JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

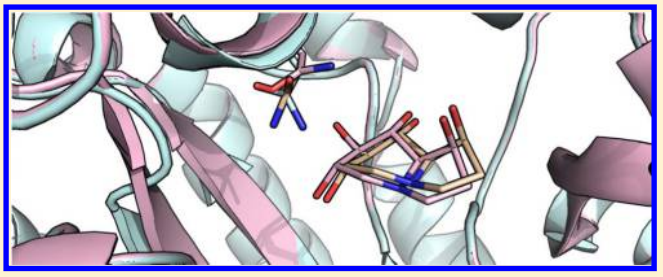
FlexAID: Revisiting Docking on Non-Native-Complex Structures

Francis Gaudreault and Rafael J. Najmanovich*

Department of Biochemistry, Faculty of Medicine and Health Sciences, University of Sherbrooke, Sherbrooke, Quebec J1H5N4, Canada

Supporting Information

ABSTRACT: Small-molecule protein docking is an essential tool in drug design and to understand molecular recognition. In the present work we introduce FlexAID, a small-molecule docking algorithm that accounts for target side-chain flexibility and utilizes a soft scoring function, i.e. one that is not highly dependent on specific geometric criteria, based on surface complementarity. The pairwise energy parameters were derived from a large dataset of true positive poses and negative decoys from the PDBbind database through an iterative process using Monte Carlo simulations. The prediction of binding poses is tested using the widely used Astex dataset as well as the HAP2 dataset, while performance in virtual screening is evaluated using a subset of the DUD dataset. We compare FlexAID to AutoDock Vina, FlexX, and rDock in an extensive number of scenarios to understand the strengths and limitations of the different programs as well as to reported results for Glide, GOLD, and DOCK6 where applicable. The most relevant among these scenarios is that of docking on flexible non-native-complex structures where as is the case in reality, the target conformation in the bound form is not known *a priori*. We demonstrate that FlexAID, unlike other programs, is robust against increasing structural variability. FlexAID obtains equivalent sampling success as GOLD and performs better than AutoDock Vina or FlexX in all scenarios against non-native-complex structures. FlexAID is better than rDock when there is at least one critical side-chain movement required upon ligand binding. In virtual screening, FlexAID results are lower on average than those of AutoDock Vina and rDock. The higher accuracy in flexible targets where critical movements are required, intuitive PyMOL-integrated graphical user interface and free source code as well as precompiled executables for Windows, Linux, and Mac OS make FlexAID a welcome addition to the arsenal of existing small-molecule protein docking methods.



INTRODUCTION

Proteins coexist as an ensemble of states in thermodynamic equilibrium[1] with the population of each state proportional to their relative free energy differences with respect to some reference state. In crystals, protein structures are subject to packing constraints that may vary among different crystals of the same protein.[2,3] At times such constraints lead to distinct arrangements of side-chains in the binding-site.[4] According to the conformational selection theory for ligand-protein binding,[5] a ligand that preferentially binds a given state will shift the equilibrium toward that state. Therefore, due to a combination of the factors above, the bound structure may not necessarily represent the most common state of the protein when unbound. The analysis of protein X-ray structures in the presence or absence of ligands show that side-chain flexibility upon ligand binding is indeed common.[6] In approximately 30% of cases, side-chain rotamer changes between apo and holo forms are essential for ligand binding.[7] Therefore, accounting for side-chain flexibility in molecular docking should play an essential role when the bound structure of the protein is unknown.

Molecular docking is the most commonly used computational method to predict the structure of ligand-protein complexes. Many popular docking algorithms such as AutoDock Vina,[8] FlexX,[9,10] DOCK6,[11] rDock,[12] and GOLD[13] use stringent geometric constraints to define molecular interactions between the two molecules, particularly H-bonding interactions. Their scoring functions are called hard scoring functions[14] as good scoring depends on meeting stringent geometric constraints. These constraints assume that the protein side-chains are already correctly oriented in the binding-site to accommodate ligand binding. If these are not properly oriented, a finer search of the terminal atoms is required to satisfy the necessary geometric criteria. With some notable exceptions,[15,16] algorithms tend to be benchmarked using sets of ligand-bound (holo form) protein structures.[12,17,18] We call such complex structures native in the sense that the holo form conformation is the one that corresponds to the one observed with a bound ligand. Using this nomenclature, a non-native conformation is one obtained from an apo (unbound) form of the protein or non-apo form, i.e. a holo form bound to a ligand other than the one present in the native form being used as reference. Thus, most algorithms are tested using proteins in rather convenient conformations and do not discuss the accuracy of their method when tested on

Received: February 11, 2015
Published: June 15, 2015

DOI: 10.1021/acs.jcim.5b00078
J. Chem. Inf. Model. 2015, 55, 1323−1336

Table 1. AUC for Different Parameter Sets as a Function of the Number of Monte Carlo Iterations on an Independent Dataset

					AUC		
		complexes	groups	decoys	null[a]	random[b]	optimized[c]
MC iterations	1	1,207	544	310,459	0.586	0.526 ± 0.056	0.846
	2	1,455	630	402,043	0.606	0.538 ± 0.046	0.842
	3	1,543	657	455,047	0.611	0.514 ± 0.052	0.832
	4	1,611	678	529,571	0.606	0.545 ± 0.061	0.820
	5	1,649	688	592,385	0.604	0.529 ± 0.053	0.816

[a]All pairwise parameters set to 0. [b]Average over a 10 random sets of parameters. [c]AUC values for the best matrix obtained in the given iteration.

non-native-complex structures. This scenario is not representative of a real-life situation considering that docking is mostly used to accelerate drug discovery or understand potential ligand-protein interactions where the protein structure considered often is a homology model or a non-native-complex structure with respect to the ligand of interest. Additionally, protein flexibility is often neglected in virtual screening experiments, thus assuming that all ligands preferentially select the same state in the conformational ensemble.[19]

Different levels of protein flexibility are observed upon ligand binding. Zavodsky et al.[20] suggested that side-chains move as minimally as possible to accommodate ligand binding, a concept they named the minimal rotation hypothesis. We recently quantified the minimal rotation hypothesis and observed that it is valid in approximately 20% of cases.[7] In about 90% of binding-sites, at least one side-chain rotamer change is observed in the bound complex;[7] and in 30% of these cases, side-chain movements are essential as severe steric clashes are observed in the holo form were it to retain the apo-form side-chain conformations. In the present study, we introduce the molecular docking algorithm FlexAID that uses a scoring function based on surface complementarity. The use of surface complementary in a docking scoring function was previously introduced in the LIGIN docking method.[21,22] Rather than using stringent geometric criteria to define interactions, the method focuses on maximizing shape complementarity between atoms of the ligand and the protein representing favorable interactions as defined by a set of atom types and a pairwise atom type interaction matrix. We investigate to what extent a smooth scoring, i.e. one whose values do not change abruptly with slight changes in the structure, can implicitly simulate protein flexibility (to accommodate minimal side-chain rearrangements) to take advantage of the higher success rates observed when soft potentials are used to mimic protein flexibility.[11] Soft docking has been previously implemented through the attenuation of the penalty associated with steric clashes.[23] Our approach differs in that soft scoring is built into the scoring function pairwise interaction terms while maintaining a hard steric clash term. We also introduce explicit side-chain rotamer changes in our method to account for the larger movements that may lead to more drastic changes in the predicted energy that may not be accounted for by a smooth scoring function.

The development of scoring functions is often based on the use of large datasets of complexes and ignores the potential contribution that negative examples may provide. For instance, empirical scoring functions such as the ones from AutoDock Vina, FlexX, and rDock calibrate their energetic terms only using structures representing native complexes. Although the combined use of negative and positive examples was observed to be detrimental in other fields such as protein−protein interactions and protein structure predictions,[24−26] it has been recently shown that it can improve the accuracy of scoring functions for protein−ligand docking.[27] The FlexAID method introduced here utilizes a soft scoring function, i.e. one that has no angular term and that does not vary abruptly with changes in geometry,[14] developed using both positive and negative examples.

RESULTS

Any optimization problem, docking included, has three interconnected components that affect the accuracy as well as usability of a method. These are representation, search, and scoring.

FlexAID uses PDB structures "as is", that is no manual curation is required. In particular, the introduction of hydrogen atoms prior to docking is not necessary, and any such atoms are ignored if present. In addition, the calculation of partial charges or protonation is not required either. FlexAID relies on genetic algorithms as its search strategy as discussed in the Methods section. The following section discusses the development of the scoring function and is followed by the validation of the method.

Development of the Scoring Function. We developed a soft scoring function for FlexAID using contact surfaces based on the complementarity function (CF) originally introduced by Sobolev et al.[22] Unlike potential functions that describe the interaction between two atoms with a functional form with a minimum around an optimal distance, in the CF the interaction energy between two atoms varies linearly with their surface area in contact. As a soft scoring function, hydrogen atoms are implicitly accounted for. While this implicit treatment prevents the use of directionality constraints for interactions, the "softer" or smoother surface helps maximize shape complementarity between protein and ligand atoms. As described in the Methods section, our implementation of the CF function is composed of three terms (eq 1), a term representing the interactions between ligand and protein atoms, a term representing interactions between ligand atoms and the implicit solvent, and last a term to prevent steric clashes.

Ligand-protein interactions are quantified with a matrix of pairwise interaction energy terms between atom types and modulated by the surface area in contact between atoms. The pairwise parameters are derived with the use of statistical potentials optimized using a decoys set generated from the PDBbind database.[28] We grouped the proteins by sequence to account for redundancy to avoid biasing the potential toward pairwise interactions frequently found in protein families more widely represented in the PDBbind.

A Monte Carlo procedure was utilized to derive the potential with 820 pairwise interactions between 40 atom types as described in the Methods section. In each successive Monte Carlo iteration, we made the decoy sets harder by injecting new

DOI: 10.1021/acs.jcim.5b00078
J. Chem. Inf. Model. 2015, 55, 1323−1336

low-energy decoys predicted with the potential from the preceding step. In addition to this increase of the number of decoys per complex, we also increased the number of complexes used to generate decoys. We maintained a high area under the receiver-operator-curve (AUC) value on an independent test set with successive iterations despite the increase in the number of complexes and the difficulty of the decoys (Table 1). The number of proteins considered in our study is smaller than what was used in previous studies,[27] but given that the PDBbind database is enriched with drug-like ligands, the energy parameters in our scoring function represent interactions relevant for drug design. We obtained an AUC around 0.6 when the parameters are set to zero (negative control), suggesting as expected that the true positives have less steric clashes than the false positives. Likewise, random values resulted in AUC values around 0.5. We observe a slight decrease in AUC values as the decoy sets are made larger and more difficult in successive MC iterations to a final value of 0.82. Despite this decrease, more complexes in the independent set can be successfully predicted (Figure 1). The parameters obtained also permit to better rank the solutions (Figure 1).

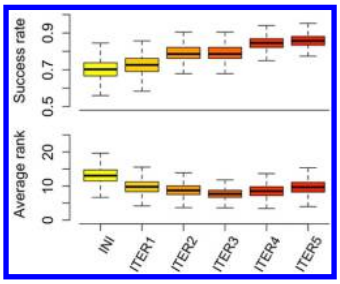


Figure 1. Bootstrapped average success rate and rank of FlexAID predictions on the independent Astex diverse set with consecutive parameter sets. The dataset comprises 84 protein−ligand complexes. The colors are used for clarity purposes only from the less optimized (yellow) to the more optimized (red) potential. The success rate is defined as the fraction of complexes where at least one successful solution (RMSD of 2.0 Å from the native pose) was found. The rank is a measure of how well our method can discriminate the true binding mode from other modes. It represents the rank at which the first success is observed when the results are ordered according to the scoring function. The results are bootstrapped over 10 000 iterations.

Given the functional form of the CF (eq 1), the energy parameters obtained through the optimization process can be compared to our qualitative understanding of known chemical properties. To do so, we analyzed the interactions between some of the most frequently occurring atom types in the training and testing datasets (Table S1). The top 14 pairwise interactions analyzed account for 28% of the total contact surfaces observed in native structures. We assigned each interaction as either favorable or unfavorable based on chemical intuition (e.g., hydrophobic-polar interactions are unfavorable). We observe an ever-increasing agreement between the expected qualitative nature of interactions and their relative strength in the derived potential as the number of Monte Carlo iterations grows (Figures S1A-F). The final parameter set naturally encodes this chemical intuition (Figure 2). For example, the interactions between positively and negatively charged atoms and other polar interactions are highly attractive, while interactions between similarly charged atoms or pairs of donors or acceptors are found to be repulsive. In addition to this qualitative agreement, the relative ordering of the strength of

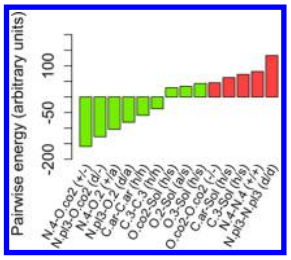


Figure 2. Pairwise interaction parameters for selected interactions. The lower a parameter, the more attractive is the interactions. Interactions that are attractive or repulsive according to chemical intuition are colored in green or red, respectively, and shown in parentheses: positive (+) or negative (−) charge, H-bond donor (d) or acceptor (a), hydrophobic (h) and hydrophilic (s). The following Sybyl atom types are shown: aromatic carbon (C.ar), aliphatic carbon (C.3), positively charged nitrogen (N.4), trigonal planar nitrogen (N.pl3), oxygen in carboxylate (O.co2), oxygen of carbonyl (O.2), oxygen of hydroxyl or ether (O.3) and solvent (Sol).

interactions also makes sense from a chemical point of view. For example, the decreasing trend in the strength of polar interactions reflects an intuitive order in which interactions involving charged groups are stronger than weaker hydrogen bonds. Moreover, the set of parameters naturally simulates the hydrophobic effect[29] as the exposure of polar atoms to solvent was observed to be less penalizing than the exposure of nonpolar atoms. It is important to stress that this agreement between the strength of pairwise interaction energies and chemical intuition is a property of the interaction matrix that emerged as a natural result of the optimization process without being part of the optimization objective function.

Interestingly, with the current approach the resulting strength of interactions between atom types and the solvent arises from the optimization process as repulsive independently of the atom type and validates previous approaches that added such solvent repulsive interactions ad-hoc.[30] In the context of molecular docking, repulsive solvent interactions force the ligand to interact with the target rather than maintain a large solvent accessible surface. It was previously found that optimizing for pose fidelity alone tends to increase the polar contributions of the scoring function.[31] Our results agree with this previous result as we also observe that nonpolar interactions are generally weaker than polar interactions.

Docking algorithms estimate the energy of thousands of unique protein−ligand conformations during a simulation. Ideally, the scoring function would easily distinguish the native conformation from others and rank it with the lowest energy. We optimized the interactions such that the native conformation represents inasmuch as possible the global minimum of the energy landscape of the complementarity function. With the nonoptimized interactions, we observe that the predicted energy of the best conformation (referred to as CF_{min}) suggested by our method is lower than the CF of the native conformation (CF_{ref}) for nearly all complexes of the Astex diverse set[16,32] that is independent from the PDBbind set (Figure S2A). This result shows that the native conformations do not represent the global minimum of the CF indicating the presence of many local minima lower in energy than the native structure. This bias toward overoptimized conformations is drastically reduced as we iteratively improve the interaction matrix (Figures S2A-F), suggesting that the native conformations are increasingly the most favorable ones. This property again emerges naturally as a result of using positive and

DOI: 10.1021/acs.jcim.5b00078
J. Chem. Inf. Model. 2015, 55, 1323−1336

negative decoys. A few outliers with overoptimized CF are notable: PDB codes 1JD0, 1MZC, 1OQ5, and 2BR1. In 3 cases out of 4 cases, the discrepancies observed between CF_{min} and CF_{ref} can be explained by the presence of a metal coordinating the ligand, a type of interaction frequent in the Astex dataset but seldom observed in the PDBbind subset used for training (Table S1).

Conformations of the complex close to that of the native complex (near-natives) share common and essential features of molecular recognition. Therefore, it is critical for the scoring function to also rank favorably these conformations to efficiently guide the search toward the true positive complexes. Conceptually, the energetic well must be smooth enough such that subtle changes in the structure do not drastically impact the overall energy. In order to do so, we assigned decoys that were sufficiently close to the native complex structure in terms of RMSD as positive examples. We included at least 2 such designed true positives for each complex in the decoys set used during the Monte Carlo optimization to build in smoothness into the parameter set. In Figure 3, we plot the relative

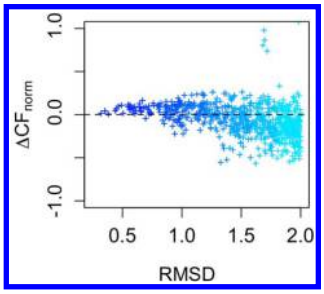


Figure 3. Smoothness of the FlexAID soft scoring function. As relative difference in CF is normalized by the CF of the reference native pose ($CF_{ref} < 0$), complexes (represented by points) above the dashed line are cases where the CF of the best solution (CF_{min}) is lower than that of the reference. The colors are used for clarity purposes only, from less (blue) to more distorted (cyan). The outliers observed above the dashed line represent cases where metals are present in the binding site.

difference in CF of all successful predictions in the Astex diverse set relative to the native CF. The relative CF difference is defined as $\Delta CF_{norm} = (CF_{min} - CF_{ref})/CF_{ref}$. As a consequence of building the decoy sets in this way, the resulting interaction matrix obtained after the successive Monte Carlo iterations displays a funnel-shaped distribution where larger deviations from the native conformation are associated with larger discrepancies in CF. This property of the CF function is notably less prominent in the nonoptimized interaction matrix (Figure S3A). The outliers observed in the lower part could be partially explained by the RMSD measure itself. Some authors previously noted that the RMSD measure can be problematic as low RMSD poses may represent significantly different interactions in comparison to the native conformation.[33,34] In fact, among the 90 complexes with $\Delta CF_{norm} < -0.25$ (representing highly similar surfaces in contact), we find 80 complexes with RMSD > 1.5 Å.

Validation of FlexAID. We validate FlexAID extensively with respect to binding mode prediction as well as in virtual screening experiments comparing the performance of FlexAID primarily to that of AutoDock Vina,[8] FlexX,[10] and rDock[12] and to reported results for DOCK6,[11] GOLD,[16] and Glide.[35] We made the pragmatic choice to focus on AutoDock Vina, FlexX, and rDock when running simulations for a number of reasons: These programs are freely available, widely used, and represent the state-of-the-art in docking simulations. Most importantly, unlike FlexAID, all these programs utilize hard scoring functions.

Binding Mode Prediction. The performance of the docking algorithms is evaluated by measuring how often the method can predict the native pose of the ligands within a margin of error for an ensemble of protein−ligand complexes. We compare the performance of FlexAID in binding mode prediction to those of AutoDock Vina, FlexX, and rDock as well as to reported results for GOLD.[16] We compare the different software in various scenarios for complexes in the Astex diverse set[16,32] and use success rate to evaluate performance. The success rate measures the number of cases in a dataset where a successful pose is found within the top 10 scoring results. At times authors include all predicted poses in the definition of success rate (called sampling success) to report the performance of docking methods.[12] The use of sampling success as a measure of quality of a docking program is informative about the performance of a method, but it cannot be expected that users inspect all generated results. We believe that reporting success rate among the top 10 results is a reasonable compromise, but we also calculate the top 1, 5, and 100 success rates (Table 2). The top 100 results are used as an approximation of sampling success. Each method is tested on its own largest possible subset of the Astex diverse set for which the particular program could be run without technical errors (called here the ideal subset) as well as the largest common subset of cases for all methods. In the following sections, we focus the discussion on results for the ideal subset; however, as the results do not differ significantly between subsets, the conclusions do not change when looking at the largest common subset. In fact, success rates are nearly identical on the Astex datasets with a maximal difference of 5.3%. Differences in success rates are more pronounced in the HAP2 dataset (maximal difference of 15.1%). Each of the two subsets has its advantages, and Table 2 should be used as a comprehensive compendium of the results.

Flexible Ligands on Native Structures. We docked flexible ligands into the native structure of proteins from the Astex diverse set. Such a scenario is somewhat artificial but would be representative of the approximately 10% of binding-sites that do not undergo side-chain conformational changes upon ligand binding.[7] The performance of FlexAID (66.7%) is closest to the performance of FlexX (78.8%) when docking in native structures (Figure 4A). AutoDock Vina and rDock achieve higher success rates with 81.8% and 89.4%, respectively, in this specific scenario where side-chain flexibility is not required. The sampling success of FlexAID is 81.0% compared to 84.7% for FlexX, 97.4% for AutoDock Vina, and 98.8% for rDock. The slight discrepancies between our results and those previously reported by Ruiz-Carmona et al.[12] may be explained in part by how success rates are calculated (in particular the use of bootstrapping) and by differences in the datasets as 11 complexes from the Astex native set were absent in their study.

Analysis of Soft and Hard Failures in FlexAID. Some of our failures can be attributed to the search procedure, previously categorized as soft failures[37] (not to be confused with the concept of soft scoring functions). We observed that ligands where FlexAID failed to find the correct solution tend to have a larger number of flexible bonds (Figure S4F). Invariably in these cases, the CF value of the best-predicted pose (CF_{min}) is

DOI: 10.1021/acs.jcim.5b00078
J. Chem. Inf. Model. 2015, 55, 1323−1336

Table 2. Success Rates on the Top 1, 5, 10, and 100 Solutions on Different Docking Scenarios[a]

				ideal subset[b]				largest common subset[c]			
		scenario	program	1	5	10	100	1	5	10	100
ASTEX	native	FLRP	FlexAID	45.2%	65.5%	66.7%	81.0%	44.2%	64.9%	66.2%	79.2%
			Vina	76.6%	79.2%	81.8%	97.4%	76.6%	79.2%	81.8%	97.4%
			FlexX	63.5%	74.1%	78.8%	84.7%	63.6%	74.0%	77.9%	84.4%
			rDock	77.6%	87.1%	89.4%	98.8%	76.6%	85.7%	88.3%	98.7%
	non-native	FLRP	FlexAID	38.5%	48.1%	57.7%	80.8%	41.7%	52.8%	61.1%	77.8%
			Vina	41.7%	45.0%	48.3%	70.0%	44.4%	47.2%	50.0%	69.4%
			FlexX	41.3%	50.0%	54.3%	65.2%	41.7%	50.0%	55.6%	63.9%
			rDock	54.7%	64.1%	68.8%	89.1%	55.6%	63.9%	69.4%	94.4%
		FLFP	FlexAID	40.4%	53.8%	61.5%	80.8%	38.9%	52.8%	63.9%	80.6%
			Vina	46.7%	51.7%	53.3%	78.3%	50.0%	52.8%	55.6%	77.8%
			FlexX	NA	NA	NA	NA	NA	NA	NA	NA
			rDock	53.1%	68.8%	76.6%	89.1%	55.6%	69.4%	77.8%	94.4%
HAP2	native	FLRP	FlexAID	22.0%	35.6%	39.0%	59.3%	22.2%	35.2%	38.9%	59.3%
			Vina	52.5%	64.4%	67.8%	89.8%	50.0%	63.0%	66.7%	88.9%
			FlexX	30.4%	37.5%	44.6%	51.8%	29.6%	37.0%	44.4%	51.9%
			rDock	44.6%	64.3%	66.1%	85.7%	42.6%	63.0%	64.8%	85.2%
	non-native	FLRP	FlexAID	20.0%	22.9%	28.6%	45.7%	16.7%	22.2%	27.8%	38.9%
			Vina	7.5%	13.2%	15.1%	34.0%	5.6%	11.1%	16.7%	27.8%
			FlexX	3.4%	13.8%	24.1%	41.4%	0.0%	11.1%	27.8%	44.4%
			rDock	8.3%	20.8%	29.2%	54.2%	5.6%	16.7%	22.2%	66.7%
		FLFP	**FlexAID**	28.6%	40.0%	42.9%	62.9%	33.3%	38.9%	38.9%	55.6%
			Vina	15.9%	31.8%	38.6%	68.2%	16.7%	38.9%	44.4%	83.3%
			FlexX	NA	NA	NA	NA	NA	NA	NA	NA
			rDock	8.3%	18.8%	22.9%	50.0%	5.6%	16.7%	22.2%	50.0%

[a]Docking Scenarios: FL: Flexible Ligand; RP: Rigid Protein; FP: Flexible Protein. Results are not available for FlexX for FP as the program considers proteins as rigid and uses multiple protein conformations to account for protein flexibility. All success rates are bootstrapped over 10 000 iterations. [b]The ideal subset is the largest subset in which a given program is able to run without technical errors; it is a different subset for each program with the following number of targets for each program: FlexAID, AutoDock Vina, FlexX, and rDock. [c]The largest common subset of the Astex diverse set contains 77 targets. The largest common subset of the Astex non-native set contains 669 structures representing 36 unique targets. The largest common subset of the HAP2 native and non-native sets contain 54 targets and 18 targets.

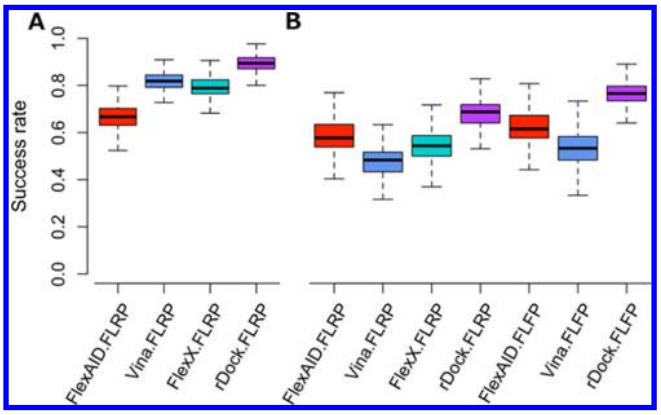


Figure 4. Comparison of the performance of FlexAID, AutoDock Vina, FlexX, and rDock for binding mode prediction. The performance of FlexAID (red), AutoDock Vina (blue), FlexX (cyan), and rDock (purple) were tested on the Astex diverse set (A) in the presence of ligand flexibility alone (denoted as FLRP) or presence of ligand and protein flexibility (denoted as FLFP). The programs were also compared on the Astex non-native set (B) in the absence (denoted as FLRP) or presence (denoted as FLFP) of protein flexibility and ligand flexibility. The cases represented are the ideal subset (see Methodology). The results are bootstrapped over 10 000 iterations.

always higher than that of the reference pose (CF_{ref}). Examples include the cases of JE2 (PDB ID 1KZK) with CF_{ref} −312.7 and CF_{min} −198.9, the case of BIR (PDB ID 1R1H) with CF_{ref} −212.3 and CF_{min} −121.8, and STI (PDB ID 1T46) with CF_{ref} −326.9 and CF_{min} −238.6.

Other failures are categorized as hard failures as they can be attributed to the scoring. For example, in the case of RQ3 (PDB ID 1G9V), the native pose of the ligand is quickly discarded with a considerably higher CF value (CF_{ref} −89.6) than the predicted pose (CF_{min} −191.2). In the case of BNE (PDB ID 1MZC) and AZM (PDB ID 1JD0), the Zn metal in coordination with the ligand is associated with large steric clashes as calculated in FlexAID in the native pose thus increasing the native CF score, a limitation for metals that we wish to address in the future in FlexAID. Another source of failed predictions are cases of poor parametrization. For example, in the case of AO5 (PDB ID 1R58), our failure could be attributed in part to the Cl atom of the ligand as well as the presence of 2 Mn^{2+} ions. The Cl atom is exposed to the solvent in the native pose and found in substantially less solvent exposed positions than the native pose (~70 Å^2 vs ~115 Å^2), as its exposure to solvent is highly penalized according to the parameter set. The native pose in the biological unit exposes to solvent a bulky hydrophobic group found buried in our predictions. In the native complex, the entropic penalty involved in exposing a hydrophobic group to the solvent is partially offset by the enthalpic gains through interactions with 2 Mn^{2+} ions in the binding-site. Unfortunately, our interaction matrix does not currently appropriately cover interactions with Mn due to a lack of examples of these interactions in the training data (Table S1). The parametrization of metals is a

DOI: 10.1021/acs.jcim.5b00078
J. Chem. Inf. Model. 2015, 55, 1323−1336

common issue when deriving scoring functions. To investigate its impact on docking performance, we calculated success rates on a subset of the Astex native set excluding cases with metals in the binding-site (PDB IDs 1GKC, 1HP0, 1HQ2, 1HWW, 1JD0, 1JJE, 1LRH, 1MZC, 1OQ5, 1R1H, 1R55, 1R58, 1UML, 1XM6, 1XOQ, 1YQY). As expected, the performance of all methods increases with success rates of 72.1%, 83.6%, 79.7%, and 91.3% (versus 66.7%, 81.8%, 78.8%, and 89.4% when including such cases) for FlexAID, AutoDock Vina, FlexX, and rDock.

To summarize, most hard failures involve atom types less frequently observed in biological molecules and are caused by the lack of training data given the nature of our approach. To account for the lack of examples in the training data, we mapped infrequent atom types to a frequent atom type with similar chemical properties or, if none was available, pairwise interactions involving the infrequent atom type were set to zero. The mapping of Zn to Mn allows the complex 1R58 to be properly docked with good ranking (Rank of 3). Predictions are also improved when interactions made with Cl atoms are made as neutral. For consistency, the success rates presented in each docking scenario were done using the original uncorrected pairwise energies and using the whole Astex dataset including cases with metals.

Considering the soft nature of our scoring function, one could expect that FlexAID would perform better on predominantly hydrophobic binding-sites where directionality is less important. We do not observe any clear correlation of the rank with the hydrophobicity index of the binding-site (Figures S5A-F).

Flexible Ligands on Rigid Non-Native Structures. We compared the performance of FlexAID to that of AutoDock Vina, FlexX, and rDock by docking into rigid non-native structures. This scenario is relevant to virtual screening in which protein flexibility is implicitly taken into account. For each method, the protein−ligand complexes that could not be successfully predicted in the native structures were not included (see Methodology). In other words, every method is only tested on cases in which it could successfully dock the ligand in the native conformation. Thus, the failures observed in this experiment are most likely attributable to the effect of conformational changes in the binding-site. When it comes to non-native-complex structures the success rate of AutoDock Vina and FlexX decrease considerably to a level below that of FlexAID that is minimally affected by the transition to rigid non-native-complex structures (four left-most bars in Figure 4B). FlexAID predicts 57.7% of the complexes when side-chain flexibility is excluded, while AutoDock Vina, FlexX, and rDock predict 48.3%, 54.3%, and 68.8%, respectively. A previous study showed that the software GOLD developed by Astex obtains a sampling success of 72% on their own Astex non-native-complex set.[16] Using the same (entire) Astex non-native-complex dataset, FlexAID obtains a top 100 bootstrapped sampling success of 67.2% when side-chain flexibility is excluded or 68.8% when included. However, as we were unable to obtain a free license for GOLD, we cannot ensure that the same methodology is being employed when comparing the two methods.

To further understand the sensitivity of the different methods, we calculated the success rate as a function of the magnitude of rearrangements observed between native and non-native forms. When docking on rigid non-native-complex structures of targets (Figure 5A), the performance of FlexX

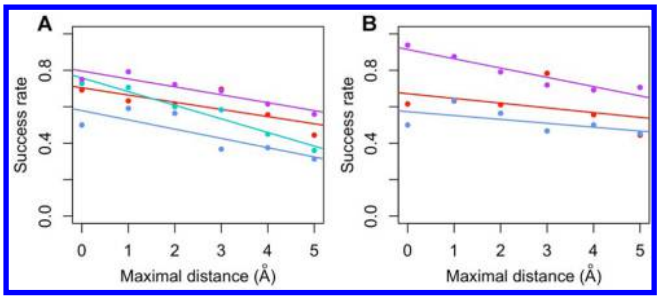


Figure 5. Impact of the magnitude of movements on the success rate. We compare the success rates of FlexAID (red), AutoDock Vina (blue), FlexX (cyan), and rDock (purple) as a function of the maximal displacement of any atom of the binding-site between apo/non-apo and holo forms of the Astex diverse sets. We compare the success rates between the methods in the absence (A) or presence (B) of protein flexibility. The following bins used maximal displacement: [0.0,1.0[, [1.0,2.0[, [2.0,3.0[, [3.0,4.0[, [4.0,5.0[, [5.0,d_{max}], where d_{max} is the maximum displacement observed across all pairs. Each point is a bootstrapped success rate over 10 000 iterations. The lines represent the linear regression of all points for a given method.

equates that of rDock for non-native structures that are closely structurally similar to the native form. However, despite the observed decrease in success rate for all programs as the magnitude of movements increase, FlexX has a steeper decreasing slope compared to those of FlexAID, AutoDock Vina, and rDock. The performance of FlexAID is always superior to that of AutoDock Vina and surpasses the one of FlexX on structures with maximal displacement greater than 1.5 Å but remains below that of rDock.

Flexible Ligands on Flexible Non-Native Structures. Next, we compared the performances of FlexAID, AutoDock Vina, and rDock when docking into flexible non-native structures. This is the most realistic of all scenarios as it is not only the situation likely to be found when the structure of a ligand-protein complex is unknown prior to the docking simulation, but one in which side-chains flexibility is included as part of the simulation. It is worth noting that the set of non-native structures includes structures that do not require side-chain movements to accommodate a ligand. FlexX considers the protein as rigid, thus we did not include it in experiments involving protein flexibility. A variant of the method called FlexE introduces protein flexibility by combining a set of up to 16 protein structures,[15] creating an ensemble that is used to combine structurally variable parts from each structure. The authors of FlexE reported using a dataset of 10 examples the same success rate as that of combining the results of FlexX from all alternative structures docked separately.[15] Therefore, for all purposes, we can consider that the results obtained in the previous section using FlexX on flexible ligands docked against rigid non-native structures reflect the results that would be obtained with FlexE.

The inclusion of side chain flexibility in FlexAID and AutoDock Vina increases success rates by approximately 5% to 61.5% and 53.3% respectively, compared to 76.6% for rDock. Both FlexAID and AutoDock Vina include side-chain flexibility and have similar decrease in success rates as a function of maximal distance, with FlexAID having better success rates throughout (Figure 5B). The slope of rDock is steeper, which could be explained in part by how rDock treats protein flexibility. This result suggests that FlexAID may have an advantage as the magnitude of protein movements required to

DOI: 10.1021/acs.jcim.5b00078
J. Chem. Inf. Model. 2015, 55, 1323−1336

accommodate the ligand increases. To test this hypothesis, we analyzed subsets of the Astex diverse set as a function of the number of side-chains with large unfavorable contributions to the CF function due to steric clashes when superimposing the holo and non-holo forms. We call such side-chains as critical as these must undergo movements to be able to accommodate the ligand in the crystallographically observed position.

When focusing on cases with at least one critical side-chain, the success rate of rDock decreases to 51.2% (on 482 apo/non-holo pairs representing 43 unique targets), while that of FlexAID is only mildly affected and higher than that of rDock at 54.5% or 57.6% (on 330 apo/non-holo pairs for 33 unique targets in both cases) when including protein flexibility either using rotamers or side-chain conformers, respectively. The differences in success rates become more apparent as the number of critical side-chains increases. Considering that rDock includes protein flexibility in a very restricted manner (rotation of polar hydrogen atoms), the high success rate in the presence of critical side-chain movements may be due to being able to accommodate the ligand in a slightly shifted position or altered conformation to avoid steric clashes with the protein atoms that should undergo movement while still maintaining an RMSD within the required cutoff for success.

The Critical Side-Chain Movement Dataset. We employ a subset of the non-redundant dataset of holo-apo protein pairs (HAP2) previously developed by our group[7] to determine if the effect observed above on cases with critical side-chains can be observed in a non-redundant dataset independent from the Astex dataset used above. Given the high level of manual curation to ensure that the dataset is non-redundant, this dataset can be directly used in the context of docking simulations. In what follows, we selected cases from the non-redundant SEQ subset of the HAP2 dataset that contains at least one critical side-chain movement for a total of 64 holo/apo protein pairs.

Similarly to the experiments described above, we docked against the native structures of HAP2 (Figure 6A). We obtain a success rate in the top 10 predictions of 39.0%, 44.6%, 67.8%, and 66.1% for FlexAID, FlexX, AutoDock Vina, and rDock, respectively. In terms of sampling success (top 100), we obtain 59.3%, 51.8%, 89.8%, and 85.7% for FlexAID, FlexX, AutoDock Vina, and rDock. These results obtained on the HAP2 dataset are considerably lower than those obtained with the Astex dataset in the equivalent scenario: (66.7%, 78.8%, 81.8%, and 89.4%) for the top 10 and (81.0%, 84.7%, 97.4%, and 98.8%) for the top 100. In particular, the sampling success of FlexAID and FlexX decreases considerably more in the HAP2 dataset than that of AutoDock Vina and rDock. It is unclear if the HAP2 dataset is an intrinsically more difficult dataset than the Astex dataset. Some of the difficulty may be due in part to the high degree of ligand flexibility (11 ligands with 10 flexible bonds or more).

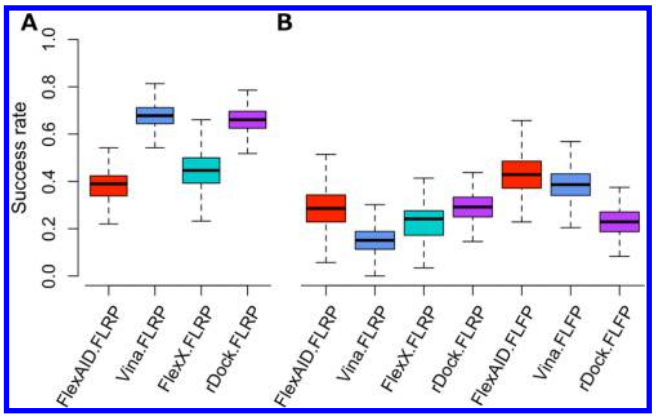


Figure 6. Comparison of the performance of FlexAID, AutoDock Vina, FlexX, and rDock for binding mode prediction. The performance of FlexAID (red), AutoDock Vina (blue), FlexX (cyan), and rDock (purple) were tested on a critical side-chains subset of the HAP2 database. We docked against holo (A) and apo (B) forms in the presence of ligand flexibility alone (denoted as FLRP) or the presence of ligand and protein flexibility (denoted as FLFP). The cases represented are the ideal subset (see Methodology). The results are bootstrapped over 10 000 iterations.

We also docked against non-native structures of the HAP2 dataset (four left-most bars in Figure 6B). We observe that FlexAID with 28.8% performs better than AutoDock Vina with 15.1% and FlexX with 24.1% in this particular scenario and equates the performance of rDock at 29.2%. FlexAID, AutoDock Vina, FlexX, and rDock obtain success rates significantly lower than those on the Astex non-native set (57.7%, 48.3%, 54.3%, and 68.8%).

The inclusion of protein flexibility (three right-most bars in Figure 6B) leads to an increase in success rates for FlexAID and AutoDock Vina to 42.9% and 38.6%, respectively. Intriguingly, the success rate of rDock in the non-native HAP2 critical subset in the presence of protein flexibility decreases to 22.9% compared to that what is observed in the absence of protein flexibility in the same dataset.

Whereas the ligands included in the Astex dataset are representative of drug-like molecules, it is unclear to what extent the Astex diverse set is representative of the true levels of critical side-chain movements. At the same time, the HAP2 subset of critical movement represents approximately 30% of cases of a non-redundant sample of binding-sites excluding nonspecific binders but not restricted to drug-like molecules. Our results in both datasets indicate that in cases where side-chain movements are present, FlexAID has a superior performance than AutoDock Vina and a clear advantage over rDock.

Virtual Screening. The analysis of a small-molecule docking program cannot be complete without the analysis of its performance on virtual screening. However, it has to be stressed that the prediction of binding poses as performed above and that of discriminating binding from non-binding small molecules are two distinct problems. Namely, it may not be possible to properly identify binders within a larger set of non-actives even if we are able to correctly predicting their binding pose. We evaluated the performance of FlexAID in virtual screening by docking against 38 targets of the DUD dataset[38] without including protein flexibility. We use the average AUC and enrichment factors at 1% of the ranked database (EF1%) as criteria of performance in virtual screening. On average, we have an AUC of 0.56 (and EF1% of 2.9) across all targets (Figures S7A-B), which closely matches results obtained from a previous study that used DOCK6.[11] Interestingly, despite the differences in the nature of the two algorithms, some of their best targets (GART, SAHH, RXR, and COX2) in terms of AUC are also targets where we had the most success where we obtained AUCs of 0.83, 0.81, 0.92, and 0.80, respectively. Using an AUC of 0.6 as a rough guideline for successful predictions over random, FlexAID and DOCK6 succeed for 16 targets of which 10 are shared. However, our

DOI: 10.1021/acs.jcim.5b00078
J. Chem. Inf. Model. 2015, 55, 1323−1336

virtual screening results are on average lower than the results reported[12] for rDock, Glide, and AutoDock Vina. These results were somehow expected considering the methodological approach used to derive the CF. As the CF was optimized for scoring appropriately native and near-native binding modes, this may come at the expense of enrichment as previously observed.[31] We therefore investigated if rescoring could improve the predictions of FlexAID in virtual screening. The rescoring of docking poses is a common practice to account for the possible caveats and biases of scoring functions.[39] We utilized the free knowledge-based scoring function RankScore as a rescoring method. On average, the AUC and EF1% increased to 0.62 and 4.9 when using RankScore[27] (Figures S6A-B) compared to 0.66 and 8.9, respectively, for AutoDock Vina, 0.69 and 11.4 for rDock, and 0.78 and 22.6 for Glide.[12] The size of the dataset used to obtain the virtual screening results for AutoDock Vina, rDock, and Glide is similar to ours, but 20 of the DUD targets were replaced in their case with DUD-E targets. Apart from this potential source of differences and given that results for FlexAID only were rescored, our results are comparable to those of AutoDock Vina and to some extent to those of rDock but still considerably lower than Glide.

Graphical User Interface. We developed a graphical user interface (GUI) called NRGsuite that is directly integrated into PyMOL to allow non-experts in the field to easily use FlexAID. Through the GUI, the users can easily set the target and ligand to be docked, define, measure, and partition the binding-site volume, view the simulation in real-time, etc. An extensive manual is provided in order to guide the users through the different steps. The NRGsuite is freely available, and its installation also installs FlexAID. The NRGsuite is compatible under the operating systems Linux, MacOS X, and Windows. A complete description of the NRGsuite features is beyond the scope of the present work and will be described elsewhere.

DISCUSSION

In this paper we present FlexAID, a docking method that uses a soft and smooth scoring function based on surfaces in contact. Our results show that FlexAID is a highly competitive docking program that performs better than several widely used methods in non-native-complex structures particularly when flexibility is crucial for binding. We compare FlexAID to commonly used (AutoDock Vina, FlexX) and state-of-the-art (rDock) software in the field as well as to reported results for Glide, GOLD, and DOCK6. The comparisons on the prediction of binding poses are carried out using the accepted *de facto* benchmarking Astex diverse set, which offers an independent dataset (except for GOLD developed by Astex) to test docking programs. Notably, comparisons to commercially available software that are widely used such as GOLD and Glide are severely restricted or entirely missing as we have no access to these software and must rely on published data where available. It would be useful if commercial software developers would at least make all data available for such benchmark datasets to allow more detailed comparisons even if independent validation is not possible.

While all current scoring functions used in major docking algorithms impose strict geometric constraints to define molecular interactions, our scoring function focuses on coarser features such as shape complementary. The simplicity of the representation is directly interconnected with the scoring methodology and offers an advantage in terms of accessibility in that PDB[40] structures can be used directly without requiring the addition of hydrogen atoms or the calculation of partial charges. Our scoring function, together with the unique way in which our pairwise energy terms were derived, contribute to produce better results in binding mode prediction than several widely used methods. As a matter of fact, in the most challenging and realistic of situations, when docking on non-native-complex structures, FlexAID performs better than the widely used programs AutoDock Vina and FlexX (and FlexE by extrapolation according to the conclusions of their authors[15]). FlexAID outperforms rDock when critical side-chain movements are necessary to accommodate ligand binding. One potential advantage of a program such as FlexAID with a soft, smooth scoring function over programs that utilize hard scoring functions is that the former are less sensitive to structural variations such as those that are common in homology models. A well-characterized phenomenon in docking programs with hard scoring functions is that the accuracy of the results is on average higher for holo form targets, followed by non-apo form targets and last homology models.[41] Given the soft and smooth scoring function in FlexAID, the program may be suited for binding mode predictions using homology models, but this remains to be tested.

Despite that more iterations were carried out to improve our scoring function (data not shown), we did not observe any further improvements in terms of success rate and ranking. We may have achieved a limit for the optimization of the scoring function with the set of atom types used. However, it may be possible to augment the number of atom types. In fact, inaccuracies in the Sybyl atom types have previously been discussed.[42] For instance, the O.3 atom type definition could be divided in at least 2 atom types as it accounts for both hydroxyl and ether groups. Such a separation is chemically meaningful as only hydroxyls can act as H-bond donors. Our scoring function would most likely be improved by extending the number of atom types as that would allow us to define more appropriately the chemical properties of atoms. Moreover, most of the hard failures observed are due to the absence or lack of sufficient training data (Table S1). Therefore, whereas in some cases it is chemically plausible to split an atom type into two, in other cases, it may be worth merging atom types with similar chemical properties due to the lack of sufficient data.

The interaction between atom types was derived employing an iterative procedure using both positive and negative information with large decoy sets. In the present study we utilize poses of a ligand with RMSD values higher than a given threshold (see Methodology) compared to its experimentally determined pose to define negative decoys. However, the combination of decoys for different ligands as negative decoys for each target may further improve virtual screening.[43] One advantage of a soft scoring function over a hard one is that it can implicitly account for protein flexibility. Despite the fact that none of the decoys included protein flexibility, our scoring function could still maintain good performance when docking into non-native proteins. Future decoys sets will include side-chain and backbone flexibility. As the values of smooth scoring functions do not drastically change with small rearrangements of the structure, achieving the global minimum of the CF in a traditional genetic algorithms docking simulation can be done within less energy evaluations than with hard scoring functions. All simulations with FlexAID were performed with 2×10^6 energy evaluations, while at least 10 times more energy evaluations were used in the case AutoDock Vina.[8] Despite that more time is required in computing surfaces in contact than simple distance calculations, the extra time required is offset by

DOI: 10.1021/acs.jcim.5b00078
J. Chem. Inf. Model. 2015, 55, 1323−1336

the smaller amount of energy evaluations required. Therefore, there appears to be a trade-off between the search and the scoring that can quickly become computationally intensive as more and more degrees of freedom are required to account for the stringent geometric constraints.

While our current implementation of side-chain flexibility has similarities to other methods, particularly to the one of AutoDock Vina,[8] the smoothness of our scoring function takes into account to some extent receptor plasticity implicitly, thus accounting for subtle side-chain movements described by the minimal rotation hypothesis. Such movements are expected to be accompanied by steeper energetic changes using hard scoring functions and would require additional flexible side-chains. Many popular docking algorithms such as FlexE[44] and ICM[44] treat protein flexibility using multiple receptor conformations (MRC). While MRC is not restricted to side-chain flexibility, the MRC approach is subjected to a combinatorial explosion of the search space. In these cases, a smooth scoring could implicitly account for subtle backbone movements and help restrict the size of the ensemble of conformations.

Given that the best RMSD obtained correlates with the number of rotatable bonds in the ligand, improvements in the methodology used to simulate ligand flexibility may increase the accuracy of FlexAID. Currently each flexible ligand bond represents a genetic algorithm variable sampling angles at 10° intervals, and intramolecular ligand interactions do not contribute to the scoring. A number of different possibilities may be explored in the future to address the gap in the loss of success rate such as finer sampling and inclusion of intramolecular ligand interactions in the CF function as well as the possibility of precomputing favorable ligand conformations.[45,46]

In terms of virtual screening, FlexAID requires rescoring to achieve comparable results to those of AutoDock Vina and rDock (although different datasets were used as noted, and a direct comparison may not be straightforward). The poses generated with FlexAID do not necessarily represent minima in the energy landscape of RankScore. Unfortunately, RankScore does not allow for performing local minimization of the structures, which could further improve the results.

Considering the differences between binding pose prediction and virtual screening and our emphasis in the development of the scoring function for the former, the results obtained are expected but may be further improved with changes in the scoring function discussed above. Lastly, it is important to keep in mind that the virtual screening experiments were performed with flexible ligands and rigid targets and improvements may be observed with the inclusion of target flexibility as it was previously observed that multiple receptor conformations lead to improvements.[47]

METHODS

Binding-Site Definition. The binding-site definition involves the detection and use of one or more cavities of the target molecule. A grid with spacing of 0.375 Å is built within the volume of each cavity to define the searchable area of the ligand where each grid vertex serves as an anchor point for the reference atom of the ligand. This discretization of space allows us to replace three translational degrees of freedom for a single variable. When the ligand is processed, a reference atom is internally defined and is normally identified as an atom within the largest rigid portion of the ligand. The ligand will not necessarily be fully contained within the volume of the binding-site, as only the reference atom is required to be anchored to a grid point. This permits to account for remodelling of the binding-site such as when side-chain flexibility is included while maintaining the originally defined grid.

Cavities are calculated using our own implementation of the flood-fill SURFNET algorithm[48] called GetCleft. Briefly, the method inserts spheres between each pair of atoms of the target and reduces their radii until there are no more clashes. A cleft is defined as an ensemble of connected spheres. This binding-site definition is similar to the definition in the program DOCK.[49] One or more cavities can be searched at the same time. The possibility to perform a global search against all cavities may be interesting when searching for binding hotspots[50] or druggable allosteric binding sites.

The Scoring Function. We use a modified form of the complementarity function to evaluate the energy of the complex.[21] A subset of 40 Sybyl atom types[51] is used to describe the chemical properties of atoms. We assign atom types using the open-source and freely accessible software Open Babel.[52] The van der Waals (VDW) radii of heavy atoms are expanded to implicitly account for hydrogen atoms.[53] The VDW radius of each atom is further expanded by the radius of a water molecule (1.4 Å) to calculate contact surfaces between atoms and solvent accessible surfaces (SAS) with the implicit solvent. We calculate surface areas in contact analytically using the constrained Voronoi procedure of McConkey et al.[54] The scoring function is given by

$$CF = \sum_{i=1}^{N}\sum_{j=1}^{M} \varepsilon_{i'j'} \times S_{ij} + \sum_{i=1}^{N} \varepsilon_{i'w} \times S_{iw} + K_{wall} \times \sum_{\substack{i=1 \\ i\neq j}}^{O}\sum_{j=1}^{O} f(i,j)\left[\left(\frac{1}{d_{ij}}\right)^{12} - \left(\frac{1}{P_e(r_i + r_j)}\right)^{12}\right] \quad (1)$$

where atoms i and j, with radii r_i and r_j loop through the M protein, N ligand and $O = M + N$ protein and ligand atoms, respectively. The interaction energy between atoms i and j of Sybyl atom types i' and j' is given by $\varepsilon_{i'j'}$ and modulated by their surface area in contact S_{ij}. Similarly to GOLD and DOCK, intramolecular interactions are omitted. The second term is used to simulate the hydrophobic effect, through an effective interaction energy between ligand atoms and the solvent $\varepsilon_{i'w}$ modulated by the solvent accessible area S_{iw} of each ligand atom. As this term involves only ligand atoms, we assume that different poses of the ligand desolvate the target equally.[55] The last term accounts for the repulsive VDW interactions. The constant $K_{wall} = 10^6$ is used to penalize steric clashes when the distance between atoms d_{ij} is smaller than the sum of their van der Waals radii. The permeability factor, $P_e = 0.9$, softens the potential to allow some receptor plasticity.[23] The repulsive VDW interactions are only calculated between atoms separated by at least 3 consecutive covalent bonds in the ligand or protein and between all ligand/protein atom pairs (defined as the function $f(i,j)$) similarly to other methods.[56] The function maximizes the area in contact between atoms with favorable interaction energies while minimizing solvent exposed areas and steric clashes.

The $\varepsilon_{i'j'}$ and $\varepsilon_{i'w}$ interaction energies are such that negative values represent favorable interactions and were obtained using an iterative optimization approach. A similar iterative approach was used to derive the ITScore potential.[57] Monte Carlo simulations were used to optimize the parameters such that

DOI: 10.1021/acs.jcim.5b00078
J. Chem. Inf. Model. 2015, 55, 1323−1336

they can discriminate, inasmuch as possible, between native and near-native binding modes (referred as true positives) from low-energy decoys (referred as false positives). During the Monte Carlo optimization, the probability of changing the value of a pairwise parameter was proportional to the frequency that it was observed in the PDBbind dataset used in training to guide the search more efficiently. We used the Area Under the Curve (AUC) of Receiver Operator Curves (ROC) as the objective optimization function. Independent datasets were used in the Monte Carlo optimizations and in the validation the resulting potentials. We iteratively used the best set of interactions to enrich the decoys set with harder false positive decoys (lower CF values) as well as increased the number of complexes (and their respective decoy sets) in each consecutive Monte Carlo optimization (Table 1).

Decoy sets for consecutive Monte Carlo iterations were generated using FlexAID with the best potential obtained in the previous iteration. We utilized the PDBbind[28] refined-set (release 2012) to predict the native binding modes of an ensemble of crystal structures representing protein−ligand complexes. The proteins were kept rigid, but ligands were fully flexible. We excluded complexes for which our method could not successfully predict the native binding mode. The remaining complexes (Table 1) had at least 2 true positives (the native pose of the ligand from the crystal and a near-native binding mode) and an unlimited number of false positives. Each complex utilized had its own decoys set representing different conformations and relative positions of the same ligand against its own receptor. We used thresholds of RMSD $\leq$ 1.5 Å for ligands with less than 15 heavy atoms[58] or RMSD $\leq$ 2.0 Å otherwise to define true positive decoys. Every decoy had to have an RMSD of at least 0.5 Å from any other decoy in its decoys set. Complexes in the PDBbind dataset that also appear in the Astex diverse or non-native sets were removed. Lastly, in order to prevent a bias toward interactions that are highly present in particular protein families, we grouped proteins by sequence identity to minimize redundancy. Each protein group was given an equal weight during the optimization. We aligned protein sequences using ClustalW.[59] Proteins belonged to the same group if they shared at least 30% sequence identity. The final set of pairwise energy terms is presented in Table S2.

General Features of FlexAID. FlexAID is a probabilistic docking program that optimizes the ligand-protein complex by minimizing the complementarity function using genetic-algorithms (GA) as search procedure. A ligand-protein conformation is referred as an individual within a population. Each individual is constituted of one chromosome. Each gene represents one optimization variable. Three genes account for the rotation of the ligand, and a single gene, that maps grid vertices of the binding-site, is used for translation. Each flexible ligand dihedral bond represents an extra gene. Furthermore, an extra gene is added for each flexible side-chain to map rotameric conformations. Briefly, the GA optimization works in 6 steps. 1. An initial population of individuals is generated randomly without any use of pre-existing coordinates/conformation of ligands. 2. The complementarity function (CF) evaluates the energy of the whole population. 3. The individuals are ranked according to a fitness function. 4. The fittest individuals, having a greater probability to reproduce, are selected pairwise to produce two new offspring. 5. A new population is created according to a reproduction technique. 6. Loop through steps 2 to 5 with the new population. The population converges toward a minimum as the generations increase. We implemented an adaptive GA to maintain diversity in the population and prevent early converge to local minima.[60] The probability of the genetic operators (mutation and crossover) increase when the population converges to a (potentially local) minimum to generate additional conformational diversity in the population and escape the minimum found. Given the properties of the selection technique (see below), a solution representing a minimun will remain in the population until a lower minimum is found.

The user can choose between linear or shared (default) fitness functions. Fitness functions are mappings that permit to weight individuals according to their CF values to further generate diversity in the population and prevent premature converge. For linear fitness, the fitness grows linearly as the CF decreases and acts as a scaling factor. The shared fitness function mimics nature, where individuals sharing a specific geographical niche (in our case similar pose) need to share resources. Internally, the fitness of individuals sharing similar search space is lowered, thus increasing the probability of sampling other, less populated regions of the space of solutions. Two different reproduction techniques are available to the user: Steady State and Population Boom (default). Briefly, Population Boom creates an entirely new population with N individuals according to the standard rules to create offspring then combines the old and new populations and selects the best fit N individuals out of the $2N$. If we denote the fitness of an individual i in generation t as f_i^t; whereas standard reproduction techniques ensure $\max(f_i^{t+1}) \geq \max(f_i^t)$, Population Boom ensures by definition that $\langle f \rangle^{t+1} \geq \langle f \rangle^t$. In plain words, instead of ensuring that a small number of the best-fit individual(s) are kept between two consecutive GA generations, Population Boom ensures that the average fitness of the whole population will not decrease through the generations.

FlexAID takes as input a configuration file and a GA parameters file. The GA file allows the user to adjust the population size and the number of generations to control the length of simulations, to control the parameters of the genetic operators to make them adaptive (default) or constant, and to change the reproduction technique and the fitness function. The configuration file is highly customizable and is used to specify the target and the ligand to be docked, define the binding-site, include extra degrees of freedom representing ligand and protein flexibility, and modify internal parameters of the program. The target and ligand files need to be processed with our auxiliary program Process_Ligand in order to derive the Sybyl atom types. PDB, MDL, and MOL2 formats are supported for the definition of the ligand and target. The processed ligand file encodes an internal coordinates system used to incrementally build the ligand into the binding-site. FlexAID generates a fixed but customizable number of results written in PDB format (default is 10). These represent a geometric clustering of the individuals sampled during the whole simulation. For example, in the hypothetical case where the entire population converges to a single solution at the end, other sampled good solutions that did not necessarily survive to the last generation are brought back to life and added to the list of the top 10 solutions according to their scoring. The user has the option to only output atoms that could move during the simulation or output the whole complex. The latter is essential if the user desires to rescore complexes with external scoring functions when side-chain flexibility is allowed, a missing feature even in modern docking suites.[12] One final character-

DOI: 10.1021/acs.jcim.5b00078
J. Chem. Inf. Model. 2015, 55, 1323−1336

istic of FlexAID is that in addition to proteins, also nucleic acids (both RNA and DNA) can be used as rigid targets.

Side-Chain Flexibility. FlexAID uses a flexibility probability scale for side-chain rotamer changes[7] (Table S3). The probability scale serves as a scaling factor to the genetic algorithm mutation operator. While the probability of undergoing conformational changes is used to bias the sampling of flexible residues, the user still needs to explicitly define the list of flexible side-chains. Our previous analysis[7] offers a number of pointers to choose the nature and number of flexible residues. Users should preferentially choose side-chains that are highly exposed to solvent and when working with X-ray structures, that have high b-factors. Moreover, we observed that a maximum of 5 flexible side-chains account for approximately 90% of binding-sites. Side-chains are modeled using conformers (actual instances of side-chain conformations) or rotamers as defined in the Penultimate Rotamer Library.[61] For any given amino acid, all side-chain rotamers are considered equally probable despite the distinct probabilities of rotamers.

Docking Experiments. We evaluate our method for the prediction of binding modes on 1) the Astex diverse set comprising 85 protein−ligand complexes, 2) the Astex non-native set comprising 1112 structures representing apo and non-apo forms of 65 proteins from the diverse set,[16,32] and on 3) a critical subset of the HAP2 dataset of 64 entries representing holo/apo protein pairs with respect to a given ligand.[7] We include side-chain flexibility when docking into the non-native forms of the proteins. AutoDock Vina and FlexAID describe each degree of freedom individually. For FlexAID and AutoDock Vina, the set of flexible side-chains in this experiment was restricted to those that are required to move to sterically accommodate ligand binding. Such critical side-chains are those where large steric repulsion was observed when the ligand is superimposed into the apo or non-apo forms as previously defined.[7] Preprocessed/prepared files ready for simulations for both datasets are available for download at our Web site (see Availability). It is possible that choosing a larger subset of flexible side-chains than exclusively those observed to be critical might affect the performance of AutoDock Vina and FlexAID. This could permit the placement of ligands within the RMSD cutoff that defines a successful docking as a result of spurious side-chain movements elsewhere. Thus, while the explicit selection of side-chains represents a bias on its own, it allows us to do a controlled experiment limiting the search space in a focused manner. The list of critical side-chains in each case can be found on our Web site (see Availability). In the case of rDock, the only type of protein flexibility built in is that of rotations of polar hydrogen atoms of binding-site side-chains. While the authors refer to these as protein flexibility, it is not equivalent in magnitude to the rotamer changes allowed in AutoDock Vina and FlexAID and represents an important limitation of rDock.

Considering that the native pose of ligands is known in the set of complexes used to validate the prediction of binding mode, the search space is reduced by restraining the volume of the binding-site to a given volume surrounding the ligand in the native pose. As previously observed, the performance of docking methods highly depends on how the binding-site is defined.[14] Ideally, one should select a binding-site definition that is equivalent, in terms of size of search space, for all methods evaluated. This is not however a straightforward task considering the differences in methodologies between the software. We tested multiple binding-site definitions on the Astex diverse set (Figure S7), and for the remaining experiments, we chose definitions that do not under- or overestimate the binding-site size as described in the next paragraph.

In the docking experiments described in this work, clefts defined with GetCleft were restricted to retain grid vertices within 5 Å of any atom of the ligand. We use the same binding-site definition when docking into the non-native forms as well as for the virtual screening experiments. AutoDock Vina defines a binding-site using a rectangular box.[8] In order to use a binding-site definition equivalent to that of AutoDock Vina, we defined the rectangular box as the smallest one that enclosed the cleft used in the FlexAID simulations. Although this binding-site definition may appear as a fair comparison in terms of search space, we obtain poor results with AutoDock Vina with this definition of binding-site considering that all ligand atoms need to be confined within the volume of the box during optimization. Therefore, we chose to expand the rectangular box by 10 Å in each dimension. With this we obtain significantly better performance, achieving results comparable to those obtained by others with AutoDock Vina.[12] Increasing the box by an extra 10 Å (20 Å in total) does not impact the success rate on the Astex diverse set but reduces the success rate on the HAP2 dataset (from 66.1% to 59.7%). The center and size of the boxes are available at our Web site (see Availability). Binding-sites in FlexX are defined as the set of protein atoms contained within 6.5 Å of the ligand. We define the binding-site for rDock using a cavity that extends up to 6.0 Å surrounding the reference ligand.

We use the threshold RMSD $\leq$ 2.0 Å as measure of success in binding mode prediction. RMSD values for FlexAID, AutoDock Vina, and FlexX were obtained from the pairwise comparison of atom IDs, while those of rDock take into account ligand topological symmetries, which could decrease their RMSD.

In general, whenever one has a sample of a population, bootstrapping[36] will generate better estimates of the population average and errors than a simple average and errors obtained from the sample itself. Ideally, the performance of a docking method should be evaluated on thousands of unique proteins. While the benchmark datasets used in this study represent an unbiased sample of proteins, they only cover a tiny fraction of biology. To derive better estimates of success rates we use bootstrapping with 10000 iterations. The error bars shown on Figures 1, 4, and 6 represent the standard errors of the estimates of bootstrapped success rates. The list of PDB entries used for bootstrapping in each of the scenario is explicitly described in the Supporting Information (Table S4) as well as on our Web site to facilitate parsing (see Availability). A program may fail on a particular target for technical reasons. Therefore, we bootstrapped the results for each program individually with its own subset of the Astex dataset to provide the best possible estimation of the success rate for each method. In the Results section, we also refer to the largest common subset, i.e. the subset comprising only cases that were docked with all programs. In the non-native docking experiments, for each method we remove from its own subset, proteins that could not be predicted in the native form. This allows us to exclude cases that may fail because of variables other than protein flexibility. We apply the latter filter as we feel that the purpose of non-native docking experiments is to assess the influence of protein flexibility. The Astex non-native set is redundant as it contains proteins that were crystallized more

DOI: 10.1021/acs.jcim.5b00078
J. Chem. Inf. Model. 2015, 55, 1323−1336

often than others. To account for this redundancy, the bootstrapping procedure is done in 2 separate steps with the selection of proteins in a first step, followed by the selection non-native conformers. The contribution of distinct numbers of non-native structures is normalized so that each protein contributes equally to the final bootstrap average.

We implemented a grid-computing framework using the BOINC infrastructure.[62] The project NRG@Home (for Najmanovich Research Group at Home) allows people around the world to contribute to our work by allowing their computers to perform FlexAID simulations using the BOINC screen saver. More information on the project and how to join is available at http://bcb.med.usherbrooke.ca/boinc.php. All docking experiments described in this work were run on NRG@Home with a fixed number of generations (2,000) and chromosomes (1,000) for a total of 2,000,000 energy evaluations. Due to the probabilistic nature of genetic algorithms as well as to ensure the return of results from BOINC users, we repeat each docking experiment for a given protein−ligand complex 10 times.

Preparation of Molecules and Execution. We processed and docked the Astex datasets as is no modification was applied to the provided structures as both proteins and ligands were manually curated by the original authors to include the assignments of protonation and tautomeric states and the addition of hydrogens as well as corrections of artifacts in the PDB.

The structures from the HAP2 critical subset were manually curated as to ensure the ligands were cognate or cognate-like with respect to their target. The apo and holo structures were processed independently using an automated procedure. The structures were preprocessed using REDUCE.[63] We used the latter to correct artifacts from X-ray crystal structures as well as to optimize the H-bond network surrounding the ligand allowing us to implicitly define the protonation and tautomeric states of the molecules. For the ligands, connectivity files including hydrogen atoms are provided in the wwPDB HET dictionary.

For all docking experiments presented in this study, the initial conformer of the proteins and ligands were used as input for all docking software evaluated. Water molecules were removed from the structures. The detailed pipeline for the preparation and processing of the molecules is shown as Supporting Information (Figure S8).

AutoDock Vina requires its input files to be in PDBQT format. The ligands and targets containing hydrogen atoms were prepared using the AutoDockTools utilities from MGLTools v1.5.6. We increased the exhaustiveness to 16 compared to the default value of 8 when docking with AutoDock Vina as previously done.[12] We ran FlexX with automatic selection of the base fragment (SELBAS a) and with standard optimization procedure (PLACEBAS 3). rDock was run with default parameters. When necessary, the molecule files were converted using OpenBabel[52] to the required format of the utilities and docking programs.

Availability. FlexAID and its accessory software (GetCleft and Process_Ligand) as well as the NRGsuite are free and open-source and are licensed under the GNU General Public License 3.0. All software is compatible under Windows, MacOS X, and Linux. The sources and precompiled bundles are available for download at http://bcb.med.usherbrooke.ca/FlexAID. Information is also provided describing all the necessary steps to execute FlexAID in a command-line manner as well as within the NRGsuite. The Web site above also contains the input files and relevant information, such as binding-site definitions, used to run the AutoDock Vina, FlexX, and rDock.

■ ASSOCIATED CONTENT

Supporting Information

Supplementary Figures S1−S8 and Tables S1−S3. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.5b00078.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rafael.najmanovich@usherbrooke.ca.

Author Contributions

F.G. and R.J.N. developed the method, designed the experiments, and wrote the manuscript. F.G. performed the calculations.

Funding

F.G. is the recipient of a doctoral Alexander Graham Bell PhD fellowship from the National Science and Engineering Research Council (NSERC) of Canada. This project was funded by NSERC Discovery Grant RGPIN-2014−05766.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are very grateful to all the BOINC users that volunteer in the NRG@Home project (http://boinc.med.usherbrooke.ca/nrg/) through the contribution of their CPU time to this work. R.J.N. is part of CR-CHUS, a member of the Institute of Pharmacology of Sherbrooke, PROTEO (the Québec network for research on protein function, structure and engineering) and GRASP (Groupe de Recherche Axé sur la Structure des Protéines).

■ REFERENCES

(1) Boehr, D. D.; Nussinov, R.; Wright, P. E. The Role of Dynamic Conformational Ensembles in Biomolecular Recognition. *Nat. Chem. Biol.* **2009**, *5*, 789−796.

(2) MacArthur, M. W.; Thornton, J. M. Protein Side-Chain Conformation: a Systematic Variation of Chi 1 Mean Values with Resolution - a Consequence of Multiple Rotameric States? *Acta Crystallogr. D* **1999**, *55*, 994−1004.

(3) Zhao, S.; Goodsell, D. S.; Olson, A. J. Analysis of a Data Set of Paired Uncomplexed Protein Structures: New Metrics for Side-Chain Flexibility and Model Evaluation. *PROTEINS: Structure, Function and Genetics* **2001**, *43*, 271−279.

(4) Gutteridge, A.; Thornton, J. Conformational Changes Observed in Enzyme Crystal Structures Upon Substrate Binding. *J. Mol. Biol.* **2005**, *346*, 21−28.

(5) Rubin, M. M.; Changeux, J. P. On the Nature of Allosteric Transitions: Implications of Non-Exclusive Ligand Binding. *J. Mol. Biol.* **1966**, *21*, 265−274.

(6) Najmanovich, R. J.; Kuttner, J.; Sobolev, V.; Edelman, M. Side-Chain Flexibility in Proteins Upon Ligand Binding. *PROTEINS: Structure, Function and Genetics* **2000**, *39*, 261−268.

(7) Gaudreault, F.; Chartier, M.; Najmanovich, R. J. Side-Chain Rotamer Changes Upon Ligand Binding: Common, Crucial, Correlate with Entropy and Rearrange Hydrogen Bonding. *Bioinformatics* **2012**, *28*, i423−i430.

(8) 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*, 455−461.
(9) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.
(10) Kramer, B.; Metz, G.; Rarey, M.; Lengauer, T. Ligand Docking and Screening with FlexX. *Medicinal Chemistry Research* **1999**, *9*, 463−478.
(11) Brozell, S. R.; Mukherjee, S.; Balius, T. E.; Roe, D. R.; Case, D. A.; Rizzo, R. C. Evaluation of DOCK 6 as a Pose Generation and Database Enrichment Tool. *J. Comput. Aided Mol. Des* **2012**, *26*, 749−773.
(12) 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*, e1003571.
(13) Jones, G.; Willett, P.; Glen, R.; Leach, A.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(14) Schulz-Gasch, T.; Stahl, M. Binding Site Characteristics in Structure-Based Virtual Screening: Evaluation of Current Docking Tools. *J. Mol. Model.* **2003**, *9*, 47−57.
(15) Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: Efficient Molecular Docking Considering Protein Structure Variations. *J. Mol. Biol.* **2001**, *308*, 377−395.
(16) Verdonk, M. L.; Mortenson, P. N.; Hall, R. J.; Hartshorn, M. J.; Murray, C. W. Protein-Ligand Docking Against Non-Native Protein Conformers. *J. Chem. Inf Model* **2008**, *48*, 2214−2225.
(17) Neves, M. A. C.; Totrov, M.; Abagyan, R. Docking and Scoring with ICM: the Benchmarking Results and Strategies for Improvement. *J. Comput. Aided Mol. Des* **2012**, *26*, 675−686.
(18) Novikov, F. N.; Stroylov, V. S.; Zeifman, A. A.; Stroganov, O. V.; Kulkov, V.; Chilov, G. G. Lead Finder Docking and Virtual Screening Evaluation with Astex and DUD Test Sets. *J. Comput. Aided Mol. Des* **2012**, *26*, 725−735.
(19) Repasky, M. P.; Murphy, R. B.; Banks, J. L.; Greenwood, J. R.; Tubert-Brohman, I.; Bhat, S.; Friesner, R. A. Docking Performance of the Glide Program as Evaluated on the Astex and DUD Datasets: a Complete Set of Glide SP Results and Selected Results for a New Scoring Function Integrating WaterMap and Glide. *J. Comput. Aided Mol. Des* **2012**, *26*, 787−799.
(20) Zavodszky, M.; Kuhn, L. Side-Chain Flexibility in Protein-Ligand Binding: the Minimal Rotation Hypothesis. *Protein Sci.* **2005**, *14*, 1104−1114.
(21) Sobolev, V.; Wade, R.; Vriend, G.; Edelman, M. Molecular Docking Using Surface Complementarity. *PROTEINS: Structure, Function and Genetics* **1996**, *25*, 120−129.
(22) Sobolev, V.; Edelman, M. Modeling the Quinone-B Binding Site of the Photosystem-II Reaction Center Using Notions of Complementarity and Contact-Surface Between Atoms. *Proteins* **1995**, *21*, 214−225.
(23) Ferrari, A. M.; Wei, B. Q.; Costantino, L.; Shoichet, B. K. Soft Docking and Multiple Receptor Conformations in Virtual Screening. *J. Med. Chem.* **2004**, *47*, 5076−5084.
(24) Samudrala, R.; Levitt, M. Decoys "R" Us: a Database of Incorrect Conformations to Improve Protein Structure Prediction. *Protein Sci.* **2000**, *9*, 1399−1401.
(25) Tobi, D.; Bahar, I. Optimal Design of Protein Docking Potentials: Efficiency and Limitations. *Proteins* **2006**, *62*, 970−981.
(26) Shoichet, B. K.; Kuntz, I. D. Protein Docking and Complementarity. *J. Mol. Biol.* **1991**, *221*, 327−346.
(27) Fan, H.; Schneidman-Duhovny, D.; Irwin, J. J.; Dong, G.; Shoichet, B. K.; Sali, A. Statistical Potential for Modeling and Ranking of Protein-Ligand Interactions. *J. Chem. Inf Model* **2011**, *51*, 3078−3092.
(28) Wang, R.; Fang, X.; Lu, Y.; Yang, C.-Y.; Wang, S. The PDBbind Database: Methodologies and Updates. *J. Med. Chem.* **2005**, *48*, 4111−4119.
(29) Tanford, C. The Hydrophobic Effect and the Organization of Living Matter. *Science* **1978**, *200*, 1012−1018.
(30) Najmanovich, R. J. Side Chain Flexibility Upon Ligand Binding: Docking Predictions and Statistical Analysis. PhD Thesis (Feinberg Graduate School, Weizmann Institute of Sciences) *arXiv:13014564* 2004.
(31) Irwin, J. J. Community Benchmarks for Virtual Screening. *J. Comput. Aided Mol. Des* **2008**, *22*, 193−199.
(32) 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*, 726−741.
(33) Cole, J. C.; Murray, C. W.; Nissink, J. W. M.; Taylor, R. D.; Taylor, R. Comparing Protein-Ligand Docking Programs Is Difficult. *Proteins* **2005**, *60*, 325−332.
(34) Kroemer, R. T.; Vulpetti, A.; McDonald, J. J.; Rohrer, D. C.; Trosset, J.-Y.; Giordanetto, F.; Cotesta, S.; McMartin, C.; Kihlén, M.; Stouten, P. F. W. Assessment of Docking Poses: Interactions-Based Accuracy Classification (IBAC) Versus Crystal Structure Deviations. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 871−881.
(35) 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.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.
(36) Davison, A. C. *Bootstrap Methods and Their Application*; Cambridge University Press: Cambridge, 1997.
(37) Verkhivker, G.; Bouzida, D.; Gehlhaar, D.; Rejto, P.; Arthurs, S.; Colson, A.; Freer, S.; Larson, V.; Luty, B.; Marrone, T.; Rose, P. Deciphering Common Failures in Molecular Docking of Ligand-Protein Complexes. *J. Comput. Aided Mol. Des* **2000**, *14*, 731−751.
(38) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, 6789−6801.
(39) Wang, R.; Lu, Y.; Wang, S. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. *J. Med. Chem.* **2003**, *46*, 2287−2303.
(40) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucl. Acids. Res.* **2000**, *28*, 235−242.
(41) McGovern, S. L.; Shoichet, B. K. Information Decay in Molecular Docking Screens Against Holo, Apo, and Modeled Conformations of Enzymes. *J. Med. Chem.* **2003**, *46*, 2895−2907.
(42) Neudert, G.; Klebe, G. DSX: a Knowledge-Based Scoring Function for the Assessment of Protein−Ligand Complexes. *J. Chem. Inf Model* **2011**, *51*, 2731−2745.
(43) Pham, T. A.; Jain, A. N. Parameter Estimation for Scoring Protein-Ligand Interactions Using Negative Training Data. *J. Med. Chem.* **2006**, *49*, 5856−5868.
(44) Cole, J. C.; Murray, C. W.; Nissink, J. W. M.; Taylor, R. D.; Taylor, R. Comparing Protein-Ligand Docking Programs Is Difficult. *Proteins* **2005**, *60*, 325−332.
(45) Lorber, D. M.; Shoichet, B. K. Flexible Ligand Docking Using Conformational Ensembles. *Protein Sci.* **1998**, *7*, 938−950.
(46) Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P.; Miller, M. D. Flexibases: a Way to Enhance the Use of Molecular Docking Methods. *J. Comput. Aided Mol. Des* **1994**, *8*, 565−582.
(47) Cavasotto, C. N.; Kovacs, J. A.; Abagyan, R. Representing Receptor Flexibility in Ligand Docking Through Relevant Normal Modes. *J. Am. Chem. Soc.* **2005**, *127*, 9632−9640.
(48) Laskowski, R. Surfnet - a Program for Visualizing Molecular-Surfaces, Cavities, and Intermolecular Interactions. *J. Mol. Graphics* **1995**, *13*, 323−330.
(49) Shoichet, B.; Bodian, D.; Kuntz, I. Molecular Docking Using Shape Descriptors. *J. Comput. Chem.* **1992**, *13*, 380−397.
(50) Landon, M. R.; Lancia, D. R.; Yu, J.; Thiel, S. C.; Vajda, S. Identification of Hot Spots Within Druggable Binding Regions by Computational Solvent Mapping of Proteins. *J. Med. Chem.* **2007**, *50*, 1231−1240.

DOI: 10.1021/acs.jcim.5b00078
J. Chem. Inf. Model. 2015, 55, 1323−1336

(51) Clark, M.; Cramer, R. D.; Van Opdenbosch, N. Validation of the General Purpose Tripos 5.2 Force Field. *J. Comput. Chem.* **1989**, *10*, 982−1012.
(52) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: an Open Chemical Toolbox. *Journal of Cheminformatics* **2011**, *3*, 33.
(53) Tsai, J.; Taylor, R.; Chothia, C.; Gerstein, M. The Packing Density in Proteins: Standard Radii and Volumes. *J. Mol. Biol.* **1999**, *290*, 253−266.
(54) McConkey, B. J.; Sobolev, V.; Edelman, M. Quantification of Protein Surfaces, Volumes and Atom-Atom Contacts Using a Constrained Voronoi Procedure. *Bioinformatics* **2002**, *18*, 1365−1373.
(55) Shoichet, B.; Leach, A.; Kuntz, I. Ligand Solvation in Molecular Docking. *PROTEINS: Structure, Function and Genetics* **1999**, *34*, 4−16.
(56) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved Protein-Ligand Docking Using GOLD. *Proteins* **2003**, *52*, 609−623.
(57) Huang, S.-Y.; Zou, X. An Iterative Knowledge-Based Scoring Function to Predict Protein-Ligand Interactions: I. Derivation of Interaction Potentials. *J. Comput. Chem.* **2006**, *27*, 1866−1875.
(58) Verdonk, M. L.; Giangreco, I.; Hall, R. J.; Korb, O.; Mortenson, P. N.; Murray, C. W. Docking Performance of Fragments and Druglike Compounds. *J. Med. Chem.* **2011**, *54*, 5422−5431.
(59) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. Clustal W and Clustal X Version 2.0. *Bioinformatics* **2007**, *23*, 2947−2948.
(60) Srinivas, M.; Patnaik, L. M. Adaptive Probabilities of Crossover and Mutation in Genetic Algorithms. *Systems, Man and Cybernetics, IEEE Transactions on* **1994**, *24*, 656−667.
(61) Lovell, S.; Word, J.; Richardson, J.; Richardson, D. The Penultimate Rotamer Library. *Proteins* **2000**, *40*, 389−408.
(62) Anderson, D. P. BOINC: a System for Public-Resource Computing and Storage. Fifth IEEE/ACM International Workshop on Grid Computing. *IEEE* **2004**, 4−10.
(63) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and Glutamine: Using Hydrogen Atom Contacts in the Choice of Side-Chain Amide Orientation. *J. Mol. Biol.* **1999**, *285*, 1735−1747.

DOI: 10.1021/acs.jcim.5b00078
J. Chem. Inf. Model. 2015, 55, 1323−1336