Bijvoet Center for Biomolecular Research, Faculty of Science - Chemistry, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Ⓢ *Supporting Information*

ABSTRACT: Understanding molecular recognition of small molecules by proteins in atomistic detail is key for drug design. Molecular docking is a widely used computational method to mimic ligand−protein association in silico. However, predicting conformational changes occurring in proteins upon ligand binding is still a major challenge. Ensemble docking approaches address this issue by considering a set of different conformations of the protein obtained either experimentally or from computer simulations, e.g., molecular dynamics. However, holo structures prone to host (the correct) ligands are generally poorly sampled by standard molecular dynamics simulations of the apo protein. In order to address this limitation, we introduce a computational

approach based on metadynamics simulations called ensemble docking with enhanced sampling of pocket shape (EDES) that allows holo-like conformations of proteins to be generated by exploiting only their apo structures. This is achieved by defining a set of collective variables that effectively sample different shapes of the binding site, ultimately mimicking the steric effect due to the ligand. We assessed the method on three challenging proteins undergoing different extents of conformational changes upon ligand binding. In all cases our protocol generates a significant fraction of structures featuring a low RMSD from the experimental holo geometry. Moreover, ensemble docking calculations using those conformations yielded in all cases native-like poses among the top-ranked ones.
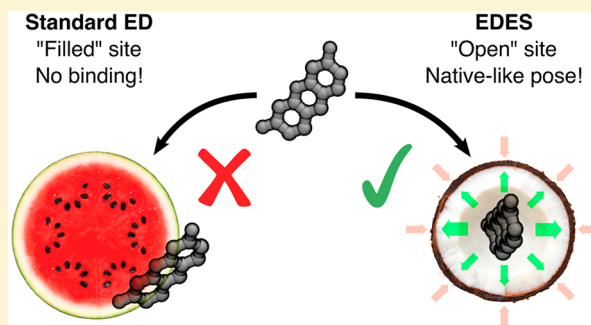
INTRODUCTION

Proteins are involved in virtually all cellular tasks, mediating physiological and pathological processes by establishing specific interactions with other biomolecules and small compounds. This feature is exploited in drug design by engineering small molecules that can interfere with pathogenic pathways. Modern drug design relies on a detailed understanding of molecular recognition processes by which biological partners such as a protein and a drug interact and bind to each other.[1−3] From a structural perspective, the rapid increase in the number of experimentally determined protein structures and the recent computational advances have fueled the development of computer-aided strategies for drug design.[4−7] In particular, protein−ligand docking[8−10] has become a well-established computational tool to complement experimental techniques such as X-ray crystallography, NMR spectroscopy, cryogenic electron microscopy, and related methods for characterizing protein−ligand complexes, often reducing the costs and improving the efficiency of high-throughput screenings.[3] However, like any computational or experimental technique, molecular docking also has its limitations and pitfalls. Dealing with the partners' flexibility is still a challenge because of the various levels of conformational changes that can occur in both the ligand and receptor upon binding.[1,3,11−17] In proteins these changes go from relatively small side-chain rearrangements to local distortions involving loops and/or confined secondary structure variations and even large-scale motions among (sub)domains (e.g., hinge-bending or shear motions).[2,18,19] Several classes of pharmaceutically relevant proteins such as kinases,[20] transferases,[21] synthases,[22] and dehydrogenases[23] undergo structural rearrangements leading to compaction upon ligand binding.[24,25]

In order to improve in silico structure-based drug design, it is crucial to account for these structural rearrangements (particularly those occurring at the binding site) when predicting drug binding and related thermodynamic and kinetic properties.[3,8,9,26] Several algorithms have been
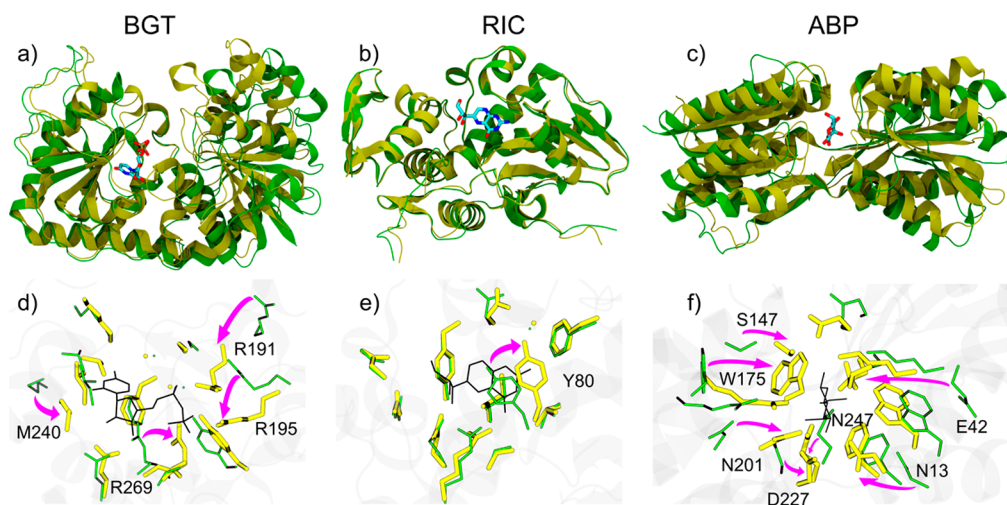
Figure 1. Comparison of the structural changes undergone by (a, d) BGT, (b, e) RIC, and (c, f) ABP upon binding of their ligands UDP, NEO, and ALL, respectively. (a−c) Overall conformational rearrangements of the proteins. The apo and holo proteins (PDB IDs 1JEJ and 1JG6, 1RTC and 1BR5, and 1GUD and 1RPJ for the BGT and BGT−UDP, RIC and RIC−NEO, and ABP and ABP−ALL systems, respectively) are shown in green and yellow ribbons, respectively, with the ligands in sticks colored by atom type. (d−f) Detailed views of the local rearrangements occurring at the binding site. The conformations of residues lining the binding site in the apo and holo forms of the proteins are shown with thin green and thick yellow sticks, respectively, while the ligands are shown with thin black sticks and the protein is shown in transparent gray ribbons. The most significant reorientations upon ligand binding are indicated by magenta arrows.

developed over the last decades to deal with protein flexibility in docking.[3,5,27,28] Among them, those termed soft docking, side-chain flexibility, and molecular relaxation still consider limited receptor flexibility either by allowing variations in the volume/shape of the binding site through rescaling of atomic interactions, by sampling a predetermined set of side-chain orientations, or by inducing minor backbone/side-chain rearrangements during structural optimization.[1,9,12,17,27−29] Unfortunately, these methods often fail in predicting protein−ligand complexes in the presence of medium to large conformational changes of the receptor upon ligand binding. Other approaches include those classified as collective degrees of freedom and ensemble docking, the latter constituting the background for our methodology.[4,5,11,12,28,30]

In a typical ensemble docking scenario, different conformations of a protein target, either interacting with different substrates or free of any ligand, are used to improve the prediction of the correct structure of the complex of interest. The method is founded on the conformational selection/population shift theory of molecular recognition, which states that proteins are able to assume drug-bound (hereafter "holo")-like conformations even in the absence of interacting ligands.[1,3,12,31] The ligand thus recognizes its target by "selecting" the most complementary conformation from an ensemble of metastable states, causing a population shift toward holo-like states (structures).

Ensemble docking has been shown to effectively enhance the performances of docking and virtual screening.[4,5,11,12,30] However, this improvement strongly depends on the ability to include in the pool of receptor structures some conformations similar to that of the true complex.[1,12,29,32−34] In particular, the inclusion of experimental structures of proteins bound to ligands similar to the one of interest has been shown to significantly increase the accuracy of docking.[5,8,12,32,35] Unfortunately, the number of targets whose three-dimensional (3D) structures have been experimentally solved remains limited compared with the druggable genome.[4,36] Furthermore, the exploration of different con-

formations in experimental structures is generally limited and biased toward (often just a few) known ligand−receptor complexes, which impacts the chemical diversity of putative lead compounds in virtual screening campaigns. Computational methods such as Monte Carlo (MC) and molecular dynamics (MD) simulations offer a relatively cheap and complementary way to sample receptor conformations.[4,5,11,12,17,37−41] The augmented conformational diversity sampled during MD simulations could in principle increase the percentage of false positives in virtual screening efforts. This issue is closely related to the limitations of current scoring functions.[3,8,9,13,42,43] Still, the significance of including MD-derived structures for discovering new active compounds has been demonstrated,[44,45] e.g., by the discovery of new (sub)pockets not yet identified by experiments.[46−50] In fact, it has been proposed that MD-derived structures could capture key interaction spots on the surface of receptors that are less biased toward one specific chemotype,[45] potentially leading to the discovery of previously unknown activities and/or mechanisms of action (binding modes) of existing drugs.[51]

In the ensemble docking framework, Lin et al.[52] introduced the concept of the relaxed complex scheme (RCS), in which a series of independent docking runs are performed from receptor conformations of the unbound (hereafter "apo") protein generated by MD simulations. Those are usually selected after clustering to capture the structural diversity of the target while keeping the number of conformers computationally tractable. Clearly, because of time scale restrictions, standard MD simulations are often unable to sample conformational states relevant to molecular recognition.[39,53] Several techniques have been proposed to enhance the sampling of rare conformations, including accelerated MD,[54] replica exchange in temperature and energy spaces,[55,56] and metadynamics,[57] which generalizes methods such as conformational flooding[58] and local elevation.[59] Several groups have demonstrated the effectiveness of these methods in improving the performance of docking and virtual screening.[60−67]
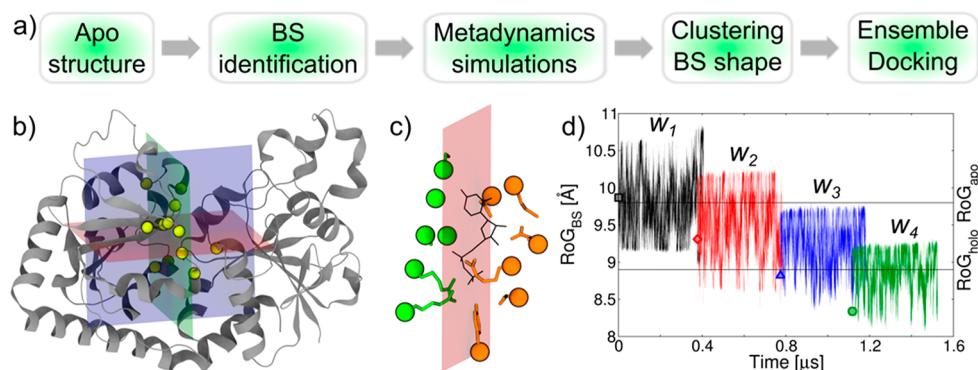
Figure 2. Overview of the EDES approach. (a) Workflow of the EDES protocol. (b) Representation of the "inertia planes" (transparent blue, red, and green) calculated at the binding site. The α-carbons of residues lining this site are shown as yellow spheres, and the protein is shown in gray ribbon. (c) Schematic view of the two groups of atoms (orange and green sticks with α-carbons as spheres) considered for the calculation of the number of contacts across one inertia plane; the ligand is also shown in black sticks. (d) Scheme of the "window" approach implemented to enhance in a controlled manner the sampling of conformations associated with different radius of gyration values of the binding site (RoG_{BS}) (the plot refers to simulations of the BGT system). The RoG_{BS} values corresponding to the initial conformation for each window are indicated by a square (w_1), diamond (w_2), triangle (w_3), and circle (w_4). The RoG_{BS} of the apo and holo experimental structures are indicated by horizontal lines.

To the best of our knowledge, no single method has yet been developed that outperforms standard ensemble docking on targets undergoing different kinds of structural distortions upon binding, unless some experimental knowledge of these conformational changes is exploited. In order to address this issue, we propose here a new approach called ensemble docking with enhanced sampling of pocket shape (EDES) that exploits relatively short metadynamics simulations of the apo protein of interest to generate a set of holo-like conformations for ensemble docking.[57,68,69] The key ingredients of our method are (i) the use of a novel set of collective variables to sample optimally and in a controlled manner different shapes of the binding site and (ii) a multistep clustering strategy allowing a large fraction of holo-like structures to be retained within the pool of cluster representatives. Notably, EDES does not exploit any a priori information about the holo structure of the protein. We assess the method on three targets that are representative of minor to very large conformational rearrangements upon ligand binding (Figures 1 and S1).

The first target is the T4 phage β-glucosyltransferase (hereafter BGT),[70] which undergoes a hinge-bending motion leading to a more closed form in its complex with uridine diphosphate (UDP) compared with the ligand-free structure (Figure 1a,d). This protein was included in the set of 10 targets selected by Seeliger and de Groot[66] to assess their workflow based on enhanced sampling using tCONCOORD[71,72] with the radius of gyration of the holo structure as a bias. While close-to-native ligand binding poses were obtained for eight out of 10 cases within the 100 top-ranked complex models, this was not the case for BGT, which makes this protein a well-suited test case for our method.

The second target is recombinant ricin (hereafter RIC),[73] representative of proteins undergoing minor but subtle conformational changes upon binding of the ligand (in this case neopterin, NEO)[74] (Figure 1b,e). RIC belongs to the Astex Diverse Data Set,[75] recently used to validate the AutoDockFR docking software, which models receptor flexibility by explicitly specifying a set of flexible side chains.[76] In cross-docking experiments using the apo conformations of the receptors, AutoDockFR outperformed AutoDock Vina[77] in terms of both number of correct poses and their ranking. However, none of the aforementioned software was able to

find any solution within a ligand root-mean-square deviation (RMSD) of 2.5 Å from the experimental structure of the complex.

The third target is the allose binding protein (ABP) (Figure 1c,f), which is also representative of targets undergoing hinge-bending motions upon binding of their ligands, in this case D-allose (ALL).[78] The conformational changes for ABP are larger than those occurring in BGT. This protein therefore represents another good test case for our approach. Motta and Bonati[63] used ABP−ALL in ensemble docking calculations performed with Glide[79,80] from conformations generated through accelerated MD simulations.[54]

In the following we demonstrate that for all of the targets considered here, EDES is able to generate native-like structures of the complexes. Using the widespread and freely available docking programs HADDOCK[81] and AutoDock4,[82] which differ in their search and scoring strategies, we identify native-like docking poses among the top-ranked ones. While being a proof of concept, this work opens the way to the automatic generation of holo-like conformations for a broad range of protein targets and as such contributes to improving in silico structure-based drug design.

■ RESULTS AND DISCUSSION

Method Workflow. Our protocol workflow is sketched in Figure 2a. First, we identify the putative binding sites on the target proteins. For the purpose of validating our methodology, these sites were defined from the structures of the BGT−UDP, RIC−NEO, and ABP−ALL bound complexes (see Figure S1g for the list of protein residues defining these sites). In real cases, binding sites could be identified using publicly available site detection packages or Web servers. For example, for the targets investigated here, there is a good agreement between the experimental binding sites and those identified by the COACH-D Web server (vide infra).[83] We then calculate the "inertia planes" at the binding site, that is, the planes orthogonal to the corresponding inertia axes and passing through the center of mass of the site (Figure 2b). Then we perform relatively short bias-exchange, well-tempered metadynamics simulations[57,68,69] of the apo protein (see Materials and Methods for details) using a set of four collective variables (CVs): three (pseudo)contacts across inertia plane (CIP)
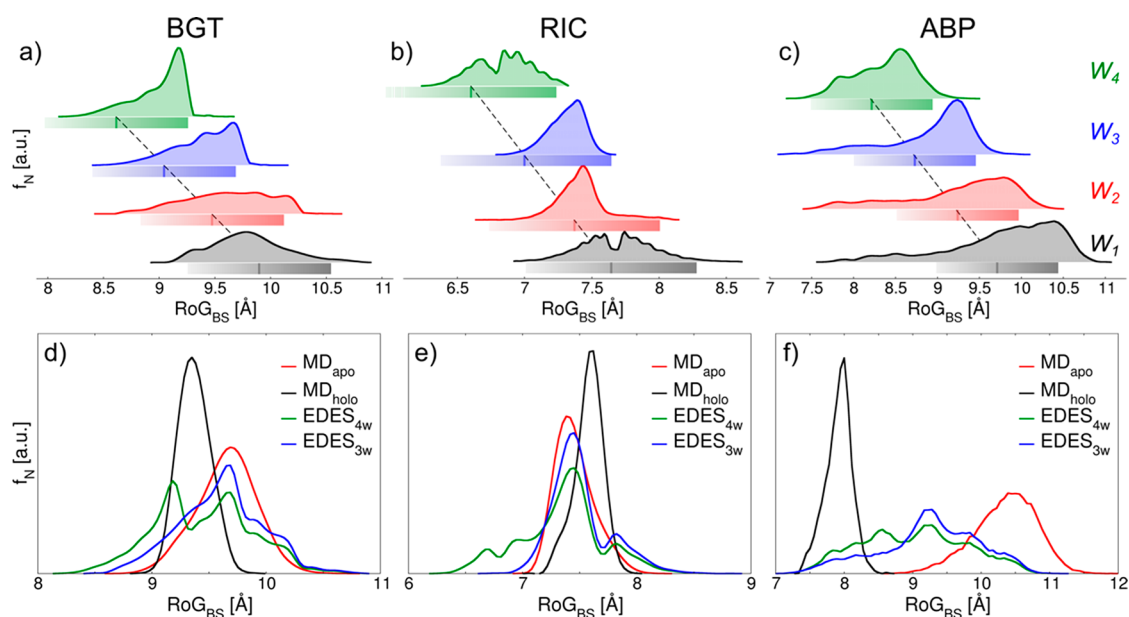
Figure 3. Distributions of RoG$_{BS}$ values. (a–c) Distributions from each EDES window for (a) BGT, (b) RIC, and (c) ABP. The colored bar below each distribution indicates the position of the lower and upper walls set for RoG$_{BS}$ in that window, and the color gradient indicates a higher force constant for the upper wall than the lower wall (the centers of the windows are indicated by darker lines within the bars and are connected by a black dashed line). (d–f) Comparison of RoG$_{BS}$ normalized distributions (area under each curve equal to 1, bin size set to 0.1 Å) obtained from the different simulations performed in this work.
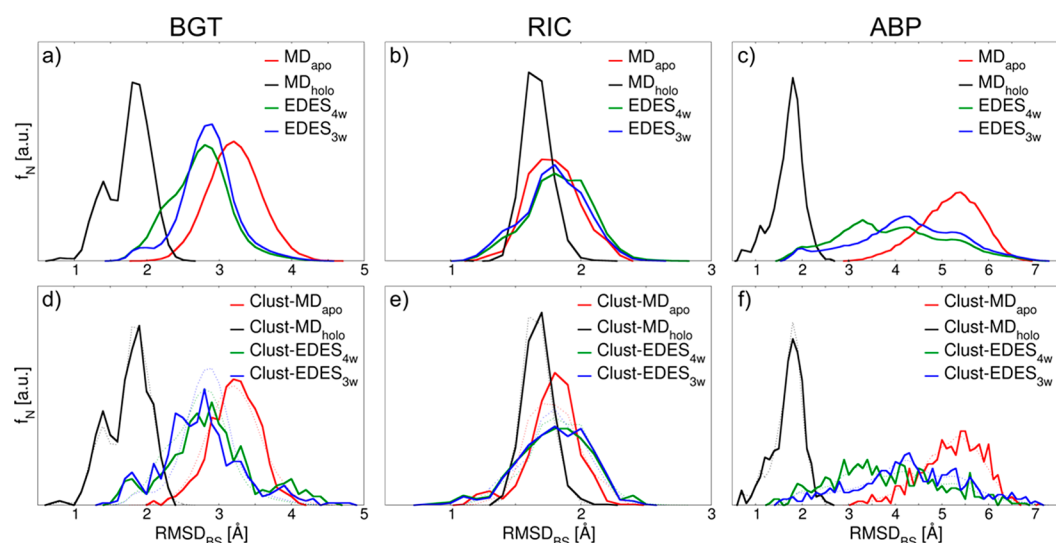
Figure 4. Normalized distributions (area under each curve equal to 1, bin size set to 0.1 Å) of the RMSD of the binding site heavy atoms (RMSD$_{BS}$) with respect to the holo structure. The upper (a–c) and lower (d–f) rows show the distributions calculated over all of the snapshots extracted from each MD simulation and over cluster representatives only, respectively. The faint dotted lines in (d–f) correspond to the distributions in (a–c).

variables, each defined as the number of contacts between residues of the binding site on opposite sides of the corresponding inertia plane (Figure 2c), and the gyration radius of the binding site (RoG$_{BS}$). We also use the last of these CVs to implement a "windows" approach (Figure 2d) aimed at sampling more effectively and in a controlled manner different shapes of the binding site (possibly mimicking conformational changes induced by ligand binding). Namely, we apply soft walls at RoG$_{BS}$ values that are 7.5% higher and lower than the value measured in the apo X-ray structure (RoG$_{X-ray}^{apo}$, corresponding to the center of window 1). Next, from the trajectory of this first window, we randomly select a

conformation of the protein whose RoG$_{BS}$ is 5% lower to initiate another MD simulation (corresponding to window 2) with walls centered at ±7.5% RoG$_{X-ray}^{apo}$ from this new center. We repeat this procedure to generate up to four windows including the first one. This leads to an overall reduction of RoG$_{BS}$ of 15% relative to the center of the first window RoG$_{X-ray}^{apo}$ (see Figure 3 and Table S1). Despite the arbitrariness of our choice, the performance of EDES is not very sensitive to the exact choice of three or four windows (and thus to the exact extent of the collapse induced at the binding site, amounting to 10% or 15% of the initial value, respectively). Moreover, although the imposed ΔRoG$_{BS}$ correspond only to

Table 1. Performance of Various MD Simulations in Reproducing Native-like Conformations of the Binding Sites of BGT, RIC, and ABP[a]

| protein | simulation | $RMSD_{BS} < 1.5$ Å [%] | | $RMSD_{BS} < 2$ Å [%] | |
		trajectory	clusters	trajectory	clusters
BGT	MD_{apo}	–	–	0.06 (1.69)	–
	MD_{holo}	23.9 (0.75)	26.6 (0.75)	85.3 (0.75)	84.8 (0.75)
	$EDES_{4w}$	0.02 (1.31)	0.2 (1.31)	5.0 (1.31)	8.4 (1.31)
	$EDES_{3w}$	0.02 (1.31)	0.4 (1.31)	3.8 (1.31)	4.0 (1.31)
RIC	MD_{apo}	9.7 (1.00)	9.6 (1.10)	89.4 (1.00)	93.2 (1.10)
	MD_{holo}	13.0 (0.91)	12.4 (0.95)	99.6 (0.91)	97.8 (0.95)
	$EDES_{4w}$	13.0 (0.77)	19.0 (0.81)	81.4 (0.77)	82.4 (0.81)
	$EDES_{3w}$	16.5 (0.77)	17.6 (0.81)	85.5 (0.77)	85.4 (0.81)
ABP	MD_{apo}	–	–	–	–
	MD_{holo}	25.8 (0.48)	25.6 (0.67)	96.4 (0.48)	88.6 (0.67)
	$EDES_{4w}$	0.6 (1.17)	1.4 (1.20)	8.6 (1.17)	9.4 (1.20)
	$EDES_{3w}$	0.5 (1.17)	2.4 (1.20)	6.1 (1.17)	8.4 (1.20)

[a]Performance is measured by the percentage of conformations with $RSMD_{BS} < 1.5$ or 2 Å with respect to the experimental structure. The headings "trajectory" and "clusters" refer to snapshots extracted from the full trajectories and to cluster representatives, respectively. The lowest value of $RMSD_{BS}$ (in Å) is reported in parentheses.
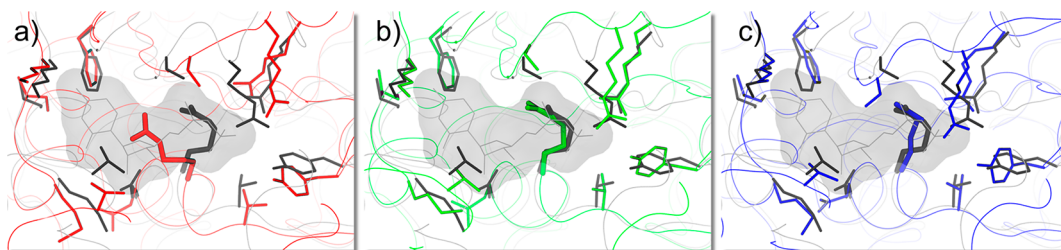
Figure 5. Binding site views of the lowest-RMSD conformations of BGT with respect to the bound complex extracted from (a) MDapo, (b) $EDES_{4w}$, and (c) $EDES_{3w}$. UDP is shown in gray lines and transparent surface. The proteins are shown as gray (holo experimental structure), red (MD_{apo}), dark green ($EDES_{4w}$), and blue ($EDES_{3w}$) thin ribbons, with side chains of residues lining the binding site represented as sticks, which are thicker for R269.

the true change seen for BGT (Figure S1g), EDES performed comparably well for all of the systems investigated here. In the following we refer to the cumulatives set of three- and four-window simulations as $EDES_{3w}$ and $EDES_{4w}$, respectively.

Sampling of Holo-like Structures. We first discuss the performance of our method in generating holo-like structures, starting with BGT. The binding of UDP induces large rearrangements at the binding site of this protein, particularly in the orientation of the three arginines (R191, R195, and R269) neutralizing the negative charge of the diphosphate group (Figure 1a,d). We first compared the performance of standard MD simulations of the apo (MD_{apo}) and holo (MD_{holo}) systems to that of EDES (simulation details are reported in Materials and Methods) using as a metric the RMSD of the binding site heavy atoms ($RMSD_{BS}$) from the geometry assumed in the holo structure. Figure 4 reveals a very poor overlap between the MD_{apo} and MD_{holo} distributions. The EDES distributions are centered somewhat in between the ones obtained from the unbiased MD simulations. In addition, most conformations sampled by EDES have $RMSD_{BS}$ lower than 2.8 Å from the experimental structure of the complex (in between the values for the apo and holo structures; see Figure S1g). The EDES distributions reveal a shoulder at lower RMSDs that increases the percentage of conformations with $RMSD_{BS} < 2$ Å compared with MD_{apo}, a feature that persists after clustering (Figure 4).

Table 1 reports the percentages of structures with low $RMSD_{BS}$ from the holo structure. As expected, this percentage

is high for MD_{holo}. While a very low number of such conformations was sampled in MD_{apo}, a large fraction was obtained by EDES using either three or four windows.

Our protocol was thus able to generate collapsed conformations of the binding site in the absence of a ligand triggering such a collapse (Figure 3 and Figure S2). This is particularly evident for R269 in the center of the binding pocket: while in MD_{apo} the side chain remains in the center of the binding site, it is displaced in the EDES simulations, making room for ligand binding (Figure 5). Our multistep cluster analysis effectively increased the percentage of structures featuring a native-like geometry of the binding site compared with the fraction sampled during MD simulations (Table 1). The enhanced sampling of holo-like conformations by EDES is also evident from the CIP metric, as shown by the improved overlap between the MD_{holo} and EDES distributions with respect to MD_{apo} (Figure 6). In particular, only EDES samples conformations with CIP values virtually identical to those of the experimental holo structure (black sphere in Figure 6). As our methodology relies on the accurate identification of the binding site(s) to enhance the conformational sampling of its shape and volume, we also investigated its performance using a slightly different definition of the binding site. Namely, we took advantage of one of the many available site-finding Web servers, COACH-D,[83] to determine consensus binding sites on BGT. Among the top sites identified by COACH-D, the top one shares 13 out of 16 residues with the experimental holo structure (Table S2).
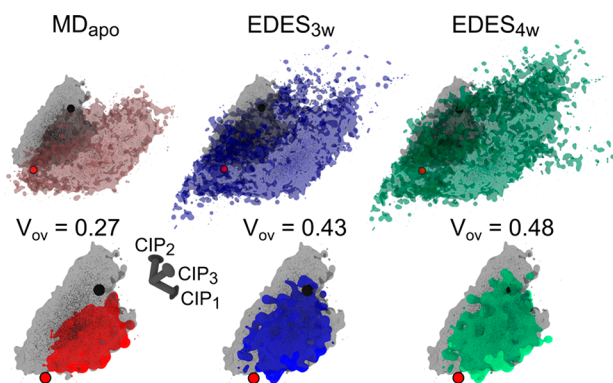
Figure 6. Sampling of the 3D space defined by CIP_1, CIP_2, and CIP_3 during the MD simulations of BGT and BGT−UDP. Top row: Comparison of the MD_{apo} (red), $EDES_{3w}$ (blue), and $EDES_{4w}$ (green) distributions with the MD_{holo} distribution (dark gray). The distributions are shown both as solid points and as transparent surfaces. The locations of the apo and holo structures are indicated by red and black spheres, respectively. Bottom row: Envelopes of the MD_{apo} (red), $EDES_{3w}$ (blue), and $EDES_{4w}$ (green) distributions overlapping with the MD_{holo} distribution (shown in dark gray as a reference). Also reported are the volumes of the overlapping distribution, V_{ov} (estimated with Voss Volume Voxelator (http://3vee.molmovdb.org) using a probe radius of 3 Å).

Furthermore, the overlap between the experimentally derived sites and those identified by COACH-D is very high for all three targets. We repeated all of the EDES simulations using the alternative definition of the binding site obtained from COACH-D. This gave virtually identical results (Table S3), indicating that our method is not sensitive to the exact binding site definition and can be used in conjunction with site detection algorithms in cases where only an apo structure is known.

Despite being primarily devised for targets undergoing rather large conformational changes, we decided to test our method also with RIC as paradigm of a protein that undergoes minor conformational changes upon ligand binding. This is important because in several interesting applications the information regarding the extent of apo-to-holo conformational rearrangement(s) it is unknown. It is therefore important to validate our method on proteins that undergo only limited conformational changes upon binding. Furthermore, previous work has already pointed out how the performance of enhanced-sampling versus standard MD

simulations in reproducing a relevant fraction of holo-like conformations is target-dependent.[63]

The subtle conformational changes of RIC upon binding of NEO (Figure 1b,e) were hardly handled by algorithms exploiting the flexibility of the binding site through side-chain torsional angles.[76] RIC has proven to be a very difficult target for both rigid and flexible docking calculations starting from the apo X-ray structures using both AutoDock Vina[77] and the recently introduced AutoDockFR[76] software (see Table 1 in ref 76). In contrast, both metadynamics and standard MD simulations were able to reproduce the geometry assumed by the binding site in the holo structure (although EDES did find conformations closer to the holo structure than those obtained from $MD_{apo/holo}$; see Table 1 and Figure 4). As expected, no clear difference was observed between EDES and standard MD in reproducing holo-like conformations of the entire protein (Figure S3), as both approaches sampled a relatively large fraction of such structures. On the basis of these results, we are confident that our approach should also be able to effectively generate holo-like structures of targets undergoing minor conformational changes upon ligand binding. As stated above, this is particularly encouraging since in a real case one might not know the extent of the conformational change in advance.

The last protein considered in this work is ABP (Figure 1c,f), which undergoes the largest conformational changes among the three targets upon binding of the neutral small molecule ALL. The difference between the performances of metadynamics and standard MD simulations in reproducing holo-like geometries of the binding site becomes evident in this case (Figures 3 and 4), as the latter approach is unable to produce any single conformation of the protein featuring an $RMSD_{BS}$ lower than 2 Å from the holo X-ray structure (Table 1). In contrast, our approach, despite enhancing the sampling of the binding site only, is able to drag the whole protein structure toward conformations close to that found in the protein−ligand complex also for this target (Figure S3).

Docking Performance. In this subsection we describe the performance of each set of structural clusters in ensemble docking calculations. The results presented for AutoDock4 were obtained after performing a cluster analysis on the top poses obtained from individual docking runs (500 for each ensemble of structures), while in HADDOCK all of the 500 conformations within each ensemble were used in a single docking run (see the captions of Tables 2 and 3 and Materials and Methods for details).

Table 2. Performance of AutoDock4 in Reproducing the Experimental Structures of the BGT−UDP, RIC−NEO, and ABP−ALL Complexes in Ensemble Docking Calculations[a]

	BGT−UDP				RIC−NEO				ABP−ALL			
	MD_{apo}	MD_{holo}	$EDES_{3w}$	$EDES_{4w}$	MD_{apo}	MD_{holo}	$EDES_{3w}$	$EDES_{4w}$	MD_{apo}	MD_{holo}	$EDES_{3w}$	$EDES_{4w}$
sampl. perf. [%]	−	84.6	1.8	2.0	3.8	10.8	2.6	3.4	−	94.0	3.0	3.2
pose rank	−	1 (1)	1 (1)	2 (2)	1 (1)	1 (1)	4 (3)	1 (1)	−	1 (1)	2 (2)	2 (7)
clus. pop.	−	48	4	4	15	20	10	13	−	225	9	5 (6)
$RMSD_{lig}$ [Å]	−	1.2 (0.7)	0.6 (0.9)	1.5 (1.2)	1.0 (0.9)	0.6 (0.6)	0.7 (0.7)	0.9 (0.8)	−	0.7 (0.2)	0.8 (0.3)	0.7 (0.2)

[a]Results refer to clusters of docking poses obtained from a cluster analysis performed on all generated complexes (500 for each ensemble of clusters of receptor structures, corresponding to the top pose from each independent docking run for that ensemble) using as the metric the distance RMSD (dRMSD) with a cutoff of 1.5 Å. The sampling performance is calculated as the percentage of poses within 2 Å from the native structure out of the top poses considered for each ensemble of receptor structures. The pose rank refers to the ranking of the first native-like pose obtained using the highest score within each cluster as sorting criterion. In parentheses, the rank of the same cluster is reported when the average score over the top three poses is used instead. The cluster population refers to the population of the corresponding cluster in the same column. The $RMSD_{lig}$ refers to the average heavy-atom RMSD of the ligand calculated for the top cluster with the standard deviation in parentheses.

Table 3. Performance of HADDOCK in Reproducing the Experimental Structures of the BGT−UDP, RIC−NEO, and ABP−ALL Complexes in Ensemble Docking Calculations[a]

	BGT−UDP				RIC−NEO				ABP−ALL			
	MD_{apo}	MD_{holo}	$EDES_{3w}$	$EDES_{4w}$	MD_{apo}	MD_{holo}	$EDES_{3w}$	$EDES_{4w}$	MD_{apo}	MD_{holo}	$EDES_{3w}$	$EDES_{4w}$
sampl. perf. [%]	1.8	37.0	14.0	8.0	23.0	45.5	15.8	20.5	2.5	49.5	21.3	13.8
pose rank	7 (6)	1 (1)	2 (3)	1 (6)	1 (1)	2 (1)	1 (1)	1 (1)	91 (346)[b]	1 (2)	1 (1)	1 (5)
clus. pop.	6	9	33	20	80	152	53	69	–	188	20	5
F_{nat}	0.72 ± 0.04 (0.75/4)	0.81 ± 0.07 (0.71/3)	0.80 ± 0.02 (0.79/1)	0.72 ± 0.04 (0.71/2)	0.93 ± 0.00 (0.93/1)	0.90 ± 0.03 (0.87/4)	0.90 ± 0.06 (1.00/3)	0.83 ± 0.06 (0.73/3)	0.47 (0.73)[b]	0.85 ± 0.06 (0.93/3)	0.80 ± 0.00 (0.80/1)	0.72 ± 0.03 (0.73/2)
$RMSD_{lig}$ [Å]	2.0 ± 0.3 (1.7/4)	0.7 ± 0.2 (0.54/3)	0.8 ± 0.2 (0.50/1)	1.3 ± 0.4 (0.87/3)	0.9 ± 0.3 (0.57/1)	1.0 ± 0.3 (0.67/4)	0.9 ± 0.3 (0.34/3)	0.9 ± 0.2 (0.62/3)	1.9 (1.5)[b]	0.9 ± 0.2 (0.59/3)	0.8 ± 0.1 (0.74/1)	2.0 ± 0.1 (1.98/2)[c]

[a]Results correspond to the statistics of the top pose in clusters obtained using ligand interface RMSD clustering with a 1 Å cutoff and a minimum of four poses per cluster. The cluster rankings are based on the average score of the top four poses (the ranking based on the score of the top pose is reported in parentheses). The first values in the last two rows refer to the average statistics of the top four poses of a cluster in the semiflexible refinement (it1 step) of HADDOCK, while values in parentheses refer to the statistics/rank of the best (smallest RMSD with respect to the reference) pose among the top four. The sampling performance was calculated as the percentage of poses within 2 Å from the experimental structure out of the 400 generated models (one docking run was performed from the ensemble of 500 MD conformations). F_{nat} indicates the fraction of native contacts recovered within a shell of 5 Å from the ligand in the experimental structures. [b]Since no acceptable clusters were obtained in this case, the reported statistics correspond to single-pose statistics for the first acceptable pose (<2 Å) and best pose (in parentheses). [c]It should be noted that a cluster was found ranking 10th with $RMSD_{lig}$ amounting to 0.86 ± 0.12 and scoring as the top pose within standard deviations (−26.8 ± 0.4 vs −25.3 ± 1.1 a.u.).

For BGT−UDP, both AutoDock4 and HADDOCK achieved improved sampling performance (defined as the percentage of docking poses displaying a value of $RMSD_{lig}$ lower than 2 Å from the holo structure) when coupled to EDES rather than MD_{apo} (Tables 2 and 3), generating consistent fractions (up to 2% and 14% for AutoDock4 and HADDOCK, respectively) of native-like ligand poses. This is much better than using clusters derived from the much longer MD_{apo} simulation (0% for AutoDock4 or 2% for HADDOCK).

Importantly, both programs were able to rank at least one native-like pose among the top two when coupled with EDES, independent of the number of windows used to generate the conformational clusters (see Tables 2 and 3 and Figure 7).

The results obtained for ABP are also very encouraging. In this case the performance of HADDOCK is overall better with $EDES_{3w}$. The top HADDOCK pose obtained with $EDES_{4w}$, while satisfying the $RMSD_{lig} \leq 2$ Å criterion, has a flipped orientation (pointing to the known limitation of using the RMSD criterion alone to evaluate the performance of docking). The 10th cluster instead virtually overlaps with the experimental one ($RMSD_{lig} = 0.9 \pm 0.1$ Å).

It should be noted that for BGT and RIC the results were virtually independent of the number of EDES windows used. Clearly, further studies will be needed to optimize the number and the width of the windows used, possibly exploiting a set of intrinsic properties of each protein so as to set up target-dependent rules.

Druggability Assessment. We used the software f-pocket[84] to assess the druggability of the binding site within the ensembles of conformations generated by EDES (see Materials and Methods). For each conformation, we evaluated the druggability score D,[85] which ranges between 0 and 1 with higher values identifying more druggable geometries. It is customary to associate scores >0.5 to putative binding sites.[85] Table 4 shows that EDES generated a much larger set of druggable structures than MD_{apo} for BGT and ABP (actually, no druggable conformation was generated from this trajectory for the latter system), while the performance of the two sets was similar for RIC, as expected. In particular, (i) the EDES-derived ensembles have a higher percentage of structures associated with $D > 0.5$ than those derived from MD_{apo}, and (ii) the percentage of structures with $D > 0.9$ is not much lower than that obtained from the MD_{holo} set for BGT and RIC.

Impact of the Clustering Strategy. As highlighted by others, no enhanced sampling algorithm would be useful in ensemble docking calculations without an effective clustering strategy to limit the number of conformations for docking while selecting druggable/holo-like structures.[11,32,63,65,86] This is of particular relevance in virtual screening experiments, where the optimal choice of the number of structures building the ensemble is crucial. To assess the sensitivity of our protocol to the number of selected clusters for docking, we performed additional cluster analysis in which we decreased the total number of clusters from 500 to 200 or 100. Using the resulting sets of conformations, we performed additional docking calculations with AutoDock4. The results reveal that (i) the fraction of conformations displaying $RMSD_{BS} \leq 2$ Å is virtually constant when going from 500 to 100 clusters (Table S4) and (ii) the docking performance is also comparable when different numbers of starting conformations are used (Table S5).

Another relevant aspect is the metric used for clustering. Indeed, it has been shown that strategies based on a 3D
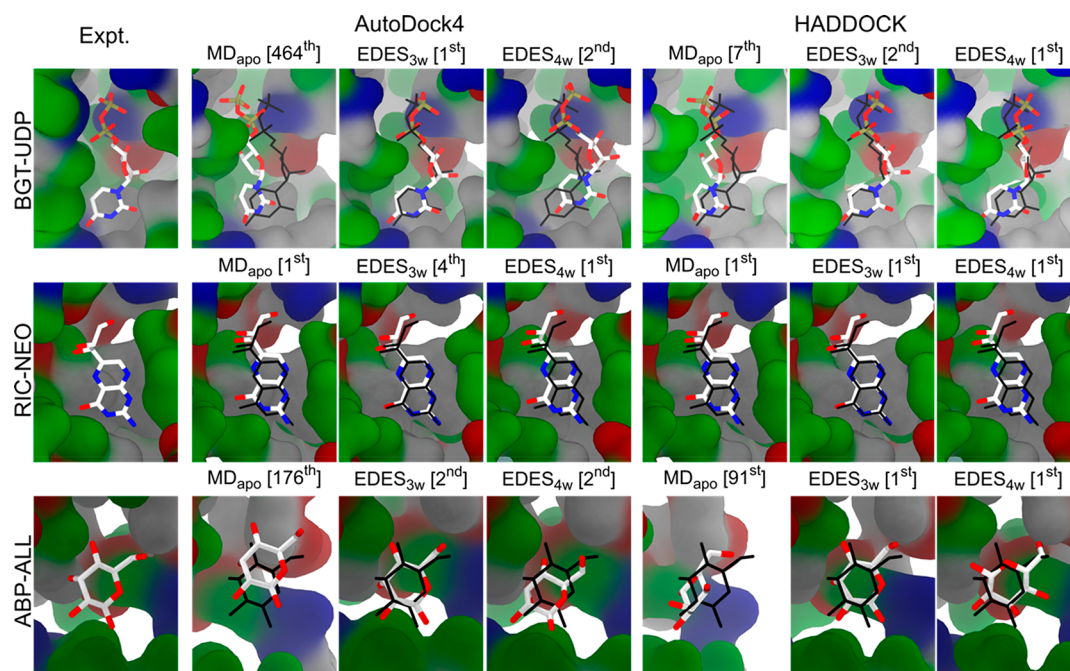
Figure 7. Docking performances of various structural ensembles in reproducing the experimental poses of BGT−UDP (top row), RIC−NEO (middle row), and ABP−ALL (bottom row). Close views of the experimental binding modes are shown in the first column, while the top-score poses within the first cluster with $RMSD_{lig} < 2$ Å, or poses with the lowest $RMSD_{lig}$ value when no native-like pose was found, are reported in the next columns, with the corresponding ranks given in square brackets. The docking was performed both using AutoDock4 and HADDOCK for comparison. In each panel, the molecular surface of the backbone and that of the C_α atoms of the protein are colored by residue type as in Figure 1, and the ligand is shown as sticks colored by atom type. In columns 2 to 7, the experimental pose is shown in black thin sticks for easy comparison.

Table 4. Performance of Various MD Simulations in Generating Druggable Conformations of the Binding Site[a]

protein	simulation	% of cluster structures with D greater than:				
		0.5	0.6	0.7	0.8	0.9
BGT	MD_{apo}	8.8	4.8	3.2	1.6	0.4
	MD_{holo}	40.0	28.6	18.0	10.6	2.4
	$EDES_{4w}$	14.2	10.8	7.2	3.6	1.8
	$EDES_{3w}$	15.8	10.2	6.8	4.6	1.4
RIC	MD_{apo}	2.6	2.2	1.2	0.4	0.2
	MD_{holo}	14.2	8.8	6.0	2.4	0.4
	$EDES_{4w}$	3.4	2.6	1.8	0.4	0.2
	$EDES_{3w}$	3.2	2.4	1.6	0.4	0.2
ABP	MD_{apo}	−	−	−	−	−
	MD_{holo}	20.4	13.0	7.2	3.6	1.6
	$EDES_{4w}$	7.4	5.4	2.6	1.2	0.2
	$EDES_{3w}$	7.6	5.2	3.0	1.8	0.4

[a]For each protein and each simulation, the percentages of structures (over the 500 cluster representatives) featuring druggability scores D larger than 0.5 to 0.9 are reported in columns 3 to 7, respectively.

descriptor linked to the pocket shape appear to be more effective in delivering maximally different conformations of the binding site.[63,65] Using BGT−UDP we compared the conformational sampling and the docking performance (evaluated with AutoDock4) of our multistep multidimensional clustering of the CVs used in metadynamics simulations (CA_{CVs}) with that of a more standard analysis based on the distance RMSD (dRMSD) of the binding site (CA_{dRMSD}). We found that the performance of our approach is superior in terms of both sampling holo-like conformations and selecting low $RMSD_{BS}$ values compared with CA_{dRMSD} (Table S6). The impact of the clustering strategy is even more evident for the

docking results with no near-native pose within the top 10 when using structures obtained from CA_{dRMSD} clustering (Table S7).

Comparison with Previous Work on the Same Targets. Seeliger and de Groot[66] used tCONCOORD[71,72] with a radius of gyration bias to enhance the sampling of holo receptor conformations on a set of 10 proteins including BGT. In nine out of 10 cases, the best models displayed an RMSD from the holo structure smaller than 2 Å at the binding site (defined as all residues within 6 Å from the ligand). The best model had an $RMSD_{BS}$ of 1.78 Å, which is higher than the lowest values obtained with our protocol (using the same definition of the binding site as in ref 66), namely, 1.51 and 1.40 Å for the clusters and the MD-derived distributions in both $EDES_{4w}$ and $EDES_{3w}$. They performed ensemble docking calculations with AutoDock Vina[77] using 5000 protein structures generated by tCONCOORD followed by a series of postdocking optimizations, filtering of models against the experimental gyration radius, further docking calculations, and finally rescoring with RosettaLigand.[87] As a result, in eight out of 10 cases, native-like ligand poses (defined there as those for which $RMSD_{lig} < 3$ Å from the experimental structure) were generated among the top 100 ones, demonstrating the general applicability of the method for blind predictions of protein−ligand complexes involving large conformational rearrangements. However, no native-like pose was found within the top 100 for BGT, and just one pose had $RMSD_{lig} < 2$ Å.

In order to understand more deeply the reasons behind the good performance of our method for BGT, we calculated the correlation between $RMSD_{lig}$ and the RMSDs of the residues within the binding site, including the arginine triad (R191, R195, and R269) and in particular R269, which in the MD_{apo} simulation often occupies the center of the pocket (Figure 5).

Despite the fact that the overall correlation between $RMSD_{UDP}$ and the RMSD of the binding site is not necessarily high, reproducing the correct orientation of the arginine triad and in particular R269 is crucial to obtain native-like poses (Figure S4). These results are in line with previous reports (see, e.g., the discussion in ref 63) pointing to the importance of reproducing, in the near-native backbone geometry, the correct orientation of the few side chains involved in key interactions with ligands rather than fairly matching the overall binding site structure. By enhancing the fluctuations in the number of contacts among two relatively small groups of atoms across three orthogonal planes (see, e.g., Figures 6 and S2), our method effectively forces the binding site to assume different volumes and shapes, thus increasing the probability of sampling native-like conformations. This should be particularly effective when dealing with long/charged side chains lining the binding site (such as R269) within an aqueous environment that favors an extended conformation because of enhanced hydration.

For RIC, as previously discussed in ref 76, neither AutoDock Vina nor AutoDockFR was able to find native-like poses. RIC proved to be a very difficult target for both rigid and flexible docking (see Table 1 in ref 76). In contrast, our protocol reproduced a high fraction of holo-like structures, which is not trivial, as recently discussed in ref 63, where the advantage of using an enhanced sampling protocol (i.e., accelerated MD[54]) versus conventional MD simulations was reported to depend on the target, in particular on the extent of conformational changes at the binding site, and on the binding specificity. Our findings for RIC are thus very encouraging considering the difference of only 0.1 Å between the RoG_{BS} of the apo and holo experimental structures (Figure 1) and the relatively large fluctuations induced by biasing protocols versus those induced by standard MD simulations (Figure 3).[41,63]

Finally, ABP was one of the two targets used in ref 63 to investigate the performance of accelerated[54] versus conventional MD simulations in reproducing near-native conformations of the ABP−ALL complex. Accelerated MD simulations (~2.5 μs in cumulative simulation time) were performed with both the single- and double-boost approaches. As expected, only the biased simulations were able to sample closed (holo-like) conformations of the binding site (defined by any residue with at least one atom within 5 Å from the ligand), reaching values of the $RMSD_{BS}$ as low as 1.42 Å. Using the same definition of the binding site, we obtained $RMSD_{BS}$ values as low as 1.20 Å for $EDES_{4w}$ and $EDES_{3w}$, with 8.5% and 9.3% of structures, respectively, having $RMSD_{BS} < 2$ Å (see Table 1). In ref 63 the docking was performed with the package Glide[79,80] on representatives of the 10 (out of the 50 extracted) most populated clusters obtained with a volumetric approach developed by the authors in order to collect structures with maximal differences in the accessible space within the binding site. Among the top 10 clusters of poses, the sixth was near-native and the only one displaying structural stability in further refinement via 50 ns-long standard MD simulations in explicit solvent. These results compare well with our findings in Tables 2 and 3.

Related Approaches. As stated in the Introduction, methods based on rescaling (scaling down) of protein−ligand interactions (originating from the soft docking approach[28,88]) generally improve docking outcomes when binding is associated with minor backbone rearrangements. Their applicability has been extended by coupling them with side-chain optimization/repacking using a variety of energy minimization, MC, or MD simulations.[28] Several publications[41,64,89] have reported interesting developments exploiting these ideas to bias sampling toward druggable conformations without relying on the identification of the putative binding site(s). Johnson and Karanicolas[41,89] implemented within the Rosetta software suite[90] an approach to enhance the sampling of protein conformations suitable for small-molecule binding by biasing the fluctuations of the volume of surface pockets. The conformations generated by their approach resemble known inhibitor-bound structures more closely than equivalent ensembles of unbiased conformations. Oleinikovas et al.[64] developed a method based on Hamiltonian replica-exchange cosolvent MD simulations that effectively enhance the discovery of cryptic pockets for different protein targets. By scaling of the nonbonded interactions between apolar atoms of the protein and water molecules, the properties of the latter shift toward ligand-like behavior, increasing cryptic site opening. These methodologies are similar to EDES in scope, although their application was focused rather on enhancing the discovery (opening) of druggable (including cryptic) sites. Our approach instead was primarily devised to enhance the sampling of holo-like conformations, although it proved to be promising also in increasing the number of druggable conformations sampled in MD simulations (Table 4). Another class of methods primarily developed to generate bound-like protein conformations is the one based on elastic networks, whereby protein flexibility is incorporated through local/global modes, exploiting in some cases additional biases.[17,61,91−94] These approaches have the advantage of being computationally cheap. Some were able to reproduce very large conformational changes associated with protein−ligand binding, up to the nanometer range. To give an example, Doruker and co-workers recently developed a protocol that, starting from apo structures ~4 to 15 Å (backbone RMSD) away from their holo counterparts, reproduced the holo conformations within ~1.5 to ~4 Å.[92,93] They also assessed the performance of these ensembles of structures (filtered to discard conformers with gyration radii larger than that of the apo protein) in docking, reporting near-native poses for three out of five targets. The two most challenging systems were those featuring the largest conformational changes upon ligand binding.

■ CONCLUSIONS AND PERSPECTIVES

We have presented a proof-of-concept study of a novel protocol allowing the generation of holo-like and druggable conformations of proteins and improving the performance of ensemble docking. Its robustness and general applicability were tested using two different docking programs against three challenging protein targets undergoing different extents of conformational changes upon ligand binding. The key points are (i) the use of gentle, adaptive biases on a carefully designed new set of CVs, enabling the generation of maximally diverse conformations of the binding site, including a relevant fraction of holo-like/druggable ones, and (ii) a multistep cluster analysis performed on the CVs and able to generate a tractable number of conformations while maintaining or even increasing (compared with the distributions from the MD simulations) the fraction of holo-like protein structures.

Looking at future developments, a straightforward way to further improve the conformational sampling of the binding site could be to couple our algorithm with cosolvent simulations,[67,95] as was done, e.g., in ref 64. Furthermore,

our method could be combined with others that enhance the sampling of orthogonal degrees of freedom, such as global protein motions,[17,61,92,93,96] rotations around torsional angles,[62,63] secondary structure changes,[97,98] or rescaled protein–ligand interactions,[55,64] just to cite a few options. In addition, experimental information from various sources could be encoded in new CVs and/or restraints. Furthermore, the number of cluster representatives used in docking calculations could be further lowered, e.g., by applying filters to the gyration radius[92] or different criteria exploiting only the information on the apo structure.

While we focused on cases for which the ligands are known, we are confident that our EDES approach should also be valuable in the context of virtual screening, with the enhanced binding pocket sampling leading to a potentially higher number of interesting lead candidates. As a long-term goal, we aim to create a database of protein structures that should help in reducing the cost associated with the generation of the structures for ensemble docking runs, allowing for a single target virtual screening of thousands of compounds in a reasonable amount of time. The ensemble of targets could also be used to reposition existing drugs for new therapeutic uses, as recently suggested.[51]

■ MATERIALS AND METHODS

Standard MD Simulations. Standard all-atom MD simulations were carried out using the *pmemd* module of the AMBER16[99] molecular modeling software. Topology files were created for each system using the *LEaP* module of AmberTools17 starting from the experimental structures available in the Protein Data Bank (PDB IDs 1JEJ[70] and 1JG6[70] for BGT and BGT–UDP, 1RTC[73] and 1BR5[74] for RIC and RIC–NEO, and 1GUD[78] and 1RPJ[78] for ABP and ABP–ALL, respectively). The ff14SB[100] and GAFF[101] force fields were used for proteins and ligands, respectively. Missing parameters for the latter were generated using the antechamber module of AmberTools17. In particular, atomic restrained electrostatic potential charges were derived after a structural optimization performed with Gaussian 09.[102] Each structure was solvated with the explicit TIP3P water model, and its net charge was neutralized with the required number of randomly placed K$^+$ or Cl$^-$ ions. The total number of atoms was ∼86 000 for BGT/BGT–UDP, ∼54 000 for RIC/RIC–NEO, and ∼62 000 for ABP/ABP–ALL. Periodic boundary conditions were employed, and long-range electrostatics was evaluated through the particle-mesh Ewald algorithm using a real-space cutoff of 12 Å and a grid spacing of 1 Å per grid point in each dimension. The van der Waals interactions were treated by a Lennard-Jones potential using a smooth cutoff (switching radius 10 Å, cutoff radius 12 Å). The initial distance between the protein and the edge of the box was set to be at least 16 Å in each direction. Multistep energy minimization with a combination of the steepest-descent and conjugate-gradient methods was carried out to relax internal constraints of the systems by gradually releasing positional restraints. Following this, the systems were heated from 0 to 310 K in 10 ns of constant-pressure heating (*NPT*) using the Langevin thermostat (collision frequency of 1 ps^{-1}) and the Berendsen barostat. After equilibration, four production runs of 2.5 μs each (for a total of 10 μs for each system) were performed for the apo systems, while a single 1 μs-long simulation was performed for each complex. A time step of 2 fs was used for preproduction runs, while equilibrium MD simulations were carried out with

a time step of 4 fs in the *NPT* ensemble (using a MC barostat) after hydrogen mass repartitioning.[103] Coordinates from production trajectories were saved every 100 and 10 ps for MD$_{apo}$ and MD$_{holo}$, respectively.

Metadynamics Simulations. Bias-exchange well-tempered metadynamics simulations[57,68,69] were performed on the three apo proteins using the GROMACS 2016.5 package[104] and the PLUMED 2.3.5 plugin.[105] The last conformation saved from the equilibration step from MD$_{apo}$ was used as the starting structure for each simulation. AMBER parameters were ported to GROMACS using the acpype parser.[106] To enhance the sampling of different binding site shapes, we used the following four CVs defined by including all heavy atoms of the residues lining the binding site itself: the radius of gyration of the binding site (RoG$_{BS}$) calculated using the *gyration* built-in function of PLUMED and the numbers of (pseudo)contacts across the "inertia planes" (CIP$_{1,2,3}$) of the binding site, defined as the planes orthogonal to the three principal inertia axes and passing through the center of mass of the binding site.

Binding site residues were defined as those within 3 Å (BGT and RIC) or 4 Å (ABP) of the ligand in the experimental structure of the complex (Figure S1). The cutoff was increased for ABP–ALL because of the low number of residues (seven) found when a 3 Å cutoff was used. Very similar definitions were found using the COACH-D Web server with the apo structures (Table S2). The CVs were calculated by an in-house tcl script based on the VMD *orient* function. Namely, residues lining the binding site were split into two lists A and B according to the positions of the geometrical centers of their backbones on each of the two sides of the inertia plane, and the overall number of pseudocontacts N_c between the two groups was calculated through the *coordination* keyword of PLUMED, which implements a switching function such as the following:

$$N_c = \sum_{i \in A} \sum_{j \in B} s_{ij}$$

where

$$s_{ij} = \left[1 - \left(\frac{r_{ij}}{r_0} \right)^n \right] \Big/ \left[1 - \left(\frac{r_{ij}}{r_0} \right)^m \right]$$

with $r_0 = 8$ Å, $n = 6$, and $m = 12$. Each replica was simulated for 100 ns (as our aim is primarily to enhance sampling of different shapes of the binding site and not to obtain converged free energy profiles), so that each window accumulated 400 ns of simulation time. Coordinates were saved every 10 ps. The height w was set to 0.6 kcal/mol for all systems, while the widths s_i of the Gaussian hills were set according to established prescriptions[107] to 0.15, 0.05, and 0.08 Å (RoG$_{BS}$), 5.4, 4.8, and 1.6 (CIP$_1$), 5.1, 3.2, and 4.9 (CIP$_2$), and 5.3, 3.1, and 6.0 (CIP$_3$) for BGT, RIC, and ABP, respectively. Hills were added every 2 ps, while the bias-exchange frequency was set to 20 ps. The bias factor for well-tempered metadynamics was set to 10. The "windows" approach briefly described in the Results and Discussion was implemented using RoG$_{BS}$ as the control parameter. Namely, we applied restraints (force constants set to 50 and 10 kcal mol^{-1} Å^{-2} for the upper and lower walls, respectively, as we seek for compression rather than enlargement of the binding site) at values of RoG$_{BS}$ that were 7.5% higher and lower than the value measured in the apo X-ray structure (RoG$_{X\text{-ray}}^{apo}$). Then, from the trajectory corresponding to this first window,

we selected a random conformation of the protein whose RoG_{BS} was 5% lower than $RoG_{X\text{-ray}}^{apo}$ and performed another simulation with walls centered at $\pm 7.5\%$ $RoG_{X\text{-ray}}^{apo}$ from this new center, repeating this procedure so as to simulate a total of four windows (see Figure 3 and Table S1). It should be noted that the walls were set to allow partial overlap between adjacent windows, which indeed occurred in all cases (Figure 3).

Cluster Analysis of MD Trajectories. The cluster analysis was performed on the four CVs defined above using in-house R scripts. We implemented a multistep strategy aimed at increasing in an unbiased manner the percentage of conformations similar to the native structure of the holo protein. Namely, the distribution of RoG_{BS} values sampled during the MD simulation was binned into 10 equally wide slices, and hierarchical agglomerative clustering (using the built-in function *hclust* and the Euclidean method to compute the distance matrix) was performed on the four CVs within each slice, setting the number of generated clusters to $x_i = (N_i/N_{tot}) \cdot N_c$, where N_i, N_{tot}, and $N_c = 500$ are the number of structures within the *i*th slice, the total number of structures, and the total number of clusters, respectively. The resulting N_c clusters were used as starting point to perform a second cluster analysis with the *K*-means method and requiring a total of N_c clusters (with the maximum number of iterations set to 10 000). Despite not making any use of specific knowledge of the structure of the complexes, our informed strategy was able to generate a larger fraction of cluster structures displaying $RMSD_{BS} < 2$ Å than that obtained from the standard application of *K*-means using randomly selected conformations as starting points (Figure S5). In fact, this latter initialization strategy is considered one of the most unreliable ones on the basis of a comparison of several alternative algorithms on a range of diverse data sets.[108] We compared our clustering protocol to a more standard calculation based dRMSD at the binding site (using the hierarchical agglomerative method implemented in the *cpptraj* module of the AMBER package). The scripts to divide the binding site residues according to the CIP scheme and perform the cluster analysis according to our ad hoc scheme are available at the webpage https://github.com/haddocking/EDES.

Molecular Docking. Molecular docking calculations were performed with AutoDock4[82] and the HADDOCK Web server version 2.2.[81,109] This choice allowed our methodology to be validated against two programs differing in their search algorithms, scoring functions, and pose selection schemes. Both programs were first validated for redocking against experimental structures (Table S8). Next, they were used to perform docking from the generated ensembles using default settings, apart from the following changes.

In AutoDock4, we used the Lamarckian genetic algorithm (LGA) to perform a hybrid global–local search of the docking poses. The grid density (*spacing* parameter changed from 0.375 to 0.25 Å) and number of energy evaluations (*ga_num_evals* increased by a factor of 10 from the default value) were both increased, with the purpose to avoid repeating each calculation several times to obtain converged results. For each set of structures, 500 rigid docking independent calculations were performed using an adaptive grid enclosing all of the residues belonging to the binding site. Next, the top poses (in total 500, one for each docking run) were clustered using the *cpptraj* module of AmberTools17 with a hierarchical agglomerative algorithm and a cutoff of 1.5 Å for the RMSD distance matrix.

In HADDOCK, a single docking run was performed per case, starting from the various ensembles of 500 conformations, with increased sampling (10000/400/400 models for rigid-body docking (it0 step), semiflexible refinement (it1 step), and final refinement in explicit solvent (wat step). The weight of the intermolecular van der Waals energy for the initial rigid-body docking stage was increased to 1.0 (from the default value of 0.01), RMSD-based clustering was selected with a cutoff of 1 Å, and the docking was guided by ambiguous distance restraints defined for the residues of the binding site and the ligand.[110] In the rigid-body stage, the protein binding site residues were defined as "active", effectively drawing the ligand into the binding site without restraining its orientation. For the subsequent stages, the restraints were such that only the ligand was active, allowing it to explore the binding site better while maintaining at least one contact with its residues.

Druggability Calculations. For each structure within the four ensembles (MD_{apo}, MD_{holo}, $EDES_{3w}$, and $EDES_{4w}$) generated for the three proteins, we used the open-source pocket detection package f-pocket[84] to identify and characterize putative protein binding sites. The algorithm implemented in f-pocket is based on Voronoi tessellation and alpha spheres and has proven to be stable, fast, and accurate, performing very well on state-of-the-art data sets.[84] Moreover, for each putative binding site identified, f-pocket also estimates its druggability through the calculation of a "druggability score", D,[85] ranging from 0 (no druggable pocket) to 1 (high probability to be druggable). To evaluate the druggability of the known binding sites on the three proteins investigated here, we recorded only D values associated with pockets whose centers of mass were found to be within 6 Å of that of the binding site identified in the experimental structures.

Figures and Graphs. Figures were generated with Maestro 10.6,[111] VMD 1.9.3,[112] and InkScape 0.91. Graphs were created with xmgrace 5.1.25.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.8b00730.

　　Supporting figures and tables (PDF)

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: a.v.vargiu@uu.nl; attilio.vargiu@gmail.com.
ORCID Ⓘ
Giuliano Malloci: 0000-0002-5985-257X
Fabio Pietrucci: 0000-0002-4892-2667
Alexandre M. J. J. Bonvin: 0000-0001-7369-1322
Attilio V. Vargiu: 0000-0003-4013-8867

Author Contributions

A.B. and A.V.V. designed research with contributions of all authors. A.B. performed MD simulations. G.M. performed druggability calculations. A.M.J.J.B. performed docking calculations with HADDOCK. A.V.V. performed docking calculations with AutoDock4. All of the authors contributed analysis tools and analyzed the data. The manuscript was written with contributions of and approved by all of the authors.

■ REFERENCES

(1) Antunes, D. A.; Devaurs, D.; Kavraki, L. E. Understanding the Challenges of Protein Flexibility in Drug Design. *Expert Opin. Drug Discovery* **2015**, *10* (12), 1301−1313.

(2) Babine, R. E.; Bender, S. L. Molecular Recognition of Protein−Ligand Complexes: Applications to Drug Design. *Chem. Rev.* **1997**, *97* (5), 1359−1472.

(3) Du, X.; Li, Y.; Xia, Y.-L.; Ai, S.-M.; Liang, J.; Sang, P.; Ji, X.-L.; Liu, S.-Q. Insights into Protein−Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* **2016**, *17* (2), 144.

(4) 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.

(5) Ganesan, A.; Coote, M. L.; Barakat, K. Molecular Dynamics-Driven Drug Discovery: Leaping Forward with Confidence. *Drug Discovery Today* **2017**, *22* (2), 249−269.

(6) Pagadala, N. S.; Syed, K.; Tuszynski, J. Software for Molecular Docking: A Review. *Biophys. Rev.* **2017**, *9* (2), 91−102.

(7) Śledź, P.; Caflisch, A. Protein Structure-Based Drug Design: From Docking to Molecular Dynamics. *Curr. Opin. Struct. Biol.* **2018**, *48*, 93−102.

(8) Ferreira, L.; dos Santos, R.; Oliva, G.; Andricopulo, A. Molecular Docking and Structure-Based Drug Design Strategies. *Molecules* **2015**, *20* (7), 13384.

(9) Forli, S. Charting a Path to Success in Virtual Screening. *Molecules* **2015**, *20* (10), 18732.

(10) Irwin, J. J.; Shoichet, B. K. Docking Screens for Novel Ligands Conferring New Biology. *J. Med. Chem.* **2016**, *59* (9), 4103−4120.

(11) Amaro, R. E.; Baudry, J.; Chodera, J.; Demir, Ö.; McCammon, J. A.; Miao, Y.; Smith, J. C. Ensemble Docking in Drug Discovery. *Biophys. J.* **2018**, *114* (10), 2271−2278.

(12) Buonfiglio, R.; Recanatini, M.; Masetti, M. Protein Flexibility in Drug Discovery: From Theory to Computation. *ChemMedChem* **2015**, *10* (7), 1141−1148.

(13) Chen, Y.-C. Beware of Docking! *Trends Pharmacol. Sci.* **2015**, *36* (2), 78−95.

(14) Grinter, S.; Zou, X. Challenges, Applications, and Recent Advances of Protein-Ligand Docking in Structure-Based Drug Design. *Molecules* **2014**, *19* (7), 10150.

(15) Kastritis, P. L.; Bonvin, A. M. J. J. On the Binding Affinity of Macromolecular Interactions: Daring to Ask Why Proteins Interact. *J. R. Soc., Interface* **2013**, *10* (79), 20120835.

(16) Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R. Towards the Development of Universal, Fast and Highly Accurate Docking/Scoring Methods: A Long Way to Go. *Br. J. Pharmacol.* **2008**, *153* (S1), S7−S26.

(17) Zacharias, M. Accounting for Conformational Changes during Protein−Protein Docking. *Curr. Opin. Struct. Biol.* **2010**, *20* (2), 180−186.

(18) Lexa, K. W.; Carlson, H. A. Protein Flexibility in Docking and Surface Mapping. *Q. Rev. Biophys.* **2012**, *45* (3), 301−343.

(19) Wei, G.; Xi, W.; Nussinov, R.; Ma, B. Protein Ensembles: How Does Nature Harness Thermodynamic Fluctuations for Life? The Diverse Functional Roles of Conformational Ensembles in the Cell. *Chem. Rev.* **2016**, *116* (11), 6516−6551.

(20) Wu, P.; Nielsen, T. E.; Clausen, M. H. FDA-Approved Small-Molecule Kinase Inhibitors. *Trends Pharmacol. Sci.* **2015**, *36* (7), 422−439.

(21) Arrowsmith, C. H.; Bountra, C.; Fish, P. V.; Lee, K.; Schapira, M. Epigenetic Protein Families: A New Frontier for Drug Discovery. *Nat. Rev. Drug Discovery* **2012**, *11*, 384.

(22) Flavin, R.; Peluso, S.; Nguyen, P. L.; Loda, M. Fatty Acid Synthase as a Potential Therapeutic Target in Cancer. *Future Oncol.* **2010**, *6* (4), 551−562.

(23) Chen, C. H.; Ferreira, J. C.; Gross, E. R.; Mochly-Rosen, D. Targeting Aldehyde Dehydrogenase 2: New Therapeutic Opportunities. *Physiol. Rev.* **2014**, *94* (1), 1−34.

(24) Amaral, M.; Kokh, D. B.; Bomke, J.; Wegener, A.; Buchstaller, H. P.; Eggenweiler, H. M.; Matias, P.; Sirrenberg, C.; Wade, R. C.; Frech, M. Protein Conformational Flexibility Modulates Kinetics and Thermodynamics of Drug Binding. *Nat. Commun.* **2017**, *8* (1), 2276.

(25) Hayward, S. Identification of Specific Interactions That Drive Ligand-Induced Closure in Five Enzymes with Classic Domain Movements. *J. Mol. Biol.* **2004**, *339* (4), 1001−1021.

(26) Fischer, M.; Coleman, R. G.; Fraser, J. S.; Shoichet, B. K. Incorporation of Protein Flexibility and Conformational Energy Penalties in Docking Screens to Improve Ligand Discovery. *Nat. Chem.* **2014**, *6*, 575.

(27) Yuriev, E.; Holien, J.; Ramsland, P. A. Improvements, Trends, and New Ideas in Molecular Docking: 2012−2013 in Review. *J. Mol. Recognit.* **2015**, *28* (10), 581−604.

(28) Guedes, I. A.; de Magalhães, C. S.; Dardenne, L. E. Receptor−Ligand Molecular Docking. *Biophys. Rev.* **2014**, *6* (1), 75−87.

(29) Baron, R.; McCammon, J. A. Molecular Recognition and Ligand Association. *Annu. Rev. Phys. Chem.* **2013**, *64* (1), 151−175.

(30) Huang, S.-Y.; Zou, X. Ensemble Docking of Multiple Protein Structures: Considering Protein Structural Variations in Molecular Docking. *Proteins: Struct., Funct., Genet.* **2007**, *66* (2), 399−421.

(31) Boehr, D. D.; Nussinov, R.; Wright, P. E. The Role of Dynamic Conformational Ensembles in Biomolecular Recognition. *Nat. Chem. Biol.* **2009**, *5*, 789.

(32) Korb, O.; Olsson, T. S. G.; Bowden, S. J.; Hall, R. J.; Verdonk, M. L.; Liebeschuetz, J. W.; Cole, J. C. Potential and Limitations of Ensemble Docking. *J. Chem. Inf. Model.* **2012**, *52* (5), 1262−1274.

(33) Moroni, E.; Paladino, A.; Colombo, G. The Dynamics of Drug Discovery. *Curr. Top. Med. Chem.* **2015**, *15* (20), 2043−2055.

(34) Wong, C. F. Flexible Receptor Docking for Drug Discovery. *Expert Opin. Drug Discovery* **2015**, *10* (11), 1189−1200.

(35) Koukos, P. I.; Xue, L. C.; Bonvin, A. M. J. J. Protein−Ligand Pose and Affinity Prediction: Lessons from D3R Grand Challenge 3. *J. Comput.-Aided Mol. Des.* **2019**, *33* (1), 83−91.

(36) Hopkins, A. L.; Groom, C. R. The Druggable Genome. *Nat. Rev. Drug Discovery* **2002**, *1*, 727.

(37) Borhani, D. W.; Shaw, D. E. The Future of Molecular Dynamics Simulations in Drug Discovery. *J. Comput.-Aided Mol. Des.* **2012**, *26* (1), 15–26.

(38) Liu, X.; Shi, D.; Zhou, S.; Liu, H.; Liu, H.; Yao, X. Molecular Dynamics Simulations and Novel Drug Discovery. *Expert Opin. Drug Discovery* **2018**, *13* (1), 23–37.

(39) Maximova, T.; Moffatt, R.; Ma, B.; Nussinov, R.; Shehu, A. Principles and Overview of Sampling Methods for Modeling Macromolecular Structure and Dynamics. *PLoS Comput. Biol.* **2016**, *12* (4), No. e1004619.

(40) Mortier, J.; Rakers, C.; Bermudez, M.; Murgueitio, M. S.; Riniker, S.; Wolber, G. The Impact of Molecular Dynamics on Drug Design: Applications for the Characterization of Ligand–Macromolecule Complexes. *Drug Discovery Today* **2015**, *20* (6), 686–702.

(41) Johnson, D. K.; Karanicolas, J. Druggable Protein Interaction Sites Are More Predisposed to Surface Pocket Formation than the Rest of the Protein Surface. *PLoS Comput. Biol.* **2013**, *9* (3), No. e1002951.

(42) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12* (1), 281–296.

(43) Totrov, M.; Abagyan, R. Flexible Ligand Docking to Multiple Receptor Conformations: A Practical Alternative. *Curr. Opin. Struct. Biol.* **2008**, *18* (2), 178–184.

(44) Amaro, R. E.; Li, W. W. Emerging Methods for Ensemble-Based Virtual Screening. *Curr. Top. Med. Chem.* **2010**, *10* (1), 3–13.

(45) Tarcsay, Á.; Paragi, G.; Vass, M.; Jójárt, B.; Bogár, F.; Keserű, G. M. The Impact of Molecular Dynamics Sampling on the Performance of Virtual Screening against GPCRs. *J. Chem. Inf. Model.* **2013**, *53* (11), 2990–2999.

(46) Hazuda, D. J.; Anthony, N. J.; Gomez, R. P.; Jolly, S. M.; Wai, J. S.; Zhuang, L.; Fisher, T. E.; Embrey, M.; Guare, J. P.; Egbertson, M. S.; et al. A Naphthyridine Carboxamide Provides Evidence for Discordant Resistance between Mechanistically Identical Inhibitors of HIV-1 Integrase. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (31), 11233–11238.

(47) Pietrucci, F.; Vargiu, A. V.; Kranjc, A. HIV-1 Protease Dimerization Dynamics Reveals a Transient Druggable Binding Pocket at the Interface. *Sci. Rep.* **2016**, *5*, 18555.

(48) Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, J. A. Discovery of a Novel Binding Trench in HIV Integrase. *J. Med. Chem.* **2004**, *47* (8), 1879–1881.

(49) Summa, V.; Petrocchi, A.; Bonelli, F.; Crescenzi, B.; Donghi, M.; Ferrara, M.; Fiore, F.; Gardelli, C.; Gonzalez Paz, O.; Hazuda, D. J.; et al. Discovery of Raltegravir, a Potent, Selective Orally Bioavailable HIV-Integrase Inhibitor for the Treatment of HIV-AIDS Infection. *J. Med. Chem.* **2008**, *51* (18), 5843–5855.

(50) Zhong, H.; Carlson, H. A. Computational Studies and Peptidomimetic Design for the Human P53–MDM2 Complex. *Proteins: Struct., Funct., Genet.* **2005**, *58* (1), 222–234.

(51) Li, Y. Y.; An, J.; Jones, S. J. A Computational Approach to Finding Novel Targets for Existing Drugs. *PLoS Comput. Biol.* **2011**, *7* (9), No. e1002139.

(52) Lin, J.-H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. Computational Drug Design Accommodating Receptor Flexibility: The Relaxed Complex Scheme. *J. Am. Chem. Soc.* **2002**, *124* (20), 5632–5633.

(53) Kuroda, D.; Gray, J. J. Pushing the Backbone in Protein-Protein Docking. *Structure* **2016**, *24* (10), 1821–1829.

(54) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120* (24), 11919–11929.

(55) Luitz, M. P.; Zacharias, M. Protein–Ligand Docking Using Hamiltonian Replica Exchange Simulations with Soft Core Potentials. *J. Chem. Inf. Model.* **2014**, *54* (6), 1669–1675.

(56) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314* (1), 141–151.

(57) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12562–12566.

(58) Grubmüller, H. Predicting Slow Structural Transitions in Macromolecular Systems: Conformational Flooding. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1995**, *52* (3), 2893–2906.

(59) Huber, T.; Torda, A. E.; van Gunsteren, W. F. Local Elevation: A Method for Improving the Searching Properties of Molecular Dynamics Simulation. *J. Comput.-Aided Mol. Des.* **1994**, *8* (6), 695–708.

(60) Antolin, A. A.; Carotti, A.; Nuti, R.; Hakkaya, A.; Camaioni, E.; Mestres, J.; Pellicciari, R.; Macchiarulo, A. Exploring the Effect of PARP-1 Flexibility in Docking Studies. *J. Mol. Graphics Modell.* **2013**, *45*, 192–201.

(61) Bolia, A.; Ozkan, S. B. Adaptive BP-Dock: An Induced Fit Docking Approach for Full Receptor Flexibility. *J. Chem. Inf. Model.* **2016**, *56* (4), 734–746.

(62) Miao, Y.; Goldfeld, D. A.; Moo, E. V.; Sexton, P. M.; Christopoulos, A.; McCammon, J. A.; Valant, C. Accelerated Structure-Based Design of Chemically Diverse Allosteric Modulators of a Muscarinic G Protein-Coupled Receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (38), E5675–E5684.

(63) Motta, S.; Bonati, L. Modeling Binding with Large Conformational Changes: Key Points in Ensemble-Docking Approaches. *J. Chem. Inf. Model.* **2017**, *57* (7), 1563–1578.

(64) Oleinikovas, V.; Saladino, G.; Cossins, B. P.; Gervasio, F. L. Understanding Cryptic Pocket Formation in Protein Targets by Enhanced Sampling Simulations. *J. Am. Chem. Soc.* **2016**, *138* (43), 14257–14263.

(65) Osguthorpe, D. J.; Sherman, W.; Hagler, A. T. Generation of Receptor Structural Ensembles for Virtual Screening Using Binding Site Shape Analysis and Clustering. *Chem. Biol. Drug Des.* **2012**, *80* (2), 182–193.

(66) Seeliger, D.; De Groot, B. L. Conformational Transitions upon Ligand Binding: Holo-Structure Prediction from Apo Conformations. *PLoS Comput. Biol.* **2010**, *6* (1), No. e1000634.

(67) Uehara, S.; Tanaka, S. Cosolvent-Based Molecular Dynamics for Ensemble Docking: Practical Method for Generating Druggable Protein Conformations. *J. Chem. Inf. Model.* **2017**, *57* (4), 742–756.

(68) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603.

(69) Piana, S.; Laio, A. A Bias-Exchange Approach to Protein Folding. *J. Phys. Chem. B* **2007**, *111* (17), 4553–4559.

(70) Moréra, S.; Larivière, L.; Kurzeck, J.; Aschke-Sonnenborn, U.; Freemont, P. S.; Janin, J.; Rüger, W. High Resolution Crystal Structures of T4 Phage β-Glucosyltransferase: Induced Fit and Effect of Substrate and Metal Binding11Edited by R. Huber. *J. Mol. Biol.* **2001**, *311* (3), 569–577.

(71) de Groot, B. L.; van Aalten, D. M. F.; Scheek, R. M.; Amadei, A.; Vriend, G.; Berendsen, H. J. C. Prediction of Protein Conformational Freedom from Distance Constraints. *Proteins: Struct., Funct., Genet.* **1997**, *29* (2), 240–251.

(72) Seeliger, D.; Haas, J.; de Groot, B. L. Geometry-Based Sampling of Conformational Transitions in Proteins. *Structure* **2007**, *15* (11), 1482–1492.

(73) Mlsna, D.; Monzingo, A. F.; Katzin, B. J.; Ernst, S.; Robertus, J. D. Structure of Recombinant Ricin A Chain at 2. *Protein Sci.* **1993**, *2* (3), 429–435.

(74) Yan, X.; Hollis, T.; Svinth, M.; Day, P.; Monzingo, A. F.; Milne, G. W. A.; Robertus, J. D. Structure-Based Identification of a Ricin Inhibitor. *J. Mol. Biol.* **1997**, *266* (5), 1043–1049.

(75) Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T. M.; Mortenson, P. N.; Murray, C. W. Diverse, High-Quality Test Set for the Validation of Protein–Ligand Docking Performance. *J. Med. Chem.* **2007**, *50* (4), 726–741.

(76) Ravindranath, P. A.; Forli, S.; Goodsell, D. S.; Olson, A. J.; Sanner, M. F. AutoDockFR: Advances in Protein−Ligand Docking with Explicitly Specified Binding Site Flexibility. *PLoS Comput. Biol.* **2015**, *11* (12), No. e1004586.

(77) 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.* **2009**, *31* (2), 455−461.

(78) Chaudhuri, B. N.; Ko, J.; Park, C.; Jones, T. A.; Mowbray, S. L. Structure of D-Allose Binding Protein from Escherichia Coli Bound to D-Allose at 1.8 Å Resolution. *J. Mol. Biol.* **1999**, *286* (5), 1519−1531.

(79) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739−1749.

(80) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *J. Med. Chem.* **2006**, *49* (21), 6177−6196.

(81) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003**, *125* (7), 1731−1737.

(82) 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.

(83) Wu, Q.; Peng, Z.; Zhang, Y.; Yang, J. COACH-D: Improved Protein−Ligand Binding Sites Prediction with Refined Ligand-Binding Poses through Molecular Docking. *Nucleic Acids Res.* **2018**, *46* (W1), W438−W442.

(84) Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An Open Source Platform for Ligand Pocket Detection. *BMC Bioinf.* **2009**, *10* (1), 168.

(85) Schmidtke, P.; Barril, X. Understanding and Predicting Druggability. A High-Throughput Method for Detection of Drug Binding Sites. *J. Med. Chem.* **2010**, *53* (15), 5858−5867.

(86) Lam, P. C.-H.; Abagyan, R.; Totrov, M. Ligand-Biased Ensemble Receptor Docking (LigBEnD): A Hybrid Ligand/Receptor Structure-Based Approach. *J. Comput.-Aided Mol. Des.* **2018**, *32* (1), 187−198.

(87) Meiler, J.; Baker, D. ROSETTALIGAND: Protein−Small Molecule Docking with Full Side-Chain Flexibility. *Proteins: Struct., Funct., Genet.* **2006**, *65* (3), 538−548.

(88) Du, X.; Li, Y.; Xia, Y.-L.; Ai, S.-M.; Liang, J.; Sang, P.; Ji, X.-L.; Liu, S.-Q. Insights into Protein−Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* **2016**, *17* (2), 144.

(89) Johnson, D. K.; Karanicolas, J. Selectivity by Small-Molecule Inhibitors of Protein Interactions Can Be Driven by Protein Surface Fluctuations. *PLoS Comput. Biol.* **2015**, *11* (2), No. e1004081.

(90) Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K. W.; Renfrew, P. D.; Smith, C. A.; Sheffler, W. Rosetta3. *Methods Enzymol.* **2011**, *487*, 545−574.

(91) Cavasotto, C. N.; Kovacs, J. A.; Abagyan, R. A. Representing Receptor Flexibility in Ligand Docking through Relevant Normal Modes. *J. Am. Chem. Soc.* **2005**, *127* (26), 9632−9640.

(92) Kurkcuoglu, Z.; Doruker, P. Ligand Docking to Intermediate and Close-To-Bound Conformers Generated by an Elastic Network Model Based Algorithm for Highly Flexible Proteins. *PLoS One* **2016**, *11* (6), No. e0158063.

(93) Kurkcuoglu, Z.; Bahar, I.; Doruker, P. ClustENM: ENM-Based Sampling of Essential Conformational Space at Full Atomic Resolution. *J. Chem. Theory Comput.* **2016**, *12* (9), 4549−4562.

(94) Ahmed, A.; Rippmann, F.; Barnickel, G.; Gohlke, H. A Normal Mode-Based Geometric Simulation Approach for Exploring Biologically Relevant Conformational Transitions in Proteins. *J. Chem. Inf. Model.* **2011**, *51* (7), 1604−1622.

(95) Bakan, A.; Nevins, N.; Lakdawala, A. S.; Bahar, I. Druggability Assessment of Allosteric Proteins by Dynamics Simulations in the Presence of Probe Molecules. *J. Chem. Theory Comput.* **2012**, *8* (7), 2435−2447.

(96) Leis, S.; Zacharias, M. Efficient Inclusion of Receptor Flexibility in Grid-Based Protein−Ligand Docking. *J. Comput. Chem.* **2011**, *32* (16), 3433−3439.

(97) Pandini, A.; Fornili, A. Using Local States To Drive the Sampling of Global Conformations in Proteins. *J. Chem. Theory Comput.* **2016**, *12* (3), 1368−1379.

(98) Pietrucci, F.; Laio, A. A Collective Variable for the Efficient Exploration of Protein Beta-Sheet Structures: Application to SH3 and GB1. *J. Chem. Theory Comput.* **2009**, *5* (9), 2197−2201.

(99) Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; et al. *Amber 16*; University of California: San Francisco, 2016.

(100) 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* (8), 3696−3713.

(101) 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.

(102) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*, revision E.01; Gaussian, Inc.: Wallingford, CT, 2009.

(103) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11* (4), 1864−1874.

(104) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.

(105) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604−613.

(106) Sousa da Silva, A. W.; Vranken, W. F. ACPYPE - AnteChamber PYthon Parser InterfacE. *BMC Res. Notes* **2012**, *5* (1), 367.

(107) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. Assessing the Accuracy of Metadynamics. *J. Phys. Chem. B* **2005**, *109* (14), 6714−6721.

(108) Celebi, M. E.; Kingravi, H. A.; Vela, P. A. A Comparative Study of Efficient Initialization Methods for the K-Means Clustering Algorithm. *Expert Syst. Appl.* **2013**, *40* (1), 200−210.

(109) van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428* (4), 720−725.

(110) Kurkcuoglu, Z.; Koukos, P. I.; Citro, N.; Trellet, M. E.; Rodrigues, J. P. G. L. M.; Moreira, I. S.; Roel-Touris, J.; Melquiond, A. S. J.; Geng, C.; Schaarschmidt, J.; et al. Performance of HADDOCK and a Simple Contact-Based Protein−Ligand Binding Affinity Predictor in the D3R Grand Challenge 2. *J. Comput.-Aided Mol. Des.* **2018**, *32* (1), 175−185.

(111) *Schrödinger Release 2015−4: Maestro*; Schrödinger, LLC: New York, 2015.

(112) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33−38.