Probing the Dynamic Nature of Water Molecules and Their Influences on Ligand Binding in a Model Binding Site

Daniel Cappel,[†,§] Rickard Wahlström,[‡] Ruth Brenk,[‡] and Christoph A. Sotriffer*[,†]

[†]Institute of Pharmacy and Food Chemistry, Julius-Maximilians University Würzburg, Am Hubland, D-97074 Würzburg, Germany

[‡]College of Life Sciences, Division of Chemical Biology and Drug Discovery, James Black Centre, University of Dundee, Dow Street, Dundee DD1 5EH, Scotland, United Kingdom

Ⓢ *Supporting Information*

ABSTRACT: The model binding site of the cytochrome *c* peroxidase (CCP) W191G mutant is used to investigate the structural and dynamic properties of the water network at the buried cavity using computational methods supported by crystallographic analysis. In particular, the differences of the hydration pattern between the uncomplexed state and various complexed forms are analyzed as well as the differences between five complexes of CCP W191G with structurally closely related ligands. The ability of docking programs to correctly handle the water molecules in these systems is studied in detail. It is found that fully automated prediction of water replacement or retention upon docking works well if some additional preselection is carried out but not necessarily if the entire water network in the cavity is used as input. On the other hand, molecular interaction fields for water calculated from static crystal structures and hydration density maps obtained from molecular dynamics simulations agree very well with crystallographically observed water positions. For one complex, the docking and MD results sensitively depend on the quality of the starting structure, and agreement is obtained only after redetermination of the crystal structure and refinement at higher resolution.
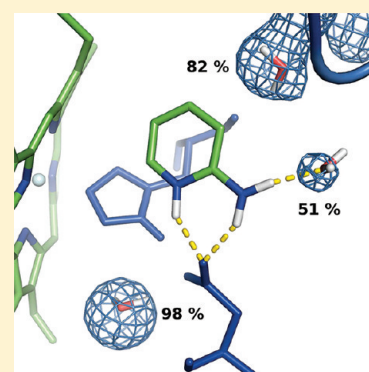
■ INTRODUCTION

Water is of utmost importance for protein structure and function and plays an essential role in protein—ligand interactions.[1-3] The hydration/dehydration balance has a major impact on complex formation and binding affinity, both in terms of enthalpic and entropic contributions.[4,5] Water molecules also affect the structure of protein—ligand complexes and influence ligand binding modes.[6-8] In structure-based drug design, this has been recognized for a long time, and strategies for addressing[9-11] or displacing[12,13] particular structural water molecules have repeatedly been followed in order to obtain high-affinity ligands. The most prominent examples in this context are given by HIV-protease inhibitors.[14-17] However, although successful examples are available and the general concepts appear to be qualitatively well understood, their application on different systems is by no means straightforward,[18,19] and a reliable quantitative prediction remains elusive. Even some fundamental issues remain a matter of dispute, as for example the question whether a ligand-free binding cavity is necessarily filled with as many water molecules as the volume permits or whether certain apolar cavities may rather be incompletely hydrated or even empty.[4,20] Not surprisingly, computational methods are also affected by the still insufficient understanding of water effects in and around protein binding sites. Although they are an essential component of the modern drug design process and many successful applications of virtual screening and computer-aided design have been reported,[21] most of today's computational drug-design methods are hardly able to reliably predict whether upon protein—ligand complex formation a water molecule is displaced or remains in the binding

pocket bridging the interaction between the binding partners. To the same extent it is challenging to accurately estimate the corresponding energetic and entropic contributions. The reasons for these short-comings are in part related to the difficulty in obtaining relevant experimental data of sufficient quality. On one hand, structural information about water molecules in proteins can have considerable limitations;[22] on the other hand, some energetical data are not even measurable since water is the solvent and, hence, separating the effects of single water molecules is virtually impossible. Nevertheless, there is a clear need for adequate handling of individual water molecules, especially in molecular docking. Improvement of the programs is only possible if the behavior of water molecules can be better understood.

Since information about water effects in protein—ligand complexes is inherently difficult to obtain, we resort to a well-suited model system showing some simplifying features for studying the role of water molecules and focus on the designed cavity of the cytochrome *c* peroxidase (CCP) W191G mutant[23-26] (Figure 1). This cavity is deeply buried, polar, negatively charged, closed on one side by heme as cofactor, and filled with five water molecules and one potassium ion in the ligand-free form. The protein is structurally well characterized, and several complexes with small, mostly heterocyclic ligands are known (currently, 53 crystal structures are available in the PDB). Various aspects of this model system have been the subject of earlier theoretical work. Docking
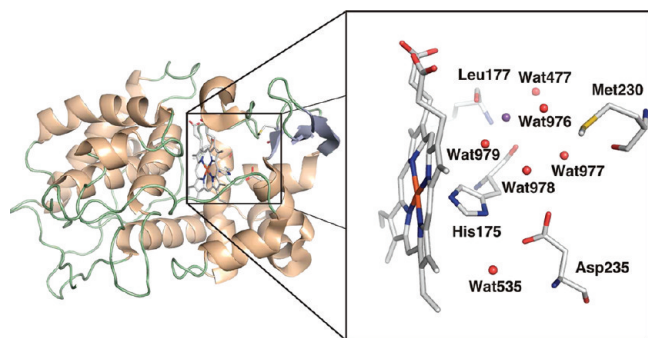
Figure 1. Cytochrome *c* peroxidase W191G crystal structure (PDB 1CPE) with a magnified view of the binding cavity. Water molecules are shown as red spheres; the purple sphere represents a potassium ion.

studies were performed, testing the ability to predict binding modes of known ligands and forecast new binders.[27−30] More recently, the inverse molecular design algorithm of Tidor and co-workers has been applied to this system.[31] The system has also been the subject of molecular dynamics (MD) simulations to investigate the flexibility of the PGGAAN 190-195 loop region and its influences on ligand binding.[32,33] The obtained dynamic data were then also used for molecular docking.[34]

In this work we investigate in detail the structural and dynamic properties of the water network buried in the CCP W191G cavity, in particular with respect to the differences between the uncomplexed state and various complexed forms, as well as the differences among the complexes. For this purpose, we focus on complexes with structurally closely related ligands and almost identical binding affinity and use comparative MD simulations based on crystal structures as starting points. We analyze the degree of mobility of the crystallographically observed water molecules and investigate whether water molecules too mobile to be crystallographically visible influence the experimentally determined ligand binding modes. In addition, we perform docking simulations in order to test the ability of current programs to correctly handle the water molecules in the investigated complexes of this valuable model system.

■ RESULTS AND DISCUSSION

System Description and Comparative Crystal Structure Analysis. The five complexes of CCP W191G with 2-aminopyridine (ligand abbreviation 2AP; PDB code 1AEO),[26] 3-aminopyridine (3AP; 1AEF),[26] 4-aminopyridine (4AP; 1AEG),[26] 2,5-diaminopyridine (DA1; 2AQD),[28] and 2-amino-5-picoline (LG4; 2EUP)[28] were investigated. All ligands are small cationic heterocycles with the pyridine nitrogen assumed to be protonated in the complexes.[28] PDB structure 1CPE[35] with a potassium ion in the binding site was used as "ligand-free" form of CCP W191G. The resolution of the six crystal structures is between 1.35 Å (2AQD) and 2.20 Å (1CPE). Only minor structural differences are observed between the ligand-free structure and the five complexes: the root-mean-square deviation (rmsd) of the C_{α} atoms with respect to 1CPE are 0.208 Å for 1AEO, 0.236 Å for 1AEF, 0.236 Å for 1AEG, 0.254 Å for 2AQD, and 0.232 Å for 2EUP, respectively (the protein coordinates of 1AEF and 1AEG are identical because only a refinement of the ligand coordinates was carried out while coordinates of all other atoms were taken from PDB structure 1AA4[25]). Also the binding pockets of all structures overlap very well.

A schematic view of the key interactions formed by the ligands in the five investigated complexes is shown in Figure 2. In all cases the ligand occupies the space between Asp235 and Leu177. One water molecule is found close to Leu177 forming hydrogen bonds with Ala176, Gly178, and Gly191 and some of the ligands. Due to the different water numbering in the crystal structures, this molecule is called water 'A' herein. In structure 1AEF a second water molecule is interacting with the ligand in the binding pocket; it is found close to Met230 and Asp235 and is called water 'B'. Finally, a third water molecule, water 'C', is coordinated to Asp235 in all structures but does not show direct interactions with any of the ligands.

The orientation of the ligands in 1AEO, 2AQD, and 2EUP is such that two hydrogen bonds are formed from the aromatic nitrogen and the amino group to one of the carboxylic oxygens of Asp235. In structure 1AEG only the ring nitrogen forms a hydrogen bond with this Asp, whereas the amino group interacts with Leu177 and water molecule A. The orientation of the amino group of the ligand in 1AEF is similar to the amino group of the ligand in 1AEG. However, due to the different topology (3-aminopyridine instead of 4-aminopyridine) the interaction to Asp235 is mediated by water molecule B in 1AEF. Notably, the three very similar molecules 2-, 3-, and 4-aminopyridine show different binding modes, in particular with respect to the involved water molecules. Furthermore, 2,5-diaminopyridine can be seen as a formal derivative of 2- as well as of 3-aminopyridine; the observed binding mode, however, resembles that of 2-aminopyridine.

Structural Comparison of Water Molecules. The WaterBase module[36−38] of Relibase+[39−41] provides a variety of descriptors for structural water molecules in protein−ligand complexes. The detailed results of the WaterBase analysis for all binding-site water molecules of the ligand-free crystal structure 1CPE are shown in Table 1.

In this analysis, the buried and highly conserved water molecule C was included for comparative purposes. The Water-Base descriptors point indeed to a low mobility, a high degree of burial, and a very favorable DrugScore value for this water molecule. Compared to C, water molecule A is slightly less buried but shows very favorable contacts and interactions with the protein. In contrast, water molecule 979 is characterized as highly mobile and shows only very weak interactions with its surrounding. Based on this characterization, it appears comprehensible why water A is always conserved, whereas water 979 is always displaced in the crystal structures of the complexes. The situation is less clear-cut for the remaining three water molecules (976, B, and 978). In particular, compared to water molecule A, they show fewer contacts with the protein but more contacts with water molecules. The DrugScore values for the interaction with the protein range between those measured for water A and 979. Based on these descriptors alone, a clear estimate whether displacement or retention is more likely to occur can hardly be given.

Water Molecular Interaction Fields and Cation-Free Model Pocket. To verify the observed positions of structural water molecules in the binding pocket of CCP W191G the program GRID[42] was used to calculate Molecular Interaction Fields (MIF), as described in Materials and Methods. The results are shown in Figure S1 of the Supporting Information. For all investigated complexes, the locations of the experimentally observed structural water molecules correspond to the energetically most favorable positions according to the GRID analysis.
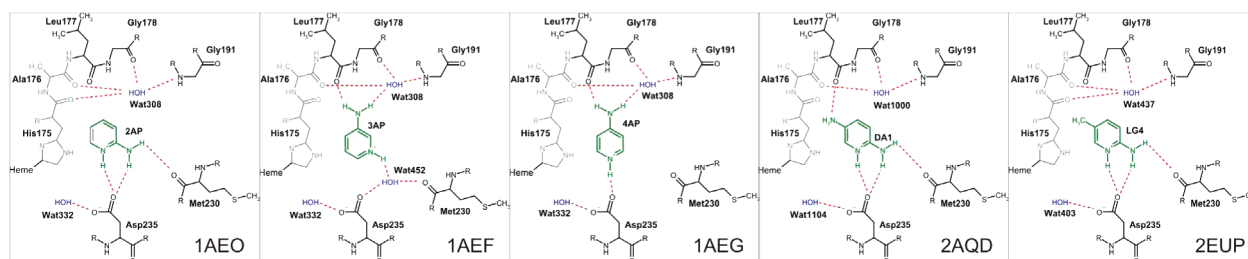
2582

dx.doi.org/10.1021/ci200052j |*J. Chem. Inf. Model.* 2011, 51, 2581–2594

Figure 2. Schematic representation of the binding modes and key polar interactions of the five investigated complexes. For the water molecules, the numbers from the corresponding PDB files are given.

Table 1. Descriptors of the Crystal Water Molecules in the 1CPE Pocket As Available from the Waterbase Module in Relibase+[a]

water molecule	477 (A)	976	977 (B)	978	979	535 (C)
mobility	0.55	0.41	0.52	0.41	1.00	0.46
protein contacts	3	2	1	1	0	2
ligand contacts	0	0	0	0	0	0
water contacts	0	1	2	2	1	0
total contacts	3	3	3	3	1	2
polarity	4.806	4.320	3.000	3.244	2.265	2.972
DrugScore - protein	−18198.92	−6084.78	−2053.03	−9564.14	2557.31	−12363.89
DrugScore - ligand	0.00	0.00	0.00	277.65	−2281.23	−6429.63
DrugScore - water	5611.19	−2216.69	−3268.77	−6505.93	−932.74	1688.15
DrugScore - total	−12587.74	−8301.46	−5321.80	−15792.42	−656.66	−17105.38
AD1	6.06	5.23	1.38	1.25	4.54	4.89
AD2	59.05	65.61	63.99	69.02	70.83	65.98
SAS	31.00%	44.00%	58.00%	58.00%	46.00%	22.00%
RMSDtet	0.15	0.24	-	-	-	-

[a] AD1 and AD2 are the atom densities in the first and second coordination sphere of the respective water molecule. SAS is the solvent accessible surface and RMSDtet gives the root mean square deviation in Å with respect to an ideal tetrahedral coordination geometry. A "-" indicates that no value was listed in the database.

For the position of water A and water C a GRID energy value of approximately −12 kcal/mol is obtained. At contour levels of −8 kcal/mol, also the MIF for position B in structure 1AEF appears. At this level, minor spots in the vicinity of position B also appear in the structures 1AEO, 1AEG, and 2AQD. Besides that, no other water hot spots are observed in the complex structures. With respect to the ligand-free structure (1CPE), the water MIF at −8 kcal/mol very well delineates and comprises the experimentally observed water locations, except for the position of water 979, which appears at significantly less favorable energy values.

The calculations for the ligand-free structure, however, are based on a cavity containing a potassium ion, which obviously influences the hydration pattern. However, knowing the water distribution and hydration state of the cavity in the absence of any ligand and any ion would be of paramount interest. Despite the negative charge of the cavity due to Asp235, such a cation-free state of the cavity may be assumed to exist, given the weak dissociation constant of the potassium ion[24] and calorimetric experiments carried out under conditions preventing competition from cations.[26] Although ligand- and ion-free CCP W191G crystal structures can be found in the PDB (such as 1AA4[25]), doubts are warranted whether indeed no ion is present in these structures. In fact, the water molecules observed in structure 1AA4 show almost identical positions as the water molecules in 1CPE, and the additional (sixth) water molecule of 1AA4 is

found at the position of the potassium ion in 1CPE. It is known that water molecules and sodium ions cannot be readily distinguished in the electron density map of protein structures.[22] In principle, the sixth water molecule could also be a H_3O^+-ion, as assumed, for example, in case of a tRNA-transglycosylase structure.[43] However, this additional water molecule shows a nearly octahedral coordination, which makes it more likely to be a sodium ion. Interestingly, an MD simulation started from structure 1AA4 with six water molecules and no ion in the binding site did not lead to a stable equilibrated state within the simulation time of 7 ns. Rather, the rmsd of the C_α protein atoms showed a continuous increase, combined with a sudden conformational change of the PGGAAN 190−195 loop after 6 ns, leading to an open access to the binding pocket and an influx of many additional water molecules (results not shown).

Given these observations and considerations, it appeared mandatory to build a model structure of an ion-free CCP W191G pocket. The construction was based on the 2AQD crystal structure, which is one of the complexes with the highest resolution. The ligand and all water molecules were deleted, and the program GRID was used to calculate a MIF for water in the empty pocket. The oxygen atom of a water molecule was then placed at the position of lowest energy according to the calculated MIF with the help of the program MOE.[44] As GRID is only taking into account the heavy atoms, there was no need to
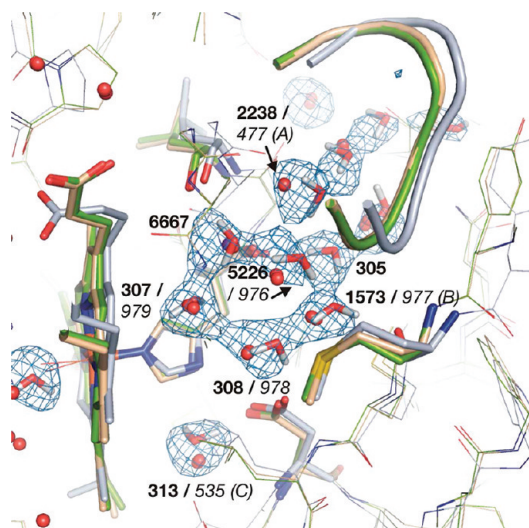
Figure 3. Comparison of the cation-free binding-pocket model of CCP W191G (gray) with the crystal structure 1AA4 (fawn) and 1CPE (green). Average hydration density maps (calculated with VolMap/VMD for the trajectory between 4 and 5 ns) are contoured at a level of 70%. The water molecules in the model system are shown as sticks, and the numbering scheme is shown with bold numerals. In cases a structurally equivalent water molecule is available in the crystal structure 1CPE these numbers are shown in italic including the position names A, B, and C.

determine the appropriate orientation of the hydrogen atoms at this step. With this water molecule in the pocket, a new MIF analysis was performed, and a second water molecule was placed at the now observed minimum position. This iterative procedure was repeated as long as no new, distinct water position could be found in the MIF map. This resulted in a total of five water molecules in the binding pocket. Finally, water molecules 445, 494, 500, 535, and 611 from the system 1CPE were added to the model. These conserved water molecules are not directly in the binding site but adjacent to it and are assumed to be of relevance for structural stability in a simulation. This final model was then subjected to a 5 ns MD simulation, following the same setup steps as in the other MD simulations of this work (cf. Materials and Methods). After equilibration, average hydration density maps (cf. below) were calculated for the binding-site water molecules for time spans of 1 ns and compared with each other. Based on these maps, seven water molecules in the binding site plus water C are highly conserved; significant alterations can only be observed near the flexible loop region. From the trajectory time period between 4 and 5 ns a structural average of the protein was calculated, and the snapshot with the smallest rmsd to this reference was taken as ligand- and cation-free binding-site model after a short relaxation. This model was used for cross-docking studies based on an ion-unbiased water distribution, as described below in the Docking section.

Interestingly, the structural differences between the cation-free model (see Figure 3) and the crystal structures 1AA4 or 1CPE are relatively small (1.20 Å and 1.22 Å rmsd, respectively, with respect to the protein atoms). Also the water positions are comparable. One additional water molecule is found near the position of the potassium ion of 1CPE. A second extra water molecule is placed next to water 977 of 1CPE. Furthermore, several water molecules could be found in the channel connecting binding site and bulk phase bordered by the PGGAAN loop

region. Here, a comparison with the crystal structures is not possible, since no water molecules for this region were reported. This is most likely caused by their high mobility, which was also observed in the MD simulation. Water molecules in the binding region, however, are more fixed at their positions and show smaller movements during the MD simulation after the initial equilibration period.

MD Simulations. To obtain a more detailed insight into the behavior of water molecules, MD simulations were performed for the five CCP W191G complexes and the ligand-free form 1CPE, as described in Materials and Methods. Details about the simulated systems are listed in Table 2.
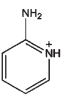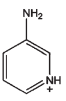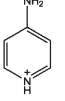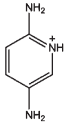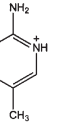
Analyzing the time course of the rmsd values for different groups of atoms and monitoring the secondary structure elements along the trajectory indicates a good overall stability of the simulated protein systems (all diagrams are provided in Figures S2 and S3 as Supporting Information). Generally, average rmsd values of the C_α atoms around or below 1.5 Å are obtained. Also within the binding site of the systems 1AEO, 1AEG, 2AQD, and 2EUP no major changes are observed, whereas larger alterations are seen in 1CPE and 1AEF, as further discussed below. Figure 4 illustrates the binding site from the simulation snapshot that is most similar to the average structure calculated from the corresponding trajectory. The similarity of these binding-pocket conformations to the crystal structure is evident. The coloring of the lines and sticks indicates the degree of mobility of the binding-site amino acids and illustrates the rather limited movements.

To capture water dynamics, an adequate metric is required. Simply tracking the occurrence of all hydrogen bonds involving cavity water molecules is virtually impractical if a network must be monitored and/or frequent exchange of water molecules occurs. Knowing the positions of interest, occupancies might be calculated. This, however, is complicated by the moving surrounding, which led McCammon and co-workers to use mean residue coordinates for the estimation of time average positions.[45−47] In the present work, an average water density map was calculated from all snapshots of a trajectory using VolMap in VMD.[48] Such a map circumvents the problem of tracking individual water molecules over time; furthermore, the water positions of interest need not to be known in advance.

Figure 4 shows such hydration density maps at a contour level of 70%. The maxima of the hydration density were extracted from the maps, and the occupancies at these positions are listed in Table 3. In the table also the sum of the occupancies is listed, which may roughly be compared to the number of water molecules observed in the crystal structure.

In all simulated complexes water molecule C is highly conserved at its position. Water molecule A of the systems 1AEG and 2EUP is also highly conserved during the simulations. For system 2AQD more than one water molecule is occupying position A during the simulation; exchange occurs with a second position in close proximity. In the simulation of 1AEO a small movement of the PGGAAN loop region in the second half of the trajectory opens a sufficiently large channel for water molecules to enter the cavity. These additional water molecules in the first place fill the space created by the loop movement. Furthermore, they come close to position A, and a few water molecules cover position B, but only for a short time span. Looking at the hydration density map at a lower contour level (45%), an additional maximum appears which is close to the position of water molecules 976 and

2584

dx.doi.org/10.1021/ci200052j |*J. Chem. Inf. Model.* 2011, 51, 2581–2594

Table 2. Overview of the Simulated Systems[a]

PDB code	1CPE[35]	1AEO[26]	1AEF[26]	1AEG[26]	2AQD[28]	2EUP[28]
		(2-aminopyridine structure)	(3-aminopyridine structure)	(4-aminopyridine structure)	(2,5-diaminopyridine structure)	(2-amino-5-picoline structure)
ligand	K^+	2-aminopyridine	3-aminopyridine	4-aminopyridine	2,5-diamino-pyridine	2-amino-5-picoline
K_d [mM]	-	0.05, 0.07	0.04, 0,07	0.04, 0.05		0.04
X-ray resolution	2.20 Å	2.10 Å	2.10 (1.25) Å	2.10 Å	1.35 Å	1.40 Å
# crystal waters	9	3	6	5	5	5
# solvent molecules	8613	8433	8762 (9295)	8820	9254	9246
# atoms in system	30488	29960	30947 (32541)	31121	32426	32426

[a] The dissociation constant K_d for the complexes 1AEO, 1AEF, and 1AEG is available from isothermal titration calorimetry (first value) and measurements of the perturbation in the heme Soret absorption (second value).[26] "# crystal waters" refers to the number of structural water molecules from the crystal structure included in the simulation system (cf. Materials and Methods), whereas "# solvent molecules" provides the total number of solvent molecules in the water box. In the column of 1AEF, the values in brackets refer to the newly crystallized complex with 3-aminopyridine, as described in the text.

977 from 1CPE. With respect to the ligand-free system, again the crystallographic water positions are conserved very well.

For the binding-site water molecules rmsd values and atomic fluctuations were calculated, as listed in Table 4. No water exchange was observed for water at position C. Hence, the water molecules at this position show very low atomic fluctuations and low rmsd values. On the other hand, at position A water exchange occurred. The values were, therefore, calculated for the water molecule with the longest residence time at this position during the trajectory. Nevertheless, significantly larger average deviations and fluctuations are observed as compared to position C. The largest deviation is observed for water B in complex 1AEF.

Compared to the other complexes, complex 1AEF shows indeed somewhat deviating characteristics. Relatively early in the trajectory, the ligand exerts an in-plane rotation around the center of the aromatic ring such that the amino group points away from the cofactor. Shortly after, the flexible loop movement enlarges the cavity. In the remainder of the trajectory the ligand turns back to its original orientation only rarely and for very limited time spans. Obviously, this also influences the water network. For most of the time, position B is not really occupied by a water molecule, as this appears no longer favorable with the altered ligand orientation. Rather, the position is significantly shifted, and another water position within hydrogen-bond distance of the ligand's amino group is observed. The alternative binding mode prevailing in the MD simulation is not in agreement with the experimentally determined binding mode in structure 1AEF. This issue has been addressed with experimental work and additional simulations and will be described in a separate subsection below.

The simulation of the ligand-free form 1CPE shows an additional water position at the place where the potassium ion is located in the crystal structure. The ion itself is on average repositioned toward the center of the pocket. In general, the binding-site water molecules appear to be much more mobile than in the complex structures, whereas the water molecule at position C shows a similar flexibility. In the cation-free model system (analyzed for the last 1 ns of the 5 ns trajectory), a relatively high mobility of the water molecules in the cavity was observed as well. It was, thus, not possible to find a single water molecule occupying position B for the majority of the simulated time span. On average seven water molecules were found in this pocket compared to the five of the system 1CPE (both in the simulation and in the crystal structure). In addition to the values shown in Table 4 the other water molecules in the CCP W191G cation-free model pocket show an average rmsd/fluctuation (in Å) of the following: Wat305 1.79/1.72, Wat307 2.20/2.17, Wat308 2.05/2.11, Wat5226 3.31/1.49, Wat6667 2.60/1.48.

Docking. *Redocking Analysis.* To investigate whether the experimental binding modes and hydration patterns (i.e., the water locations) can be reproduced by docking calculations, the ligands of all five complexes were extracted from the respective crystal structures and first redocked into the corresponding protein structure using the docking programs GOLD and FlexX, as described in the Materials and Methods section. The structural water molecules A and B (if present) were included in the docking process in different ways: either preset as switched on (i.e., present and available for interaction) or switched off (i.e., removed from the binding site) or in toggling/displaceable mode (i.e., the decision about formation of interactions or displacement is made by the docking program). The obtained docking results for the runs in toggling/displaceable mode with GOLD/Goldscore and FlexX are summarized in the upper half of Figure 5 and discussed in the following. The details for all docking variants (including GOLD/Chemscore) are listed in Tables S2, S3, and S4 of the Supporting Information.

The binding mode of 2AP (structure 1AEO) is reproduced very well with GOLD and FlexX (rmsd of top-ranked result with respect to crystal structure ≤ 1.0 Å). Water A was correctly toggled on in both the GOLD/Goldscore and FlexX docking runs. The correct prediction of binding mode and water presence for 3AP (1AEF) seems to be more challenging when looking at the different variants with explicitly switched on or off water molecules (cf. Tables S2, S3, and S4 of the Supporting Information). Nevertheless, if both water molecules (A and B) are set in toggling or displaceable mode (which is the most relevant case for most pratical purposes), GOLD/Goldscore and FlexX correctly keep both water molecules in the top-ranked
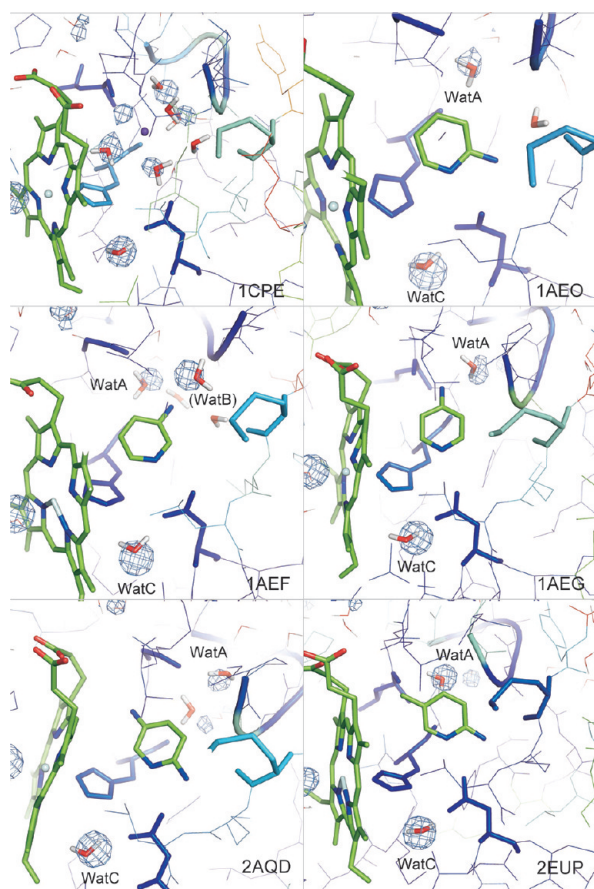
2585

dx.doi.org/10.1021/ci200052j |*J. Chem. Inf. Model.* 2011, 51, 2581–2594

Figure 4. Trajectory snapshots showing the lowest rmsd compared to the average structure calculated from the simulation. The chosen snapshots are (rmsd values given in parentheses) as follows: 4286 ps for 1CPE (0.56 Å), 4244 ps for 1AEO (0.59 Å), 3412 ps for 1AEF (0.57 Å), 4582 ps for 1AEG (0.54 Å), 2007 ps for 2AQD (0.45 Å), 3092 ps for 2EUP (0.45 Å). The binding-site amino acids are colored according to the B factors calculated from the trajectory, with "blue" referring to low and "red" to high values. In addition, average hydration density maps (calculated with VolMap/VMD) are shown, using a contour level of 70%.

Table 3. List of the Occupancies at Local Maxima in Hydration Density Maps for the CCP W191G Binding Pocket Calculated from MD Simulations[a]

system	maxima	sum	X-ray
1CPE	0.95*, 0.86, 0.85, 0.81, 0.80, 0.79, 0.71, 0.51	5.33	5
1AEO	0.99*, 0.80, 0.70, 0.45, 0.38,	2.33	1
1AEG	0.99*, 0.86	0.86	1
1AEF	0.99*, 0.90, 0.83, 0.69	2.42	2
1AEF (new)	1.00*, 0.99, 0.90, 0.56, 0.33, 0.33	3.11	2
2AQD	0.99*, 0.83, 0.76	1.59	1
2EUP	0.99*, 0.93	0.93	1
CCP W191G model	0.99*, 0.95, 0.94, 0.92, 0.90, 0.90, 0.83	5.44	

[a] Water occupancy near Asp235 (i.e., position C, which is slightly outside the pocket) is marked with a *. The sum refers to the positions in the pocket only and does not include position C. The last column reports the number of water molecules in the crystal structure next to the ligand.

Table 4. Average RMSD Values with Respect to the First Frame of the Investigated Time Span and Fluctuations (Second Value) for Water Molecules Calculated from the Trajectories[a]

	rmsd/fluctuation for water molecule		
system	A	B	C
1CPE	3.17/2.24	-	1.20/0.62
1AEO	2.38/1.30	-	1.09/0.60
1AEF	1.49/0.88	4.70/0.62	1.14/0.58
1AEF (new)	1.87/0.96	0.95/0.62	1.00/0.57
1AEG	1.45/1.18	-	1.29/0.56
2AQD	4.25/1.14	-	1.12/0.58
2EUP	3.10/0.88	-	1.11/0.56
CCP W191G model	3.69/2.04	-	1.08/0.60

[a] The values are given in Å. Data were calculated for the water molecules occupying the positions A, B, and C over the entire trajectory if no exchange occurred; otherwise, only the water molecule occupying that position for the longest time span was followed. This was the case for position A in complex 1AEO (values calculated for water molecule present from 0-4.75 ns), 1AEF (3.6-6.25 ns), 1AEG (0-6.75 ns), 2AQD (4.35-6.15 ns), and 2EUP (1-7 ns).

solutions. FlexX also provides the correct binding mode on rank 1 (rmsd 0.88 Å), whereas GOLD yields a deviating orientation (rmsd 2.23 Å) as best-scored solution, although in the same cluster a result with 1.40 Å rmsd and somewhat less favorable score is found.

For 4AP (PDB 1AEG), DA1 (PDB 2AQD), and LG4 (PDB 2EUP) top-ranked results with rmsd values below 1 Å are obtained with GOLD/Goldscore in all tested variants. In all cases, water molecule A is correctly switched on in toggling mode. It should be noted, though, that prediction of the correct binding mode was found to be independent of the presence of water A in these three complexes (cf. Table S2 of the Supporting Information). Nevertheless, compared to the docking modes obtained in its absence, higher Goldscore values are observed for all three complexes if water A is toggled on, indicating a stabilizing contribution. Similarly, redocking of DA1 and LG4 with FlexX provides the same correct binding mode (with rmsd values below 0.50 Å) regardless of the water presence; however, in these cases, also the score is unaffected, yielding virtually identical values for the different variants.

Cross-Docking Analysis. A more realistic scenario for testing the ability of a docking tool to predict binding modes and water displacement is given by docking to a hydrated but ligand-free binding pocket. Thus, the five investigated ligands were docked to the pocket of PDB-structure 1CPE, which contains five water molecules in the binding site (cf. Figure 5). The potassium ion was removed, but the water molecules were included and the docking programs were run in the toggling or displaceable mode to decide about water displacement or retention. The results are summarized in Table 5 and visualized in Figure 5.

Focusing on the top-ranked results, fairly good predictions could be obtained overall. FlexX provided results with rmsd ≤1.0 Å and correct water pattern in four of the five cases. Water molecule 979 was not literally "displaced" but rather excluded by

2586

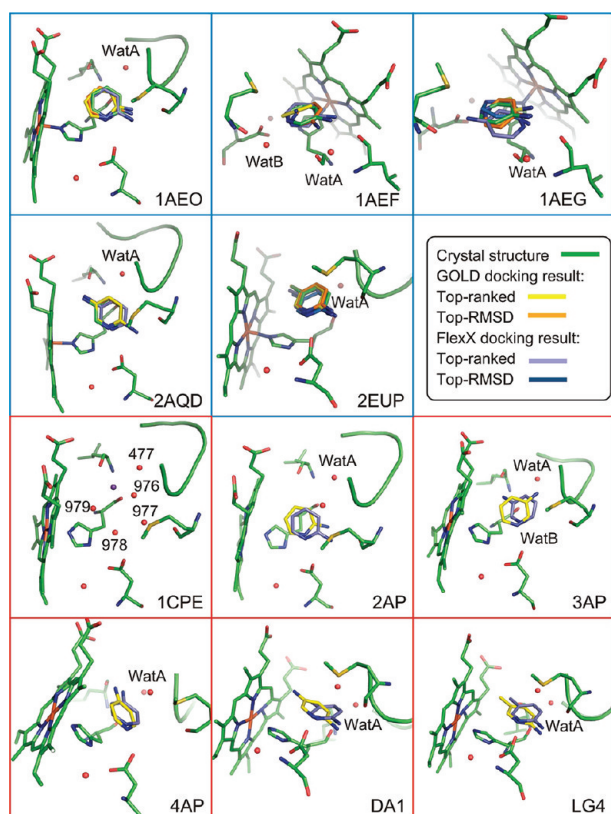dx.doi.org/10.1021/ci200052j |*J. Chem. Inf. Model.* 2011, 51, 2581–2594

Figure 5. Docking analyses of the CCP complexes 1AEO, 1AEF, 1AEG, 2AQD, and 2EUP. Images with blue border (upper half) show results from redocking with binding-site water molecules A and B set in toggling or displacable mode of GOLD or FlexX, respectively. Images with red border (lower half) show top-ranked results from cross-docking of ligands (2AP, 3AP, 4AP, DA1, LG4) from CCP W191G complexes 1AEO, 1AEF, 1AEG, 2AQD, and 2EUP into the binding pocket of the ligand-free form of CCP W191G 1CPE (which is also shown here). Binding-site water molecules were set to toggling or displacable mode of GOLD or FlexX, respectively, and the potassium ion was removed. All presented GOLD docking solutions were obtained with Goldscore as scoring function. Only water molecules present in at least one of the shown docking solutions are displayed.

FlexX from all calculations because it shows no direct interactions with the protein. For GOLD with Goldscore, the results were somewhat less precise but still between 1.0 and 1.5 Å rmsd for the same four complexes (2AP, 4AP, DA1, and LG4). In all these cases, besides water molecule A, GOLD retained an additional water molecule (976), which is not found in the crystal structures of the complexes. Both docking variants failed to correctly predict the orientation of 3AP. Although an rmsd value of 2 Å could be reached, the binding modes differ from the crystal structure, as the critical amino group points into another direction. Moreover, water molecule B was erroneously displaced by both programs. GOLD with Chemscore, finally, predicted three binding modes correctly (with rmsd <0.4 Å) and two completely off (rmsd around 3.5 Å); all water molecules were consistently toggled off by this docking variant (the GOLD/Chemscore results are, therefore, not further discussed herein).

In summary, although the cross-docking results with FlexX and GOLD/Goldscore are quite satisfying, some problems remain. First of all, 3AP is not well predicted by any of the programs. Interestingly, however, the obtained docking modes

are similar to the alternative orientation observed in the MD simulation. This is further analyzed in a separate section dedicated to complex 1AEF below. Second, with respect to water, the water molecules A, 978, and 979 are correctly switched on (A) or off (978, 979) in all five cases. On the other hand, water molecule B is always toggled off (although it should be on in the case of 3AP), and water molecule 976 is set on by GOLD in the four aforementioned cases (although it should always be off according to the crystal structures). Why are three water molecules consistently well handled, but the others not? Based on the Water-Base analysis and on the MD simulations, it was to be expected that water A is always conserved and water 979 is always replaced. The replacement of water 978 is also not too surprising, as otherwise binding of any of the ligands would hardly be possible for simple steric reasons. For water B, the results are at least consistent with the MD simulation for complex 1AEF, where it is not conserved, even though this is not in line with the experimental structure (cf. below). Water molecule 976, in turn, is not observed in the crystal structures, but significant occupancy is seen close to this position in the MD of complex 1AEO. Together with the docking observations, this could point to an essentially favorable water position with, however, somewhat too low occupancy to be observed in crystal structures.

Cross-Docking to the Binding-Site Model with Unbiased Water-Distribution. The "ligand-free" structure used for the above cross-docking is not truly devoid of a ligand because it actually contains a potassium ion, which certainly strongly influences the position of the water molecules in the cavity. Accordingly, the constructed binding-site model with unbiased water-distribution may constitute a more realistic reference structure for cross-docking. As described above, it contains a total of seven water molecules in the binding pocket. When switching only those water molecules "on" which are closest to the position of the structural water molecules observed in the corresponding complex structures, the binding modes were fairly well reproduced with FlexX (rmsd of top ranked result ≤1.25 Å in all five cases) and Gold/Goldscore (rmsd ≤1.5 Å in four cases; 2.39 Å for 3AP) (cf. Table S5 of the Supporting Information). In contrast, leaving all seven water molecules in the binding pocket and performing docking in full toggling/displaceable mode to predict the hydration pattern is clearly more challenging. Nevertheless, FlexX predicts the binding mode almost correctly (rmsd of top ranked result ≤1.25 Å; in case of 3AP, the orientation of the amino group deviates significantly from the experimental binding mode) in four cases and fails only to place DA1 at the correct position on rank 1 (the result on rank 3, however, shows an rmsd of only 0.86 Å). With respect to the water locations, the water molecule corresponding to water A is always retained by FlexX. In the case of 3AP, also the water molecule at position B is retained. In addition, however, water 305 is always retained, which does not correspond to the crystallographic observations. The same is true for water 5226 in case of the 3AP complex. All other water molecules are displaced, in agreement with the expectations. GOLD performs less well under this scenario (details of all results are shown in Table S6 as Supporting Information). Apparently, the combinatorial burden precludes a sufficient convergence at the chosen search settings. Certainly, a preselection of at least some water molecules for retention or displacement could be advantageous. For this purpose, the results from MD simulations may be of help.

Cross-Docking to the Binding-Site Model with Preselected Water Molecules. Using the cation-free model system, a 7 ns MD

2587

dx.doi.org/10.1021/ci200052j |*J. Chem. Inf. Model.* 2011, 51, 2581–2594

Table 5. Results (Top-Ranked and Lowest-RMSD Solutions) from Docking Using GOLD 4.0 and FlexX 3.1.4 into Potassium-Free Binding Pocket of 1CPE[a]

ligand	taken from complex	program	rank	CR	GS/CS/FS	rmsd (Å)	water molecules: 477(A)/976/ 977(B)/978/979
2AP	1AEO	GOLD/GS	1	1	49.91	1.42	**on**/on/off/off/off
			43	1	46.58	1.10	**on**/off/off/off/off
		GOLD/CS	1	1	23.53	0.37	**off**/off/off/off/off
			39	1	23.28	0.23	**off**/off/off/off/off
		FlexX	1	1	−16.45	0.36	**yes**/d/d/d/-
			2	2	−15.18	0.34	**yes**/d/d/d/-
3AP	1AEF	GOLD/GS	1	1	45.73	2.10	**on**/off/**off**/off/off
			47	1	44.52	1.69	**on**/off/**off**/off/off
		GOLD/CS	1	1	23.23	3.57	**off**/off/**off**/off/off
			46	1	22.92	3.52	**off**/off/**off**/off/off
		FlexX	1	1	−14.49	2.01	**yes**/d/**d**/d/-
			35	25	−9.82	0.74	**yes**/d/**d**/d/-
4AP	1AEG	GOLD/GS	1	1	47.71	1.01	**on**/on/off/off/off
			40	1	46.39	0.79	**on**/on/off/off/off
		GOLD/CS	1	1	23.02	3.51	**off**/off/off/off/off
			31	1	22.79	3.51	**off**/off/off/off/off
		FlexX	1	1	−15.55	0.58	**yes**/yes/d/d/-
			22	14	−10.26	0.50	**yes**/d/d/d/-
DA1	2AQD	GOLD/GS	1	1	46.83	1.39	**on**/on/off/off/off
			39	1	44.86	1.01	**on**/off/off/off/off
		GOLD/CS	1	1	23.03	0.36	**off**/off/off/off/off
			46	1	22.68	0.26	**off**/off/off/off/off
		FlexX	1	1	−17.27	0.95	**yes**/d/d/d/-
			12	12	−15.08	0.60	**yes**/d/d/d/-
LG4	2EUP	GOLD/GS	1	1	47.98	1.10	**on**/on/off/off/off
			50	1	44.70	0.67	**on**/off/off/off/off
		GOLD/CS	1	1	25.65	0.37	**off**/off/off/off/off
			37	1	25.04	0.29	**off**/off/off/off/off
		FlexX	1	1	−16.91	0.71	**yes**/d/d/d/-
			7	7	−13.31	0.70	**yes**/d/d/d/-

[a] Water molecules can be switched "on" or "off" by GOLD and be kept ("yes") or be displaced ("d") by FlexX. All solutions were clustered according to their mutual RMSD difference. The clusters were ranked according to the order of the highest scored solution belonging to an individual cluster. The cluster rank is given in the column labeled "CR". "GS" stands for "Goldscore", "CS" for "Chemscore", and "FS" for "FlexX Score". RMSD values were calculated between the docking solution and the crystal structure reference. In the corresponding reference crystal structures only water 477 ("A") is present, except for 3-aminopyridine (3AP) where also water 977 ("B") is observed. Water molecules which should be present in a complex are marked by bold letters. Wat979 is not displaced by FlexX; instead, the program skipped its handling because of no sufficient interaction with the protein.

simulation was started to more thoroughly probe the water locations in this system. In the hydration density maps thus obtained, besides water C only four water positions appeared as conserved. These correspond to the locations of water 308, 2238 (i.e., water A), and 6667 in the starting structure. The fourth position was not occupied by a water molecule in the starting structure and is rather a consequence of the partial opening of the PGGAAN loop in the simulation. This opening also led to a frequent exchange and larger movements of the other water molecules (including water B!), resulting in only low occupancies in all other positions.

Accordingly, for a further docking experiment, only those three water molecules in the starting structure that correspond to conserved water positions in the simulation were included and set in toggling/displaceable mode. The detailed results are reported in Table S7 of the Supporting Information. For FlexX, the results are virtually identical with those obtained with all

seven water molecules. Apparently, the four additional water molecules did hardly influence the FlexX results. In contrast, the GOLD/Goldscore results are affected by the preselection. A significant improvement is observed for 2AP (rmsd 1.27 Å) and 4AP (0.65 Å), where in both cases water A is correctly retained, and the other two water molecules are correctly displaced. For the other three ligands, the correct result could still not be obtained on the top rank.

Taken together, this experiment shows that some improvement can be obtained with an appropriate preselection based on MD hydration density maps, but it also highlights the problems of this endeavor. In particular, the water distribution obtained from a simulation of the apo state may not yield all positions that are favored in the presence of a particular ligand. In our case, this happened for water B, which was not conserved in the simulation of the model binding site. Accordingly, its presence cannot be considered by the docking runs. Similarly, shifts of water

2588

dx.doi.org/10.1021/ci200052j |*J. Chem. Inf. Model.* 2011, 51, 2581–2594

Table 6. Crystallographic Data and Refinement Statistics for "1AEF New"[e]

	details of data collection
PDB code	2Y5A
space group	P212121
unit cell dimensions (Å)	$a = 50.853$, $b = 75.362$, $c = 106.884$
resolution range (Å)	25−1.25 (1.29−1.25)
observations	417457
unique observations	113329
redundancy	3.7
completeness (%)	99.2 (97.5)
$\langle I/\sigma(I) \rangle$	42.3 (4.5)
R_{merge} (%)[a]	3.2 (18.5)

	refinement statistics
resolution range (Å)	25−1.25
R-factor (R_{work}/R_{free})[b]	14.9/16.4
number of atoms[c]	2343/440/7/43
mean B-factor(^{2})[d]	14.5/31.4/14.1/11.7
rms bond length deviation (Å)	0.007
rms bond angle deviation (°)	1.104

[a] $R_{merge} = \Sigma|I - \langle I \rangle|/\Sigma \langle I \rangle$. [b] R-factor = $\Sigma|F_o - F_c|/\Sigma F_o$. [c] Number of atoms of protein, water molecules, ligand, and cofactor, respectively. [d] Mean B-factor for protein, water molecules, ligand, and cofactor, respectively. [e] Values in brackets are for the highest resolution shell.

positions in the MD of the apo state may influence the docking results and lead to deviations compared to the experimental complex structure. In essence, the problem is related to the dynamic nature of the binding process. As long as docking relies on switchable but preset and fixed water locations (and a rigid protein) it will remain difficult or impossible in some cases to obtain the correct binding mode and hydration pattern by just relying on the hydrated apo state.

Focusing on System 1AEF. In most of the docking calculations and in the MD simulation, complex 1AEF showed somewhat deviating characteristics. The alternative binding mode observed in the MD does not agree with the experimental binding mode in structure 1AEF. Interestingly, the deviating binding modes produced by the docking calculations are similar to the ligand orientation obtained from MD. Hence, the question arose whether there are any shortcomings in the structure which lead to these conflicting results.

Since the 1AEF crystal structure was incompletely refined (cf. above), a new crystal structure was determined for this complex as described in Materials and Methods. The structure was fully refined to a resolution of 1.25 Å (Table 6); it is referred to as "1AEF new" throughout the text (PDB code 2Y5A). This new structure did not show an alternative binding mode but essentially confirmed the binding mode observed in the original 1AEF structure. Overall, only minor and apparently insignificant differences were detected, except for a second side-chain conformation of Arg48, as illustrated in Figure 6.

Despite the differences, which are at first sight insignificant with respect to the old structure, docking calculations and an MD simulation based on the "1AEF new" structure led to improved results. Details of the docking experiments are provided in Table S8 of the Supporting Information. Briefly, redocking with

GOLD/Goldscore was able to reproduce the binding mode of the ligand exactly even if both water molecules were handled in the "toggling" mode. Once more, GOLD with Chemscore provided no correct prediction of the binding mode and the hydration pattern on the top rank. In contrast, FlexX was able to precisely identify the correct binding mode and hydration pattern of 3AP in this redocking scenario. So overall, both docking programs performed much better on the basis of the new 1AEF structure. The most likely reason is the somewhat different position of the PGGAAN loop (rmsd between loop atoms: 0.52 Å), which seems to significantly impact the binding mode detection. As can be seen in Figure 6, the loop is slightly shifted toward the cofactor, thereby reducing the pocket diameter. As a consequence, also the ligand and water molecules A and B are shifted by a few tenth of an Angstrom. Such structural differences can already lead to significant differences in the docking score and, as a consequence, to different binding-mode preferences. In fact, whereas with 1AEF the best-ranked GOLD/Goldscore solution obtained a score of 44.62 (rmsd 2.23 Å), the best-scored solution in "1AEF new" obtained a value of 48.61 (rmsd 0.47 Å), indicating improved interactions.

Also a further MD simulation with the same settings as before was started from the "1AEF new" structure, with Arg48 in the new side-chain conformation. In this trajectory the alternative binding mode of 3-aminopyridine was no longer dominant. Although it was also observed, it did not constitute the major conformation. Rather, the experimentally observed binding mode was considerably more stable and prevailed overall. This also led to different water dynamics compared to the original simulation. Positions obtained from average hydration density maps (Figure 6) are well-defined and maxima (see Table 3) for positions A, B, and C are at 90% or above, indicating a high occupancy of these regions. In particular, the water molecule at position B is highly conserved at this place, which is in agreement with the fact that it can be clearly observed in the crystal structure. Consequently, rmsd values and fluctuations (Table 4) differ for the two trajectories of system 1AEF and "1AEF new". Whereas the position of water B changed in the simulation of the "old" 1AEF structure, but was then rather stable (higher rmsd, low fluctuation), in the new 1AEF simulation all three water positions are well conserved (low rmsd values). Only around position A additional maxima in the hydration density map could be found which reflect some movement of the water molecule at this position.

These observations obviously rise the question, whether longer sampling would lead to converging results for the 1AEF system, regardless whether the MD is started from the original or the newly determined structure. To address this question, both simulations were extended by a factor of 3, giving total simulation times of 21 ns for each system. Interestingly, the two systems seem to "stabilize" into different conformations (cf. Figures S4 and S5 of the Supporting Information): For "1AEF new", the native binding mode becomes more and more dominant, and transitions to the alternative binding mode occur only rarely as the simulation proceeds (cf. bottom diagram of Figure S5). In contrast, in the 1AEF simulation, the alternative binding mode dominates even more clearly in the second half of the simulation than it does in the first. Visual inspection of the two trajectories does not provide a clear reason as to why the two systems do not converge to the same binding mode. We hypothesize that the initial drifting away of water molecule B in 1AEF (caused by the small inaccuracies in the starting
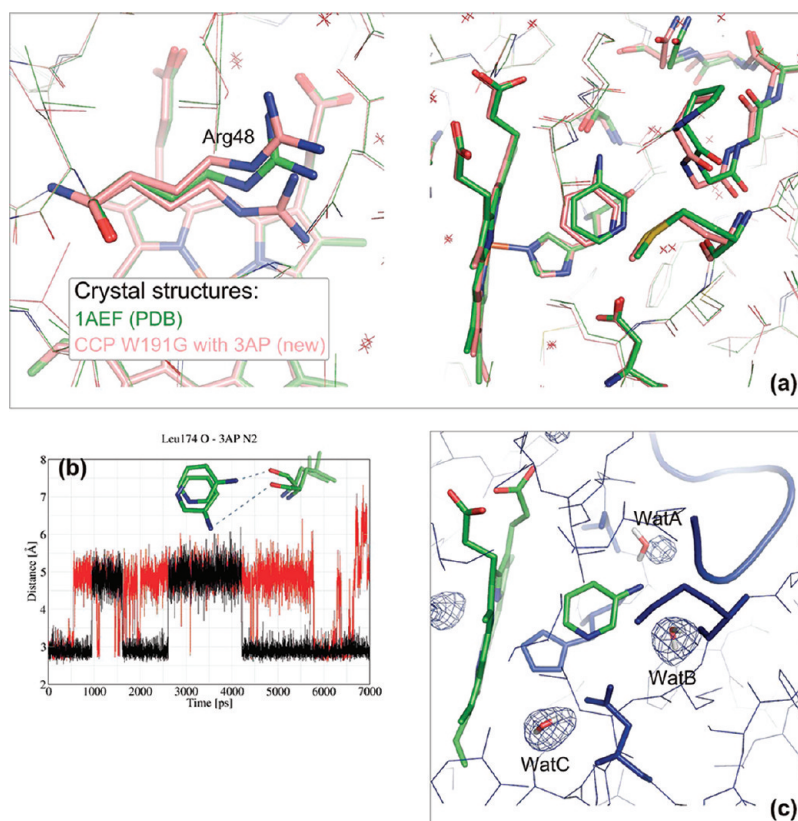
2589

dx.doi.org/10.1021/ci200052j |*J. Chem. Inf. Model.* 2011, 51, 2581–2594

Figure 6. Comparison of the newly determined crystal structure of the 3AP-CCP W191G complex ("1AEF new") with the original PDB structure 1AEF and the corresponding MD simulations. (a) Only minor differences are observed in the binding pocket, as evident from the right half of the figure. The most significant difference arises for Arg48 at the opposite side of the cofactor: In the new structure, two alternate side chain conformations are found, one of which differs significantly from the conformation of the original structure. (b) Distance between the orientationally conserved Leu174 backbone oxygen and the ligand, serving as metric for the rotation of the ligand within the binding site. The red line refers to the MD started from the original 1AEF structure, whereas the black line is based on the trajectory started from the new crystal structure. (c) Snapshot of the "1AEF new" trajectory with lowest rmsd (0.49 Å) compared to the calculated average structure of the simulation. The binding-site amino acids are colored according to the B factors calculated from the trajectory, with "blue" referring to low and "red" to high values. In addition, average hydration density maps (calculated with VolMap/VMD) are shown, using a contour level of 70%.

structure) renders a stabilization into the native binding mode more difficult. Most importantly, however, this seems to be related to movements of the PGGAAN loop. As mentioned above, early on in the 1AEF simulation the loop adopts a conformation which enlarges the cavity and facilitates the drift of water B. Subsequently, the original loop orientation is hardly revisited. In contrast, in the "1AEF new" simulation, the loop remains fairly stable in its initial position. This is reflected by the 21-ns average rmsd of the PGGAAN C_α atoms with respect to the starting structure, which is only 1.17 Å (± 0.28 Å) for the "1AEF new" simulation, but 2.44 Å (± 0.50 Å) for the original simulation. Also the B-factors calculated from the trajectories indicate much lower fluctuations for the new simulation (14.6 Å^2 for the loop C_α atoms in the "1AEF new" simulation versus 48.5 Å^2 in the 1AEF simulation). Apparently, once the loop is destabilized and water B has lost its original position, a restabilization into the native structure is an (entropically) unfavored event which does not occur within the simulated time span.

■ CONCLUSION

Despite the relative simplicity of the small cavity of the CCP W191G mutant, a complete characterization and a correct (retrospective) prediction of the water interaction network in ligand complexes of this model binding site is not trivial. Although an overall good agreement between computational results and crystallographic data is obtained, some discrepancies remain. The following conclusions about the applied methods and the investigated systems can be drawn:

Based on single crystal structures as input, water interaction fields calculated with GRID agree very well with crystallographically observed water positions in the investigated structures. All experimental water positions are found at local energy minima of the interaction fields, and no additional energetically favorable positions are obvious for any of the complexes. Analyzing the water molecules in the ligand-free structure 1CPE with Water-Base descriptors distinguishes the most conserved from the most weakly bound water molecule. However, for molecules with intermediate WaterBase-descriptor values, predicting the likelihood of conservation or replacement upon ligand binding appears hardly possible.

The main goal of water handling in structure-based drug design is the correct prediction of ligand binding modes *and* water hydration pattern from an apo structure or a given complex structure. This ability was tested here with the docking programs FlexX and GOLD. As far as redocking of the ligands into the corresponding crystal structure and consideration of the crystal

water molecules in toggling/displaceable mode is concerned, very good results are obtained with GOLD/Goldscore and FlexX (for four of the five complexes), both in terms of the ligand binding mode and the water molecules. With both programs, the small structural inadequacies in the 1AEF crystal structure lead to deviating results, which are not observed if the newly solved crystal structure is used. Also cross-docking of the five ligands to the 1CPE structure (without potassium ion, but with five water molecules in toggling mode) yields fairly good results. For all but one case the correct binding mode is obtained on rank 1 with FlexX; the results are somewhat less precise with GOLD/Gold-score. Water molecule A is always correctly retained; two other water molecules (978 and 979) are always correctly switched off. In contrast, water molecule B is never toggled on (although it should be in one case), and one additional water molecule (976) is recurrently suggested to be present by GOLD/Goldscore. As a complicating factor, a monovalent cation is present in 1CPE, which obviously influences the water positions. To obtain an unbiased hydration pattern of the ligand-free binding pocket, an ion-free model pocket was built, which contains seven water molecules in the binding site. Cross-docking to this apo-structure with all seven water molecules in toggling mode yields acceptable binding modes and hydration patterns only with FlexX. The binding modes are somewhat less precise than in the results mentioned previously, and at least one additional experimentally not observed water molecule is kept in all complexes. With seven water molecules to handle, the result is respectable, but, clearly, better results can be obtained if some preselection based on crystal structure information is carried out; this is particularly evident for GOLD. Preselection of water molecules based on the MD simulation of the ion-free binding-site model provided also some improvement for docking with GOLD. Nevertheless, even with preselection the problem of different water locations in the binding-site model (representing the apo state) and the complex structures cannot be overcome.

The MD simulations of the investigated complexes show stable binding modes and fairly high agreement of the calculated hydration density maps with the water locations in the crystal structure. Water positions of the crystal structures are occupied over major parts of the trajectory, and, with a few exceptions, no additional water positions are observed. The water density maps for position A are not always as well-shaped as for position C and the maps cover a smaller volume at a given contour level, because some exchange between position A and bulk phase occurs. At a lower contour level a local density maximum at a location not seen in the crystal structure is found for complex 1AEO, which corresponds to a position suggested to be occupied also in different docking calculations. This could be an indication for the partial presence of a water molecule not detectable by the crystallographic experiment.

Larger discrepancies between docking and MD simulations on one hand and the experimental structure on the other were initially observed for complex 1AEF. Therefore, a crystal structure of this complex was newly determined at higher resolution and fully refined; it showed an alternative side-chain conformation and a slightly shifted position of the PGGAAN loop. With this structure as a starting point for the calculations, much better agreement of the computational results with the experimental observations was obtained.

In summary, the CCP W191G cavity and its water network can be fairly well modeled by current computational tools. Through the combined use of molecular interaction fields, MD simulations and docking calculations, a mostly correct picture of the binding modes and hydration patterns can be obtained. Unexpected findings which at first sight disagree with crystallographic observations may also be due to limitations of the crystal structures. Accordingly, the combined use of crystallographic analysis and computational methods for the investigation of water interactions in protein—ligand complexes is generally recommended. Ultimately, the pure structural considerations addressed in this work need to be complemented by an energetic/thermodynamic analysis. Such work is currently under way and will be presented elsewhere.

■ MATERIALS AND METHODS

The crystal structures used in this study were obtained from the Protein Databank (PDB),[49] with the exception of the structure determined on purpose for this work, as described in detail below. A complete list of the investigated PDB structures along with further details about the different systems is given in Table 2.

Binding-site analyses by means of molecular interaction fields (MIF) were carried out with the program GRID,[42,50−52] version 22a obtained from Molecular Discovery. Ligand atoms in the PDB files were renamed according to the atom type definitions in the GRUB table. All further modifications were made with the GUI GREATER. The water probe was chosen, and a grid spacing of 0.2 Å was used in a box with a minimum distance of 10 Å between heme cofactor and box wall. All other parameters were kept at standard settings. PyMOL[53] was used for visualization of the MIFs.

For the MD simulations the crystal structures of the protein—ligand complexes were used as starting points. Structural water molecules in contact with the ligand or adjacent to these water molecules were included as well as the water molecule bound to Asp235 and the one coordinated to the heme iron. A detailed list of all included water molecules is provided in Table S1 as Supporting Information.

AMBER 9[54] and AMBER 10[55] were used to perform the simulations. Protonation patterns of histidines were carefully assigned. His96 was protonated at N_ε, whereas the proximal His175 was protonated in δ-position. Atomic charges for the ligands and the heme cofactor were calculated with the RESP[56] methodology from electrostatic potentials obtained at the HF/6-31G* level with Gaussian 03.[57] Missing force field parameters of the ligands were determined using the PARMCHK[58] module in AMBER according to the GAFF force field.[59] The missing force field parameters for the cofactor were taken from the AMBER parameter database.[60,61] The correct GAFF atom and bond types of the ligand were determined with the ANTECHAMBER program.[58] An explicit bond between the heme Fe and the N_ε of His175 was defined. For the protein atoms the Parm99SB force field of AMBER[62] was used. Missing hydrogens were added to the protein (and the explicitly included crystal water oxygens) with LEaP, which was also used for all further preparation steps. All complexes were then minimized with the module SANDER, using 200 steps of steepest descent and a modified generalized Born implicit solvent model.[63] Sodium ions were added to the minimized structures to ensure neutrality. The complexes were then solvated with the TIP3P solvent model[64] in a rectangular box with a minimum solute to box wall distance of 8.0 Å. Depending on the complex, this resulted in approximately 8400−9300 water molecules in the box. The solvated systems

2591

dx.doi.org/10.1021/ci200052j |*J. Chem. Inf. Model.* 2011, 51, 2581–2594

were heated from 100 to 300 K in 20 ps and then cooled down again to 100 K in 5 ps in the *NVT* ensemble with fixed atom positions except for water molecules and ions. Subsequently, the system was heated in the *NPT* ensemble to 300 K over 25 ps without restraining forces on the atoms. After that, a 7 ns trajectory was calculated under periodic boundary conditions, at a pressure of 1 atm and a temperature of 300 K. Pressure and temperature were controlled with the weak-coupling algorithm.[65] A time step of 2 fs was used for all simulations, applying the SHAKE[66] algorithm to all bonds involving hydrogen atoms. The particle mesh Ewald method[67] was used to treat the long-range electrostatic interactions, and an 8.0 Å cutoff was applied to the van-der-Waals interactions. Coordinates were saved every 1 ps and energy data every 10 time steps. For analysis, the generated trajectories were all centered on the protein, projected back to their initial solvent boxes, and rms fitted to the C_α-atom-positions of the first frame using the program PTRAJ. Structural descriptors such as interatomic distances, rmsd values, or the occurrence of secondary structure elements were calculated with PTRAJ. The trajectories were visualized with VMD,[48] and figures were made with PyMOL.[53] Hydration density maps were calculated with the Volmap-plugin in VMD with a resolution of 0.5 Å. Local maxima in the hydration density were extracted with the help of a python script using the APBS Python tools.[68] For this purpose, a minimum distance of 2.5 Å between two maxima was assumed (which is close to the minimum distance between two hydrogen-bonded heavy atoms).

Molecular docking experiments were carried out with GOLD 4.0[69,70] and FlexX 3.1.4.[71,72] If not stated otherwise, standard settings (i.e., for GOLD: population size 100, selection pressure 1.1, number of operations 100000, number of islands 5, niche size 2, crossover frequency 95, mutation frequency 95, migration frequency 10) were applied. Hydrogen atoms were added to the crystal structures with SYBYL 8.0.[73] Water molecules outside the binding pocket were deleted. Further setup steps were done with the GUIs of GOLD and FlexX, respectively. The protonation pattern and orientation of binding-site amino acids was carefully checked. The pocket was defined with the help of the corresponding reference ligand. The behavior of the binding-site water molecules was probed with each of the three available options in the docking programs. In GOLD 4.0 the interaction of structural water molecules can be switched "on", "off", or set in the "toggling" mode.[74] In version 3.1.4 of FlexX, water molecules can be set in an "on", "off" or "displacable" state. For the GOLD calculations the number of GA runs was set to 50 and the clustering threshold to 1.0 Å. In FlexX calculations no early termination was allowed, thus keeping all solutions. For each of the five systems (1AEO, 1AEG, 1AEF, 2AQD, 2EUP) a redocking was performed. Additionally, the ligands of these five complexes were docked into the ligand-free structure 1CPE (after deletion of the potassium ion), with all binding-site water molecules set in the "toggle" or "displacable" state, respectively.

Crystal Structure Determination. CCP W191 was purified and crystallized as described previously.[23,75] Crystals were soaked for 1 h in 50 mM ligand stock made up of 125 mM acetic acid, 25% MPD buffer (pH 4.5 with BTP as counterion) and flash frozen in liquid nitrogen. Diffraction data were collected at the ID14-1 beamline at the European Synchrotron Radiation Facility (ESRF). Data were integrated and scaled using the HKL software suite.[76] Phases were calculated by molecular replacement using MOLREP[77] with the PDB structure 1AC4 as starting model. The structure was refined using REFMAC5.[78]

Manual alterations (including side-chain adjustments and water molecule additions) were carried out in COOT.[79] The 3-aminopyridine refinement parameters were obtained from the CCP4 library.[80]

Accession Numbers. Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 2Y5A.

ASSOCIATED CONTENT

Ⓢ Supporting Information. Additional data tables of docking results, figures of molecular interaction fields, list of crystallographic water molecules included in the simulated systems, and further MD analysis data. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*E-mail: sotriffer@uni-wuerzburg.de.

Present Addresses
§Schrödinger GmbH, Dynamostrasse 13, D-68165 Mannheim.

ACKNOWLEDGMENT

REFERENCES

(1) 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.

(2) 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.

(3) Li, Z.; Lazaridis, T. Water at biomolecular binding interfaces. *Phys. Chem. Chem. Phys.* **2007**, 9, 573–581.

(4) Homans, S. W. Water, water everywhere - except where it matters? *Drug Discovery Today* **2007**, 12, 534–539.

(5) 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.

(6) Levitt, M.; Park, B. H. Water: now you see it, now you don't. *Structure* **1993**, 1, 223–226.

(7) Takano, K.; Funahashi, J.; Yamagata, Y.; Fujii, S.; Yutani, K. Contribution of water molecules in the interior of a protein to the conformational stability. *J. Mol. Biol.* **1997**, 274, 132–142.

(8) Fischer, S.; Verma, C. S. Binding of buried structural water increases the flexibility of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, 96, 9613–9615.

(9) Tame, J. R.; Murshudov, G. N.; Dodson, E. J.; Neil, T. K.; Dodson, G. G.; Higgins, C. F.; Wilkinson, A. J. The structural basis of sequence-independent peptide binding by OppA protein. *Science* **1994**, 264, 1578–1581.

(10) Sleigh, S. H.; Tame, J. R. H.; Dodson, E. J.; Wilkinson, A. J. Peptide Binding in OppA, the Crystal Structures of the Periplasmic Oligopeptide Binding Protein in the Unliganded Form and in Complex with Lysyllysine. *Biochemistry* **1997**, 36, 9747–9758.

(11) Leftheris, K.; et al. The Discovery of Orally Active Triaminotriazine Aniline Amides as Inhibitors of p38 MAP Kinase. *J. Med. Chem.* **2004**, 47, 6283–6291.

(12) Weber, P. C.; Pantoliano, M. W.; Simons, D. M.; Salemme, F. R. Structure-Based Design of Synthetic Azobenzene Ligands for Streptavidin. *J. Am. Chem. Soc.* **1994**, *116*, 2771–2724.

(13) 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.

(14) Lam, P. Y.; 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.; Erickson-Viitanen, S. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **1994**, *263*, 380–384.

(15) Grzesiek, S.; Bax, A.; Nicholson, L. K.; Yamazaki, T.; Wingfield, P.; Stahl, S. J.; Eyermann, C. J.; Torchia, D. A.; Hedge, C. N.; Lam, P. Y. S.; Jadhav, P. K.; Chang, C.-H. NMR Evidence for the Displacement of a Conserved Interior Water Molecule in HIV Protease by a Non-Peptide Cyclic Urea-Based Inhibitor. *J. Am. Chem. Soc.* **1994**, *116*, 1581–1582.

(16) Hodge, C. N.; et al. Improved cyclic urea inhibitors of the HIV-1 protease: synthesis, potency, resistance profile, human pharmacokinetics and X-ray crystal structure of DMP 450. *Chem. Biol.* **1996**, *3*, 301–314.

(17) Lam, P. Y. S.; et al. Cyclic HIV Protease Inhibitors: Synthesis, Conformational Analysis, P2/P2′ Structure-Activity Relationship, and Molecular Recognition of Cyclic Ureas. *J. Med. Chem.* **1996**, *39*, 3514–3525.

(18) Mikol, V.; Papageorgiou, C.; Borer, X. The Role of Water Molecules in the Structure-Based Design of (5-Hydroxynorvaline)-2-cyclosporin: Synthesis, Biological Activity, and Crystallographic Analysis with Cyclophilin A. *J. Med. Chem.* **1995**, *38*, 3361–3367.

(19) Clarke, C.; Woods, R. J.; Gluska, J.; Cooper, A.; Nutley, M. A.; Boons, G.-J. Involvement of Water in Carbohydrate-Protein Binding. *J. Am. Chem. Soc.* **2001**, *123*, 12238–12247.

(20) Matthews, B. W.; Liu, L. A review about nothing: Are apolar cavities in proteins really empty? *Protein Sci.* **2009**, *18*, 494–502.

(21) Kubinyi, H. In *Computer Applications in Pharmaceutical Research and Development*; Ekins, S., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2006; Chapter 16, pp 377−424.

(22) 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.

(23) Fitzgerald, M. M.; Churchill, M. J.; McRee, D. E.; Goodin, D. B. Small molecule binding to an artificially created cavity at the active site of cytochrome c peroxidase. *Biochemistry* **1994**, *33*, 3807–3818.

(24) Fitzgerald, M. M.; Trester, M. L.; Jensen, G. M.; Mcree, D. E.; Goodin, D. B. The role of aspartate-235 in the binding of cations to an artificial cavity at the radical site of cytochrome c peroxidase. *Protein Sci.* **1995**, *4*, 1844–1850.

(25) Fitzgerald, M. M.; Musah, R. A.; McRee, D. E.; Goodin, D. B. A ligand-gated, hinged loop rearrangement opens a channel to a buried artificial protein cavity. *Nat. Struct. Biol.* **1996**, *3*, 626–631.

(26) Musah, R. A.; Jensen, G. M.; Bunte, S. W.; Rosenfeld, R. J.; Goodin, D. B. Artificial Protein Cavities as Specific Ligand-binding Templates: Characterization of an Engineered Heterocyclic Cation-binding Site that Preserves the Evolved Specificity of the Parent Protein. *J. Mol. Biol.* **2002**, *315*, 845–857.

(27) Rosenfeld, R. J.; Goodsell, D. S.; Musah, R. A.; Morris, G. M.; Goodin, D. B.; Olson, A. J. Automated docking of ligands to an artificial active site: augmenting crystallographic analysis with computer modeling. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 525–536.

(28) Brenk, R.; Vetter, S.; Boyce, S. E.; Goodin, D. B.; Shoichet, B. K. Probing molecular docking in a charged model binding site. *J. Mol. Biol.* **2006**, *357*, 1449–1470.

(29) Graves, A. P.; Shivakumar, D. M.; Boyce, S. E.; Jacobson, M. P.; Case, D. A.; Shoichet, B. K. Rescoring Docking Hit Lists for Model Cavity Sites: Predictions and Experimental Testing. *J. Mol. Biol.* **2008**, *377*, 914–934.

(30) Deng, W.; Verlinde, C. L. M. J. Evaluation of Different Virtual Screening Programs for Docking in a Charged Binding Pocket. *J. Chem. Inf. Model.* **2008**, *48*, 2010–2020.

(31) Huggins, D. J.; Altman, M. D.; Tidor, B. Evaluation of an inverse molecular design algorithm in a model binding site. *Proteins: Struct., Funct., Bioinf.* **2009**, *75*, 168–186.

(32) Baron, R.; McCammon, J. A. Dynamics, Hydration, and Motional Averaging of a Loop-Gated Artificial Protein Cavity: The W191G Mutant of Cytochrome c Peroxidase in Water as Revealed by Molecular Dynamics Simulations. *Biochemistry* **2007**, *46*, 10629–10642.

(33) Baron, R.; McCammon, J. A. (Thermo)dynamic Role of Receptor Flexibility, Entropy, and Motional Correlation in Protein-Ligand Binding. *ChemPhysChem* **2008**, *9*, 983–988.

(34) Amaro, R. E.; Baron, R.; McCammon, J. A. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput.-Aided Mol. Des.* **2008**, 693–705.

(35) Miller, M. A.; Han, G. W.; Kraut, J. A cation binding motif stabilizes the compound I radical of cytochrome c peroxidase. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11118–11122.

(36) 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.

(37) Nayal, M.; Cera, E. D. Valence Screening of Water in Protein Crystals Reveals Potential Na+ Binding Sites. *J. Mol. Biol.* **1996**, *256* 228–234.

(38) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-Based Scoring Function to Predict Protein-Ligand Interactions. *J. Mol. Biol.* **2000**, *295*, 337–356.

(39) Bergner, A.; Günther, J.; Hendlich, M.; Klebe, G.; Verdonk, M. Use of Relibase for Retrieving Complex 3D Interaction Patterns Including Crystallographic Packing Effects. *Biopolymers (Nucleic Acid Sci.)* **2002**, *61*, 99–110.

(40) Hendlich, M.; Bergner, A.; Günther, J.; Klebe, G. Relibase - Design and Development of a Database for Comprehensive Analysis of Protein-Ligand Interactions. *J. Mol. Biol.* **2003**, *326*, 607–620.

(41) Günther, J.; Bergner, A.; Hendlich, M.; Klebe, G. Utilising Structural Knowledge in Drug Design Strategies - Applications Using Relibase. *J. Mol. Biol.* **2003**, *326*, 621–636.

(42) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.

(43) Ritschel, T.; Kohler, P. C.; Neudert, G.; Heine, A.; Diederich, F.; Klebe, G. How to Replace the Residual Solvation Shell of Polar Active Site Residues to Achieve Nanomolar Inhibition of tRNA-Guanine Transglycosylase. *ChemMedChem* **2009**, *4*, 2012–2023.

(44) Molecular Operating Environment (MOE). version 2008.10; Chemical Computing Group: Quebec, Canada, 2008.

(45) Henchman, R. H.; McCammon, J. A. Extracting hydration sites around proteins from explicit water simulations. *J. Comput. Chem.* **2002**, *23*, 861–869.

(46) Henchman, R. H.; McCammon, J. A. Structural and dynamic properties of water around acetylcholinesterase. *Protein Sci.* **2002**, *11*, 2080–2090.

(47) Henchman, R. H.; Tai, K.; Shen, T.; McCammon, J. A. Properties of Water Molecules in the Active Site Gorge of Acetylcholinesterase from Computer Simulation. *Biophys. J.* **2002**, *82*, 2671–2682.

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

(49) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(50) Wade, R. C.; Clark, K. J.; Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds. *J. Med. Chem.* **1993**, *36*, 140–147.

(51) Wade, R. C.; Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. *J. Med. Chem.* **1993**, *36*, 148–156.

(52) Carosati, E.; Sciabola, S.; Cruciani, G. Hydrogen Bonding Interactions of Covalently Bonded Fluorine Atoms: From Crystallographic Data to a New Angular Function in the GRID Force Field. *J. Med. Chem.* **2004**, *47*, 5114–5125.

(53) The PyMOL Molecular Graphics System. version 0.99; DeLano Scientific: San Carlos, CA, 2002.

(54) Case, D. A. et al. AMBER 9; University of California: San Francisco, 2006.

(55) Case, D. A. et al. AMBER 10; University of California: San Francisco, 2008.

(56) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(57) Frisch, M. J. et al. Gaussian 03, Revision D.01. 2004; Gaussian, Inc.: Wallingford, CT.

(58) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.

(59) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(60) Amber Parameter Database. http://pharmacy.man.ac.uk/amber/ (accessed November 30, 2006).

(61) Giammona, D. A. Ph.D. Thesis, University of California, Davis, 1984.

(62) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049–174.

(63) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 383–394.

(64) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(65) Harvey, S. C.; Tan, R. K.; T. E. Cheatham, I. The flying ice cube: Velocity rescaling in molecular dynamics leads to violation of energy equipartition. *J. Comput. Chem.* **1998**, *19*, 726–740.

(66) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(67) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(68) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037–10041.

(69) Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245*, 45–53.

(70) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1994**, *267*, 727–748.

(71) 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.

(72) Rarey, M.; Kramer, B.; Lengauer, T. The Particle Concept: Placing Discrete Water Molecules During Protein-Ligand Docking Predictions. *Proteins: Struct., Funct., Bioinf.* **1999**, *34*, 17–28.

(73) Sybyl. version 8.0; Tripos Inc.: St. Louis, MO, 2007.

(74) 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.

(75) Musah, R. A.; Jensen, G. M.; Bunte, S. W.; Rosenfeld, R. J.; Goodin, D. B. Artificial protein cavities as specific ligand-binding templates: characterization of an engineered heterocyclic cation-binding site that preserves the evolved specificity of the parent protein. *J. Mol. Biol.* **2002**, *315*, 845–857.

(76) Otwinowsky, Z.; Minor, W. In *Methods in Enzymology*; Carter, C. W. J., Sweet, R. M., Eds.; Academic Press: New York, NY, 1997; Vol. 276, pp 307–326.

(77) Vagin, A.; Teplyakov, A. Molecular replacement with MOL-REP. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 22–25.

(78) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1997**, *53*, 240–255.

(79) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126–2132.

(80) CCP4, The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994**, *50*, 760–763.