Binding Mode Prediction of Cytochrome P450 and Thymidine Kinase Protein−Ligand Complexes by Consideration of Water and Rescoring in Automated Docking

Chris de Graaf,[†,§] Pavel Pospisil,[‡,§] Wouter Pos,[†] Gerd Folkers,[‡] and Nico P. E. Vermeulen*[,†]

Leiden/Amsterdam Center for Drug Research, Division of Molecular Toxicology, Vrije Universiteit Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands, and Department of Chemistry and Applied Biosciences, Institute of Pharmaceutical Sciences, ETH Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

Received May 10, 2004

The popular docking programs AutoDock, FlexX, and GOLD were used to predict binding modes of ligands in crystallographic complexes including X-ray water molecules or computationally predicted water molecules. Isoenzymes of two different enzyme systems were used, namely cytochromes P450 ($n = 19$) and thymidine kinases ($n = 19$) and three different "water" scenarios: i.e., docking (i) into water-free active sites, (ii) into active sites containing crystallographic water molecules, and (iii) into active sites containing water molecules predicted by a novel approach based on the program GRID. Docking accuracies were determined in terms of the root-mean-square deviation (RMSD) accuracy and, newly defined, in terms of the ligand catalytic site prediction (CSP) accuracy. Consideration of both X-ray and predicted water molecules and the subsequent pooling and rescoring of all solutions (generated by all three docking programs) with the SCORE scoring function significantly improved the quality of prediction of the binding modes both in terms of RMSD and CSP accuracy.

Introduction

Automated molecular docking methods have frequently been used to predict energetically favorable conformations and orientations of ligands in the interior structure of proteins. These methods combine search algorithms to generate different poses (docking), and scoring functions to determine the tightness of protein−ligand interactions.[1] Several docking algorithms and scoring functions have been described in the past few years, but docking (prediction of binding orientation) and scoring (prediction of binding free energy) accuracies of docking−scoring combinations still vary with selected protein target and physicochemistry of protein−ligand interactions.[2,3] Currently, there are several 'open issues' in automated docking, such as the inclusion or omission of explicit water molecules in the ligand binding pocket.[4,5] Despite the fact that water molecules can play an essential role in ligand−protein binding,[6−8] concrete water molecules are usually not taken into account in docking studies. Although many scoring functions used for automated docking include an energy term accounting for the free energy of desolvation of a ligand upon binding to a protein and occlusion of the ligand binding site from solvent (hydrophobic effect),[9−12] most docking methods ignore water-mediated interactions between proteins and ligands. FlexX has been extended with an algorithm for integration and placement of water molecules during docking,[13,14] but the average improvement of docking accuracy was small. With the SLIDE docking program, it is possible to predict conserved binding site water molecules, which are displaced during docking if collisions with a ligand cannot be resolved by iterative translations.[15] To include solvation effects, Glide docks explicit water molecules into the binding site for each energetically competitive ligand pose and employs scoring terms that measure the exposure of various functional groups to the explicit waters.[16] Structural water heterogeneity was recently successfully incorporated into the program AutoDock for predicting binding modes of HIV-1 protease inhibitors by using two weighted average methods of combining multiple target structures within a single grid of interaction energies.[17]

Despite the lack of effective automated algorithms for including waters 'on-the-fly' during the docking process, fixed explicit water molecules have been used in several docking programs. However, only a few docking studies are reported which evaluate the effects of fixed water molecules in ligand−protein binding sites. Moreover, very few studies show significant effects of these water molecules on docking accuracy. When two specific water molecules present in the crystal structure of Factor Xa were taken into account during the docking of synthetic inhibitors, it was observed that all inhibitors formed an essential hydrogen bond with one of the two active site water molecules.[18] In contrast, the inclusion of explicit water molecules in crystal structures of influenza virus neuraminidase did not significantly improve the docking performance of GOLD for ligands forming water-mediated contacts, but it did decrease docking accuracy of ligands displacing waters.[19] Validation of the GOLD docking program[20] with the CCDC/Astex test-set without consideration of discrete water molecules showed that docking accuracies for subsets of structures without water-mediated protein−ligand interactions were sig-

* To whom correspondence should be addressed. Phone: +31 20 5987590. Fax: +31 20 5987610. E-mail: npe.vermeulen@few.vu.nl.
† Vrije Universiteit Amsterdam.
‡ ETH Zürich.
§ These authors equally contributed to the work.

10.1021/jm049650u CCC: $30.25
Published on Web 02/17/2005

nificantly higher than accuracies for subsets with water-mediated interactions.[21] The program GRID[22] has been successfully used to select X-ray water molecules for the docking of carbohydrate derivatives to heat-labile enterotoxin with an earlier version of AutoDock.[23] Structural effects of water molecules on the docking performance were also evaluated for EUDOC.[24] It was concluded that EUDOC may fail to make correct predictions of ligand−receptor complexes when the information of structural water molecules in the binding site is not available.

Automated docking can be used to select potential protein-specific substrates from large chemical databases and to predict ADME properties of new drugs and drug candidates.[25,26] In the present study, automated docking strategies will be applied to two pharmacologically relevant biotransformation systems, namely cytochromes P450 and thymidine kinases. Cytochromes P450 (Cyt P450s) are hemoproteins which catalyze the oxidation and reduction of a wide variety of endogenous and xenobiotic compounds.[27,28] They generally detoxify potentially hazardous compounds, but in a number of cases nontoxic parent compounds are bioactivated into toxic metabolites, and procarcinogens into their ultimate carcinogens.[29] Automated docking has primarily been applied to refine and validate Cyt P450 pharmacophore models and protein homology models, as was, for example, recently done for Cyt P450 2D6.[30,31] Regarding Cyt P450s, methodological research on automated docking approaches has mainly been focused on FlexX, including cross-docking[32] and the use of different FlexX-scoring combinations.[33] These docking studies, however, did not consider the effect of water molecules. Caffeine and 2-amino-3-methylimidazo[4,5-*f*]quinoline (MeIQ) together with water molecules placed in the vicinity of their molecular electrostatic potential (MEP) minima were docked into a water-free binding site of a Cyt P450 1A2 homology model using AutoDock.[34] This study, however did not demonstrate important effects of considering "ligand-anchored" water molecules. Thymidine kinases (TKs) phosphorylate thymidine and other nucleic acid bases that are subsequently phosphorylated by other enzymes and incorporated into the DNA. The Herpes simplex virus type 1 thymidine kinase (HSV1 TK), for example, phosphorylates pyrimidine and purine analogues that are triphosphorylated to inhibit cellular DNA polymerase or to cause toxic effects when incorporated into DNA.[35] Most antiviral-based therapies exploit the large substrate acceptance of HSV1 TK relative to the human isoenzyme. An extensive evaluation of different docking programs with respect to their docking accuracy and different docking−scoring combinations with respect to their scoring accuracy, and ability to select HSV1 TK substrates from a chemical database was described previously.[2] The same enzyme was recently also considered for the evaluation of the docking and scoring performance of Glide.[16,36] These studies did not consider the effect of water molecules, however. Recently, the presence of active site water molecules was shown to have a large impact on the results of virtual screening for typical HSV1 TK substrates with a sequential DOCK−FlexX combination.[37] A water-free active site was more suitable for purine-like structures, whereas an active site filled with two essential X-ray water molecules was more suitable for pyrimidine-like compounds.

The primary aim of the present study was to find optimal docking strategies for binding mode prediction of ligands of two pharmacologically relevant enzyme systems, by evaluating the performance of various docking−scoring combinations, and considering the presence and absence of active site water molecules. The enzyme systems chosen for this purpose were cytochrome P450s (Cyt P450s) and thymidine kinases (TKs). Concretely, 19 crystallized P450− and 19 TK−ligand complexes were selected for docking. To test the incorporation of water molecules in the docking process, three scenarios were pursued based on whether docking was performed: into (i) a water free protein active site (N), (ii) an active site containing X-ray water molecules as a solid part of the protein (X), and (iii) an active site containing water molecules whose positions were unarbitrarily predicted by a novel GRID-based protocol (P). All ligand−protein complexes were predicted by three selected docking programs, namely AutoDock, FlexX, and GOLD. Moreover, the resulting docking poses were scored and ranked using the native scoring function implemented in the respective docking programs and the stand-alone scoring function SCORE. Docking accuracy was examined as root-mean-square deviation (RMSD) of heavy atoms of ligand docking poses from the reference X-ray structure and as a newly defined criterion for docking accuracy, namely catalytic site prediction (CSP) accuracy. This new endpoint of binding mode prediction can be used to reflect whether and how a putative ligand will be enzymatically transformed and thus as an analysis tool in virtual screening applications. Beside the use of three water scenarios (including not only X-ray water molecules, but also computationally predicted waters), three different docking algorithms, (re)scoring, and two cases of hard docking targets (presence of waters, various ligand acceptance, and presence of prosthetic group (heme) or cofactor (ADP) in the active site), this study presents one of the first comprehensive evaluations of the effects of the incorporation of active site water molecules on automated molecular docking.

Results

Active Site Water Molecules. Docking simulations were performed using three water scenarios: without water molecules (N), with water molecules present in the X-ray structure (X), and with predicted water molecules (P). Numbers of X-ray water molecules (x), predicted water molecules (p), and correctly predicted water molecules (c) in the vicinity ($\leq$ 4.0 Å) of ligand atoms and their interactions with reference ligand structure were analyzed to examine the correctness of the new water prediction method applied in this study, both for Cyt P450s and TKs (Figure 1A and B).

Prediction of Active Site Water Molecules. Cyt P450. In many cases, the locations of X-ray active site water molecules and their interactions with Cyt P450 ligands (filled balls, see Figure 1A) were well predicted ("c" $\geq$ 1), concretely with 2-phenylimidazole (pdb code: 1phe, Cyt $P450_{cam}$), 6-deoxyerythronolide B (1jio, Cyt $P450_{EryF}$), dimethylsulfophenazole derivative (1n6b, Cyt P450 2C5), diclofenac (1nr6, Cyt P450 2C5), and epothilone D (1pkf, Cyt $P450_{EpoK}$) (see Figure 1A). In the case of

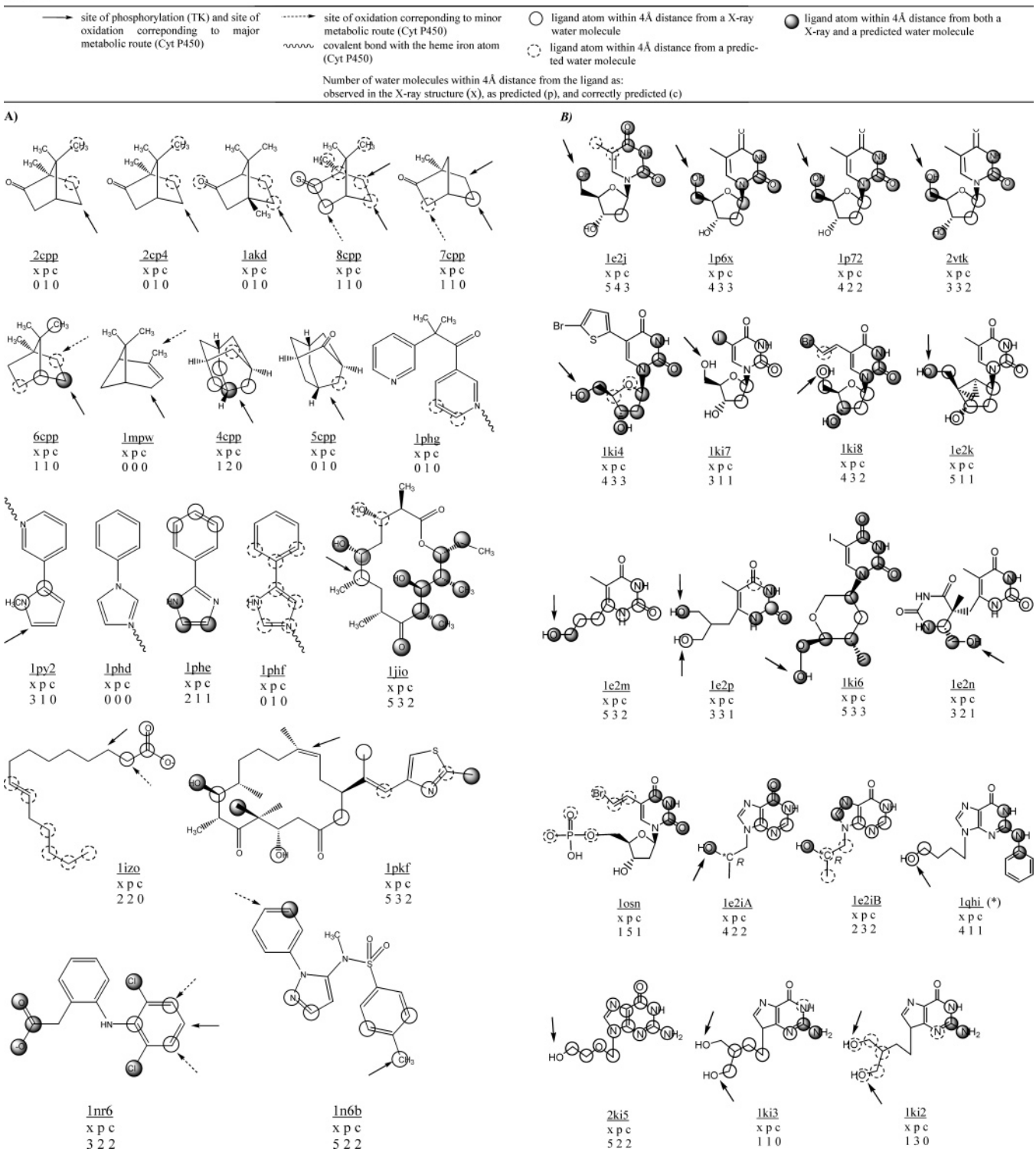

Figure 1. Molecular structures of ligands, their water-mediated interactions to the protein, and correctness of water location prediction. Symbols are explained in the legend below. (A) P450 pdb codes correspond to the following compounds: Cyt $P450_{cam}$: 2cpp/2cp4(T252A mutant), 1*R*-camphor; 1akd, 1*S*-camphor; 8cpp, thiocamphor; 7cpp, camphene; 6cpp, 1R-norcamphor; 4cpp, adamantine; 5cpp adamantanone; 1p2y, 1S-nicotine; 1phd, 1-phenylimidazole; 1phe, 2-phenylimidazole; 1phf, 4-phennylimidazole; 1phg, metyrapone; Cyt $P450_{EryF}$: 1jio, 6-deoxyerythronolide B; Cyt $P450_{EpoK}$: 1pkf, epothilone D; Cyt $P450_{BS\beta}$:1izo, palmitoleic acid; Cyt P450 2C5: 1n6b, 4-methyl-*N*-methyl-*N*-(2-phenyl-2*H*-pyrazol-3-yl)benzenesulfonamide (the atom interacting with both X-ray and predicted water and atoms only interacting with crystal waters belong to the A and B chain binding modes, respectively); 1nr6, diclofenac. (B) TK pdb codes correspond to the following compounds: TK EHV4: 1p6x and1p72, thymidine; HSV1 TK: 1e2j and 2vtk, thymidine; 1e2i, 9-hydroxypropyladenine; 1e2k, (North)-methanocarba-thymidine; 1e2m, 6-hydroxypropylthymidine; 1e2n, 6-{[4-hydroxymethyl]-5-methyl-2,6-dioxohexahydropyrimmethyl}-5-methyl-2,3(1*H*,3*H*)-pyrimidinedione; 1e2p, 6-[3-hydroxy-2-(hydroxymethyl)propyl]-5-methyl-2,4(1*H*,3*H*)-pyrimidinedione; 1ki2, ganciclovir; 1ki3, penciclovir; 1ki4, 5-bromothienyldeoxyuridine; 1ki6, 1′,5′-anhydro-2′,3′-dideoxy-2′-(5-iodouracil-1-yl)-D-arabino-hexitol; 1ki7, 5-iododeoxyuridine; 1ki8, 5-bromovinyldeoxyuridine; 1qhi, 9-(4-hydroxybutyl)-*N*2-phenylguanine ((*)inhibitor, is not phosporylated); 2ki5, aciclovir; TK VZV: 1osn, (*E*)-5-(2-bromovinyl)-2′-deoxyuridine-5′-monophosphate.

Table 1. Docking Results for Each Protein−Ligand Complex in Terms of RMSD Accuracy[a]

	Cyt P450										TK								
	AutoDock			FlexX			GOLD				AutoDock			FlexX			GOLD		
cam										HSV1									
1akd										1e2iA		X	P		X^S		N^S	•	P^S
2cp4			•	N^I	X^I	P^I	N	X	P	1e2iB		•							
2cpp				N^S	X^S	P^S	N	X	P	1e2j	•	X^S	P^S	•			•	X^I	
4cpp		X^S					N^S	X	P	1e2k	N	X	P	N	X	P	N	X	P
5cpp				N^S	X^S	P^S	N	X	P	1e2m	N	X	P	N	X	P^S	N	X	P
6cpp	N^S	X^S	P^S	•	•	•		•		1e2n	N	X	P						P
7cpp		•		N	X	P		X		1e2p	N	X	P	N	X	P	N	X	P
8cpp		•	P^S	•		•				1ki2	N	X	P	N^I	X	P^S	N	X	P
1mpw	N^S		P^I	N	X	P				1ki3	N	X	P	N	X	P	N	X	P
1phd			•	•	X^I	P^I	N^S	X^I		1ki4	N	X	P				N	X	P
1phe			P	N^S	X	P				1ki6	N	X	P	N^I	X	P	N	X	P
1phf			•	•	•	P^I			P^s	1ki7	N	X	P				N	X	P
1phg			P				N	X	P	1ki8	N^S	X	P		X	P	N	X^S	P
1p2y	•	•	•	N	X	P	N	X	P	2vtk	N	X	P	N	X	P	•	X	P
EryF										1qhi	N	X	P	N	X	P	N	X	P
1jio	N^S	X	P		X^I	P^S	N	X	P	2ki5	N	X	P	N	X	P	N	X	P
βSB										VZV									
1pkf	N	X	P	N		P	N	X	P	1osn	N	X	P	N	X	P	N	X	P
EpoK										EHV4									
1izo	N	X	P	N^S	•	P	N	X	P	1p72	N	X	P	N	X	P	N	X	P
2C5										1p6x			P					X	P
1n6b	N^S	X	P		X^S		N^I	X	P^S										
1nr6	N^I	X	P			•	•	•	•										

[a] Letters N (no water), X (X-ray water), and P (predicted water) indicate the different scenarios concerning the presence or absence of water molecules. Cases in which RMSD values are below 2 Å of only no. 1 ranked solutions according to the program implemented scoring functions (^{I}); only no. 1 ranked solutions according to the stand-alone SCORE scoring function (^{S}); no. 1 ranked solutions according to both the program implemented and stand-alone SCORE scoring functions (**bold**); any of the docking solutions, but not considered as no. 1 ranked solutions by either the program implemented or stand-alone SCORE scoring functions (●).

almost all Cyt $P450_{cam}$−terpene structures (2cpp, 2cp4, 4cpp, 6cpp, 7cpp, 8cpp), up to two water molecules were predicted to be close to the respective ligand, while not present in the X-ray structure ("p − c" = no. of false positives). Water molecules ligated to the heme Fe-atom according to crystallographic studies (4cpp, 6cpp, 7cpp, 1phe) are not predicted ("x − c" = no. of false positives). However, preliminary docking studies with a water molecule positioned at 1.8 Å above the heme iron atom did not have an overall positive effect on docking accuracy (data not shown here).

Prediction of Active Site Water Molecules. TK. In almost all TK complexes including thymidine derivatives, the locations of X-ray active site water molecules and their interactions with pyrimidine rings and sugar mimicking groups were correctly predicted ("c" ≥ 1) (see Figure 1B). For purines, less X-ray active site water locations were correctly predicted and the occurrence of false positives ("p − c") and false negatives ("p − x") was shown to be highly compound specific. Interestingly, up to 4 false positives were predicted in case of the phosphorylated thymidine metabolite (1osn, VZV TK). No water molecules were accurately predicted for therapeutically used ganciclovir (1ki2, TK HSV1) and penciclovir (1ki3, TK HSV1) (false negatives).

Binding Mode Prediction. Binding modes of Cyt P450− and TK−ligand complexes were predicted using the docking programs AutoDock, FlexX, and GOLD in combination with their native scoring function and the SCORE scoring function, while considering the three different water scenarios (*N*, *X*, and *P*, respectively). These 18 (3 (water scenarios) × 3 (docking algorithms) × 2 (scoring functions)) docking strategies were tested by assessing their ability to reproduce the experimental binding orientations of protein−ligand complexes of five Cyt P450 isoenzymes and three TK isoenzymes in terms of RMSD accuracy and CSP accuracy. Concrete results for each individual protein−ligand complex in terms of RMSD accuracy are presented in Table 1, while the statistics of RMSD and CSP accuracy are presented in Figures 2A and B, and Figures 2C and D, respectively.

Incorporation of Active Site Water Molecules in Automated Docking Studies. Cyt P450. The incorporation of X-ray water molecules increased the docking accuracy of the various docking−scoring combinations. The relative averaged increase (RAI) of the RMSD accuracy was increased (19% for AutoDock (AD), 11% for FlexX (F), and 22% for GOLD (G)), considering successfully docked complexes (RMSD < 2 Å) (Figure 2A). In addition, the RAI of the CSP accuracy was improved (7% for AD, 16% for F, and 13% for G, Figure 2C). The presence of predicted water molecules during docking studies was shown to strongly improve RMSD accuracy (RAI of 70% for AD, 32% for F, and 7% for G),

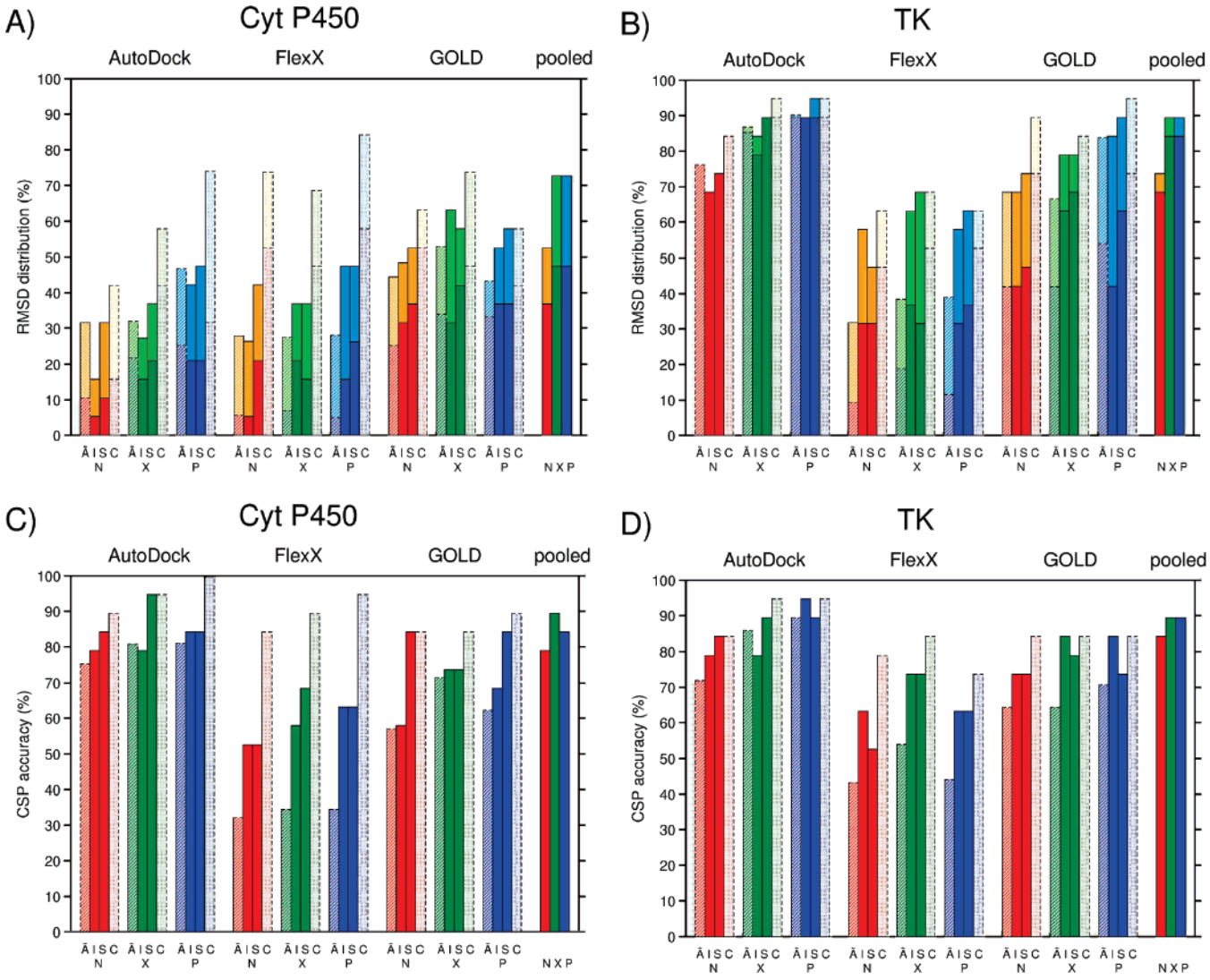

Figure 2. Docking accuracy of different automated docking approaches for Cyt P450 and TK in terms of RMSD (A and B) and CSP accuracy (C and D), considering different scenarios with respect to the presence of water N: no water; X: X-ray water; P: predicted water). Abbreviations on *x*-axis correspond to: Ā, Average RMSD distributions and CSP accuracy for all solutions of each docking study; I, RMSD distributions and CSP accuracy for no. 1 ranked solutions according to the program implemented scoring function; S, RMSD distributions and CSP accuracy for no. 1 ranked solutions according to SCORE; C, RMSD distributions and CSP accuracy for poses closest to the experimentally determined structure or accurately predicting the catalytic site, whatever its ranking. Color schemes of the bars: red (N), dark green (X), dark blue (P): RMSD ≤ 1.0 Å or correctly predicted catalytic site; orange (N), light green (X), light blue (P): RMSD > 1.0 Å, ≤ 2.0 Å.

and CSP accuracy of various docking−scoring combinations (RAI of 5% for AD, 13% for F, and 9% for G). The RMSD accuracy of the optimal docking−scoring combination for Cyt P450 without water (GOLD-SCORE) was increased with 16% and 10% by respectively including X-ray and predicted water molecules, while the CSP accuracy (AutoDock-SCORE) was increased with 11% and 0% respectively. Table 1 shows that the performance of AutoDock in finding reliable solutions for several Cyt $P450_{cam}$ terpene (4cpp, 8cpp) and imidazole (1phe, 1phg) analogues, and the ligands of Cyt $P450_{EryF}$ (1jio, shown in Figure 3) and Cyt P450 2C5 (1n6b, 1nr6), improved by including X-ray and/or predicted active site water molecules. For several ligand molecule types and different Cyt P450 isoenzymes, RMSD accuracy of FlexX (1phd, 1phe, 1phf, 1jio, 1izo, 1n6b) and GOLD (4cpp, 7cpp, 1phd, 1phf, 1n6b) was also improved by the incorporation of water.

Incorporation of Active Site Water Molecules in Automated Docking Studies. TK. Compared to the "no water" scenario (N), RMSD accuracy for thymidine kinases in the presence of X-ray waters resulted in RAI of 17% for AD, 35% for F, and 0% for G (Figure 2B). The RAI of the CSP accuracy was also increased (8% for AD, 27% for F, and 8% for G) (Figure 2D). The presence of predicted water molecules during docking simulations was also found to improve RMSD accuracy (RAI of 23% for AD, 12% for F, and 23% for G), as well as CSP accuracy of various docking−scoring combinations (RAI of 17% for AD, 7% for F, and 8% for G). The RMSD accuracy of the optimal docking−scoring combinations for TK without water (AutoDock-SCORE) was increased with 15% by including either X-ray or predicted water molecules, while the CSP accuracy (AutoDock-SCORE) was increased with 11% and 0% by including X-ray and predicted water molecules, respectively. The performance of AutoDock in finding reliable docking solutions of the pyrimidine ligands thymidine in TK HSV1 (1e2j) and EHV-4 TK (1p6x), 5-bromovinyldeoxyuridine (1ki8, TK HSV1, shown in Figure 3),

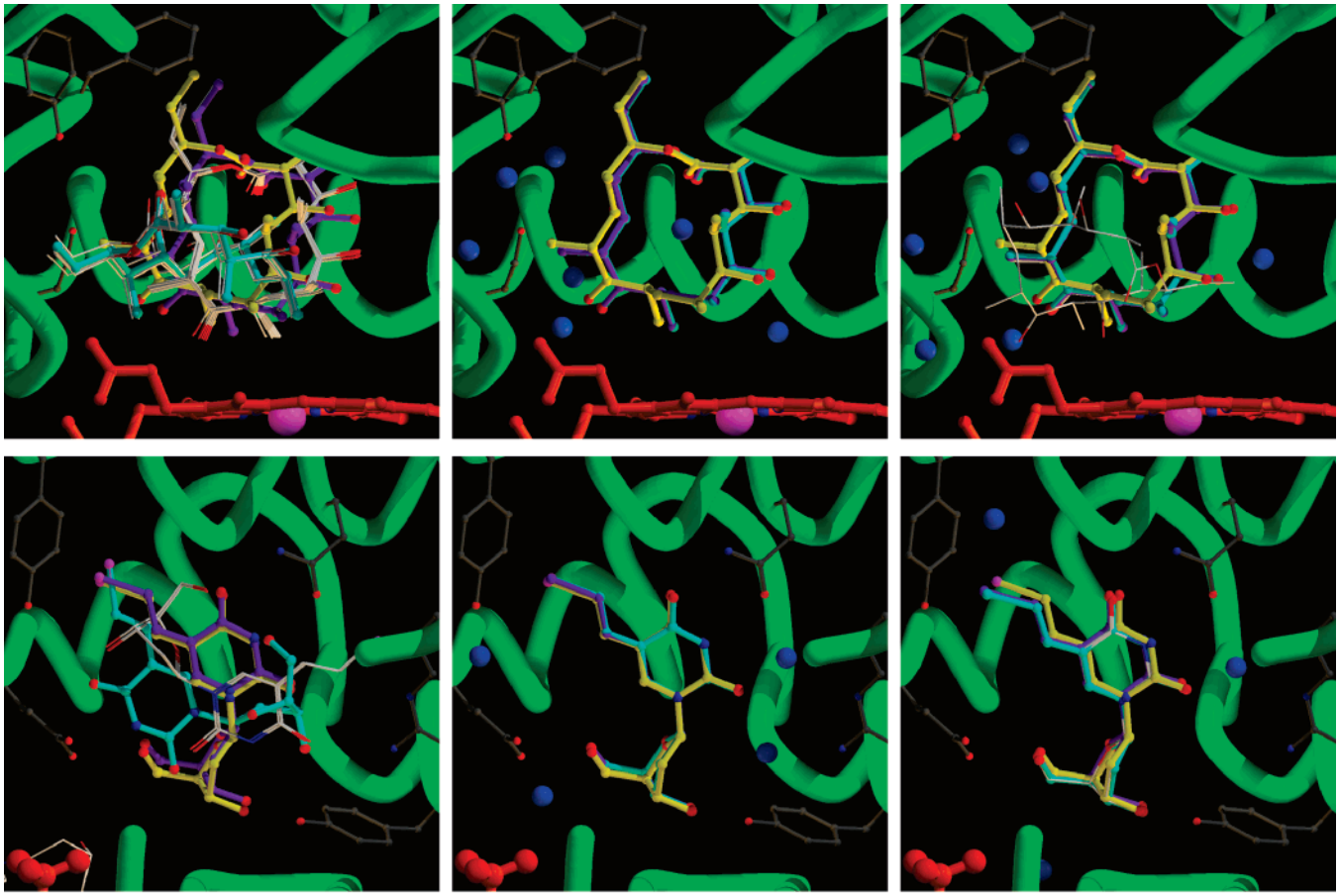

Figure 3. The effect of water and rescoring on docking accuracy. The role of water in predicting binding modes of 6-deoxyerythronolide B in the binding pocket of Cyt $P450_{EryF}$ (1jio) (top) and 5-bromovinyldeoxyuridine in HSV1 TK (1ki8) (bottom), using AutoDock and different water scenarios (no water (N, left panel), X-ray water (X, middle panel), predicted water (P, right panel)). Orientations of docking solutions (white carbon atoms) ranked as no. 1 with the scoring function of AutoDock (cyan carbon atoms) and SCORE (purple carbon atoms) are compared with those observed in the crystal structure (yellow). Water oxygen atoms are depicted in blue. The figures were prepared using Molscript[55] and Raster3D.[56]

and the *R*-stereoisomer of the acyclic purine ligand 9-hydroxypropyladenine (1e2iA, TK HSV1) is increased by incorporating X-ray and/or predicted water molecules in the protein active site (Table 1). The RMSD accuracy of FlexX (1e2iA, 1ki8) and GOLD (1e2j, 1p6x, 1e2n, 2vtk) also increased for at least two of these ligands by the consideration of explicit X-ray and/or predicted active site water molecules.

RMSD vs CSP Accuracy. Cyt P450. Generally, the different docking approaches were much more accurate in catalytic site prediction (CSP, correct indication of site(s) of oxidation and occurrence of covalent binding (Figure 2C)) than in reproducing actual X-ray binding conformations of Cyt P450-ligand complexes (Figure 2A). For example, in water scenario *P*, AutoDock was shown to be the best docking algorithm with respect to CSP accuracy (AD (up to 84%) > G ≫ F), while in only up to 48% of the Cyt P450 test cases, AutoDock was able to rank as top solution a docking pose within 2 Å RMSD of the X-ray structure. GOLD was found to be superior to the AutoDock and FlexX when the RMSD was the criterion of docking accuracy (G ≫ F > AD).

RMSD vs CSP Accuracy. TK. Different docking approaches were equally successful in predicting sites of phosphorylation (CSP) in TK substrates (Figure 2D) as in reproducing X-ray binding conformations (Figure 2B). AutoDock was shown to be the best in predicting TK-ligand binding modes, both with respect to RMSD (AD > G ≫ F) as well as to CSP accuracy (AD > G ≫ F).

Rescoring. Cyt P450. Reranking of the poses generated by each of the three different docking algorithms with the scoring function SCORE improved the docking performance of almost all docking-water scenario combinations (docking strategies). This is reflected by an increase in docking accuracy both in terms of RMSD (relative increase of up to 58%) and CSP accuracy (relative increase of up to 45%) when compared to reranking with the program implemented scoring function only. Rescoring of all pooled poses generated by AutoDock, GOLD, and FlexX with SCORE (S) yielded higher docking accuracies than those obtained with single docking−scoring combinations (Figure 2A): 53% vs 16% (AD−AD) (AutoDock docking algorithm in combination with the AutoDock scoring function) to 53% (G−S) for the single docking−scoring combinations without water; 74% vs 26% (AD-AD) to 64% (G−G) with X-ray water molecules; and 74% vs 41 (AD−AD) to 58% (G−S) with predicted waters. Rescoring 'pooled' solutions of the different docking algorithms yielded CSP accuracies comparable to those obtained with the AutoDock−SCORE combination. Concerning the RMSD accuracies obtained with AutoDock and FlexX, and the

CSP accuracy obtained with FlexX, there is a big difference between the chance of finding a reliable solution ($\bar{A}$) and the propensity of the docking algorithm to find a reliable solution, whatever its ranking (C). In contrast, this difference was smaller for the CSP accuracy of AutoDock and both the CSP and RMSD accuracy of GOLD.

Rescoring. TK. Reranking generated by each of the three different docking algorithms with the scoring function SCORE was not observed to improve docking performance neither in terms of RMSD nor CSP accuracy of the various docking-water scenario combinations. However, rescoring of all pooled poses generated by AutoDock, GOLD, and FlexX with SCORE yielded docking accuracies comparable to those obtained with the most accurate single docking−scoring combination (Figure 2B): 74% vs 47% (F−S) (FlexX docking algorithm in combination with the SCORE scoring function) to 74% (AD−S) for the single docking−scoring combinations for docking without water; 90% vs 63% (F−F) to 90% (AD−S) with X-ray water molecules; and 90% predicted water molecules (vs 58% (F−S) to 90% (AD−S)). Rescoring 'pooled' solutions of all three docking algorithms yielded CSP accuracies comparable to those obtained with the AutoDock-SCORE approach. Concerning the RMSD accuracies obtained with AutoDock and FlexX and the CSP accuracy obtained with FlexX, a big difference between $\bar{A}$ and C was observed. In the case of the RMSD accuracy obtained with GOLD, and the CSP accuracy obtained with AutoDock and GOLD, this difference was much smaller.

Discussion

The primary aim of the present study was to develop optimal docking strategies for binding mode and catalytic site prediction of ligands of two pharmacologically relevant biotransformation enzyme systems, namely Cyt P450 and TK, by evaluating the performance of various docking−scoring combinations, and considering the effects of active site water molecules.

The Effect of Active Site Water Molecules on Docking Accuracies. Despite the fact that water molecules can play an essential role in ligand−protein binding,[6−8] only few studies show that consideration of concrete active site water molecules improves the accuracy of automated docking. The novel GRID-based water prediction protocol used in this work was shown to strongly improve the docking accuracy of the three docking programs AutoDock, FlexX, and GOLD. The increase in docking accuracy due to the incorporation of X-ray water molecules was comparable to the effect of predicted water molecules. Results presented in Figure 1A and B show that many water locations and water-mediated protein−ligand interactions are correctly predicted. In the case of Cyt P450, false positive water molecules are predicted as well. It must be stated, however, that crystal structure determinations are normally unable to detect disordered or mobile water molecules.[38] This leaves the possibility open that there are water molecules involved in protein−ligand binding which are not resolved in crystal structures. Indeed, it is shown that these additional predicted water molecules placed on energetically favorable locations in binding pockets (false positives) improve the docking accuracy of many Cyt P450 complexes by mediating protein−ligand interactions and 'fixating' ligand molecules close to the center of the protein active site (Table 2). The present study shows that for five Cyt P450 isoenzymes and three TK isoenzymes prediction of water locations, based on the energy landscape of the active site surface, was an appropriate method leading to satisfactory results. Docking into binding pockets with predicted water molecules generally yielded higher docking accuracies than binding mode predictions using the X-ray water molecules. The positions of X-ray waters in the active site are highly ligand−protein complex specific, while the predicted water scenario is largely ligand-independent.

In principle, there are still some caveats associated with the use of the current predicted water docking strategy. If ligands of variable size are to be docked, one might use several differently shaped and sized solvated protein binding pockets for a target, containing different numbers and configurations of water molecules. Such a procedure will significantly increase the computational effort, however. Moreover, this strategy will still not take the replacement of water molecules by functional groups of ligands fully into account, as ligands can possess different functional groups and can be differently positioned in the active site. Docking poses in differently solvated binding pockets should be compared based on scoring functions; however, at this stage it is not fully clear yet how docking scores obtained from structures containing different numbers of bound waters should be compared. Future evaluation of the present predicted water approach on a wider range of ligands and protein targets will show whether such an extended predicted water scenario is generally applicable and useful for virtual screening approaches, including docking into apo-enzymes.

Protein−Ligand Complex Specific Docking Accuracy. Docking accuracies of AutoDock, FlexX, and GOLD were found to be protein−ligand complex specific. The performance of all docking−scoring combinations to predict top ranked docking poses within 2 Å from the X-ray structure, was much higher for TK−ligand complexes, than for Cyt P450−ligand complexes. This might be explained by the fact that the binding pockets of the respective Cyt P450 isoenzymes are larger than those of the TK isoenzymes. Concerning Cyt P450 docking, GOLD resulted in a significantly higher RMSD accuracy than the two other docking programs. AutoDock was superior with respect to RMSD accuracy and catalytic site prediction (CSP) accuracy in the case of TK, while AutoDock RMSD accuracy was rather low for Cyt P450-ligand binding modes. Previous docking studies comparing the RMSD docking accuracy of FlexX and GOLD showed GOLD to give superior results,[2,3] while two recent comparative studies also demonstrated that the relative performance of AutoDock, FlexX, and GOLD[39,40] varied with the selected protein target. As protein-dependent docking accuracies were also observed in the present study, it is suggested that, prior to docking studies determining unknown binding modes of protein−ligand complexes, one should find the best docking algorithm for a specific protein target by using a test set of protein−ligand X-ray structures. For most proteins, however, a sufficiently large set of experimentally determined structures is not yet available.

It is known from literature that the docking accuracy of docking−scoring combinations may not only vary with the protein target, but also with the physicochemistry of protein−ligand interactions.[2,3] The Cyt P450 ligands included in this work are small and rigid apolar terpene substrates and phenylimidazole inhibitors covalently bound to the heme Fe-atom of Cyt $P450_{cam}$, as well as larger and more polar macrocycles of Cyt $P450_{EryF}$ and Cyt $P450_{EpoK}$, flexible long-chained aliphatic palmitoleic acid substrate of Cyt $P450_{BS\beta}$, and moderately flexible aromatic substrates of Cyt P450 2C5. Selected TK complexes were nucleic acid bases of two general classes: pyrimidine and purine derivatives. Generally, all selected TK isoenzymes have a polar active site including an ADP cofactor or SO_4^{2-} ion. In comparison to the Cyt P450s, this active site of TKs is rather small. The AutoDock docking algorithm and scoring function were not capable of accurately reproducing the binding conformations of the relatively small and rigid, apolar camphor and phenylimidazole derivatives in the Cyt $P450_{cam}$ (Table 1). In contrast, AutoDock was very successful in reproducing the binding conformations of rigid, big polar macromolecules 6-deoxyerythronolide B and epothilone D, the flexible palmitoleic acid, the moderately flexible sulfaphenazole derivative, and diclofenac in other P450 isoenzymes, especially when taking active site water molecules into consideration. AutoDock was also able to predict correct binding modes of almost all TK ligands (both pyrimidine and purine derivatives). The GOLD algorithm and scoring function was not very suitable for some camphor and imidazole derivatives but was appropriate for all other Cyt P450 and TK ligands. For both enzyme systems, the docking accuracy of the FlexX algorithm and scoring function was more docking-case specific than ligand type specific.

Influence of Input Ligand Geometry and Convergence of Docking Simulations on Docking Accuracy. Several protein−ligand complexes used in this study were also incorporated in single and comparative evaluations of AutoDock, FlexX, or GOLD described by others for Cyt P450,[10,20,21,32,40,41] as well as for TK,[2,39] although without taking the presence of explicit water molecules into account. The RMSD accuracy results presented in these studies are not always consistent with RMSD accuracies obtained in the present study. Most likely this is due to the effect of input ligand geometries and orientations[16,32] and the number of independent docking runs used (convergence of the docking simulation)[42] on the performance of the docking programs. In the present study, the ligands were energy minimized and unarbitrarily translated to the midpoint of the protein active site, rather than the usual approach in which the X-ray conformation is used as starting structure. The latter method may yield biased results, as the ligand is already in an energetically favorable conformation and orientation (energy minimum). Regarding TK, small conformational changes in the TK HSV1 binding sites used in this work, compared to the (lower resolution) protein structure (pdb-code 1kim) used in other studies, might explain differences in docking accuracies as well.

Catalytic Site Prediction Accuracy. Prediction of metabolite formation is increasingly seen as essential for the discovery and development of new drugs and drug candidates.[31] To predict whether and how a putative ligand will be enzymatically transformed, we have introduced a novel docking accuracy criterion, called catalytic site prediction (CSP). No matter what docking approach was used, the qualitative predictions of catalytic sites in ligands were found to be very accurate compared to the more quantitative criterion of RMSD accuracy, even though a nonlinear relationship between both parameters is undisputed. Consequently, the CSP accuracy could be used as a powerful new analysis tool in virtual screening applications to select hits (i.e. substrates, products, and inhibitors) from chemical databases without manual and visual inspection. The need for computational approaches to predict likely sites of metabolism, or 'soft spots', is also expressed by the recent development of other molecular modeling methods for this purpose.[43]

Rescoring. The Cyt P450 and TK docking performance of each of the three docking programs showed a significant improvement by reranking the ligand poses with the scoring function SCORE. The positive effect of rescoring was most pronounced in the case of Cyt P450. Docking accuracy (RMSD and CSP) was found to be even more improved (in the case of Cyt P450), or comparable to the most accurate docking−scoring combination (in the case of TK) by rescoring all 'pooled' AutoDock, GOLD, and FlexX docking runs with SCORE, no matter which water scenario was taken into account. These remarkable findings show that scoring/rescoring is an essential aspect of automated docking and even predominates docking, a conclusion in agreement with previously published comparisons.[2,42,44] By using different docking algorithms, one can search conformational space more extensively, and, by inference, one is more likely to find 'correct' binding conformations among the large ensembles of docking solutions. The success of this proposed strategy of using multiple docking programs was recently also demonstrated by the evaluation of ConsDock[3], a program developed for the consensus analysis of all possible poses generated by DOCK,[45] FlexX, and GOLD. It is therefore advised to rescore the 'pooled' solutions of multiple docking algorithms with a robust separate scoring function and to include active site water molecules in docking simulations. Future studies still have to determine whether better results can be achieved by using other scoring functions or consensus scoring approaches.[44] The present study proved SCORE to be very suitable for this purpose.

Conclusions

The present study concerns a first comprehensive evaluation of the effects of explicit active site water molecules on molecular docking based binding mode prediction with three different docking programs, namely AutoDock, FlexX, and GOLD. Scenarios considering crystallographically determined and computationally predicted active site water molecules tested on two sets of therapeutically important protein−ligand systems (i.e. cytochromes P450 (Cyt P450s) and thymidine kinases (TKs)) turned out to perform better than docking approaches that omit water molecules in terms of RMSD accuracy. The scenario including predicted water molecules was comparable or better than the scenario

including X-ray water molecules. Future research should be focused on a more diverse description of multiple water configurations in the active site, combined with a robust scoring function properly taking solvent configurations into account. A newly defined endpoint of docking accuracy was the catalytic site prediction (CSP) accuracy, in the present systems corresponding to sites of oxidation, covalent binding, and/or phosphorylation. As the catalytic sites were accurately predicted by the (water incorporated) docking strategies, the CSP accuracy criterion can also be used for virtual screening applications. Rescoring of poses with the stand-alone scoring function SCORE significantly improved the docking accuracy of each of the three docking algorithms, especially when rescoring was performed on 'pooled conformations' produced by all three docking programs. Pooled scoring is advised to be used, especially when the computational chemist has no appropriate protein−ligand test set available to find an optimal docking strategy. The presented docking strategies, considering active site water molecules and (pooled) rescoring, mark an alternative development in ligand−protein binding mode and catalytic site prediction. They can be used not only to select potential protein-specific substrates and to predict probable sites of catalysis, but also to refine and evaluate homology models, and to generate energetically favorable starting structures for advanced molecular dynamics simulations.

Computational Methods

Preparation of Ligand and Target Molecules. Ligand input files have been generated with Sybyl 6.8 (TRIPOS Inc., St. Louis, MO). First, ligand structures were extracted from the Protein Databank (pdb) file (containing only non-hydrogen atoms). After proper assignment of Tripos atom types,[46] hydrogen atoms were added to the ligands (assuming a physiological pH of 7.4) and partial atomic charges were calculated using the Gasteiger−Marsili method. Subsequently, ligands were energy-minimized in vacuo using the Tripos force field and translated to the center of the active site as determined by PASS.[47] This procedure guarantees that the ligand input structure for docking has no "X-ray information" of the pdb structure. X-ray structures served as reference structures for the calculation of the RMSD and CSP accuracy (see below). Figure 1A and B show the molecular structures of all ligands of the Cyt P450 and TK enzyme systems. Nineteen structures of Cyt P450−ligand complexes were selected from 45 complexes deposited at the Protein Databank. The excluded complexes are either variations on Cyt P450−terpene or Cyt P450−imidazole and Cyt P450−alkyl cyanide complexes or Cyt P450 complexes containing large covalently bound inhibitors or two ligands simultaneously. Complexes in which the orientation of the ligand did not correspond to experimentally determined biotransformation products were also excluded from this study to be able to make a clear comparison between the RMSD and CSP accuracies of the docking strategies used. From 25 TK complexes deposited at the Protein databank, analogously, 19 structures were selected excluding complexes with lower resolutions and complexes with large polyphosphate ligands resolved to study the intermediate state.[48] Since many TKs are dimers, those monomers with more X-ray waters present in the active site were selected. In the case of 1e2i, with stereoisomers each binding in two different binding modes, one stereoisomer-specific mode from each monomer was selected. To have also an equine protein−ligand complex in the TK test set, VZV TK with ((E)-5-(2-bromovinyl)-2′-deoxyuridine-5′-monophosphate) (pdb-code 1osn) was included in the analysis, even though this is a structure with a monophosphorylated product. In both enzyme systems, heme moiety (Cyt P450), cofactor ADP (TK) or sulfate ions (TK) were considered as a part of the protein and Tripos atom types were defined accordingly. Following the standard program protocols,[10,13,20] protein structures were protonated using the Biopolymer module in Sybyl. For the AutoDock studies, Kollman united atom and Gasteiger-Marsili partial charges were assigned to the protein and ligand structures, respectively.

Consideration of Active Site Water Molecules. To test the incorporation of water molecules in the docking process, three scenarios were pursued based considering the presence or absence of water in the protein active site: without water molecules (N), with water molecules solved in the X-ray structure (X), and with water molecules predicted using a protocol based on the program GRID[22] and a ligand-based cut off (P).

To predict the energetically favorable positions of water molecules in the interior structure of the protein target (scenario P), a rectangular grid box of $21.75 \times 21.75 \times 21.75$ Å^3 with grid points separated by 0.333 Å, centered on the midpoint of the ligand binding pocket, was automatically hydrated with 25 water molecules (using an energy cutoff value of 5 kcal/mol) within the GRID21 version of GREATER (www.moldiscovery.com/docs/grid21). An AutoDock 3.0 tool, pdb-volume, was used to calculate the dimensions of a 'minimal' box for the different ligands. Predicted water molecules situated within at least 4.5 Å, or half the length of the largest ligand box-dimension from the active site center, were excluded from the docking studies. The positions of the hydrogen atoms of both crystallized (scenario X) and predicted water (scenario P) molecules were optimized using DOWSER.[49] Due to the fact that this program only carries out calculations on buried water molecules with an energy cut off of 12 kcal/mol, the DOWSER algorithm had to be adapted for this purpose (the modified code for DOWSER is available upon request). It was essential to modify the AutoGrid program (part of the AutoDock package) to enable docking in the presence of water. This program was changed so that it considered an oxygen atom bound to two hydrogen atoms as a potential hydrogen bond acceptor, as was done in an earlier AutoDock study[23] (the modified code for AutoGrid is available upon request).

Docking Studies. Automated docking studies were performed with three different docking algorithms, AutoDock 3.0 ('Lamarckian' genetic algorithm),[10] FlexX 1.10 (incremental construction algorithm,[13] as implemented in Sybyl 6.8), and GOLD 1.2 ('Darwinian' genetic algorithm).[20] As scoring is a very important second aspect of automated docking methodologies, it was decided to investigate the effect of rescoring: the process of reprioritization of docking solutions (primarily ranked by the 'native' scoring function implemented in the docking program) with an additional stand-alone scoring function. Preliminary docking studies in Cyt P450 and TK showed that among different "stand-alone" scoring functions (SCORE,[11] scoring functions included in the Sybyl CScore module (TRIPOS Inc.) and X-SCORE suite[12]), SCORE was either the best one (for Cyt P450s) or performed very well (for TKs) in selecting docking conformations with the lowest RMSD values from all docking poses generated by the three different docking algorithms (data not shown here). To limit the amount of data on docking accuracies to be presented, only the rescoring results of SCORE were therefore shown. SCORE parameters were developed for the heme residue in Cyt P450s[50] as well as AutoDock, FlexX, and SCORE parameters for ADP and SO_4^{2-} in TKs, consistent with the heme, $NADP^+$, and SO_4^{2-} Tripos atom types (used by GOLD), respectively (see Supporting Information). For all docking parameters, standard values were used as described for AutoDock,[10] FlexX,[51] and GOLD ("standard default settings"),[20] except the amount of independent docking runs performed for each docking simulation. Multiple docking runs can increase the performance of docking programs,[5] as was shown specifically in the case of AutoDock.[42] To meet aspects of calculation time and data size on one hand, and convergence criteria and statistical relevance on the other hand, 50 independent docking runs were performed for each docking case. The active site center determined

by PASS is taken as AutoDock affinity grid center, probe location for FlexX studies, and the starting position of the GOLD flood fill algorithm. AutoDock affinity grid calculations were carried out with the same grid box as used for the prediction of positions of active site water molecules. The FlexX binding pocket is defined as the amino acid residues within 7 Å from the 'translated' ligand structure (see above). In this way, active sites of approximately the same size were used in all three docking programs.

Analysis of Docking Results. Three different criteria were used for determining the docking accuracy of different docking approaches:

• Root-mean-square deviation (RMSD) of heavy atoms of ligand docking poses from the reference X-ray structure, referred to as RMSD accuracy. RMSD values were calculated comparing all non-hydrogen atoms of the ligands with respect to the experimental X-ray structure binding modes using the *g_rms* tool of the Gromacs package.[52] Test cases were considered to be successfully docked and very accurately docked when RMSD values were lower than 2.0 and 1.0 Å, respectively.[3,21,32,53,54]

• Percentage of docking solutions with binding modes corresponding to experimentally determined major biotransformation products, referred to as catalytic site prediction (CSP) accuracy. Ligand atoms were considered to be potential sites of oxidation (Cyt P450 substrates) or covalent binding (Cyt P450 phenylimidazole inhibitors) when they were within 5.5 Å from the Cyt P450 heme Fe-atom, and potential sites of phosphorylation (TK substrates) within 5.0 Å from HSV1 TK Glu-83, VZV TK Glu-48, or EHV-4 TK Glu-60 carboxylate oxygen atoms (see also Figure 1A and B).

RMSD and CSP accuracies are presented as:

• Average RMSD distributions and average percentages of successful catalytic site predictions for all solutions of each docking study (illustrating the chance of finding a reliable solution) ($\bar{A}$).

• RMSD distributions and CSP accuracy for no. 1 ranked solutions according to the program implemented scoring function (reflecting the ability of the program implemented scoring functions to properly rank poses after the docking procedure) (I).

• RMSD distributions and CSP accuracy for no. 1 ranked solutions according to the scoring function SCORE (reflecting the ability of SCORE to properly rank poses after the docking procedure) (S).

• RMSD distributions and CSP accuracy for poses closest to the experimentally determined structure (the propensity of the docking algorithms to find a reliable solution, whatever its ranking) (C).

In the case of ligands experimentally determined to bind in two orientations in the same protein binding pocket (e.g., Cyt P450: 1akd, 8cpp, 1p2y, and 1n6b, and TK: 1e2i), RMSD and CSP accuracies were determined by comparing docking poses closest to the X-ray binding mode. To compare the 'overall' difference in performance of the different docking−scoring combinations and water scenarios (referred to as docking strategies), the term relative averaged increase (RAI) is introduced:

$$\mathrm{RAI} = \frac{((W_\mathrm{A} - N_\mathrm{A})/W_\mathrm{A}) + ((W_\mathrm{I} - N_\mathrm{I})/W_\mathrm{I}) + ((W_\mathrm{S} - N_\mathrm{S})/W_\mathrm{S})}{3}$$

where $\bar{A}$, I, and S indicate the RMSD or CSP accuracies defined above for the scenario without active site water N, and a water scenario *W* (either with X-ray waters (X) or predicted waters (P)).

Supporting Information Available: SCORE "RESIDUE" parameter files. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A review of protein-small molecule docking methods. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 151−166.

(2) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759−4767.

(3) Paul, N.; Rognan, D. ConsDock: A new program for the consensus analysis of protein- ligand interactions. *Proteins-Struct. Funct. Genet.* **2002**, *47*, 521−533.

(4) Carlson, H. A.; McCammon, J. A. Accommodating protein flexibility in computational drug design. *Mol Pharmacol* **2000**, *57*, 213−218.

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

(6) Poornima, C. S.; Dean, P. M. Hydration in drug design. 1. Multiple hydrogen-bonding features of water molecules in mediating protein−ligand interactions. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 500−512.

(7) 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.

(8) Poornima, C. S.; Dean, P. M. Hydration in drug design. 3. Conserved water molecules at the ligand-binding sites of homologous proteins. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 521−531.

(9) Bohm, H. J. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 593−606.

(10) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E. et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.

(11) Wang, R. X.; Liu, L.; Lai, L. H.; Tang, Y. Q. SCORE: A new empirical method for estimating the binding affinity of a protein−ligand complex. *J. Mol. Model.* **1998**, *4*, 379−394.

(12) Wang, R. X.; Lai, L. H.; Wang, S. M. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11−26.

(13) 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.

(14) Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: Efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **2001**, *308*, 377−395.

(15) Schnecke, V.; Kuhn, L. A. Virtual screening with solvation and ligand-induced complementarity. *Perspect. Drug Discovery Des.* **2000**, *20*, 171−190.

(16) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J. et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(17) Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* **2002**, *46*, 34−40.

(18) Rao, M. S.; Olson, A. J. Modelling of factor Xa-inhibitor complexes: a computational flexible docking approach. *Proteins* **1999**, *34*, 173−183.

(19) 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*, 855−869.

(20) 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.

(21) Nissink, J. W. M.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C. et al. A new test set for validating predictions of protein−ligand interaction. *Proteins Struct. Funct. Genet.* **2002**, *49*, 457−471.

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

(23) Minke, W. E.; Diller, D. J.; Hol, W. G. J.; Verlinde, C. The role of waters in docking strategies with incremental flexibility for carbohydrate derivatives: Heat-labile enterotoxin, a multivalent test case. *J. Med. Chem.* **1999**, *42*, 1778−1788.

(24) Pang, Y. P.; Perola, E.; Xu, K.; Prendergast, F. G. EUDOC: A computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases. *J. Comput. Chem.* **2001**, *22*, 1750−1771.

(25) Clark, D. E.; Grootenhuis, P. D. J. Progress in computational methods for the prediction of ADMET properties. *Curr. Opin. Drug Discovery Dev.* **2002**, *5*, 382−390.

(26) Hou, T. J.; Xu, X. J. Recent development and application of virtual screening in drug discovery: An overview. *Curr. Pharm. Des.* **2004**, *10*, 1011−1033.

(27) Guengerich, F. P. Reactions and significance of cytochrome P-450 enzymes. *J. Biol. Chem.* **1991**, *266*, 10019−10022.

(28) Goeptar, A. R.; Scheerens, H.; Vermeulen, N. P. Oxygen and xenobiotic reductase activities of cytochrome P450. *Crit. Rev. Toxicol.* **1995**, *25*, 25−65.
(29) Vermeulen, N. P. E. Role of metabolism in chemical toxicity. *Cytcochromes P450: Metabolic and Toxicological Aspects*; CRC Press: Boca Raton, FL, 1996; pp 29−53.
(30) Venhorst, J.; ter Laak, A. M.; Commandeur, J. N.; Funae, Y.; Hiroi, T. et al. Homology modeling of rat and human cytochrome P450 2D (CYP2D) isoforms and computational rationalization of experimental ligand-binding specificities. *J. Med. Chem.* **2003**, *46*, 74−86.
(31) Vermeulen, N. P. Prediction of drug metabolism: the case of cytochrome P450 2D6. *Curr. Top. Med. Chem.* **2003**, *3*, 1227−1239.
(32) Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein−ligand docking. *Proteins* **1999**, *37*, 228−241.
(33) Keseru, G. M. A virtual high throughput screen for high affinity cytochrome P450cam substrates. Implications for in silico prediction of drug metabolism. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 649−657.
(34) Lozano, J. J.; Lopez-de-Brinas, E.; Centeno, N. B.; Guigo, R.; Sanz, F. Three-dimensional modelling of human cytochrome P450 1A2 and its interaction with caffeine and MeIQ. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 395−408.
(35) Blaese, R. M. Gene therapy for cancer. *Sci. Am.* **1997**, *276*, 111−115.
(36) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L. et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750−1759.
(37) Pospisil, P.; Kuoni, T.; Scapozza, L.; Folkers, G. Methodology and problems of protein−ligand docking: case study of dihydroorotate dehydrogenase, thymidine kinase, and phosphodiesterase 4. *J. Recept. Signal Transduct. Res.* **2002**, *22*, 141−154.
(38) Ernst, J. A.; Clubb, R. T.; Zhou, H. X.; Gronenborn, A. M.; Clore, G. M. Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. *Science* **1995**, *267*, 1813−1817.
(39) Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of docking performance: comparative data on docking algorithms. *J. Med. Chem.* **2004**, *47*, 558−565.
(40) Bursulaya, B. D.; Totrov, M.; Abagyan, R.; Brooks, C. L. Comparative study of several algorithms for flexible ligand docking. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 755−763.
(41) Vigers, G. P. A.; Rizzi, J. P. Multiple active site corrections for docking and virtual screening. *J. Med. Chem.* **2004**, *47*, 80−89.
(42) Wang, R. X.; Lu, Y. P.; Wang, S. M. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287−2303.
(43) Zamora, I.; Afzelius, L.; Cruciani, G. Predicting drug metabolism: a site of metabolism prediction tool applied to the cytochrome P450 2C9. *J. Med. Chem.* **2003**, *46*, 2313−2324.
(44) Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. Consensus scoring for ligand/protein interactions. *J. Mol. Graph. Model.* **2002**, *20*, 281−295.
(45) Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 411−428.
(46) Clark, M.; Cramer, R. D.; Vanopdenbosch, N. Validation of the General-Purpose Tripos 5.2 Force-Field. *J. Comput. Chem.* **1989**, *10*, 982−1012.
(47) Brady, G. P.; Stouten, P. F. W. Fast prediction and visualization of protein binding pockets with PASS. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 383−401.
(48) Gardberg, A.; Shuvalova, L.; Monnerjahn, C.; Konrad, M.; Lavie, A. Structural basis for the dual thymidine and thymidylate kinase activity of herpes thymidine kinases. *Structure (Camb)* **2003**, *11*, 1265−1277.
(49) Zhang, L.; Hermans, J. Hydrophilicity of cavities in proteins. *Proteins: Struct., Funct., Genet.* **1996**, *24*, 433−438.
(50) Wang, R. X. Personal communication, 2003.
(51) Rarey, M.; Kramer, B.; Lengauer, T. Docking of hydrophobic ligands with interaction-based matching algorithms. *Bioinformatics* **1999**, *15*, 243−250.
(52) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306−317.
(53) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein−ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337−356.
(54) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in molecular recognition: The effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.* **2004**, *47*, 45−55.
(55) Kraulis, P. J. Molscript a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **1991**, *24*, 946−950.
(56) Merritt, E. A.; Murphy, M. E. P. Raster3d Version-2.0 − a Program for Photorealistic Molecular Graphics. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **1994**, *50*, 869−873.

JM049650U