Ligand−Protein Docking with Water Molecules

Benjamin C. Roberts[†] and Ricardo L. Mancera*[,†,‡]

Western Australian Biomedical Research Institute, School of Pharmacy, and School of Biomedical Sciences, Curtin University of Technology, GPO Box U1987, Perth WA 6845, Australia

Received August 3, 2007

The presence of water molecules plays an important role in the accuracy of ligand−protein docking predictions. Comprehensive docking simulations have been performed on a large set of ligand−protein complexes whose crystal structures contain water molecules in their binding sites. Only those water molecules found in the immediate vicinity of both the ligand and the protein were considered. We have investigated whether prior optimization of the orientation of water molecules in either the presence or absence of the bound ligand has any effect on the accuracy of docking predictions. We have observed a statistically significant overall increase in accuracy when water molecules are included during docking simulations and have found this to be independent of the method of optimization of the orientation of water molecules. These results confirm the importance of including water molecules whenever possible in a ligand−protein docking simulation. Our findings also reveal that prior optimization of the orientation of water molecules, in the absence of any bound ligand, does not have a detrimental effect on the improved accuracy of ligand−protein docking. This is important, given the use of docking simulations to predict the binding modes of new ligands or drug molecules.

INTRODUCTION

When a protein molecule is in solution, its entire surface is covered by water molecules with properties different to those in the bulk.[1] Cryogenic X-ray determinations and molecular dynamics simulations reveal the existence of large networks of water molecules around the surface of proteins.[2−4] Most of the water molecules in the vicinity of a protein are loosely bound to it, remain mobile, have short interaction times, and are not readily observed via conventional X-ray crystallography. Although some water molecules that are observed in crystal structures are artifacts of the determination,[5] others are clearly tightly bound to the protein surface,[6−8] particularly in clefts on the protein surface, such as ligand binding sites.[9]

Water molecules observed in the crystal structures of proteins have a tendency to occupy conserved positions in structurally related proteins,[10−22] as well as in structures obtained under different conditions[2,23] and/or different bound ligands.[24−29] The most frequent structural change among structurally related ligands bound to the same protein seems to involve different arrangements of water molecules.[30]

The importance of water molecules found in the binding site of a protein lies in their ability to mediate the interactions between the ligand and the protein and form hydrogen-bonded networks that can stabilize a protein−ligand complex in solution.[10,31−39] Such a hydrogen-bonded network of water molecules may stabilize the complex formed with one ligand but not another, thus contributing to the specificity of ligand recognition.[33,40] Water molecules may also help to stabilize the conformation of the active sites of enzymes.[41] Water molecules have also been used to improve the predictive ability and rationalization of three-dimensional quantitative structure−activity relationship (QSAR) models[42,43] provide a structural rationale for ligand-derived pharmacophore models of binding sites,[44−47] and improve the performance of virtual screening.[48,49]

The binding of a ligand to a protein receptor often involves the thermodynamically favorable release of water molecules from the protein surface to the bulk solvent. However, the retention of water molecules that are bound tightly to the protein surface upon ligand binding may be associated with an entropic penalty that is outweighed enthalpically, through favorable hydrogen-bonding interactions to both the protein and the ligand.[50−57] Computer simulations have been used successfully to compute the free-energy changes associated with hydrating binding sites and displacing water molecules upon ligand binding,[53−55,57−63] as well as the binding of tightly bound water molecules.[64] In addition, several approaches have been taken to predict hydration sites,[65−68] conserved water mediated and polar ligand interactions,[69−71] water occupancy,[72] and the displacement of tightly bound water molecules.[62,73,74]

The importance of water molecules is now recognized in structure-based drug design, where the displacing, mimicking, and/or targeting of bound water molecules is performed to improve the binding affinity of ligand molecules.[75] The displacement and mimicking of tightly bound water molecules may result in increased binding affinity through the entropy gain of releasing such ordered water molecules.[76,77] However, this does not always seem to be true,[78] and, in some cases, the recruitment of an additional tightly bound water molecule that can bridge the interactions between the ligand and the protein has been determined to decrease the

* Author to whom correspondence should be addressed. Phone: +61 8 92661017. Fax: +61 8 92662342. E-mail: R.Mancera@curtin.edu.au.

[†] School of Pharmacy.

[‡] Western Australian Biomedical Research Institute and School of Biomedical Sciences.

10.1021/ci700285e CCC: $40.75
Published on Web 01/23/2008

binding affinity of ligands.[79] It has also been observed that natural substrates[80] and designed inhibitors[81] may not necessarily displace tightly bound water molecules but rather preserve water-mediated interactions. In some cases, water molecules have been observed not to mediate any ligand−protein interactions but simply to better define the steric shape of the binding site.[82] The consideration of tightly bound water molecules in de novo drug design methods has also demonstrated their role in modulating the binding modes and chemical diversity of designed ligands by imposing steric and hydrogen-bonding constraints.[83−85]

A few methods have been developed to place water molecules during ligand−protein docking simulations, with some success.[86,87] However, various studies have provided no conclusive evidence that the inclusion of tightly bound water molecules improves docking accuracy,[45,48,49,88−92] mostly because not enough protein targets were investigated to achieve statistical significance. Nonetheless, the first comprehensive docking studies with large datasets of ligand−protein complexes were contradictory, either revealing that including water molecules did not increase accuracy[93] or showing that it did.[94]

Recent studies attempted to assess the accuracy of ligand−protein docking in the presence of tightly bound water molecules.[95] The binding modes of various ligands of cytochrome P450 and thymidine kinase were predicted using three different docking programs, with and without water molecules. The positions of water molecules were obtained directly from the known crystal structures and from predictions using a novel GRID-based method.[95] Docking accuracy improved in the presence of crystallographic water molecules, with a larger improvement measured when predicted water molecules were included. Improvements were also detected in the accuracy of virtual screening.[96] Water molecules were observed not only to mediate the interactions between ligands and proteins, but also to help in the placement of ligands close to the center of the active sites.[96]

A simple entropy penalty term has been introduced to account for the unfavorable loss of rotational and translational entropy that accompanies the tight binding of a water molecule to a protein surface.[97] This penalty term was used to predict the displacement of tightly bound water molecules upon ligand binding. An improvement in the accuracy of ligand−protein docking was observed for a large dataset of ligand−protein complexes.[97] This effect was most noticeable in those cases where water molecules were determined to mediate the ligand−protein interaction.

In another study with a large set of protein−ligand complexes, the inclusion of all crystallographic water molecules within 6.0 Å of any ligand atom resulted in a large increase in the docking accuracy.[98] However, as the authors acknowledged themselves, by indiscriminately including all water molecules in the binding site, the search space may have been drastically biased, virtually leaving the correct binding mode as the only possible low-energy solution.[98]

We have performed a comprehensive survey of the role of water molecules on the accuracy of ligand−protein docking simulations by expanding the number of ligand−protein complexes considered to include all those in the original CCDC/Astex test set[93] that contain water molecules in their crystal structures. This ensures a more-thorough examination of the effect of including water molecules on the accuracy of predictions of binding modes by a standard docking/scoring strategy. We also focus on two particular issues: (1) the influence of the method chosen to optimize the orientation of water molecules prior to docking, and (2) the inclusion of only those water molecules that are in the immediate vicinity of both the ligand and the protein.

MATERIALS AND METHODS

Dataset and Preparation of Ligand−Protein Structures. The original CCDC/Astex dataset[93] defines a subset of "clean" ligand−protein complexes, after identifying and removing structures that have factual and structural errors, ligands with unlikely conformations or that have been determined inconsistently with respect to their electron density, severe steric clashes between any ligand and protein atoms, and ligand contacts to crystallographically related protein residues. The resulting "clean" subset contains 224 ligand−protein complexes and has been subsequently filtered based on the crystallographic resolution (R) of the protein structures. Two new subsets were defined: the first contained 180 structures that have a resolution of 2.5 Å or better, and the second subset contained 92 structures that have a resolution of 2.0 Å or better.[93]

All ligand−protein complexes from the CCDC/Astex dataset that contain water molecules in the binding site were considered. To select these water molecules, a 2.5 Å cutoff from any ligand atom and a 3.0 Å distance cutoff from any protein atom were jointly applied. These criteria excluded not only those water molecules that do not interact with the ligand, but also those water molecules located in the vicinity of the ligand molecule but too far from the protein surface to interact with both. Such water molecules "capping" the ligand are not deemed to be tightly bound to the protein surface and directly participating in the ligand−protein interaction. The above criteria also ensure that only those water molecules in the immediate vicinity of a ligand molecule are considered, with the vast majority of them forming putative hydrogen bonds with the ligand.

A total of 242 of the 305 entries in the CCDC/Astex dataset contain water molecules in the binding site that meet the aforementioned criteria. Of these 242 entries, 180 fall within the "clean" category, 154 have a crystallographic resolution of 2.5 Å or better, and 79 have a crystallographic resolution of 2.0 Å or better. Table 1 provides an incremental list of PDB codes in each subset.

Several ligand−protein complexes (PDB codes: 1bgo, 1tpp, and 3gch) have covalently bound ligands. In these cases, the mediating atom in the covalent bond was retained in both the ligand and the protein molecule, because any steric effects that might affect docking accuracy would equally affect docking with and without water molecules.

Five ligand−protein complexes (PDB codes: 1htf, 2mip, 4phv, 5abp, and 6abp) featured alternative binding modes for their ligands. The ligands in structures 1htf and 2mip exhibited notably different binding modes and were not included in our study, because it is unclear with which ligand binding mode the water molecules found in the crystal structure should be considered. On the other hand, while each of the ligands in structures 4phv, 5abp, and 6abp have two binding modes, they are in very close proximity. Hence, these structures were retained in our study and only one ligand

Table 1. List of Protein Data Bank (PDB) Codes of Ligand−Protein Complexes That Contain Water Molecules in the Binding Site in Each Subset Considered[a]

complex	PDB codes
Resolution ≤ **2.0 Å subset** (79 entries)	1a28 1a4q 1a6w 1abe 1abf 1aec 1aoe 1apt 1apu 1aqw 1atl 1b58 1b59 1bma 1byb 1c5c 1c5x 1c83 1cbs 1cil 1coy 1d0l 1d3h 1ejn 1f3d 1flr 1glp 1glq 1hfc 1hpv 1hsb 1hsl 1hyt 1ida 1jap 1lcp 1lic 1lna 1lst 1mld 1mmq 1mrg 1mrk 1mts 1nco 1ppc 1pph 1rds 1rnt 1rob 1slt 1snc 1srj 1tmn 1tng 1tnh 1tni 1tnl 1tpp 1ukz 1vgc 1wap 1xid 1xie 2ak3 2cmd 2ctc 2fox 2gbp 2qwk 2tmn 2tsc 3cla 3ert 3tpi 4dfr 5abp 6rnt 7tim
Resolution ≤ **2.5 Å subset** (154 entries)	All above plus: 1a42 1a4g 1aco 1ai5 1azm 1b9v 1bbp 1bgo 1blh 1byg 1cbx 1cdg 1ckp 1cle 1com 1cps 1cvu 1d4p 1dd7 1dg5 1dhf 1dog 1dr1 1dy9 1ebg 1eed 1eil 1eoc 1epo 1etr 1ets 1f0r 1f0s 1fkg 1fl3 1frp 1hiv 1hos *1htf* 1imb 1lah 1lpm 1mdr 1mup 1ngp 1nis 1okl 1okm 1pdz 1poc 1ppi 1pso 1ptv 1qcf 1qpe 1qpq 1rne 1tlp 1trk 1ydr 1ydt 1yee 25c8 2aad 2ack 2ada 2ifb 2pcp 3erd 3gpb 4aah 4lbd 4phv 5erl 6rsa
Clean subset (180 entries)	All above plus: 1a9u 1aaq 1acj 1acl 1acm 1bl7 1cl2 1dbj 1did 1dwb dwc 1dwd 1ett 1fgi 1hak 1ivq 1lyb 1lyl 1rt2 1tdb 1uvs 1uvt 1phh 2ypi 4cts 4fbp
Parent set (242 entries)	All above plus: 1a07 1a1b 1a1e 1a4k 1aha 1ake 1b6n 1c2t 1cf8 1cin 1ctt 1die 1ela 1elb 1elc 1eld 1ele 1fbl 1fig 1ghb 1gpy 1hef 1icn 1ivd 1jao 1lkk 1lmo 1ml1 1mmb 1mnc 1mtw 1nsd 1pgp 1pha 1ppl 1qh7 1ql7 1rbp 1srf 1srg 1srh 1stp 1tph 1xkb 1yds 2cgr 2er7 *2mip* 2plv 2r04 2sim 3gch 3mth 3nos 3ptb 3er2 4tpi 5p2p 6abp 6cpa 7cpa 8gch

[a] Each subset is listed incrementally. Entries shown in italics (1htf and 2mip) were subsequently removed due to the presence of two alternative binding modes (see text).

binding mode was arbitrarily selected, because interactions with the water molecules would be almost identical. Overall, the total number of ligand−protein complexes with water molecules considered was thus reduced from 242 to 240, the number of complexes considered in the clean subset was reduced from 180 to 179, the number of complexes considered in the subset with $R < 2.5$ Å was reduced from 154 to 153, and the number of complexes considered in the subset with $R < 2.0$ Å remained the same at 79.

All selected structures were prepared for the docking simulations. Hydrogen atoms were added using Sybyl 7.2 (Tripos, Inc.) and considering a pH value of 7.0. The protonation state of histidine residues in the vicinity of the binding site of each protein was checked for consistency with the original CCDC/Astex dataset, which had these assigned manually. The orientations of all selected water molecules in the binding site of each ligand−protein complex were optimized by energy minimization using the Tripos force field in SYBYL 7.2 (Tripos, Inc.), with an initial simplex minimization followed by the Powell method until the energy gradient was ≤ 0.001 kcal mol^{-1} Å^{-1}. The positions of the water oxygen atoms, as observed in the crystal structures, were kept fixed, and only water hydrogen atoms were allowed to move after being assigned random orientations. The optimization of the orientations of water molecules was performed in two different ways. In the first approach, the ligand molecule was left in the binding site in its crystallographic position during energy minimization, hence biasing the orientations of water molecules to form optimal hydrogen bonds with both the ligand and the protein. In the second approach, the ligand was removed from the binding site prior to the optimization, hence allowing only the characteristics of the protein surface to determine the optimal orientations of water molecules.

Ligand−Protein Docking and Scoring. The docking of ligands to their protein receptors was performed using an implementation of the stochastic tunneling method,[99,100] with the Piecewise Linear Potential (PLP)[101] scoring function being used to represent the potential energy surface of the ligand−protein interaction. A bounding simulation box was defined around the binding site of each protein by considering a distance of 3.0 Å away from the outermost atoms of the bound ligand in each the *x*-, *y*-, and *z*-directions in the bound conformation of the ligand. A total of 300 simulations of each ligand−protein complex were performed, and all resulting binding poses were re-scored using the ScreenScore scoring function,[102] which is consistent with our previous docking studies using the CCDC/Astex dataset.[100]

The predictive ability of the docking simulations was assessed by determining the root mean square deviation (RMSD) of the best energy of binding obtained (the top ranked solution). Consistently with the original CCDC/Astex dataset[93] and our own previous work,[100] the following qualitative classification of the accuracy of the docking solutions was applied: a good solution had RMSD ≤ 1.0 Å, a close solution had 1.0 Å < RMSD ≤ 2.0 Å, a solution with errors had 2.0 Å < RMSD ≤ 3.0 Å, and a wrong solution had RMSD > 3.0 Å.

Three sets of ligand−protein docking simulations were conducted: a set of benchmark/control simulations were done in the absence of any water molecules, and two sets of simulations were done including water molecules optimized by each of the two approaches previously described. The free energy of binding of the top-ranked solution and its associated RMSD (with respect to the crystallographic binding mode), as well as the best RMSD observed for a binding pose during the simulations and its associated free energy of binding, were recorded for each ligand−protein complex. Full tables of results for each ligand−protein complex are available as Supporting Information.

A statistical analysis was then conducted to compare the performance of the docking simulations in the absence and presence of water molecules. Both the average best RMSD achieved and the average RMSD of the top-ranking solution of all ligand−protein complexes were determined for each set of simulations and compared across each set for statistical performance using a *t*-test.

RESULTS AND DISCUSSION

Table 2 summarizes the differences in the efficacy of the docking search between the three scenarios previously described: no water molecules (denoted as NoW), water molecules optimized in the presence of the ligand (denoted as WL), and water molecules optimized without the ligand (denoted as W). The results show a small but significant

Table 2. Statistical Testing for the Difference between Mean Best RMSD Achieved for the Entire Dataset ($n = 240$) in the Absence of Water (NoW), in the Presence of Water Molecules Optimized with the Ligand Present (WL), and in the Presence of Water Molecules Optimized without the Ligand (W)

parameter	NoW	WL	W
mean RMSD (Å)	0.85	0.70	0.72
T statistic		6.67	5.76
$P (T \leq t)$		1.78×10^{-10}	2.55×10^{-8}
T(critical)		2.60	2.60

Table 3. Mean Best RMSD Achieved for Each Subset in the Absence of Water (NoW), in the Presence of Water Molecules Optimized with the Ligand Present (WL) and in the Presence of Water Molecules Optimized without the Ligand (W)

	Mean Best RMSD (Å)		
	NoW	WL	W
Parent set ($n = 240$)	0.85	0.70	0.72
Clean subset ($n = 179$)	0.77	0.64	0.62
Resolution ≤ **2.5 Å subset (**$n = 153$)	0.78	0.63	0.63
Resolution ≤ **2.0 Å subset (**$n = 79$)	0.78	0.57	0.57

Table 4. Statistical Testing for Difference between Mean RMSD of Top Energy-Ranked Solutions for the Entire Dataset ($n = 240$) in the Absence of Water (NoW), in the Presence of Water Molecules Optimized with the Ligand Present (WL), and in the Presence of Water Molecules Optimized without the Ligand (W)

parameter	NoW	WL	W
mean RMSD (Å)	2.49	1.83	1.91
T statistic		5.38	4.08
$P (T \leq t)$		1.77×10^{-7}	6.25×10^{-5}
T(critical)		2.60	2.60

Table 5. Mean RMSD of Top Energy-Ranked Solutions for Each Subset in the Absence of Water (NoW), in the Presence of Water Molecules Optimized with the Ligand Present (WL) and in the Presence of Water Molecules Optimized without the Ligand (W)

	Mean RMSD (Å)		
	NoW	WL	W
Parent set ($n = 240$)	2.49	1.83	1.91
Clean subset ($n = 179$)	2.27	1.66	1.71
Resolution ≤ **2.5 Å subset (**$n = 153$)	2.20	1.47	1.62
Resolution ≤ **2.0 Å subset (**$n = 79$)	1.95	1.20	1.42

improvement in the average best RMSD achieved from a value of 0.85 Å (NoW) to values of 0.72 Å (WL) and 0.70 Å (W). This demonstrates that the presence of water molecules improves the efficacy of the docking search algorithm by reducing, on average, the best RMSD observed from all binding poses, with only a small difference between the two water optimization sets. The results of paired *t*-testing for the difference between the average best RMSD observed are also shown in Table 2. Variance equivalences were checked and confirmed using an *F*-test. Actual *P* values and critical values for the *t*-statistic for a two-tailed test are also included. The α level was set to the stringent value of 0.01.

The efficacy of the search method was very high, and this is also revealed in each water scenario by the low proportion of ligand−protein complexes (out of 240) where the best RMSD observed was >2.0 Å: 10/240 (NoW), 8/240 (WL), and 5/240 (W). (See the Supporting Information for detailed results for each ligand−protein complex in each water scenario.) Thus, the corresponding percentages of cases where the correct solution was found by the search method were 96% (NoW), 97% (WL), and 98% (W), which is consistent with previous reports for the entire CCDC/Astex dataset.[100] These statistics indicate that the search method was not the limiting factor when determining the accuracy of the ligand−protein simulations (see further below), because the correct binding mode was almost always found.

Table 3 summarizes the differences in the efficacy of the docking search for each subset in the dataset. The results show that the performance of the docking search method improved in the presence of water molecules (both WL and W) for all sets, with the improvement becoming more noticeable for the clean subsets as the resolution of the crystal structures of the proteins improved. Importantly, no significant difference could be observed between the two water optimization scenarios. It might be expected that optimizing the orientation of water molecules in the presence of the bound ligand (WL) would best "encourage" finding the native binding mode by the docking simulations, because water molecules would already be oriented to optimally interact with the ligand. However, the observation that including water molecules optimized in the absence of the ligand (W) is sufficient to obtain similar improvements in docking search efficacy is encouraging, because it suggests that it is not necessary to optimize the orientations of water molecules that may interact with a ligand prior to docking it. This has clear advantages, because ligand−protein docking is commonly used to predict the binding mode of a ligand where the crystal structure of the bound complex is not available.

Table 4 summarizes the differences in the accuracy of the docking simulations between the three previously described scenarios. The results show that the accuracy of the docking simulations, as measured by the average RMSD of the top-energy-ranked solutions improves in the presence of water molecules, from a value of 2.49 Å to values of 1.83 Å (WL) and 1.91 Å (W). This improvement may be partially due to the improvement in the efficacy of the docking search, as described previously. However, the magnitude of the improvement in the mean RMSD of the top-energy-ranked solutions is significantly greater than the magnitude of the improvement in the mean best RMSD found (∼0.6 Å, compared to 0.15 Å, respectively). Therefore it can be concluded that the inclusion of water molecules improves the predictions of the scoring function, predominantly because of additional water−ligand interactions. Below, we further discuss many specific examples to rationalize these observations. The results of paired *t*-testing for the difference between the average RMSD of the top-energy-ranked solutions also can be also observed in Table 4. In addition, as observed previously, there was no statistically significant difference in the accuracy of the docking simulations obtained by either of the two water optimization methods.

Table 5 summarizes the differences in the accuracy of the docking simulations for each subset in the dataset. The results show that the accuracy of the docking simulations improved in the presence of water molecules (both WL and W) for all sets, with the improvement becoming more noticeable for the clean subsets as the resolution of the crystal structure of the proteins improved. Although these improvements become slightly more pronounced in the WL optimization case, the difference in results between the two water optimization methods is not statistically significant.

Table 6 summarizes the overall qualitative accuracy of the docking simulations for each of the water optimization

Table 6. Overall Quality of Top Ranked Solutions in All Water Optimization Cases and Ligand−Protein Subsets Reported by Absolute Numbers and Percentages

	Good		Close		Errors		Wrong	
water optimization protocol	absolute number	percentage	absolute number	percentage	absolute number	percentage	absolute number	percentage
			NoW					
Parent set ($n = 240$)	80	33%	77	32%	18	8%	65	27%
Clean subset ($n = 179$)	68	38%	57	32%	12	7%	42	23%
Resolution ≤ 2.5 Å subset ($n = 153$)	56	37%	51	33%	11	7%	35	23%
Resolution ≤ 2.0 Å subset ($n = 79$)	30	38%	32	40%	3	4%	14	18%
			WL					
Parent set ($n = 240$)	118	49%	60	25%	23	10%	39	16%
Clean subset ($n = 179$)	99	55%	43	24%	16	9%	21	12%
Resolution ≤ 2.5 Å subset ($n = 153$)	89	58%	36	24%	14	9%	14	9%
Resolution ≤ 2.0 Å subset ($n = 79$)	45	57%	22	28%	8	10%	4	5%
			W					
Parent set ($n = 240$)	107	44%	69	29%	19	8%	45	19%
Clean subset ($n = 179$)	87	49%	53	30%	13	7%	26	14%
Resolution ≤ 2.5 Å subset ($n = 153$)	75	49%	47	31%	12	8%	19	12%
Resolution ≤ 2.0 Å subset ($n = 79$)	37	47%	30	38%	5	6%	7	9%

protocols and each of the ligand−protein subsets. If we consider a successful docking simulation to be one whose top-energy-ranked solution is within a 2.0 Å RMSD of the crystallographic binding mode (in other words, the combination of solutions is classified as good and close), then a success rate of 65.8% was achieved for the parent set. This success rate improved when docking was performed in the presence of water molecules, to 73.4% (WL) and 74.2% (W). As mentioned previously, it is important to note that the improvement in the accuracy of the docking simulations due to the inclusion of water molecules was not dependent on the method used to optimize the orientation of these water molecules.

The results shown in Table 6 also reveal the improvements in the overall qualitative accuracy of the docking simulations that were observed for each subset in the dataset. As reported earlier,[100] an improvement in qualitative docking accuracy, with respect to the parent set, can be observed in the absence of water molecules for the "clean" subset (∼70% of the good and close solutions), which becomes distinctly noticeable when the crystallographic resolution of the protein structure is <2.0 Å (∼78.5%). The inclusion of water molecules results in very similar improvements for each subset, with qualitative docking accuracies of ∼78% (WL) and ∼79% (W) for the "clean" subset, improving to ∼84% (WL) and ∼85% (W) for proteins with a crystallographic resolution of <2.0 Å. It is noteworthy that the improvements obtained by including water molecules in the docking simulations were greatest when poorer quality structures were included. This suggests that although docking a ligand to a high-resolution protein structure is likely to result in improved accuracy of the predicted binding mode, the inclusion of tightly bound water molecules can have a larger impact on docking accuracy in lower-resolution structures of proteins.

We now discuss in detail a number of specific ligand−protein complexes, to provide a rationale for the various effects that the inclusion of water molecules can have on both the efficacy of the docking search and the accuracy of the docking simulations. Table 7 summarizes the results obtained for these selected representative cases. Structures 1a1e and 2fox are examples where the efficiency of the docking search (as measured by the best RMSD achieved) showed an improvement upon the inclusion of water molecules (more substantial in 2fox than in 1a1e) and also resulted in major improvements in the RMSD of the top energy-ranked solutions (more substantial in 1a1e than in 2fox). On the other hand, structures 1c12, 1dd7, and 1ett are examples where the docking search was successful in finding the correct solution both with and without water molecules, but which were, nonetheless, not energy-ranked highly. However, the addition of water molecules resulted in three types of effects on the accuracy of the docking simulations (as measured by the RMSD of the top energy-ranked solutions): a large improvement (1dd7), a substantial worsening (1c12), and no significant effect (1ett).

Table 7. Docking Results for Selected Ligand−Protein Complexes

	Best RMSD Found (Å)			RMSD of Top Energy-Ranked Solution (Å)		
PDB code	NoW	WL	W	NoW	WL	W
1a1e	1.46	0.55	0.64	7.52	1.05	0.65
1c12	0.81	0.81	0.79	1.06	5.88	6.01
1dd7	0.48	0.57	0.51	6.97	0.72	0.62
1ett	0.66	0.41	0.53	3.95	4.93	4.96
2fox	2.88	0.43	0.47	3.03	0.56	0.59

Structure 1a1e is a complex of human c-src tyrosine kinase with a substituted butylpiperidine ligand.[103] Docking in the absence of water molecules (NoW) did not produce the correct solution: the RMSD of the top-ranked solution was >7.5 Å, with respect to the crystallographic binding mode. This occurred despite the fact that the best RMSD achieved by the search algorithm was 1.46 Å. The inclusion of water molecules improved the best RMSD achieved to 0.55 Å (WL) and 0.64 (W). Importantly, the RMSD of the top energy-ranked solution was greatly improved to 1.05 Å (WL) and 0.65 (W). A visual comparison of the crystallographic binding mode, the top energy-ranked solution predicted in the presence of water molecules (WL) and the top energy-ranked solution predicted in the absence of water molecules (NoW) can be observed in Figure 1. In this case, a major contributor to the improvement in the docking predictions seems to be a water molecule that clashes sterically with the ligand when it adopts the best binding pose predicted in the absence of water molecules. After this water molecule is present in the binding site, certain binding modes are no longer possible, because of steric repulsion. This water molecule is within hydrogen-bonding distance of the ligand

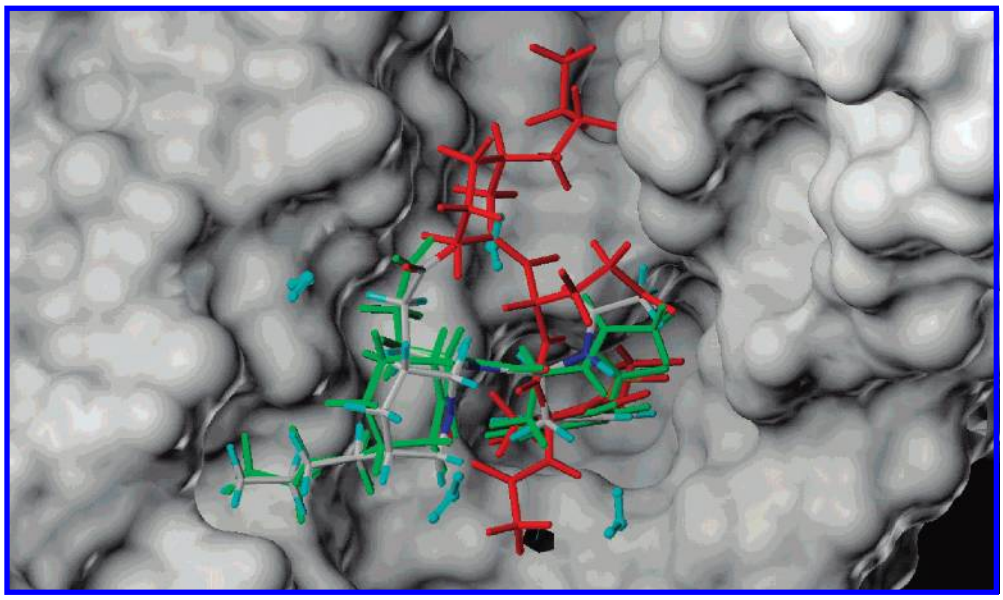

Figure 1. Docking simulation results for human tyrosine kinase in complex with a substituted butylpiperidine ligand (PDB code 1a1e). The ligand in its native crystallographic binding mode is colored according to atom type: the top energy-ranked binding pose predicted in the presence of water molecules (W) is shown in green, and the top energy-ranked binding pose predicted in the absence of water molecules (NoW) is shown in red. Water molecules are shown in cyan.

in its native binding mode. The favorable energy of this interaction, along with other hydrogen bonds formed between other water molecules and the ligand atoms, contributed to the improvement in the binding energy score rankings.

Structure 2fox is a complex of flavodoxin with a flavin mononucleotide ligand.[104] The ligand is a fused rigid heteronuclear aromatic tricycle with a polar flexible side chain. In the native binding mode, numerous polar interactions anchor the phosphate group of the ligand to a hydrophilic region formed by threonine, serine, and glycine residues. The hydrophilic portion of the flavin moiety is hydrogen-bonded to glutamate and glycine residues with several water molecules closely associated. Docking in the absence of water molecules (NoW) did not produce the correct solution: the RMSD of the top-ranked solution was >3.03 Å, with respect to the crystallographic binding mode. In this case, this was not surprising, because the best RMSD observed was 2.88 Å. The inclusion of water molecules significantly improved the best RMSD achieved to 0.43 Å (WL) and 0.47 Å (W). The RMSD of the top energy-ranked solution was also significantly improved to 0.56 Å (WL) and 0.59 Å (W).

Figure 2 illustrates and compares the crystallographic binding mode, the top energy-ranked solution predicted in the presence of water molecules (WL), and the top energy-ranked solution predicted in the absence of water molecules (NoW). The phosphate-containing side chain is similarly placed in the native and top energy-ranked binding poses, whereas deviations in the binding mode are due to the displacement of the flavin ring system. The top energy-ranked binding poses obtained in the absence of water molecules placed the pteridine portion of the flavin ring system interacting with a different set of polar side chains than those in the crystallographic binding mode. When the docking simulations were performed in the presence of water molecules, the native binding mode was reproduced fairly closely and energy-ranked at the top. Water molecules participate in the formation of a ligand−protein complex by establishing hydrogen bonds with the ligand and the protein, thus enthalpically stabilizing the complex. In this case, several water molecules establish a network of hydrogen bonds that drive the correct placement of the flavin ring system in the binding site.

Structure 1c12 is a complex of traseolide, which is a musk odorant, with an antibody directed against it (only the Fab portion of the antibody was crystallized).[105] Docking in the absence of water molecules (NoW) correctly reproduced the native binding mode: the RMSD of the top-ranked solution was 1.06 Å, with respect to the crystallographic binding mode. The inclusion of water molecules drastically worsened the results: the RMSD of the top energy-ranked solution was 5.88 Å (WL) and 6.01 Å (W). This is a remarkable result, given that the best RMSD achieved by the docking search did not change significantly either in the absence of water molecules (0.81 Å) or in the presence of water molecules: 0.81 Å (WL) and 0.79 Å (W). This demonstrates that the correct binding pose was only scored highly in the absence of any water molecules.

Figure 3 illustrates and compares the crystallographic binding mode and the top energy-ranked solutions predicted in the presence of water molecules (W and WL). The ligand is buried deeply in a cleft formed between two protein domains, and the one water molecule that satisfied the distance criteria for inclusion is also deeply buried. In the native conformation of the bound ligand, a methyl group is closest to the water molecule. Consequently, strong interactions (such as a hydrogen bond) cannot be expected between the water molecule and this hydrophobic group and thus would not be favored by the scoring function, resulting in an alternative binding pose being ranked higher.

It is important to note that the hybrid scoring function ScreenScore is inferior to the PLP scoring function for modeling hydrophobic interactions, because of its additional penalty terms for "ambiguous" interactions (such as lipophilic−polar interactions) and steric clashes. The RMSD of the top energy-ranked pose was greatly improved when PLP

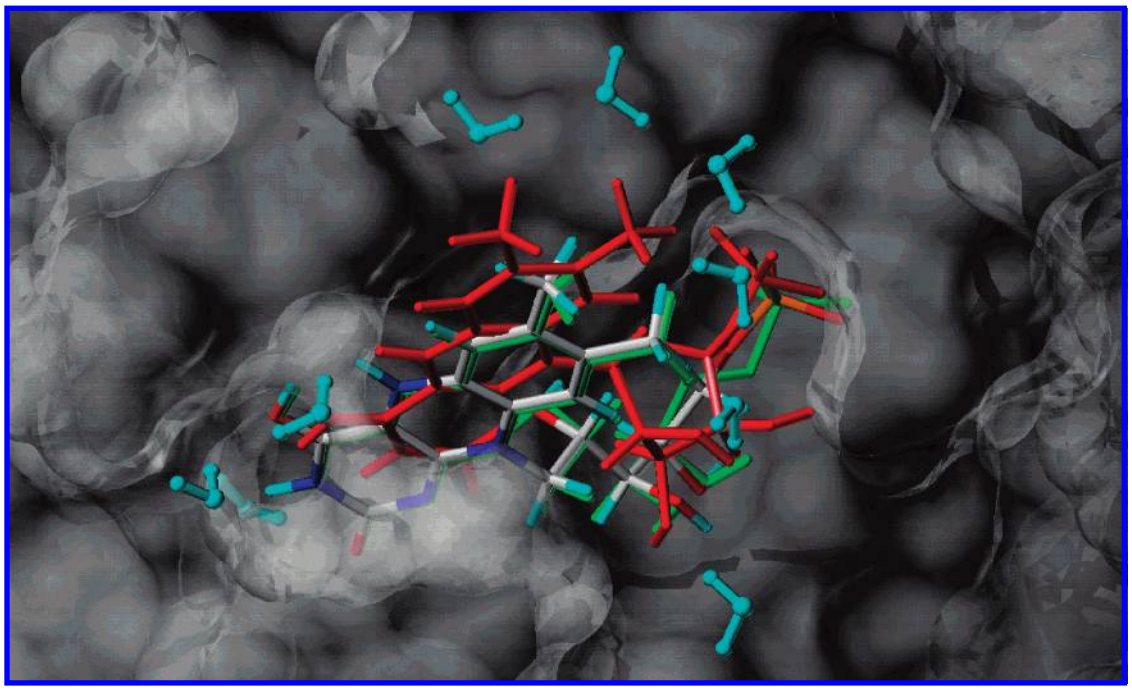

Figure 2. Docking simulation results for flavodoxin in complex with a flavin mononucleotide (PDB code 2fox). The ligand in its native crystallographic binding mode is colored according to atom type: the top energy-ranked binding pose predicted in the presence of water molecules (W) is shown in green, and the top energy-ranked binding pose predicted in the absence of water molecules (NoW) is shown in red. Water molecules are shown in cyan.

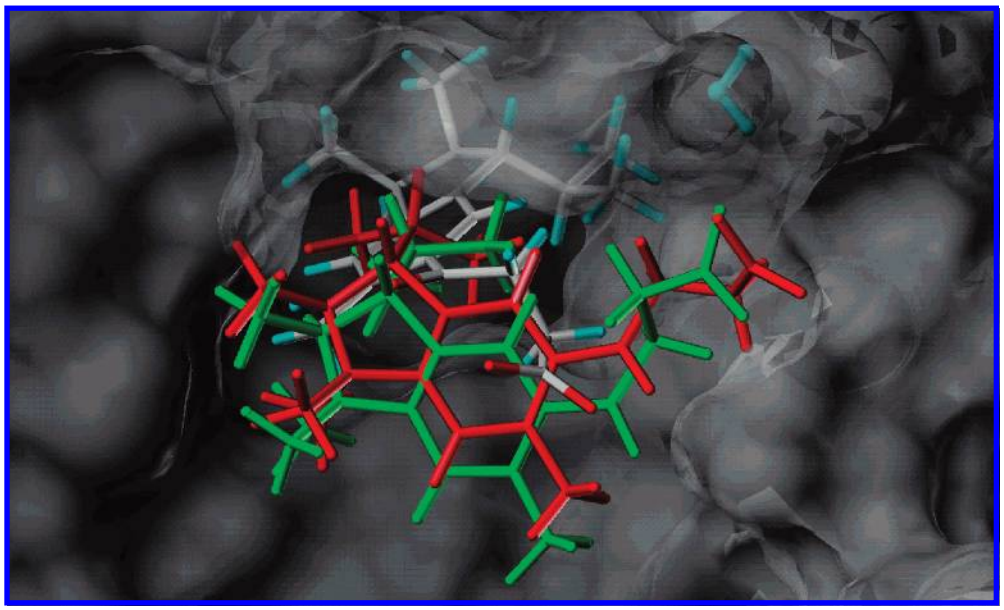

Figure 3. Docking simulation results for traseolide in complex with an antibody fragment (PDB code 1c12). The ligand in its native crystallographic binding mode is colored according to atom type: the top energy-ranked binding poses predicted in the presence of water molecules are shown in green (W) and red (WL). The only water molecule present is shown in cyan (above and right from the ligand molecules).

was used to score the interactions of the ligand with the protein. In the absence of water molecules, the PLP top ranked binding mode had an RMSD of 1.64 Å (a slight worsening compared to ScreenScore); however, in this case, the inclusion of water molecules did not result in the wrong prediction of the binding mode, because the RMSD measured was 1.95 Å (WL) and 1.91 Å (W). This is a good example of the role that the choice of scoring function can have when predicting ligand−protein interactions in the presence of water molecules that may not directly interact with the ligand but where hydrophobic interactions are more important. Admittedly, structure 1c12 could have been excluded from this study, because the only water molecule in the immediate vicinity of the ligand does not interact through hydrogen bonding with it. In cases such as this one, there may indeed be no reason to include water molecules for docking purposes.

Structure 1dd7 is a complex of the oxygenase domain of a murine inducible nitric oxide synthase with a polyheterocyclic acetamide inhibitor.[106] Docking in the absence of water

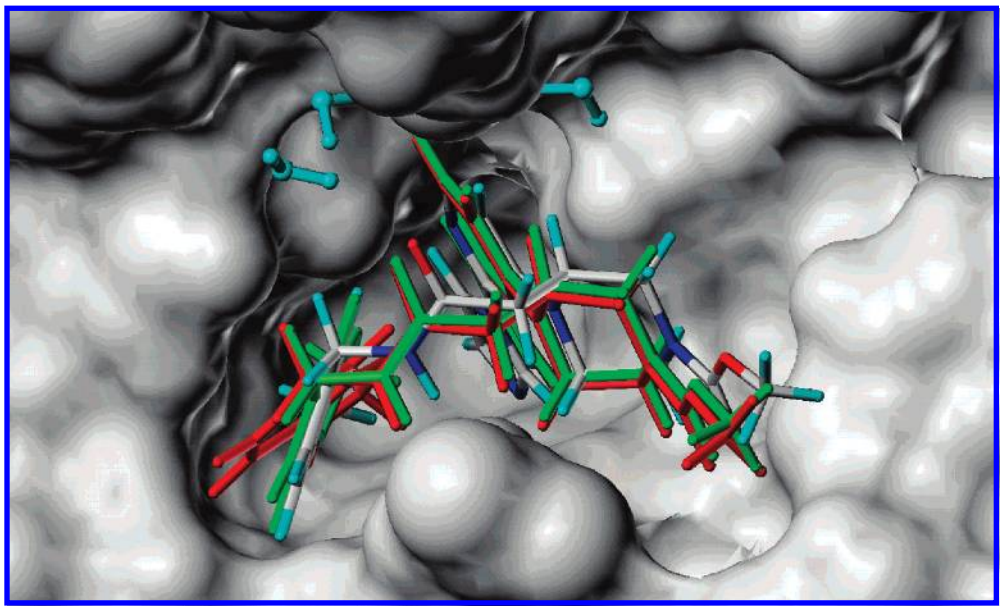

Figure 4. Docking simulations results for a polyheterocyclic acetamide inhibitor in complex with the oxygenase domain of a murine inducible nitric oxide synthase (PDB code 1dd7). The ligand in its native crystallographic binding mode is colored according to atom type: the top energy-ranked binding pose predicted in the presence of water molecules (W) is shown in green, and the top energy-ranked binding pose predicted in the absence of water molecules (NoW) is shown in red. Water molecules are shown in cyan.

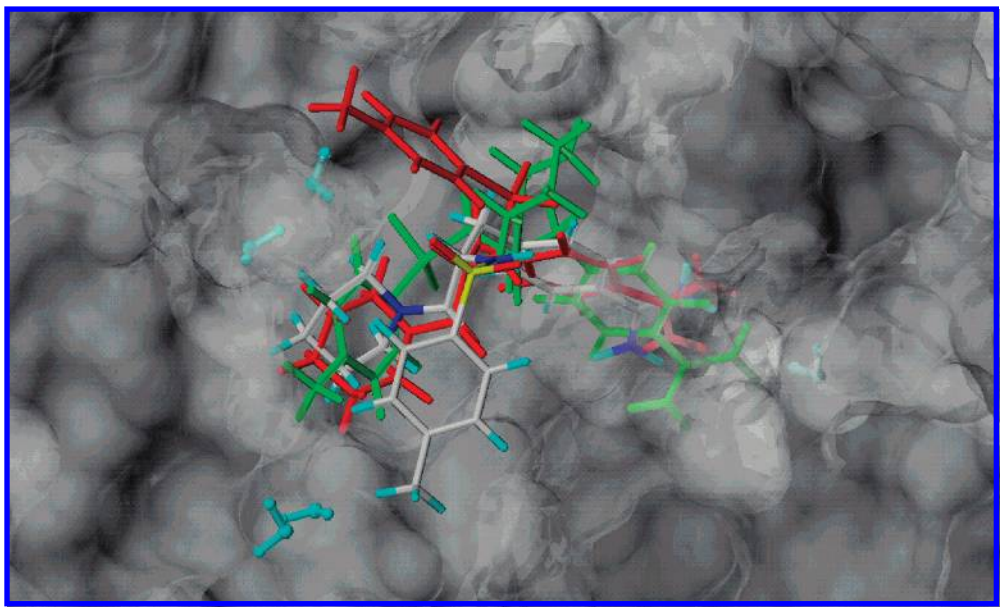

Figure 5. Docking simulations results for a small arginine-containing benzamidine inhibitor in complex with a bovine serine protease (PDB code 1ett). The ligand in its native crystallographic binding mode is colored according to atom type: the top energy-ranked binding pose predicted in the presence of water molecules (W) is shown in green, and the top energy-ranked binding pose predicted in the absence of water molecules (NoW) is shown in red. Water molecules are shown in cyan.

molecules (NoW) did not reproduce the native binding mode: the RMSD of the top-ranked solution was 6.97 Å, with respect to the crystallographic binding mode. This occurred despite the fact that the docking program was able to sample the correct binding mode: the best RMSD achieved was 0.48 Å. The inclusion of water molecules greatly improved the results: the RMSD of the top energy-ranked solution decreased to 0.72 Å (WL) and 0.62 Å (W), primarily because of changes in the orientation of the benzodioxol ring. Figure 4 illustrates and compares the crystallographic binding mode, the top energy-ranked solution predicted in the presence of water molecules (WL), and the top energy-ranked solution predicted in the absence of water molecules (NoW). In this case, several hydrogen-bonding interactions between the ligand and the water molecules present in the binding site reduced the predicted free energy of binding sufficiently so that a near-native binding pose was energy-ranked at the top.

Structure 1ett is a complex between a bovine serine protease and both a tripeptide (Phe−Pro−Arg) and a small arginine-containing benzamidine inhibitor.[107] We considered the benzamidine-based molecule as the ligand to be docked in our simulations. Docking in the absence of water molecules (NoW) did not reproduce the native binding

mode: the RMSD of the top-ranked solution was 3.95 Å, with respect to the crystallographic binding mode. This occurred despite the fact that the docking search was able to find the correct binding mode: the best RMSD observed was 0.66 Å. The inclusion of water molecules did not make any difference (in fact, the results were somewhat worse): the RMSD of the top-ranked solution was 4.93 Å (WL) and 4.96 Å (W).

Figure 5 illustrates and compares the crystallographic binding mode, the top energy-ranked solution predicted in the presence of water molecules (WL), and the top energy-ranked solution predicted in the absence of water molecules (NoW). None of the water molecules present in the binding site are within hydrogen-bonding distance of any polar group in the ligand, and the most closely placed ligand atoms are all nonpolar atoms. This suggests that the presence of these water molecules did not provide any interactions that would make the correct binding mode more favorable, in terms of its energy score. The scoring function (ScreenScore) was simply unable to assign a sufficiently low score to the correct binding mode.

CONCLUSIONS

Water has a key role in ligand−protein interactions. However, the treatment of water molecules in ligand−protein docking has remained a challenge, because of the need to know how many water molecules mediate the interactions between the ligand and the protein and where and how they should be included in a docking simulation. In this study, the efficacy of docking searches and the accuracy of docking predictions were determined to be significantly improved via the inclusion of crystallographic water molecules in the binding site of numerous protein structures.

The magnitude of these improvements (measured by computing the root mean square deviation (RMSD) of predicted binding poses, with respect to the crystallographic binding mode) was not significantly dependent on the method chosen to optimize the orientation of the water molecules. This is an important finding, revealing that a bound ligand does not need to be present (or even available) when optimizing the hydrogen-bonding interactions that water molecules may have with the protein surface. This means that it should be possible to apply this approach when trying to predict the binding mode of a different similarly sized ligand to a water-containing binding site in a target protein. Clearly, in the case of ligand docking to protein structures determined in the absence of a bound ligand (the 'apo' form), which we have not investigated here, there would be an initial requirement to remove water molecules for a ligand to fit in the binding site, for which methods to predict the displacement of tightly bound water molecules could be used.

Many representative ligand−protein complexes were investigated in more detail, to exemplify the various effects that the presence of water molecules may have on the efficacy and accuracy of docking simulations. The first scenario that was observed was an improvement of the docking efficacy when water molecules were present. The analysis of representative examples revealed that more-accurate search results were obtained because the presence of water molecules sterically excluded certain binding modes that would otherwise be incorrectly energy-ranked higher, and/or because water molecules provided additional hydrogen-bonding interactions that made the energy of interaction of the ligand more favorable in the correct binding mode.

The second scenario that was observed involved a high level of efficacy in the docking search but mixed results during scoring. An analysis of representative examples revealed that the presence of water molecules seemed to have mixed effects. When an improvement in the docking accuracy was observed, this was due to the additional hydrogen-bonding interactions with water molecules, which decreased the energy of interaction of the ligand. When no improvements in the docking accuracy were observed, this seemed to occur because those water molecules present in the vicinity of the ligand did not interact to any significant extent with it. Finally, when a worsening of the docking accuracy was observed, this seemed to be due to the inability of the scoring function to account properly for hydrophobic interactions.

Although our work adds to the growing evidence in favor of including water molecules in ligand−protein docking, there are several methodological limitations that must be overcome. The first problem is that scoring functions may need to be recalibrated and/or parametrized to account for ligand−water (and even protein−water) interactions. Most empirical scoring functions have been parametrized to reproduce free energies of binding; however, the presence of explicit water molecules may result in certain interactions (hydration and hydrophobic effects) being effectively taken into account twice, albeit it in an incomplete manner. In addition, it is not really possible to compare energy scores obtained after docking the same ligand in the presence and absence of water molecules, because the presence of water molecules will often result in a larger number of hydrogen bonding interactions with the ligand. Furthermore, it is not clear how energy scores for ligands docked in the presence of different numbers of water molecules should be compared, given that a full hydration shell is to be expected in aqueous solution. Finally, it is clear that the application of this approach to true predictions of the binding mode of a ligand will still require an initial decision in regard to which water molecules to retain (because some may be displaced by the ligand) and face the limitation imposed by the possibility of missing water molecules in the crystal structure being used.

It is also important to recognize that further investigations are required to determine the utility of including crystallographic water molecules in practical applications. Typically cross-docking studies with proteins that have been crystallized with several ligands should be performed to establish whether improvements in docking accuracy and virtual screening enrichment can be obtained if water molecules are included.

ACKNOWLEDGMENT

Part of this work was funded by a Curtin Strategic Grant.

Supporting Information Available: Docking results for each ligand−protein complex for each water scenario. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Purkiss, A.; Skoulakos, S.; Goodfellow, J. M. The protein-solvent interface: a big splash. *Philos. Trans. R. Soc. London Ser. A* **2001**, *359*, 1515−1527.

(2) Nakasako, M. Large-scale networks of hydration water molecules around bovine beta-trypsin revealed by cryogenic X-ray crystal structure analysis. *J. Mol. Biol.* **1999**, *289*, 547−564.
(3) Higo, J.; Nakasako, M. Hydration structure of human lysozyme investigated by molecular dynamics simulation and cryogenic X-ray crystal structure analyses: on the correlation between crystal water sites, solvent density, and solvent dipole. *J. Comput. Chem.* **2002**, *23*, 1323−1336.
(4) Yokomizo, T.; Higo, J.; Nakasako, M. Patterns and networks of hydrogen-bonds in the hydration structure of human lysozyme. *Chem. Phys. Lett.* **2005**, *410*, 31−35.
(5) Davis, A. M.; Teague, S. J.; Kleywegt, G. J. Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew. Chem., Int. Ed.* **2003**, *42*, 2718−2736.
(6) Poornima, C. S.; Dean, P. M. Hydration in drug design 1. Multiple hydrogen-bonding features of water molecules in mediating protein-ligand interactions. *J. Comput-Aided Mol. Des.* **1995**, *9*, 500−512.
(7) Hendlich, M.; Bergner, A.; Günter, J.; Klebe, G. Relibase: Design and development of a database for comprehensive analysis of protein-ligand interactions. *J. Mol. Biol.* **2003**, *326*, 607−620.
(8) Lu, Y.; Wang, R.; Yang, C-Y.; Wang, S. Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. *J. Chem. Inf. Model.* **2007**, *47*, 668−675.
(9) Poornima, C. S.; Dean, P. M. Hydration in drug design. 2. Influence of local site surface shape on water binding. *J. Comput-Aided Mol. Des.* **1995**, *9*, 513−520.
(10) Chung, E.; Henriques, D.; Renzoni, D.; Zvelebil, M.; Bradshaw, J. M.; Waksman, G.; Robinson, C. V.; Ladbury, J. E. Mass spectrometric and thermodynamic studies reveal the role of water molecules in complexes formed between SH2 domains and tyrosyl phosphopeptides. *Struct. Fold. Des.* **1998**, *6*, 1141−1151.
(11) Baker, E. N.; Hubbard, R. E. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97−179.
(12) Sreenivasan, U.; Axelsen, P. H. Buried water in homologous serine proteases. *Biochemistry* **1992**, *31*, 12785−12791.
(13) Loris, R.; Stas, P. P.; Wyns, L. Conserved waters in legume lectin crystal structures. The importance of bound water for the sequence-structure relationship within the legume lectin family. *J. Biol. Chem.* **1994**, *269*, 26722−26733.
(14) Shaltiel, S.; Cox, S.; Taylor, S. S. Conserved water molecules contribute to the extensive network of interactions at the active site of protein kinase A. *Proc. Natl. Acad. Sci., U.S.A.* **1998**, *95*, 484−491.
(15) Sanschagrin, P. C.; Kuhn, L. A. Cluster analysis of consensus water sites in thrombin and trypsin shows conservation between serine proteases and contributions to ligand specificity. *Protein Sci.* **1998**, *7*, 2054−2064.
(16) Krem, M. M.; Di Cera, E. Conserved water molecules in the specificity pocket of serine proteases and the molecular mechanism of Na+ binding. *Proteins* **1998**, *30*, 34−42.
(17) Loris, R.; Langhorst, U.; De Vos, S.; Decanniere, K.; Bouckaert, J.; Maes, D.; Transue, T. R.; Steyaert, J. Conserved water molecules in a large family of microbial ribonucleases. *Proteins* **1999**, *36*, 117−134.
(18) Ogata, K.; Wodak, S. J. Conserved water molecules in MHC class-I molecules and their putative structural and functional roles. *Protein Eng.* **2002**, *15*, 697−705.
(19) Bottoms, C. A.; Smith, P. E.; Tanner, J. J. A structurally conserved water molecule in Rossmann dinucleotide-binding domains. *Protein Sci.* **2002**, *11*, 2125−2137.
(20) Prasad, B. V. L. S.; Suguna, K. Role of water molecules in the structure and function of aspartic proteinases. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58*, 250−259.
(21) Bottoms, C. A.; Schuermann, J. P.; Agah, S.; Henzl, M. T.; Tanner, J. J. Crystal structure of rat α-parvalbumin at 1.05 Å resolution. *Protein Sci.* **2004**, *13*, 1724−1734.
(22) Bottoms, C. A.; White, T. A.; Tanner, J. J. Exploring structurally conserved solvent sites in protein families. *Proteins* **2006**, *64*, 404−421.
(23) Zhang, X.-J.; Matthews, B. W. Conservation of solvent-binding sites in 10 crystal forms of T4 lysozyme. *Protein Sci.* **1994**, *3*, 1031−1039.
(24) Carrell, H. L.; Glusker, J. P.; Burger, V.; Manfre, F.; Tritsch, D.; Biellmann, J.-F. X-ray analysis of D-xylose isomerase at 1.9 Å: native enzyme in complex with substrate and with a mechanism-designed inactivator. *Proc. Natl. Acad. Sci., U.S.A.* **1989**, *86*, 4440−4444.
(25) Faerman, C. H.; Karplus, P. A. Consensus preferred hydration sites in six FKBP12-drug complexes. *Proteins* **1995**, *23*, 1−11.
(26) Poornima, C. S.; Dean, P. M. Hydration in drug design. 3. Conserved water molecules at the ligand-binding sites of homologous proteins. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 521−531.
(27) Babor, M.; Sobolev, V.; Edelman, M. Conserved positions for ribose recognition: importance of water bridging interactions among ATP, ADP and FAD−protein complexes. *J. Mol. Biol.* **2002**, *323*, 523−532.
(28) Powers, R. A.; Shoichet, B. K. Structure-based approach for binding site idenfication on AmpC beta-lactamase. *J. Med. Chem.* **2002**, *45*, 3222−3234.
(29) Mustata, G.; Briggs, J. M. Cluster analysis of water molecules in alanine racemase and their putative structural role. *Protein Eng.* **2004**, *17*, 223−234.
(30) Boström, J.; Hogner, A.; Schmitt, S. Do structurally similar ligands bind in a similar fashion. *J. Med. Chem.* **2006**, *49*, 6716−6725.
(31) Huang, K.; Lu, W.; Anderson, S.; Laskowski, M.; James, M. N. G. Water molecules participate in proteinase-inhibitor interactions: crystal structure of Leu18, Ala18 and Gly18 variants of turkey ovomucoid inhibitor third domain complexed with *Streptomyces griseus* proteinase B. *Protein Sci.* **1995**, *4*, 1985−1997.
(32) Engh, R. A.; Brandstetter, H.; Sucher, G.; Eichinger, A.; Baumann, U.; Bode, W.; Huber, R.; Poll, T.; Rudolph, R.; von der Saal, W. Enzyme flexibility, solvent and 'weak' interactions characterize thrombin-ligand interactions: implications for drug design. *Structure* **1996**, *4*, 1353−1362.
(33) Rejto, P. A.; Verkhivker, G. M. Mean field analysis of FKBP12 complexes with FK506 and rapamycin: implications for a role of crystallographic water molecules in molecular recognition and specificity. *Proteins* **1997**, *28*, 313−324.
(34) Rutenber, E. E.; Stroud, R. M. Binding of the anticancer drug ZD1694 to E. coli thymidylate synthase: assessing specificity and affinity. *Structure* **1996**, *4*, 1314−1324.
(35) Finley, J. B.; Atigadda, V. R.; Duarte, F.; Zhao, J. J.; Brouillette, M. J.; Air, G. M.; Luo, M. Novel aromatic inhibitors of influenza virus neuraminidase make selective interactions with conserved residues and water molecules in the active site. *J. Mol. Biol.* **1999**, *293*, 1107−1119.
(36) Palomer, A.; Perez, J. J.; Navea, S.; Llorens, O.; Pascual, J.; Garcia, L.; Mauleon, D. Modeling cyclooxygenase inhibition. Implication of active site hydration on the selectivity of ketoprofen analogues. *J. Med. Chem.* **2000**, *43*, 2280−2284.
(37) Vogt, J.; Perozzo, R.; Pautsch, A.; Prota, A.; Schelling, P.; Pilger, B.; Folkers, G.; Scapozza, L.; Schulz, G. E. Nucleoside binding site of Herpes simplex type 1 thymidine kinase analyzed by X-ray crystallography. *Proteins* **2000**, *41*, 545−553.
(38) Ni, H.; Sotriffer, C. A.; McCammon, J. A. Ordered water and ligand mobility in the HIV-1 integrase-5CITEP complex: a molecular dynamics study. *J. Med. Chem.* **2001**, *44*, 3043−3047.
(39) Lemieux, R. U. How water provides the impetus for molecular recognition in aqueous solution. *Acc. Chem. Res.* **1996**, *29*, 373−380.
(40) Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dijols, S.; Dansette, P. M.; Mansuy, D.; Stout, C. D. Structure of mammalian cytochrome P4502C5 complexed with diclofenac at 2.1 angstrom resolution: Evidence for an induced fit model of substrate binding. *Biochemistry* **2003**, *42*, 9335−9345.
(41) Pujadas, G.; Palau, J. Molecular mimicry of substrate oxygen atoms by water molecules in the β-amylase active site. *Protein Sci.* **2001**, *10*, 1645−1657.
(42) Pastor, M.; Cruciani, G.; Watson, K. A. A strategy for the incorporation of water molecules present in a ligand binding site into a three-dimensional quantitative structure−activity relationship analysis. *J. Med. Chem.* **1997**, *40*, 4089−4102.
(43) Wang, T.; Wade, R. C. Comparative binding energy (COMBINE) analysis of influenza neuraminidase−inhibitor complexes. *J. Med. Chem.* **2001**, *44*, 961−971.
(44) Anstead, G. M; Carlson, K. E.; Katzenellenbogen, J. A. The estradiol pharmacophore: ligand structure-estrogen receptor binding affinity relationships and a model for the receptor binding site. *Steroids* **1997**, *62*, 268−303.
(45) Grüneberg, S.; Stubbs, M. T.; Klebe, G. Successful virtual screening for novel inhibitors of human carbonic anhydrase: Strategy and experimental confirmation. *J. Med. Chem.* **2002**, *45*, 3588−3602.
(46) Brenk, R.; Naerum, L.; Grädler, U.; Gerber, H.-D.; Garcia, G. A.; Reuter, K.; Stubbs, M. T.; Klebe, G. Virtual screening for submicromolar leads of tRNA-guanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. *J. Med. Chem.* **2003**, *46*, 1133−1143.
(47) Lloyd, D. G.; García-Sosa, A. T.; Alberts, I. L.; Todorov, N. P.; Mancera, R. L. The effect of tightly bound water molecules on the structural interpretation of ligand-derived pharmacophore models. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 89−100.
(48) Schnecke, V.; Kuhn, L. A. Virtual screening with solvation and ligand-induced complementarity. *Perspect. Drug Discov. Des.* **2000**, *20*, 171−190.
(49) Pospisil, P.; Kuoni, T.; Scapozza, L.; Folkers, G. Methodology and problems of protein-ligand docking: Case study of dihydroorotate dehydrogenase, thymidine kinase, and phosphodiesterase. 4. *J. Recept. Signal Transduct. Res.* **2002**, *22*, 141−154.

(50) Dunitz, J. D. The entropic cost of bound water in crystals and biomolecules. *Science* **1994**, *264*, 670−671.
(51) Dunitz, J. D. Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. *Chem. Biol.* **1995**, *2*, 709−712.
(52) Ladbury, J. E. Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem. Biol.* **1996**, *3*, 973−980.
(53) Li, Z.; Lazaridis, T. Thermodynamic contributions of the ordered water molecule in HIV-1 protease. *J. Am. Chem. Soc.* **2003**, *125*, 6636−6637.
(54) Fornabaio, M.; Spyrakis, F.; Mozzarelli, A.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 3. The free energy contribution of structural water molecules in HIV-1 protease complexes. *J. Med. Chem.* **2004**, *47*, 4507−4516.
(55) Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Free energy of ligand binding to protein: evaluation of the contribution of water molecules by computational methods. *Curr. Med. Chem.* **2004**, *11*, 3093−3118.
(56) Kříž, Z.; Otyepka, M.; Bartová, I.; Koča, J. Analysis of CDK2 active site hydration: a method to design new inhibitors. *Proteins* **2004**, *55*, 258−274.
(57) Lu, Y.; Yang, C.-Y.; Wang, S. Binding free energy contributions of interfacial waters in HIV-1 protease/inhibitor complexes. *J. Am. Chem. Soc.* **2006**, *128*, 11830−11839.
(58) Helms, V.; Wade, R. C. Thermodynamics of water mediating protein-ligand interactions in cytochrome P450CAM. A molecular dynamics study. *Biophys. J.* **1995**, *69*, 810−824.
(59) Helms, V.; Wade, R. C. Hydration energy landscape of the active site cavity of cytochrome P450. *Proteins* **1998**, *32*, 381−396.
(60) Helms, V.; Wade, R. C. Computational alchemy to calculate absolute protein-ligand binding free energy. *J. Am. Chem. Soc.* **1998**, *120*, 2710−2713.
(61) Hamelberg, D.; McCammon, J. A. Standard free energy of releasing a localized water molecule from the binding pockets of proteins: double-decoupling method. *J. Am. Chem. Soc.* **2004**, *126*, 7683−7689.
(62) Li, Z.; Lazaridis, T. The effect of water displacement on binding thermodynamics: concanavalin A. *J. Phys. Chem. B* **2005**, *109*, 662−670.
(63) Li, Z.; Lazaridis, T. Thermodynamics of buried water clusters at a protein-ligand interface. *J. Phys. Chem. B* **2006**, *110*, 1464−1475.
(64) Barillari, C.; Taylor, J.; Viner, R.; Essex, J. W. Classification of water molecules in protein binding sites. *J. Am. Chem. Soc.* **2007**, *129*, 2577−2587.
(65) Ehrlich, L.; Reczko, M.; Bohr, H.; Wade, R. C. Prediction of protein hydration sites from sequence by modular neural networks. *Protein Eng.* **1998**, *11*, 11−19.
(66) Henchman, R. H.; McCammon, J. A. Extracting hydration sites around proteins from explicit water simulations. *J. Comput. Chem.* **2002**, *23*, 861−869.
(67) Kortveylesi, T.; Dennis, S.; Silberstein, M.; Brown, L.; Vajda, S. Algorithms for computational solvent mapping of proteins. *Proteins* **2003**, *51*, 340−351.
(68) Schymkowitz, J. W. H.; Rousseau, F.; Martins, I. C.; Ferkinghoff-Borg, J.; Stricher, F.; Serrano, L. Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proc. Natl. Acad. Sci., U.S.A.* **2005**, *102*, 10147−10152.
(69) Raymer, M. L.; Sanschagrin, P. C.; Punch, W. F.; Venkataraman, S.; Goodman, E. D.; Kuhn, L. A. Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *J. Mol. Biol.* **1997**, *265*, 445−464.
(70) Carugo, O. Correlation between occupancy and B factor of water molecules in protein crystal structures. *Protein Eng.* **1999**, *12*, 1021−1024.
(71) Carugo, O.; Bordo, D. How many water molecules can be detected by protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999**, *55*, 479−483.
(72) Carugo, O.; Argos, P. Accessibility to internal cavities and ligand binding sites monitored by protein crystallographic thermal factors. *Proteins* **1998**, *31*, 201−213.
(73) Amadasi, A.; Spyrakis, F.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Mapping the energetics of water-protein and water-ligand interactions with the "natural" HINT forcefield: predictive tools for characterizing the roles of water in biomolecules. *J. Mol. Biol.* **2006**, *358*, 289−309.
(74) García-Sosa, A. T.; Mancera, R. L.; Dean, P. M. WaterScore: a novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein−ligand complexes. *J. Mol. Model.* **2003**, *9*, 172−182.
(75) Marrone, T. J.; Briggs, J. M.; McCammon, J. A. Structure-based drug design: Computational advances. *Annu. Rev. Pharmacol.* **1997**, *37*, 71−90.
(76) Lam, P. Y. S.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C. H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.; Ericksonviitanen, S. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **1994**, *263*, 380−384.
(77) Chen, J. M.; Xu, S. L.; Wawrzak, Z.; Basarab, G. S.; Jordan, D. B. Structure-based design of potent inhibitors of scytalone dehydratase: displacement of a water molecule from the active site. *Biochemistry* **1998**, *37*, 17735−17744.
(78) Mikol, V.; Papageorgiou, C.; Borer, X. The role of water molecules in the structure-based design of (5-hydroxynorvaline)-2-cyclosporine−synthesis, biological activity and crystallographic analysis with cyclophilin-A. *J. Med. Chem.* **1995**, *38*, 3361−3367.
(79) Holdgate, G. A.; Tunnicliffe, A.; Ward, W. H. J.; Weston, S. A.; Rosenbrock, G.; Barth, P. T.; Taylor, I. W. F.; Pauptit, R. A.; Timms, D. The entropic penalty of ordered water accounts for weaker binding of the antibiotic novobiocin to a resistant mutant of DNA gyrase: a thermodynamic and crystallographic study. *Biochemistry* **1997**, *36*, 9663−9673.
(80) Cherbavaz, D. B.; Lee, M. E.; Stroud, R. M.; Koschl, D. E. Active site water molecules revealed in the 2.1 angstrom resolution structure of a site-directed mutant of isocitrate dehydrogenase. *J. Mol. Biol.* **2000**, *295*, 377−385.
(81) Finley, J. B.; Atigadda, V. R.; Duarte, F.; Zhao, J. J.; Brouillette, W. J.; Air, G. M.; Luo, M. Novel aromatic inhibitors of influenza virus neuraminidase make selective interactions with conserved residues and water molecules in the active site. *J. Mol. Biol.* **1999**, *293*, 1107−1119.
(82) Pickett, S. D.; Sherborne, B. S.; Wilkinson, T.; Bennett, J.; Borkakoti, N.; Broadhurst, M.; Hurst, D.; Kilford, I.; McKinnell, M.; Jones, P. S. Discovery of novel low molecular weight inhibitors of IMPDH via virtual needle screening. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1691−1694.
(83) Mancera, R. L. De novo ligand design with explicit water molecules: an application to bacterial neuraminidase. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 479−499.
(84) García-Sosa, A. T.; Mancera, R. L. The effect of tightly-bound water molecules on scaffold diversity in the computer-aided de novo ligand design of CDK-2 inhibitors. *J. Mol. Model.* **2006**, *12*, 422−431.
(85) García-Sosa, A. T.; Firth-Clark, S.; Mancera, R. L. Including tightly-bound water molecules in de novo drug design. Exemplification through the in silico generation of poly(ADP-ribose)polymerase ligands. *J. Chem. Inf. Model.* **2005**, *45*, 624−633.
(86) Rarey, M.; Kramer, B.; Lengauer, T. The particle concept: Placing discrete water molecules during protein−ligand docking predictions. *Proteins* **1999**, *34*, 17−28.
(87) 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.
(88) Rao, M. S.; Olson, A. J. Modelling of factor Xa-inhibitor complexes: a computational flexible docking approach. *Proteins* **1999**, *34*, 173−183.
(89) Minke, W. E.; Diller, D. J.; Hol, W. G.; Verlinde, C. L. The role of waters in docking strategies with incremental flexibility for carbohydrate derivatives: heat-labile enterotoxin, a multivalent test case. *J. Med. Chem.* **1999**, *42*, 1778−1788.
(90) Österberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* **2002**, *46*, 34−40.
(91) Floriano, W. B.; Vaidehi, N.; Zamanakos, G.; Goddard, W. A., III. HierVLS hierarchical docking protocol for virtual ligand screening of large-molecule databases. *J. Med. Chem.* **2004**, *47*, 56−71.
(92) Bellocchi, D.; Macchiarulo, A.; Constantino, G.; Pellicciar, R. Docking studies on PARP-1 inhibitors: insights into the role of a binding pocket water molecule. *Bioorg. Med. Chem.* **2005**, *13*, 1151−1157.
(93) Nissink, J. W. M.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. A new test set for validating predictions of protein−ligand interaction. *Proteins* **2002**, *49*, 457−471.
(94) Yang, J.-M.; Chen, C.-C. GEMDOCK: A generic evolutionary method for molecular docking. *Proteins* **2004**, *55*, 288−304.
(95) de Graaf, C.; Pospisil, P.; Wouter, P.; Folkers, G.; Vermeule, N. P. E. Binding mode prediction of cytochrome P450 and thymidine kinase protein−ligand complexes by consideration of water and rescoring in automated docking. *J. Med. Chem.* **2005**, *48*, 2308−2318.
(96) de Graff, C.; Oostenbrink, C.; Keizers, P. H. J.; van der Wijst, T.; Jongejan, A.; Vermeulen, N. P. E. Catalytic site prediction and virtual screening of cytochrome P450 2D6 substrates by consideration of water and rescoring in automated docking. *J. Med. Chem.* **2006**, *49*, 2417−2430.

(97) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W. M.; Taylor, R. D.; Taylor, R. Modeling water molecules in protein−ligand docking using GOLD. *J. Med. Chem.* **2005**, *48*, 6504−6515.

(98) Hatshorn, 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.

(99) Todorov, N. P.; Mancera, R. L.; Monthoux, P. H. A new quantum stochastic tunneling method for ligand protein docking. *Chem. Phys. Lett.* **2003**, *369*, 257−263.

(100) Mancera, R. L.; Källblad, P. K.; Todorov, N. P. ligand−protein docking using a quantum stochastic tunneling optimization method. *J. Comput. Chem.* **2004**, *25*, 858−864.

(101) Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem. Biol.* **1995**, *2*, 317−324.

(102) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **2001**, *44*, 1035−1042.

(103) Charifson, P. S.; Shewchuk, L. M..; Rocque, W.; Hummel, C. W.; Jordan, S. R.; Mohr, C.; Pacofsky, G. J.; Peel, M. R.; Rodriguez, M.; Sternbach, D. D.; Consler, T. G. Peptide ligands of pp60(c-src) SH2 domains: a thermodynamic and structural study. *Biochemistry* **1997**, *36*, 6283−6293.

(104) Ludwig, M. L.; Pattridge, K. A.; Metzger, A. L.; Dixon, M. M.; Eren, M.; Feng, Y.; Swenson, R. P. Control of oxidation-reduction potentials in flavodoxin from Clostridium beijerinckii: the role of conformation changes. *Biochemistry* **1997**, *36*, 1259−1280.

(105) Langedijk, A. C.; Spinelli, S.; Anguille, C.; Hermans, P.; Nederlof, J.; Butenandt, J.; Honegger, A.; Cambillau, C.; Plückthun, A. Insight into odorant perception: the crystal structure and binding characteristics of antibody fragments directed against the musk odorant traseolide. *J. Mol. Biol.* **1999**, *292*, 855−869.

(106) McMillan, K.; Adler, M.; Auld, D. S.; Baldwin, J. J.; Blasko, E.; Browne, L. J.; Chelsky, D.; Davey, D.; Dolle, R. E.; Eagen, K. A.; Erickson, S.; Feldman, R. I.; Glaser, C. B.; Mallari, C.; Morrissey, M. M.; Ohlmeyer, M. H. J.; Pan, G.; Parkinson, J. F.; Phillips, G. B.; Polokoff, M. A.; Sigal, N. H.; Vergona, R.; Whitlow, M.; Young, T. A.; Devlin, J. J. Allosteric inhibitors of inducible nitric oxide synthase dimerization discovered via combinatorial chemistry. *Proc. Natl. Acad. Sci., U.S.A.* **2000**, *97*, 1506−1511.

(107) Brandstetter, H.; Turk, D.; Hoeffken, H. W.; Grosse, D.; Stürzebecher, J.; Martin, P. D.; Edwards, B. F.; Bode, W. Refined 2.3 A X-ray crystal structure of bovine thrombin complexes formed with the benzamidine and arginine-based thrombin inhibitors NAPAP, 4-TA-PAP and MQPA. A starting point for improving antithrombotics. *J. Mol. Biol.* **1992**, *226*, 1085−1099.

CI700285E