Role of Water in Molecular Docking Simulations of Cytochrome P450 2D6

Rita Santos, Jozef Hritz, and Chris Oostenbrink*

Leiden-Amsterdam Center for Drug Research, Section of Molecular Toxicology, Department of Chemistry and Pharmaceutical Sciences, Vrije Universiteit, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands

Received August 7, 2009

Active-site water molecules form an important component in biological systems, facilitating promiscuous binding or an increase in specificity and affinity. Taking water molecules into account in computational approaches to drug design or site-of-metabolism predictions is currently far from straightforward. In this study, the effects of including water molecules in molecular docking simulations of the important metabolic enzyme cytochrome P450 2D6 are investigated. The structure and dynamics of water molecules that are present in the active site simultaneously with a selected substrate are described, and based on this description, water molecules are selected to be included in docking experiments into multiple protein conformations. Apart from the parent substrate, 11 similar and 53 dissimilar substrates are included to investigate the transferability of active-site hydration sites between substrates. The role of water molecules appears to be highly dependent on the protein conformation and the substrate.

INTRODUCTION

Water is a highly versatile component at the interface of biomolecular complexes because of its unique physical chemical properties: it can act both as a hydrogen-bond donor and as a hydrogen-bond acceptor; it imposes few steric constraints on bond formation; and it is able to form hydrogen-bond networks, occupying less space than the polar side chains of a protein. Therefore, water can offer a high level of adaptability to a surface, allowing promiscuous binding, or at the same time, it can provide increased specificity and affinity to an interaction.[1] The role of water molecules can be as a solvent, involved in the stabilization of a biomolecular complex; as a bridge between two molecules; as a reagent, being a substrate or product of many enzymatically catalyzed reactions; as a lubricant, through the formation of a network linking distant residues or by promoting the flexibility required for the events of the catalytic cycle;[2] and as building block of macromolecules, being part of the receptor binding site and, thereby, altering the topological surface and recognition determinants.[3]

Because of the importance of water molecules in biological systems, discussion has arisen in the literature as to whether these molecules should or should not be included in computational approaches for structure-based drug design, mainly in molecular docking simulations.[4−10] However, no consensus has been reached about the role of water molecules in such simulations. For instance, Huang et al. reported that the docking accuracy of 12 targets from the DUD database was improved by including water molecules,[11] whereas Birch et al. described the opposite in a study of neuraminidase structures.[12]

Cytochromes P450 are a family of heme-containing oxygenases and are considered to be highly important in phase I biotransformation.[13,14] There are several human subfamilies of cytochromes P450, mainly involved in the synthesis of critical signaling molecules for homeostasis of the organism and in the metabolism of endogenous (fatty acids, steroids, prostaglandins) and exogenous (natural products contained in plants, drugs, environmental pollutants) compounds. By rendering the compounds more water-soluble, their excretion is facilitated.[13]

Human cytochrome P450 2D6 (CYP2D6) is considered to be an important drug-metabolizing enzyme, despite the fact that it corresponds to only approximately 2% of the cytochrome P450 liver content. CYP2D6 is responsible for the metabolism of approximately 15−20% of the current drugs on the market, such as beta blockers, neuroleptics, antidepressants, and others.[15] Genetic polymorphisms of CYP2D6 are known. For instance, 7% of the Caucasian population does not possess this functional enzyme, resulting in the defective metabolism of many important drug molecules.[16] Therefore, the assessment of the metabolism of drugs under development is desirable as early as possible to prevent adverse drug effects. Metabolism prediction in silico can speed up the identification of compounds that might be relevant for further investigation. We distinguish methods that are based on the molecular structure and reactivity of substrates solely from methods that take the protein structure explicitly into account.[17] The commercial program MetaSite combines both approaches.[18] As an alternative, molecular docking simulations are most commonly used to predict the orientation of substrates in the enzyme active site.

The apo crystal structure of human cytochrome P450 2D6 was recently resolved by Rowland et al.,[19] and so far, this is the only available crystal structure of CYP2D6. Eleven water molecules can be found in the crystal structure of CYP2D6; however, they are positioned more than 1.0 nm away from the heme iron. A previous study using a homology model of CYP2D6 indicated that strategically placed static water molecules in the active site can significantly influence the binding pose of 65 substrates.[20] On the

* Corresponding author phone: +31 20 5987606; fax: +31 20 5987610; e-mail c.oostenbrink@few.vu.nl.

10.1021/ci900293e CCC: $40.75
Published on Web 11/09/2009

other hand, we showed recently that active-site flexibility and plasticity could account for similar effects.[21] It is unclear what the role of active-site water molecules is in the orientation of the substrates in the active site. Should water molecules be included explicitly in molecular docking simulations, or is their effect negative? Is this role substrate-dependent, and can the selection of water molecules in the active site be automated?

In this study, we tried to address some of these questions by analyzing the behavior of water molecules in the active site of cytochrome P450 2D6 in a complex with *R*-3,4-methylenedioxy-N-ethylamphetamine (MDEA) using molecular dynamics (MD) simulations. The dynamic properties of these water molecules were determined for a few selected protein conformations, and the water molecules that seemed to influence the position of the ligand most were identified. It was our aim to identify the role of these waters and their influence on the reliability of predictions of the site of metabolism (SOM) by molecular docking simulations for MDEA, along with sets of similar and dissimilar compounds. MDEA was chosen as a representative because, in our previous work, we found that MD-generated structures with this substrate can accommodate many substrates while still leaving room for improvement of the SOM predictions.[21]

MATERIALS AND METHODS

All docking experiments were performed using protein structures that were extracted from a molecular dynamics (MD) simulations of cytochrome P450 2D6 in complex with the substrate (*R*)-3,4-methylenedioxy-N-ethylamphetamine (MDEA). Initial coordinates of the protein were obtained from the protein databank (www.pdb.org) entry code 2F9Q.[19] For details concerning the preparation of the complex structure and simulation settings, we refer to our earlier publication.[21] In short, the protein was solvated in 20292 SPC water molecules[22] and 7 Na^+ ions. Using the GROMOS simulation package,[23] a 10-ns simulation was performed at a constant temperature of 300 K and a pressure of 1 atm. The temperature and pressure were kept constant by weak coupling, using relaxation times of 0.1 and 0.5 ps, respectively.[24] The isothermal compressibility was set to 4.575×10^{-4} (kJ mol^{-1} nm^{-3})$^{-1}$. All bond lengths were constraint to their optimal values using the SHAKE algorithm[25] with a relative geometric accuracy of 10^{-4}. All interactions were calculated according to the GROMOS force field, parameter set 45A4.[26,27] Nonbonded interactions were calculated using a triple-range cutoff scheme. At every time step (2 fs), nonbonded interactions at distances shorter than 0.8 nm were calculated using a pair list that was constructed every five steps. Upon pair-list construction, interactions at distances up to 1.4 nm were also calculated and kept constant between pair-list updates. A reaction-field contribution[28] was added to the electrostatic interactions and forces to account for a homogeneous medium outside the cutoff sphere, using a relative dielectric permittivity of 61.[29] No additional constraints were added to the simulation.

Coordinates were stored every 10 ps. Starting at 2 ns, eight structures (A−H) were selected every 1 ns to comprise the test set for docking experiments. For each of these structures, the water behavior was analyzed by considering the 500 ps before and the 500 ps directly following the snapshot. All protein structures were superposed based on the backbone atoms prior to analysis. The distribution of water molecules in the active site was obtained by placing the protein on a regular grid with a spacing of 0.05 nm and monitoring the number of water molecules closest to the grid points. For every static protein conformation, the water molecules occupying clusters of grid points with a probability at least 30 times larger than the probability of finding a water molecule at a similar grid point in bulk water (0.4% for SPC at a density of 970 g/L) were selected initially. Even though the affinity of the water molecules was not calculated explicitly,[30,31] the high probability of occurrence at specific sites can be considered a thermodynamic measure. Hydrogen bonds were analyzed for these water molecules using a geometric criterion. A hydrogen bond was defined as having a minimum donor−hydrogen−acceptor angle of 135° and a maximum hydrogen−acceptor distance of 0.25 nm. The diffusion of water molecules relative to the protein structure, *D*, was calculated from the mean-square displacement using the Einstein equation

$$D = \lim_{t \to \infty} \frac{|\mathbf{r}_0 - \mathbf{r}(t)|^2}{2N_d t} \quad (1)$$

where $\mathbf{r}_0$ and $\mathbf{r}(t)$ are the water positions in a reference configuration and at time t, respectively. N_d is the number of dimensions that are being considered, which is 3 here.

From the protein structures A−H, the substrate, ions, and water molecules that were not selected were removed. Coordinates for hydrogen atoms that were implicitly treated in the simulations were added, and the resulting files were converted into mol2 file format using the standard Tripos atom and bond types for the amino acids and the heme group. Docking was performed using GOLD (Genetic Optimization for Ligand Docking),[32] version 3.3.1, in combination with the Chemscore scoring function[33] parametrized for heme-containing proteins.[34] The center point for docking was placed in the middle of the cavity between residues Phe120 and Phe483. The radius from this point was set to 1.8 nm to include the solvent channel in the accessible volume for the docking. At most, 100000 operations in the genetic algorithm were performed using a population of 100 genes. At least five independent docking simulations were performed to reach statistical significance. At least 50 poses were stored from every docking simulation, except if the best three docking poses had root-mean-square deviations smaller than 0.15 nm. To prevent an early convergence, a relative pressure of 1.1 was specified, and to account for diversity, the number of niches was set to 2. Ligand flexibility was specified as follows: The flipping of the free corners of ligand rings, the flipping of amide bonds, and the flipping of planar nitrogens were allowed. Intramolecular hydrogen bonds were also allowed. GOLD offers the possibility of automatically determining whether a specific water molecule should be bound or displaced by turning its interactions ON or OFF during the docking simulation.[35] In addition, the water molecules can be allowed to spin around their three principle axes. Because this increases the number of degrees of freedom, GOLD developers advise that it be done for at most three water molecules at a time (private communication). For structures in which more than three waters were selected based on local densities, preliminary docking experiments

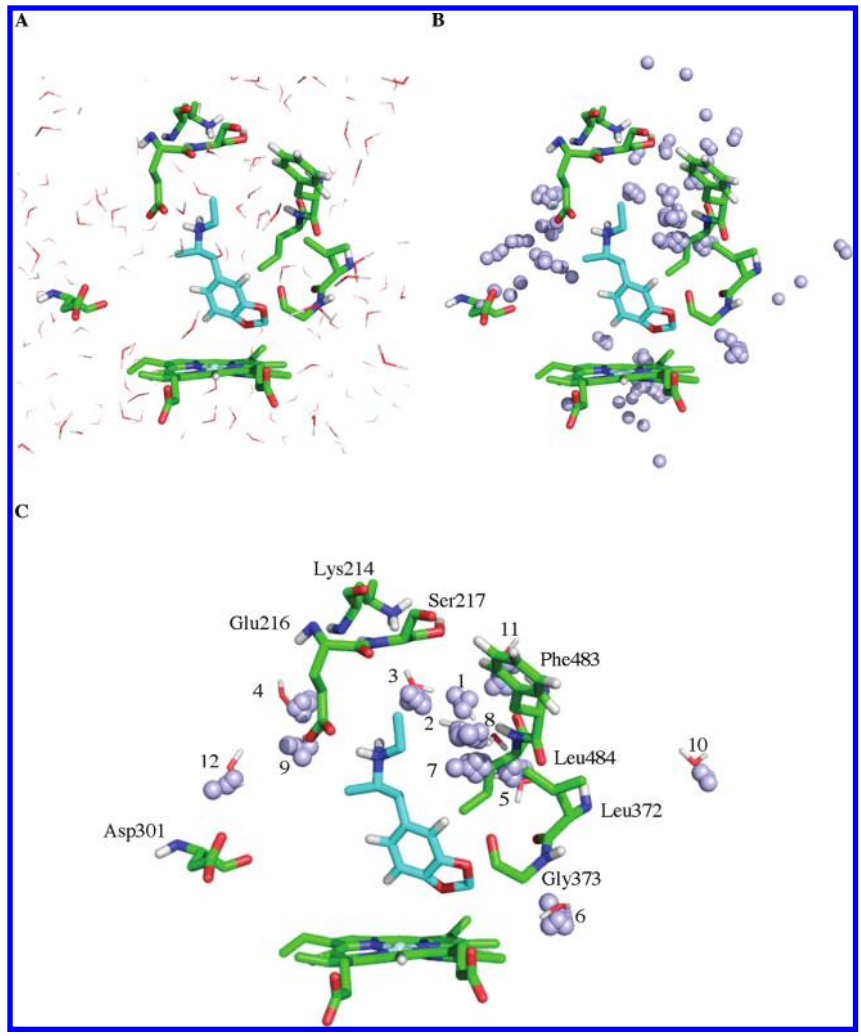

Figure 1. (A) Snapshot (denoted conformation A) of the molecular dynamics simulations of the complex of cytochrome P450 2D6 with MDEA. (B) Identification of hydration sites in the cavity of cytochrome P450 2D6 by determining the probability of occurrence over a grid. (C) Selection of the most static hydration sites in protein conformation A. Selected residues in the binding site are highlighted, and the hydration sites are numbered in accordance with Table 1.

were performed using MDEA, to determine the three water molecules for which the possibility of toggling between ON and OFF was most crucial for finding the correct binding pose. A water molecule was, therefore, considered to be critical if it increased the reliability of the docking prediction when determined to be ON and decreased the reliability prediction when determined to be OFF. In subsequent docking simulations, the effects of these water molecules on the binding poses for 11 substrates that are similar to MDEA and for 53 substrates that are dissimilar to MDEA were explored.[21] For each of these substrates, the site of metabolism (SOM) was determined experimentally; for complete references, see ref 20. As before, we considered a docking pose to be incorrect if the substrate site of metabolism was farther from the heme iron atom than 0.6 nm.[20,21] The final SOM prediction was subsequently based on the binding mode with the highest score among the five independent docking simulations.

RESULTS

Identification of Hydration Sites. Analysis of the trajectory of CYP2D6 with MDEA (Figure 1a) revealed hydration sites in the active site of CYP2D6 close to the substrate (Figure 1b). To discriminate between genuine hydration sites and false-positive hydration sites, the regions identified for protein conformation A were analyzed by visual inspection with VMD.[36] A genuine hydration site was defined as a position where a water molecule would remain for at least 80% of the time, which corresponds to an occurrence at least 30 times larger than that for a similar site in bulk water. Water molecules that were within 0.15 nm of a grid point belonging to a hydration site were initially selected for each protein conformation. In this way, dynamic changes in the shape and size of the hydration site were taken into consideration. This increase in the hydration site region also accounts for the dynamics of the water molecules, as even static water molecules still move considerably around a certain position (0.05−0.1 nm).

Twelve hydration sites were identified in protein conformation A (Figure 1c and Table 1). Between 10 and 17 hydration sites with dynamics similar to those described for protein conformation A could be identified in the other protein conformations.

Selection of Water Molecules for Docking. Despite our attempt to select only the most static water molecules from the molecular dynamics simulations of CYP2D6, the number of water molecules selected was still too high to reliably include in the docking experiments. To reduce the number

Table 1. Parameters That Reflect the Dynamics of the Water Molecules Identified in Protein Conformation A

hydration site	N_{HS}[a] (%)	V_{HS}[b] (10^{-4} nm^3)	r_{HS}[c] (nm)	HB_{pro}[d](%)	HB_{lig}[e](%)	HB_{HOH}[f] (%)	f_p[g]	HOH[h]	D[i] (10^{-3} nm^2 ps^{-1})
1	14	3.75	0.054	100	0	0	11.7	466	0.094
2	18	5.00	0.060	5.6	5.6	100	11.3	467	0.148
3	26	7.50	0.068	0	96.2	3.8	10.8	468	0.238
4	27	7.50	0.068	81.5	0	50	11.3	469	0.098
5	22	6.25	0.064	0	0	100	11.0	470[j]	0.224
6	39	10.00	0.075	100	0	0	12.2	471	0.064
7	32	8.75	0.072	0	0	90.6	11.4	472[j]	0.154
8	18	3.75	0.054	0	16.7	100	15.0	473	0.048
9	16	5.00	0.060	6.3	0	100	10.0	474	0.082
10	16	3.75	0.054	81.3	0	56.3	13.3	475	1.37
11	35	8.75	0.072	91.4	0	88.6	12.5	476[j]	0.156
12	15	3.75	0.054	0	0	100	12.5	477	0.338
bulk water	0.4	1.25	–	–	–	–	1.00	–	2.3[k]

[a] Hydration site occupancy (N_{HS}), defined as the sum of all occupancies of the grid points belonging to the hydration site. [b] Hydration site volume (V_{HS}), defined as the volume of one grid point (1.25×10^{-4} nm^3) multiplied by the number of grid points contributing to the hydration site. [c] Hydration site radius (r_{HS}), defined as the radius of a sphere with the same volume as the hydration site. [d] Occurrence of hydrogen bonds between the water molecules in the cluster and the protein. [e] Occurrence of hydrogen bonds between the water molecules in the cluster and the ligand. [f] Occurrence of water−water hydrogen bonds. [g] Spot density ratio, $f_p = \rho_{HS}/\rho_B$, where ρ_{HS} is the spot density for the hydration site, $\rho_{HS} = N_{HS}/V_{HS}$, and ρ_B is the bulk water density in molecules per nm^3 (32 nm^{-3}, calculated from the density of SPC water, 970 g/L). [h] Sequence numbers of the water molecule occupying the site in protein conformation A. [i] Relative diffusion constant (D) determined using eq 1. [j] Selected water molecules to be used for molecular docking simulations. [k] Value taken from ref 41.

of waters to be use for docking, preliminary docking studies were performed to identify the 3 water molecules that most influenced in a positive way the docking poses of MDEA. Therefore, water molecules that have a tendency to be ON when the SOM is further than 0.6 nm from the iron and at the same time have a tendency to be OFF when the SOM is within 0.6 nm from the iron, were removed from the complex, and in the end, only the water molecules responsible for improved docking poses remained. For some protein conformations, it was not clear which water molecules to remove from the initial selection, in these cases a visual inspection of the preliminary docking poses was carried out to identify the water molecule(s) responsible for a pose where the SOM was far from the iron. The improvement observed in the preliminary docking results for protein conformation A, could not be achieved by using the water molecules from the 3 hydration sites with the highest occupancy (data not shown).

The selected waters for protein conformation A were HOH470, HOH472, HOH476 belonging to cluster 5, 7, and 11 respectively (Figure 1c and Table 1).

Dynamic Properties of Selected Water Molecules. On the 1-ns time scale, the hydration sites identified are quite static positions, as can be seen for protein conformation A in Table 1. Between different protein conformations, the identified hydration sites could differ though. The spot density ratios are at least 10 times larger than for bulk water. The same trend can be observed for the diffusion constants of the water molecules within the hydration sites. A more thorough analysis of bulk water simulations revealed that no hydration site could be identified with a density greater than 640 molecules per cubic nanometer, and the hydration sites in the active site of CYP2D6 were found to be even more pronounced than that (minimum of 960 molecules/nm^3). However, only a few hydration sites revealed a high probability to form hydrogen bonds between the water molecules contributing to the hydration sites and the protein, and even fewer for the ligand. This indicates the possibility of a hydrogen-bond network between water molecules, or the existence of trapped water molecules in the cavity of CYP2D6 with a lubricant role. Because of the low resolution of the sampling times in MD, the average residence time can not be calculated with precision, but it was observed that the average residence time was above 10 ps only in some cases, being below 10 ps in the majority of the cases. For 70% of the hydration sites, it was found that more than one water molecule contributed to it over 1 ns. On average, three water molecules contributed to the same hydration site, with a maximum of five water molecules per site. Even though the density of water molecules at the hydration sites was quite high, the water molecules themselves appeared to be quite mobile.

Table 2. Comparison of Molecular Docking Results for MDEA, 11 MDEA-like Compounds, and 53 Non-MDEA-like Compounds in Protein Conformation A, When the Water Molecules Were Specified as OFF, Toggle, or ON

compound	water scenario	correctly docked[a] (%)
MDEA	HOH OFF	0
	HOH toggle	100
	HOH ON	100
MDEA-like	HOH OFF	90.9
	HOH toggle	90.9
	HOH ON	81.8
non-MDEA-like	HOH OFF	52.8
	HOH toggle	56.6
	HOH ON	54.7

[a] Percentage of correctly docked substrates based on the highest-ranked pose over the five independent simulations.

ON vs Toggle. Preliminary docking studies with MDEA and protein conformation A were done in order to determine the best treatment for water molecules during the simulations. The results are displayed in Table 2 and indicate that allowing the constant presence of water molecules during the docking runs (HOH ON) does not alter the docking results for MDEA, but it decreases the performance for MDEA-like and non-MDEA-like compounds when compared to the possibility of displacing these water molecules during the run (HOH toggle). Therefore, in this project, we allowed GOLD to determine during the docking run whether the water molecules would be displaced by the ligand.

Table 3. Molecular Docking Simulation Results for MDEA, MDEA-like and Non-MDEA-like Compounds Per Protein Conformation[a]

	MDEA		MDEA-like		non-MDEA-like	
protein conformation	HOH OFF	HOH toggle	HOH OFF	HOH toggle	HOH OFF	HOH toggle
A	0	100	90.9	90.9	52.8	56.6
B	100	100	81.8	90.9	62.3	54.7
C	0	0	36.4	72.7	45.3	50.9
D	0	0	45.5	63.6	41.5	49.1
E	0	100	90.9	63.6	62.3	58.5
F	0	0	45.5	90.9	45.3	41.5
G	100	0[b]	45.5	54.6	58.5	54.7
H	100	100	100	100	56.6	54.7

[a] Percentage of correctly docked substrates based on the highest-ranked pose over the five independent simulations in the (possible) presence and absence of water molecules. [b] No water was decided to be ON. The inconsistency arises because of a threshold problem; see text and Figure 2 for details.

Docking Results for MDEA. Docking results for MDEA back into the active site of CYP2D6, in the (possible) presence and absence of water molecules (Table 3), showed that, overall, the presence of water molecules improved the reliability of the docking prediction for some protein conformations and did not decrease the reliability for others, except in protein conformation G. In this conformation, the results apparently worsen. However, this is due to a pitfall of the 0.6 nm rule, which only monitors the distance between the SOM and the iron and not the geometrical similarity between docking poses. In this protein conformation, no water molecule was switched ON by GOLD, and when analyzing the docking poses, we observed that this pose was just above the limit (0.6 vs 0.61 nm), but that the poses were quite similar (Figure 2).

A distribution of the effect of water molecules on the docking results of MDEA over all protein conformations with different degrees of confidence can be found in Figure 3A. A positive value on the *x* axis indicates an improvement of the reliability of SOM predictions when water is included. Larger values indicate a better reproducibility between individual docking simulations and, thus, a larger statistical significance. A negative *x* value indicates a worsening of the results when water is included, again with more statiscal confidence for larger values. A slightly higher frequency toward the positive side of the graph can be seen, meaning that the presence of water molecules might increase the accuracy of the docking results for MDEA or at least not alter them, given that the highest frequency is at 0.

Docking Results for MDEA-like and Non-MDEA-like Compounds. Analysis of the docking results of similar and dissimilar compounds in the active site of CYP2D6 (Table 3) shows that there is a greater effect for MDEA-like compounds and almost no effect for non-MDEA-like compounds. Further analysis reveals that there is a slight tendency for an improvement of the docking results for MDEA-like compounds when water molecules are included, because the distribution is slightly shifted to the positive side of the histogram (Figure 3B, C).

Effect of Water Molecules on the Reliability of SOM Predictions. As can be seen from the results above, the effect of water molecules in the docking results is ambiguous, being highly dependent on the protein conformation and substrate.

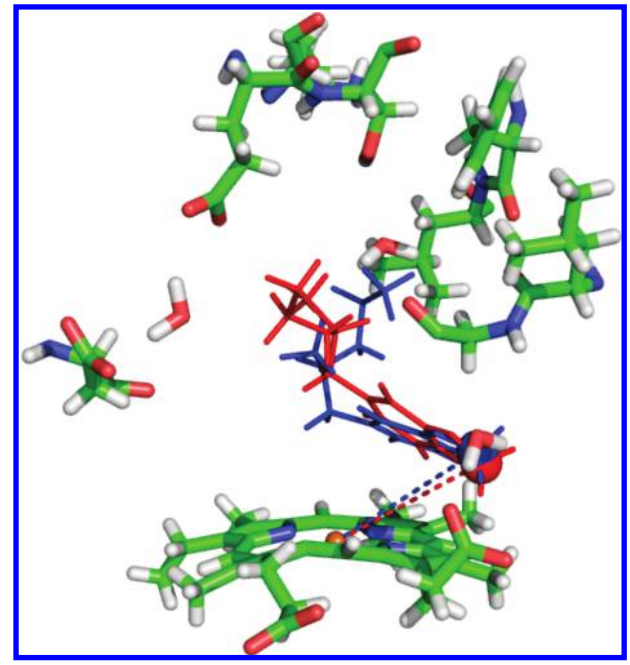

Figure 2. MDEA in frame G is a good example that the 0.6-nm rule might not be ideal for monitoring the correctness of the docking poses. By monitoring only the distance and not the poses, it seems that the presence of water worsens the SOM prediction (going from 2/5 correct to 0/5), whereas, in reality, the poses are very similar but the Fe−SOM distance increases just slightly over the threshold (0.61 nm with water vs 0.60 nm without water). Red and blue poses are achieved without water and with water, respectively. The sphere represents the SOM.

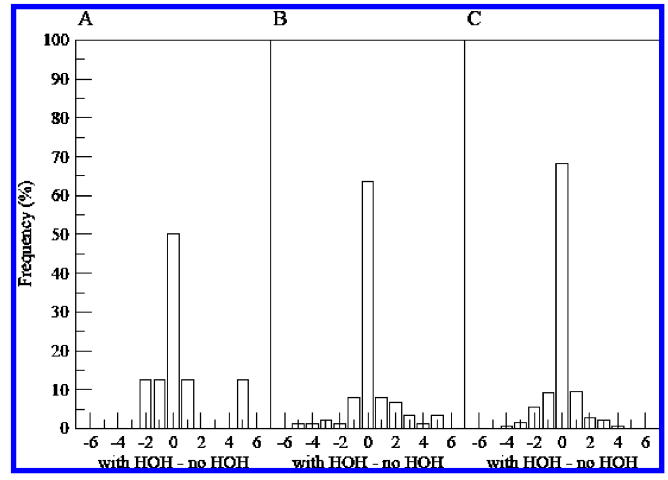

Figure 3. Distribution of reliability difference over all protein conformations between docking simulations both including possible water molecules and excluding water for (A) MDEA and (B) MDEA-like and (C) non-MDEA-like compounds. The scale of the *x* axis is as follows: A value of +5 indicates that, in five independent simulations, the substrate always docked in the first ranked pose with the SOM within 0.6 nm of the iron in the presence of water but always farther from the iron in the absence of water; −5 indicates that, in five independent simulations, the substrate always docked in the first ranked pose with the SOM far from the iron in the presence of water but within 0.6 nm of the iron in the absence of water; and 0 indicates no difference in the reliability of the docking prediction, regardless of whether it was docking with the SOM far or near the iron in the absence of water. On the *y* axis, the frequency is represented.

In Figure 4, we identify four effects of water molecules: improvement, worsening, no effect, and no improvement. In Figure 4A, there is an improvement of the docking results for MDEA, as water molecules occupy a position that

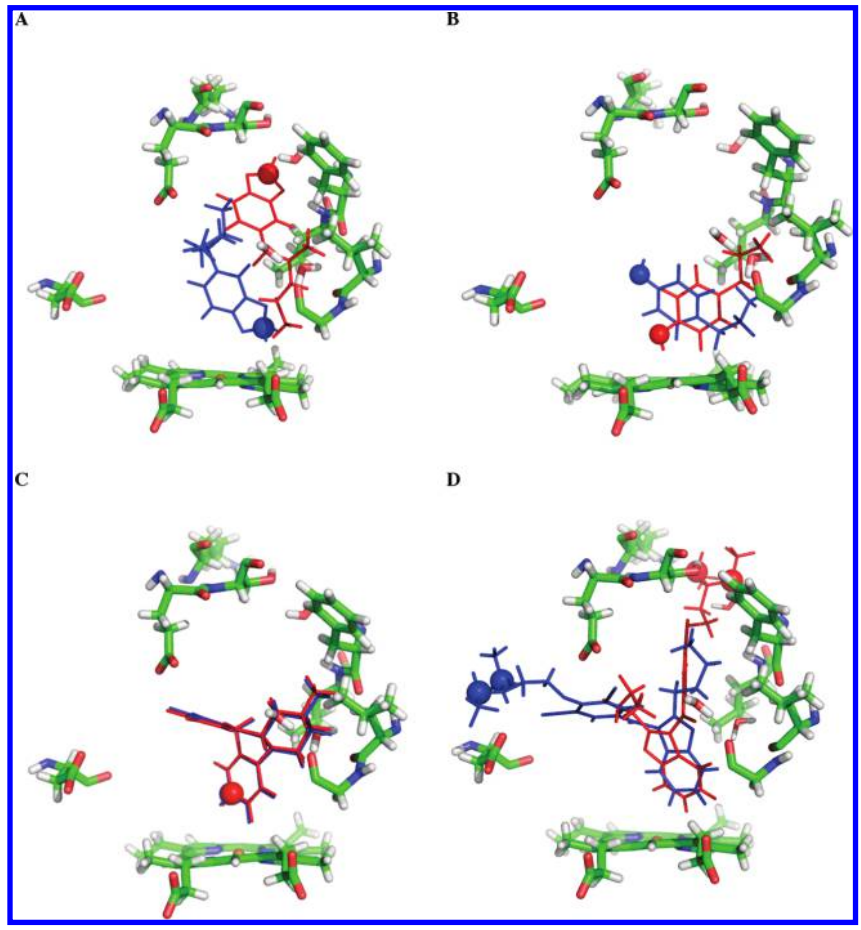

Figure 4. Effect of water molecules on the docking poses. Red poses and blue were achieved without water and with water, respectively. The sphere represents the experimentally determined SOM. (A) Improvement, MDEA in frame A: water molecules are preventing the substrate from binding in a region far above the heme group, leading to an improvement of 100%. (B) Worsening, TRP in frame A: water molecules are forming hydrogen bonds with the substrate, which leads to incorrect poses. (C) No effect, MRP in frame A: the presence of water molecules does not influence the poses. (D) No improvement, AMI in frame A: the presence of water molecules leads to different poses, but not to any with the SOM within 0.6 nm of the iron.

prevents the ligand to dock in a subpockect between Ser217 and Phe483. HOH476 is the main water molecule responsible for this improvement and is involved in a hydrogen bond with the protein 91% of the time. In Figure 4B, the substrate TRP no longer finds a catalytically active pose because water molecules are occupying a position that prevents the ligand from forming a hydrogen bond essential for GOLD to find a correctly docked pose. In Figure 4C, the docking of MRP is not affected by the (possible) presence of water molecules, as the poses are perfectly superimposed. Finally, in Figure 4D, the best-ranked pose of AMI is altered in the presence of water molecules, but the new poses still do not bring the SOM closer to the iron. Similar cases were observed for other protein conformations as well.

DISCUSSION

Water molecules play an important role in biological systems. Studies of the water molecules in the ligand-binding cavities of cytochromes P450 indicate that their high mobility facilitates the movement of the substrates and products into and out of the active site.[37] In this study, we identified hydration sites in the cavity of human CYP2D6 with an occupation probability at least 30 times larger than that in bulk water. However, water molecules contributing to it have an average residence time below 10 ps, being quite mobile, and having a low probability of forming hydrogen bonds with the protein or ligand. Rather, they are involved in hydrogen-bonded networks with other water molecules in the vicinity, occupying the empty space in the cavity. The dynamic nature of water molecules might add to the promiscuity of CYP metabolism and to the reported malleability of the active site.[38] The position of the hydration sites changes quite a lot from one protein conformation to the other, supporting the idea that the water molecules are quite mobile and considerably change the hydration shell over the span of 1 ns. However, regions with high probability of finding water molecules can be identified, and inclusion of these water molecules in a docking protocol leads to an improvement of the reliability of the SOM prediction for the compound that was used to optimize the selection of water molecules (MDEA) in several protein conformations and does not alter the results for the others. For compounds that are similar to MDEA, an effect on the pose prediction is also observed when water molecules are included, leading to a slight improvement in the results. There is virtually no effect when water molecules are included for compounds that are unlike MDEA. The water molecules seem to offer yet another way to allow CYPs to catalyze the metabolism of a wide range of structurally diverse compounds.[21,38]

It is also clear from the results that the effect of including water molecules in the docking protocol is rather dependent on the protein conformation and substrate. For instance, in protein conformations A and E, there is a clear improvement of the docking results for MDEA, but for the other protein conformations, there are no alterations in the results, except for protein conformation G. In protein conformation G, no water molecule was decided to be ON by GOLD, and therefore, the inconsistency of the results arises as a result of boundary effects of the 0.6-nm rule used to discriminate between correct and incorrect docking poses. This approach monitors only the SOM−Fe distance and does not account for the geometrical dissimilarity between poses. Because the metabolic site of the substrates is the only available experimental indication of the binding orientation, a root-mean-square geometrical measure cannot be used to distinguish "correct" and "incorrect' poses".

There are, however, a few pitfalls in this approach, that might influence the accuracy of the results and, consequently, our conclusions: (i) the GOLD parametrization to determine whether a water molecule should be displaced or bound, (ii) the fairness of comparing simulations in the presence and absence of water molecules using the same maximum number of genetic algorithm operations when the number of degrees of freedom differs, and (iii) comparison of docking results with different numbers of water molecules.

According to the literature, a water molecule is determined to be ON by GOLD if the intrinsic binding affinity of the water molecule outweighs the free energy associated with the loss of rigid-body entropy upon binding to the target.[35] The free energy associated with the loss of entropy upon binding to the target varies for different water binding sites, as tightly binding water molecules will loose more rigid-body entropy than loosely binding water molecules. However, this term in GOLD is treated as a constant for simplification. This constant is optimized for a set of water molecules in 58 crystal structures such as HIV-1, FXa, TK, and OppA, which include both highly structured (e.g., HIV-1) and rather promiscuous hydration sites (e.g., OppA). It might be too high to be used accurately with relatively mobile water molecules in the promiscuous CYP2D6, selected from molecular dynamics simulations. A more accurate determination of the water binding thermodynamics[31] is too computationally expensive to perform within a docking approach.

When docking in the possible presence of water molecules, the number of degrees of freedom increases (water ON/OFF, rotation around the principal axes), so to ensure sufficient sampling, the maximum number of genetic algorithm operations (maxops) needs to be increased. This does not hold when docking without water, because the number of degrees of freedom is smaller. However, no reliable estimate could be made of the value of maxops that would lead to a fair comparison because the degrees of freedom are very ligand-dependent and the exact relation between the number of configurations and the number of degrees of freedom is unknown.

Finally, it has been mentioned in the literature that an accurate comparison of the energy scores seems to rely on an equally homogeneous environment.[8] Therefore, the unequal numbers of water molecules among the docking poses might introduce some noise in the comparison, because it does not mimic the full hydration shell that is expected in aqueous solution. However, currently, no solution has been proposed for this problem.[8,39]

Taking these pitfalls into account, it still seems clear that water molecules do influence the docking orientations of the substrates in the CYP2D6 active site. The relevant water molecules and their presence, however, seem to be strongly dependent on the protein conformation and the substrate under consideration.

In contrast, de Graaf et al.[20] came to a different conclusion. The reliability of the SOM predictions for CYP2D6 based on docking simulations were improved by including explicit water molecules (specified as ON) for all 65 compounds (MDEA, MDEA-like, and non-MDEA-like). The positions of the water molecules were based on grid-based energy calculations, in which a region surrounding the heme iron atom was disallowed for water molecules. Because of the constant presence of water molecules during the docking simulations, the search space was restricted, and the substrate was pushed toward the heme, thereby increasing the chance of a successful SOM prediction. The improvement that was observed might not be due only to the presence of water molecules and their favorable interactions with the substrate or the protein, but might also be caused by space restrictions. This hypothesis is supported by our preliminary docking results for MDEA in protein conformation A of CYP2D6, where we see a decrease in the performance if water molecules cannot be displaced by the MDEA-like and non-MDEA-like sets of substrates. Overall, the role of water molecules in a promiscuous protein such as CYP2D6 does not appear to be straightforward. From the results for the MDEA and MDEA-like compounds, it is clear that water molecules should not be excluded from the calculations completely. On the other hand, for new substrates, different water molecules might be relevant. Therefore, we do not recommend to use a single set of water molecules for all substrates.

Analysis of the docking poses allowed us to classify the role of water molecules in four groups: improvement, worsening, no effect, and no improvement. For instance, water molecules can lead to improvement by preventing the substrate from being docked in subpockets or by forming hydrogen bonds that are involved in the stabilization of the correct pose. However, water molecules can also lead to worsening by forming hydrogen bonds that destabilize the correct pose or by forming hydrogen bonds with the protein that subsequently prevent a favorable hydrogen bond with the substrate that is essential for the correct pose to occur. However, water molecules can also not have any impact on the results, or they can simply not lead to an improvement of the docking results, meaning new poses can occur, but they do not bring the SOM closer to the heme. This effect of water molecules has been observed earlier. Some docking studies in proteases and kinases have also pointed out that, indeed, the inclusion of water molecules can lead to improvement, worsening, or no effect depending on the protein under study and its conformation.[8] We observed similar effects previously for a crystallographic water molecule in cytochrome P450 1A2.[40]

CONCLUSIONS

In this study, we demonstrated that hydration sites can be found in the cavity of CYP2D6, by monitoring the positions of water molecules in molecular dynamics simulations of CYP2D6 with MDEA. Even though the probability of finding a water molecule at these sites was at least 30 times larger than expected for bulk water, the water molecules themselves were rather mobile and dynamic, possibly adding to the well-known promiscuity of CYPs. Inclusion of selected water molecules in molecular docking simulations of CYP2D6 had an effect on the reliability of the site of metabolism prediction. However, their role is not unique, sometimes leading to a slight improvement or to no overall alterations depending on the protein conformation and the substrate. The larger effect was observed for the substrate that was used to select the water molecules. The same waters seemed to be transferable to similar substrates, still leading to a small overall improvement. However, for dissimilar compounds, no net effect could be observed. Our study sheds light on the relevant yet highly versatile role of water molecules in the CYP2D6 active site.

Abbreviations. AMI, amiodarine [(2-{4-[(2-butyl-1-benzofuran-3-yl)carbonyl]-2,6-diiodophenoxy}-ethyl)diethylamine]; CYP, cytochrome P450; CYP2D6, cytochrome P450 isoform 2D6; MDEA, *R*-3,4-methylenedioxy-N-ethylamphetamine; MD, molecular dynamics; MRP, *R*-mianserine (2-methyl-1,2,3,4,10,14b-hexahydrodibenzo[*c*,*f*]pyrazino[1,2-*a*]azepine); SOM, site of metabolism; TRP, *p*-tyramine (4-hydroxyphenethylamine).

REFERENCES AND NOTES

(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) Helms, V. Protein Dynamics Tightly Connected to the Dynamics of Surrounding and Internal Water Molecules. *ChemPhysChem* **2007**, *8*, 23–33.

(3) Cozzini, P.; Fornabaio, M.; Mozzarelli, A.; Spyrakis, F.; Kellogg, G. E.; Abraham, D. J. Water: How to Evaluate Its Contribution in Protein−Ligand Interactions. *Internat. Quant. Chem.* **2006**, *106* (3), 647–651.

(4) 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.

(5) 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.

(6) McConkey, B. J.; Sobolev, V.; Edelman, M. The performance of current methods in ligand−protein docking. *Curr. Sci.* **2002**, *83* (7), 845–856.

(7) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein−ligand docking: Current status and future challenges. *Proteins* **2006**, *65* (1), 15–26.

(8) Roberts, B. C.; Mancera, R. L. Ligand−Protein Docking with Water Molecules. *J. Chem. Inf. Model.* **2008**, *48* (2), 397–408.

(9) Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R. Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go. *Br. J. Pharmacol.* **2008**, *153* (S1), S7–S26.

(10) Corbeil, R. C.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 3. Impact of input ligand conformation, protein flexibility, and water molecules on the accuracy of docking programs. *J. Chem. Inf. Model.* **2009**, *49*, 997–1009.

(11) Huang, N.; Shoichet, B. K. Exploiting Ordered Waters in Molecular Docking. *J. Med. Chem.* **2008**, *51* (16), 4862–4865.

(12) Birch, L.; Murray, C. W.; Hartshorn, M. J.; Tickle, I. J.; Verdonk, M. L. Sensitivity of molecular docking to induced fit effects in influenza virus neuraminidase. *J. Comput.-Aided Mol. Des.* **2002**, *16* (12), 855–869.

(13) Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I. Structure and Chemistry of Cytochrome P450. *Chem. Rev.* **2005**, *105*, 2253–2277.

(14) Sono, M.; Roach, M. P.; Coulter, E. D.; Dawson, J. H. Heme-Containing Oxygenases. *Chem. Rev.* **1996**, *96* (7), 2841–2888.

(15) Williams, J. A.; Hyland, R.; Jones, B. C.; Smith, D. A.; Hurst, S.; Goosen, T. C.; Peterkin, V.; Koup, J. R.; Ball, S. E. Drug−drug interactions for UDP-glucuronosyltransferase substrates: A pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios. *Drug Metab. Dispos.* **2004**, *32*, 1201–1208.

(16) Ghoneim, M. M.; Korttila, K.; Chiang, C. K.; Jacobs, L.; Schoenwald, R. D.; Mewaldt, S. P.; Kayaba, K. O. Diazepam effects and kinetics in Caucasians and Orientals. *Clin. Pharmacol. Ther.* **1981**, *29*, 749–756.

(17) Stjernschantz, E.; Vermeulen, N. P. E.; Oostenbrink, C. Computational prediction of drug binding and rationalisation of selectivity towards cytochromes P450. *Exp. Opin. Drug Metab. Toxicol.* **2008**, *4* (5), 513–527.

(18) Cruciani, G.; Carosati, E.; DeBoeck, B.; Ethirajulu, K.; Mackie, C.; Howe, T.; Vianello, R. MetaSite: Understanding Metabolism in Human Cytochromes from the Perspective of the Chemist. *J. Med. Chem.* **2005**, *48* (22), 6970–6979.

(19) Rowland, P.; Blaney, F. E.; Smyth, M. G.; Jones, J. J.; Leydon, V. R.; Oxbrow, A. K.; Lewis, C. J.; Tennant, M. G.; Modi, S.; Eggleston, D. S.; Chenery, R. J.; Bridges, A. M. Crystal Structure of Human Cytochrome P450 2D6. *J. Biol. Chem.* **2006**, *281* (11), 7614–7622.

(20) de Graaf, 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* (8), 2417–2430.

(21) Hritz, J.; de Ruiter, A.; Oostenbrink, C. Impact of Plasticity and Flexibility on Docking Results for Cytochrome P450 2D6: A Combined Approach of Molecular Dynamics and Ligand Docking. *J. Med. Chem.* **2008**, *51* (23), 7469–7477.

(22) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction models for water in relation to protein hydration. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981; pp 331−342.

(23) Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem.* **2005**, *26*, 1719–1751.

(24) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690.

(25) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341.

(26) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; vdf Hochschulverlag AG an der ETH Zürich: Zürich, 1996.

(27) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.* **2001**, *22* (11), 1205–1218.

(28) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.* **1995**, *102* (13), 5451–5459.

(29) Heinz, T. N.; van Gunsteren, W. F.; Hünenberger, P. H. Comparison of four methods to compute the dielectric permittivity of liquids from molecular dynamics simulations. *J. Chem. Phys.* **2001**, *115* (3), 1125–1136.

(30) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817–2831.

(31) Michel, J.; Tirado-Rives, J.; Jorgenson, W. J. Prediction of the water content in protein binding sites. *J. Phys. Chem. B* **2009**, *113*, 13337–13346.

(32) 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.* **1997**, *267*, 727–748.

(33) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11* (5), 425–445.

(34) Kirton, S. B.; Murray, C. W.; Verdonk, M. L.; Taylor, R. Prediction of binding modes for ligands in the cytochromes P450 and other heme-containing proteins. *Proteins* **2005**, *58*, 836–844.

(35) 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* (20), 6504–6515.

(36) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33–38.
(37) Rydberg, P.; Rod, T. H.; Olsen, L.; Ryde, U. Dynamics of Water Molecules in the Active-Site Cavity of Human Cytochromes P450. *J. Phys. Chem. B* **2007**, *111* (19), 5445–5457.
(38) Guengerich, F. P. A malleable catalyst dominates the metabolism of drugs. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (37), 13565–13566.
(39) de Graaf, C.; Pospisil, P.; Pos, W.; Folkers, G.; Vermeulen, 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* (7), 2308–2318.
(40) Vasanthanathan, P.; Hritz, J.; Taboureau, O.; Olsen, L.; Steen Jorgensen, F.; Vermeulen, N. P. E.; Oostenbrink, C. Virtual Screening and Prediction of Site of Metabolism for Cytochrome P450 1A2 Ligands. *J. Chem. Inf. Model.* **2009**, *49* (1), 43–52.
(41) Harris, K.; Woolf, L. Pressure and temperature dependence of the self diffusion coefficient of water and oxygen-18 water. *J. Chem. Soc., Faraday Trans. 1* **1980**, *76*, 377–385.

CI900293E