Catalytic Site Prediction and Virtual Screening of Cytochrome P450 2D6 Substrates by Consideration of Water and Rescoring in Automated Docking

Chris de Graaf,[†] Chris Oostenbrink,[†] Peter H. J. Keizers,[†] Tushar van der Wijst,[†] Aldo Jongejan,[‡] and Nico P. E. Vermeulen*[,†]

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

Received August 29, 2005

Automated docking strategies successfully applied to binding mode predictions of ligands in Cyt P450 crystal structures in an earlier study (de Graaf et al. *J. Med. Chem.* **2005**, *7*, 2308−2318) were used for the catalytic site prediction (CSP) of 65 substrates in a CYP2D6 homology model. The consideration of water molecules at predicted positions in the active site and the rescoring of pooled docking poses from four different docking programs (AutoDock, FlexX, GOLD-Goldscore, and GOLD-Chemscore) with the SCORE scoring function enabled the successful prediction of experimentally reported sites of catalysis of more than 80% of the substrates. Three docking algorithms (FlexX, GOLD-Goldscore, and GOLD-Chemscore) were subsequently used in combination with six scoring functions (Chemscore, DOCK, FlexX, GOLD, PMF, and SCORE) to assess the ability of docking-based virtual screening methods to prioritize known CYP2D6 substrates seeded into a drug-like chemical database (in the absence and presence of active-site water molecules). Finally, the optimal docking strategy in terms of virtual screening accuracy, GOLD-Chemscore with the consideration of active-site water (60% of known substrates recovered in the top 5% of the ranked drug-like database), was verified experimentally; it was successfully used to identify high-affinity CYP2D6 ligands among a larger proprietary database.

Introduction

Automated molecular docking has become an important computational method for predicting protein−ligand interactions, guiding lead finding, and optimization in drug discovery.[1,2] It combines search algorithms to generate multiple conformations and orientations of ligands within the binding site of proteins with scoring functions to determine the tightness of protein−ligand interactions.[3] Several docking algorithms and scoring functions have been described in the past few years, and very recently, several comparative studies of available docking tools have been reported.[4−17] Accuracies of the docking (prediction of binding orientation), scoring (prediction of absolute binding free energy), and ranking (discrimination of active from random compounds) of the different combinations of docking algorithms and scoring functions still depend on the protein target and the physicochemistry of the protein−ligand interactions.[4,18] This suggests that a docking−scoring strategy should be specifically optimized for the system under study. Other unresolved issues in automated docking are the consideration of protein flexibility and the inclusion or omission of explicit water molecules in the ligand binding pocket.[19,20]

In the present study, automated docking strategies that were successfully applied to the binding mode prediction of the Cyt P450-ligand crystal structures in an earlier study[20] will be used for the binding mode prediction and structure-based virtual screening of substrates of one of the most relevant drug metabolizing cytochrome P450 (CYP) isoenzymes, human CYP2D6. Cytochrome P450 isoenzymes are hemoproteins which catalyze the oxidation and reduction of a wide variety of endogenous and xenobiotic compounds.[21,22] They generally detoxify potentially hazardous compounds. In a number of cases, nontoxic parent compounds are bioactivated into toxic metabolites, or procarcinogens are bioactivated into their ultimate carcinogens.[23] Although the expression levels of CYP2D6 are only 2% of all hepatic CYPs, it is the second most important drug metabolizing enzyme after CYP3A4, and it is involved in the metabolism of about 30% of the currently marketed drugs.[24,25] Large interindividual differences exist in CYP2D6 activity because of gene multiplicity and polymorphisms, thus further increasing its clinical importance.[26,27] Early identification of potential CYP2D6 substrates and the prediction of their metabolism is therefore advantageous in the discovery and development of new drugs. CYP2D6 is one of the CYP isoforms studied most extensively using molecular modeling.[28] Several homology model structures of CYP2D6 have been built (e.g.[29−31]), refined, and validated experimentally, by site-directed mutagenesis studies and NMR spin lattice relaxation rate measurements, and are consistent with pharmacophore models (notably 3D-QSAR) of inhibitors and substrates.[32,33]

Automated docking approaches have been successfully applied to the prediction of the site of catalysis in substrates, the refinement and validation of CYP homology models, and the construction of pharmacophore models.[28] Very recently, improved docking strategies for the binding mode prediction of crystallized CYP-ligand complexes were described, which considered the active-site water and rescoring of pooled conformations from different docking programs.[20] In another recent study, an optimized scoring function describing heme−ligand interactions was used.[34] Although automated docking has been frequently used for the binding mode prediction of CYP-ligand complexes, structure-based virtual screening studies of chemical databases against cytochrome P450 are scarce. Experimentally determined binding affinities of 11 different CYP101(*cam*)-ligand complexes showed no clear correlation with values from different CScore scoring functions on CYP101(*cam*)-ligand crystal structures or on complexes produced by the automated docking program FlexX.[35] Nevertheless, docking

* To whom correspondence should be addressed. Phone: +31-20-5987590. Fax: +31-20-5987610. E-mail: npe.vermeulen@few.vu.nl.
[†] Division of Molecular Toxicology.
[‡] Division of Pharmaceutical Sciences.

10.1021/jm0508538 CCC: $33.50
Published on Web 03/17/2006

and scoring could still be used to prioritize virtual screening hits of a chemical database against a CYP101(*cam*) crystal structure and a CYP3A4 homology model.[35] Early virtual screening studies were also performed with the docking program DOCK to identify selective substrates[36] and imidazole inhibitors[37] for wildtype and L244A mutant CYP101 (*cam*) from 20 000 and 3508 compound databases, respectively. Very recently, no significant correlations were found between two training sets of 21 and 30 experimental K_i values and the GOLD-Chemscore docking scores on a CYP2D6 homology model.[38] In the same study, a weak correlation was found between experimental IC_{50} values and the docking scores of a small database of 33 compounds, but the docking scores could be successfully used to discriminate between weak ($IC_{50} > 10\ \mu M$) and strong inhibitors ($IC_{50} < 10\ \mu M$).[38]

Here, we present an extensive automated docking study for the catalytic site prediction (CSP) of human CYP2D6 substrates as well as the first automated docking-based virtual screening for high-affinity substrates of CYP2D6 in a large chemical database. The primary aim of the present study was to find optimal docking strategies for the binding mode prediction and virtual screening of CYP2D6 substrates by evaluating the performance of various docking−scoring combinations and considering the presence and absence of active-site water molecules. The automated docking approaches recently applied to the binding mode prediction of Cyt P450-ligand crystal structures[20] were used for the catalytic site prediction (CSP) of 65 substrates of CYP2D6 (Table 1) and tested with respect to their ability to prioritize 20 known CYP2D6 substrates seeded into a chemical database of 980 drug-like compounds.[4] Docking was performed with four different docking algorithms: AutoDock[39] (Lamarckian genetic algorithm, AD), FlexX[40] (incremental construction algorithm, F), GOLD-GOLD (GG), and GOLD-Chemscore (GC) (in which the GOLD[41] (Darwinian genetic algorithm) docking simulation is guided by the Goldscore[41] and Chemscore[42] scoring functions, respectively[14]). The active site of our carefully refined and validated CYP2D6 homology model[31,43,44] was considered to be either water-free (N) or containing water molecules whose positions were unarbitrarily predicted by a novel GRID-based[45] protocol (W).[20] The resulting docking poses were scored and ranked using the scoring function implemented in the docking program and a number of stand-alone scoring functions (those implemented in the CScore module (Chemscore,[42] D-Score,[46] F-Score,[40] G-Score,[41] PMF[47]) and SCORE[48]). Finally, the optimal docking strategy was verified experimentally. It was used to identify high-affinity CYP2D6 ligands in a larger proprietary database (19 619 entries) and to discriminate between high-affinity and medium-affinity ligands.

Results

Catalytic Site Prediction. The percentage of docking solutions with binding modes corresponding to experimentally determined major biotransformation products, referred to as catalytic site prediction (CSP) accuracy (Materials and Methods), was used as the criterion for determining the docking accuracy of different docking approaches. Catalytic sites of 65 known CYP2D6 substrates were predicted using the docking algorithms AutoDock (AD), FlexX (F), and GOLD (in which the docking simulation is guided by either the Goldscore (GG) or the Chemscore (GC) scoring function). Docking algorithms were used in combination with their native scoring functions and the SCORE (S) scoring function, with (W) and without (N) the consideration of water molecules at predicted positions in the active site of our refined and validated CYP2D6 homology model.[31,43,44] In addition, docking poses generated by all four docking algorithms were pooled and rescored with SCORE. CSP results of each ligand−protein complex by the 18 docking approaches (i.e., 4 docking algorithms × 2 water scenarios × 2 scoring functions + pooled conformations with and without active-site water) are listed in Table 1. Also indicated are the relative average increase percentages (RAI, see Materials and Methods) of CSP accuracy due to the consideration of water and rescoring with SCORE for these approaches. The statistical CSP accuracy is graphically summarized in Figure 1. The presence of predicted water molecules in the docking studies was shown to strongly improve the CSP accuracy of all docking−scoring combinations (Table 1 and Figure 1). CSP accuracies using FlexX ($RAI_{water} = 43\%$) and AutoDock (30%) were more improved by the consideration of water than were the docking performances of GOLD-Goldscore (20%) and GOLD-Chemscore (14%). The CSP accuracy of the optimal docking−scoring combination for CYP2D6 without water (AutoDock-SCORE) was increased by 33% by including predicted active-site water molecules. An illustrative example of the effect of active-site water molecules on CSP is shown in Figure 2.

The CSP accuracies of the AutoDock, FlexX, GOLD-Goldscore, and GOLD-Chemscore programs were more docking-case specific than ligand-type specific. No strong correlations could be found between CSP accuracy and the molecular weight or the number of rotatable bonds. Docking results were in many cases sensitive toward the small differences in the chemical structure of the substrate. This is most clearly exemplified by the fact that for all docking programs stereoisomer-dependent (amiflamine, metoprolol, fluoxetine, tolterodine, venlafaxine, carteolol, propranolol, bufuralol, MDMA, MDEA, MDPA, promethazine, citalopram, and azelastine), regioisomer-dependent (tyramine), and MDMA- and MAMC-derivative-dependent CSP accuracies were observed (Table 1). Re-ranking 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 CSP accuracy (RI_{score} of up to 27%, see Figure 1 and Table 1) compared to that when ranking with the scoring function implemented in the respective docking program. Rescoring of all pooled poses generated by AutoDock (AD), FlexX (F), GOLD-Goldscore (GG), and GOLD-Chemscore (GC) using the SCORE scoring function (S) yielded equal or higher CSP accuracies than those obtained with single docking−scoring combinations (Figure 1): 68% (pooled conformations rescored with SCORE) versus 20% (FlexX docking algorithm in combination with the FlexX scoring function; F-F) to 69% (GC-S) for the single docking−scoring combinations without water; 80% (pooled-SCORE) versus 33% (F-F) to 82% (AD-S and GC-S) with predicted water molecules.

Evaluation of Virtual Screening Strategies. The virtual screening accuracies of three different automated docking strategies was subsequently evaluated in terms of hit rate and yield (Materials and Methods) of 20 known CYP2D6 substrates (Table 1) present in the top-ranked docking solutions of a chemical database of 1000 drug-like compounds. Three docking algorithms (FlexX, GOLD-Goldscore, and GOLD-Chemscore) were used in combination with six different scoring functions (Chemscore, DOCK, FlexX, Goldscore, PMF, and SCORE) with and without the consideration of active-site water molecules, yielding in total $3 \times 6 \times 2 = 36$ different docking strategies.

Table 1. Catalytic Site Prediction (CSP) of 65 CYP2D6 Substrates.[a]

Compound	AD	F	GG	GC	pooled
m-tyramine	**N W**		N^I **W**	**N W**	$\bullet^S$ W^S
2-methoxyphenamine (*R*)	$\bullet$ **W**	**W**	**N W**	**N W**	N^S W^S
amiflamine (*R*)	**N W**	N^S **W**	**N W**	**N W**	N^S W^S
mexelitine	**N W**	**N W**	**N W**	**N W**	N^S W^S
metoprolol (*R*)	**N W**	**W**	**N W**	**N W**	$\bullet^S$ W^S
phenphormine	**N W**	W^S	**W**	N^S **W**	N^S W^S
diphenhydramide*	$\bullet$ W^S	**N** W^I	$\bullet$ W^S	**N** $\bullet$	N^S W^S
fluoxetine (*R*)		$\bullet$		**W**	$\bullet^S$ $\bullet^S$
tolterodine (*R*)*			W^S **N**	W^S	N^S W^S
venlafaxine (*R*)	$\bullet$ **W**		$\bullet$ W^S	$\bullet$ W^S	$\bullet^S$ W^S
MAMC	N^I **W**	**W**	N^S W^S	**N W**	N^S W^S
EMAMC*	N^S **W**	$\bullet$		**N W**	N^S W^S

compound	AD	F	GG	GC	pooled
p-tyramine	**N W**	$\bullet$	**N W**	**N W**	N^S W^S
2-methoxyphenamine (*S*)	N^S W^S	$\bullet$	**N W**	**N W**	N^S W^S
amiflamine (*S*)	N^I **W**	N^S **W**	**N W**	**N** W^S	N^S W^S
bunitrolol	$\bullet$ $\bullet$	W^S	$\bullet$ $\bullet$	**N W**	N^S W^S
metoprolol (*S*)	$\bullet$ **W**	**W**	**N W**	**N W**	N^S W^S
debrisoquine*	N^I W^S	**W**	N^S **W**	**N W**	$\bullet^S$ W^S
MPTP	**N** W^S	$\bullet$ $\bullet$	N^I **W**	**N W**	N^S W^S
fluoxetine (*S*)	W^S	N^I		**W**	$\bullet^S$ $\bullet^S$
tolterodine (*S*)*	**W**		**W**	**N** W^S	N^S W^S
venlafaxine (*S*)		**N**	N^I W^I	N^I **W**	$\bullet^S$ W^S
timolol	$\bullet$ **W**		N^S W^S	**W**	$\bullet^S$ W^S
diMAMC	**N W**	$\bullet$	$\bullet$ W^S	N^I **W**	$\bullet^S$ W^S

Table 1 (Continued)

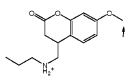
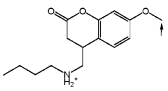
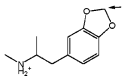
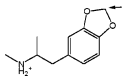
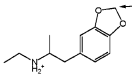
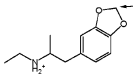
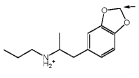
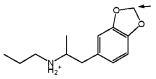
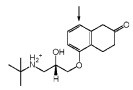
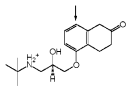
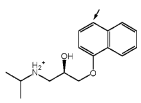
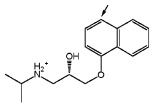
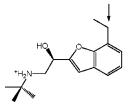
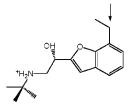
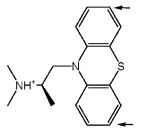
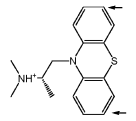
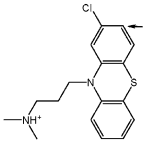
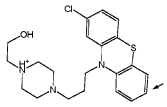
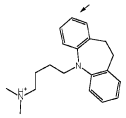
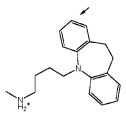
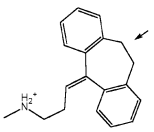
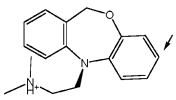
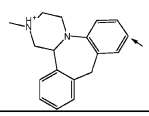
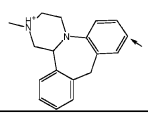
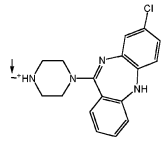
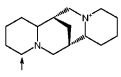
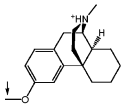
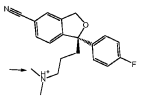
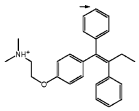
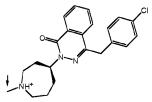
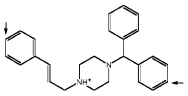
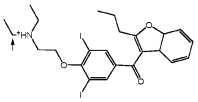
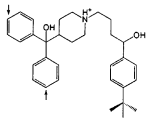
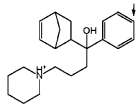
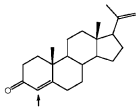
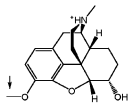
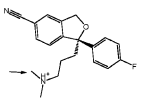
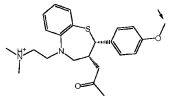
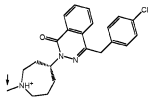
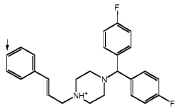
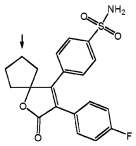
Compound	AD		F		GG		GC		pooled		compound	AD		F		GG		GC		pooled	
PMAMC 	**N**	**W**	**N**	**W**			N^S	**W**	N^S	W^S	BMAMC 	**N**	**W**			N^I	**W**	N^S	**W**	N^S	W^S
MDMA (*R*) 	**N**	**W**		**W**	**N**	**W**	**N**	**W**	N^S	W^S	MDMA (*S*) 	**N**	**W**		•	**N**	**W**	**N**	**W**	$•^S$	W^S
MDEA (*R*) 	**N**	**W**	N^S	**W**	**N**	**W**	**N**	**W**	N^S	W^S	MDEA (*S*) 	**N**	W^I	•	W^S	N^I	**W**	N^I	W^I	$•^S$	$•^S$
MDPA (*R*)* 	N^S	W^S	N^S	**W**	N^S	**W**	**N**	W^I	N^S	$•^S$	**MDPA (*S*)*** 	N^S	W^S	•	•	**N**	**W**	**N**	W^I	$•^S$	$•^S$
carteolol (*R*) 		W^S		**W**		**W**	**N**	**W**	$•^S$	W^S	carteolol (*S*) 	N^S	**W**		**W**	**N**	**W**	**N**	**W**	N^S	W^S
propranolol (*R*)* 	•	**W**		**W**	•	**W**	**N**	**W**	N^S	W^S	**propranolol (*S*)*** 	N^S	**W**		•	•	**W**	**N**	**W**	N^S	W^S
bufuralol (*R*)* 	N^I	•		W^S	•	**W**	**N**	**W**	$•^S$	W^S	**bufuralol (*S*)*** 	N^S	W^S		•	•	**W**	**N**	**W**	N^S	W^S
promethazine (*R*) 	N^S	**W**	**N**		**N**	**W**	**N**	**W**	N^S	W^S	promethazine (*S*) 	-	**W**			**N**	**W**	**N**	**W**	N^S	W^S
chloropromazine* 	N^S	**W**	**N**	**W**		•	**N**	**W**	N^S	W^S	**pherphenazine*** 					N^I	W^I		W^S	$•^S$	$•^S$
imipramine* 	**N**	**W**	**N**		•	**W**	•	**W**	N^S	W^S	desipramine 	N^S	**W**	•	**W**	N^S	**W**	**N**	**W**	N^S	W^S
nortryptiline* 	**N**	**W**		•	**N**	W^S		**W**	N^S	W^S	**E-doxepin*** 	•	W^S	N^S	**W**		**N**	**W**	**W**	N^S	W^S
mianserine (*R*)* 	**N**	**W**			**N**	**W**	**N**	**W**	N^S	W^S	**mianserine (*S*)*** 							**N**	**W**	N^S	W^S

Table 1 (Continued)

Compound	AD	F	GG	GC	pooled
clozapine 	• **W**		**N** **W**	N^{I} **W**	N^{S} •^{S}
sparteine 	N^{S} W^{S}	W^{I}	**N** **W**	N^{S} •	N^{S} W^{S}
dextromethorphan* 	**N** **W**	**N** **W**	**N** **W**	N^{S} W^{S}	N^{S} W^{S}
citalopram (*R*) 	• W^{S}			• •^{S}	•^{S}
tamoxifen 	• W^{S}		**N** W^{S}	N^{S} **W**	•^{S} W^{S}
azelastine (*R*) 	N^{S} **W**	•	N^{I} **W**	W^{S}	N^{S} W^{S}
cinnarizine* 	**N** **W**	**W**	N^{I} W^{S}		•^{S} •^{S}
amiodarone 	N^{I} **W**		**N** **W**	•	N^{S} W^{S}
terfenadine 	• •		• W^{I}	• •	•^{S} •^{S}

compound	AD	F	GG	GC	pooled
biperiden 	• •	W^{S}	•		•^{S} •^{S}
progesterone 	**N** **W**		**N** **W**	**W**	N^{S} W^{S}
codeine 	**N** W^{I}	**N** •	**N** **W**	**N** **W**	N^{S} W^{S}
citalopram (*S*) 	• W^{S}		N^{S} **W**	N^{I} W^{S}	•^{S} W^{S}
diltiazem 	• •	**N**	**N** **W**	**N** **W**	N^{S} W^{S}
azelastine (*S*) 	•				•^{S}
flunarizine 	• **W**	•	**N** •	N^{S} •	N^{S} •^{S}
spirosulfonamide 	• W^{I}		• •	**N** **W**	•^{S} W^{S}

	AD	F	GG	GC	pooled
RAI$_{water}$ (%)	**30**	**43**	**20**	**14**	**20**
RI$_{score}$ (%)	**27**	**22**	**6**	**4**	

[a] Letters AD (AutoDock), F (FlexX), GG (GOLD guided by Goldscore), and GC (GOLD guided by Chemscore) indicate the docking programs used, and pooled indicates that all pooled poses generated by the four docking programs are rescored with SCORE. Letters N (no water) and W (predicted active-site water) indicate the different scenarios concerning the presence or absence of water molecules. Cases in which catalytic sites of the major metabolite are only correctly predicted by the docking pose ranked as number 1 by the program-implemented scoring functions (superscript I); are only correctly predicted by the docking pose ranked as number 1 by the stand-alone SCORE scoring function (superscript S); are correctly predicted by both the docking pose ranked as number 1 by the program-implemented scoring function and the docking pose ranked as number 1 by the stand-alone SCORE scoring function (**bold**); are correctly predicted by any of the docking solutions, but not considered as number 1 ranked solutions by either the program implemented or stand-alone SCORE scoring functions (•). The names of 20 substrates added to the 980 drug-like compound database used for virtual screening studies studies are depicted in bold and indicated with an asterisk (*). For each docking program the relative increase by rescoring with SCORE (RI_{SCORE}) and the relative averaged increase by the consideration of water (RAI_{water}) are calculated as defined in the *Materials and Methods* section.

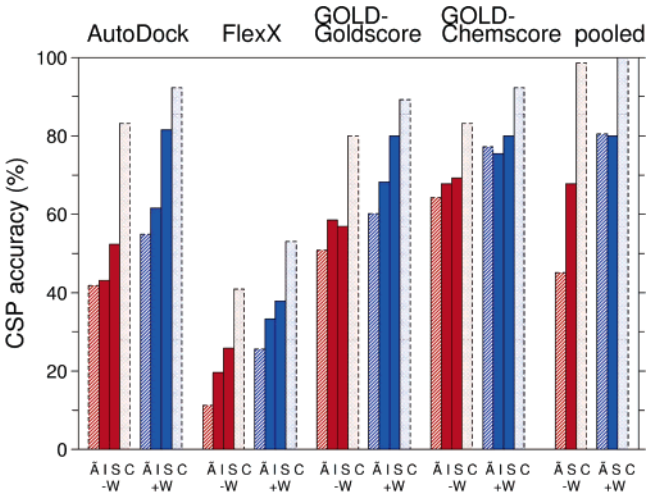


Figure 1. Cataltytic site prediction accuracies of different automated docking approaches for 65 CYP2D6 substrates, considering different scenarios with respect to the presence of water (N: no active-site water (red); W: water at predicted positions in the active site (blue)). Abbreviations on the *x*-axis correspond to: *Ā*, average CSP accuracy for all solutions of each docking study; *I*, CSP accuracy for solutions ranked as number 1 by the scoring function implemented in the docking program; *S*, CSP accuracy for solutions ranked as number 1 by SCORE; *C*, CSP accuracy for poses accurately predicting the experimentally determined catalytic site, whatever their ranking.

AutoDock was not applied in our virtual screening studies because this docking algorithm is not fast enough for this purpose. For clarity, only the results of the three docking programs in combination with the scoring function implemented in the program and the best performing stand-alone scoring function (at top 5%) are shown in Table 2 and Figure 3. The results for all docking−scoring combinations are available in the Supporting Information. Virtual screening for high-affinity CYP2D6 substrates was consistently improved by including active-site water, but the accuracy of virtual screening depended strongly on the docking−scoring combination. Hit rates and yields of FlexX docking were unsatisfactory low, and even by rescoring with DOCK and considering active-site water, the virtual screening accuracy of FlexX docking resulted in hit rates and yields of only 12 and 30%, respectively (considering the top 5% of ranked compounds). The virtual screening accuracy of GOLD-Goldscore docking was even lower, resulting in a maximum yield and hit rate of 4 and 10%, respectively, including active-site water and using either the Goldscore or the PMF scoring function (at top 5%). GOLD-Chemscore docking in combination with the Chemscore scoring function, however, clearly outperformed all other docking−scoring combinations with respect to virtual screening accuracy represented by a yield and hit rate of 24 and 60%, respectively (at top 5%), when active-site water molecules were included.

Virtual Screening for High-Affinity CYP2D6 Substrates. An initial proprietary database of 5760 structures was enriched by generating stereoisomers and tautomers for each compound, yielding 19 619 entries in total. The most successful virtual screening−docking strategy (GOLD-Chemscore, with consideration of active-site water molecules) was used to screen and rank this database against our CYP2D6 homology model[31,43,44] (Figure 4). The affinities (reflected by IC_{50}'s) of four typical CYP2D6 inhibitors and substrates (Table 3), eight top 0.1% (Table 4), and eight top 10% (Table 5) of the ranked compounds were determined experimentally. The compounds were selected from their respective bins in Figure 4. No knowledge of the CYP2D6 pharmacophore[32,33] was used to select these test compounds, but the selection procedure was rather guided by structural diversity. Stereoisomers (BS7840, BS7581, and BS7565) and tautomers (GBR30111) of the selected hits were found to have approximately the same docking scores. Among the eight top 0.1% of the ranked compounds, four compounds had an inhibitory capacity for CYP2D6 close to that of quinidine (IC_{50} < 0.3 μM), one compound comparable to that of dextromethorphan and quinine (1−10 μM), and two comparable to that of sparteine (10−100 μM), whereas one compound had negligible inhibitory capacity for CYP2D6 (>200 μM). Among the eight top 10% ranked compounds, no compounds had an inhibitory capacity for CYP2D6 close to that of quinidine, one compound had an inhibitory capacity comparable to that of dextromethorphan and quinine, three comparable to that of sparteine, and four had negligible inhibitory capacity for CYP2D6. Using an IC_{50} < 10 μM to define a CYP2D6 inhibitor (as was done in another study[38]), hit rates of the experimentally tested samples at the top 0.1% and top 10% of the ranked scorers were 63 and 13%, respectively.

Discussion

The primary aim of the present study was to find optimal docking strategies for catalytic site prediction and virtual

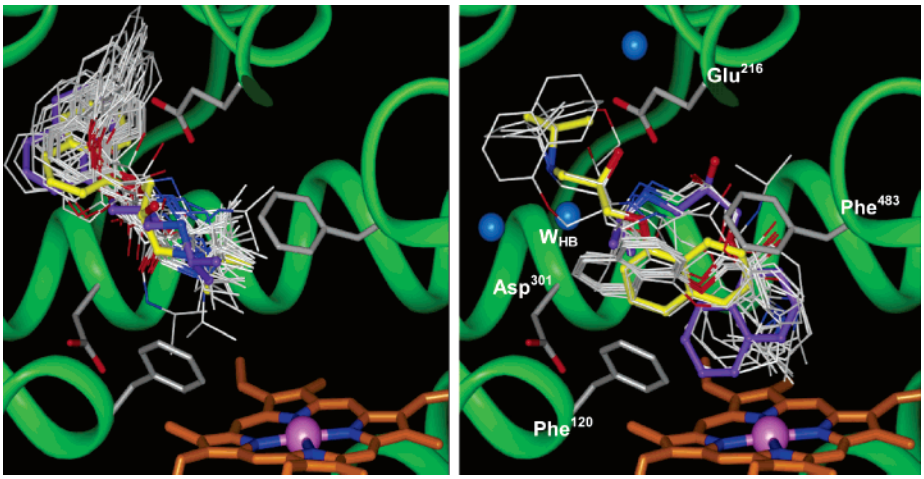


Figure 2. *R*-propranolol docked in the binding pocket of CYP2D6 using FlexX in the absence (left panel) and presence (right panel) of active-site water. Orientations of 50 docking solutions (white) top ranked by the scoring function of FlexX (yellow) and SCORE (purple) are compared. Amino acid residues involved in substrate binding (Phe^{120}, Glu^{216}, Phe^{483}, and Asp^{301}) are shown. Water oxygen atoms are depicted in blue ,and the active-site water molecule observed to mediate protein−ligand hydrogen bond is also indicated (W_{HB}).

Table 2. Validation of Virtual Screening Strategies for the Selection of High-Affinity CYP2D6 Substrates.[a]

			top 2.5%		top 5%		top 10%	
docking program	scoring function	water scenario	r_H (%)	y (%)	r_H (%)	y (%)	r_H (%)	y (%)
FlexX	FlexX	N	0	0	0	0	0	0
	DOCK[b]	N	0	0	4	10	14	70
	FlexX	W	0	0	2	5	5	25
	DOCK[b]	W	0	0	12	30	15	75
GOLD-Goldscore	Goldscore	N	0	0	0	0	0	0
	DOCK[b]	N	0	0	2	5	6	30
	Goldscore	W	0	0	4	10	3	15
	PMF[b]	W	8	10	4	10	6	30
GOLD-Chemscore	Chemscore	N	12	15	10	25	7	35
	Goldscore[b]	N	0	0	4	10	3	15
	Chemscore	W	36	45	24	60	15	75
	PMF[b]	W	12	15	8	20	7	35

[a] Description of CYP2D6 hit lists generated by three docking programs in combination with the program-implemented scoring function and the best performing stand-alone scoring function (at top 5%) with (W) and without (N) the consideration of active-site water molecules. A hit list is generated from the top-scoring compounds selected at a given treshold. Hit rate (r_H) and yield (y) are calculated as described in the Materials and Methods section. [b] As implemented in Cscore.

screening of CYP2D6 substrates by evaluating the performance of various docking−scoring combinations, and considering the presence and absence of active-site water molecules.

The Effect of Water on CSP Accuracy and Virtual Screening Accuracy. Despite the fact that water molecules can play an essential role in ligand−protein binding,[49−51] explicit water molecules are usually not taken into account in automated docking studies. Many scoring functions used for automated docking include an energy term accounting for the hydrophobic effect,[39,48,52,53] but only few docking programs allow for explicit water-mediated interactions between proteins and ligands during the docking simulation.[7,54−57] Furthermore, only a few of the docking studies reported evaluate the effects of water molecules at specific locations in ligand−protein binding sites.[20,58−66] More and more of these studies show significant effects of active-site water molecules on docking accuracy and virtual screening accuracy. Very recently, a novel method for dealing with active-site water molecules in automated docking was implemented into the GOLD docking program. This method allows water molecules to rotate around their three principal axes and correctly predicts water mediation and displacement.[57] The current study shows that water molecules placed on energetically favorable locations in the CYP2D6 binding pocket, improve the CSP and virtual screening accuracies of automated docking. The active-site water molecules were observed to both mediate protein−ligand interactions and fixate ligand molecules close to the center of the protein active site (Figure 2). In an earlier article, we discussed the possible caveats associated with the use of the GRID-algorithm to predict water positions in the current docking strategy: the need for differently solvated binding pockets of a protein target (containing different numbers and configurations of water molecules) to accommodate ligands of variable size and topology.[20] Linked to this are the needs for increased computational efforts and the use of scoring functions for comparisons of docking scores obtained from ligand−protein complexes containing different numbers of bound waters.

Substrate- and Docking-Strategy-Specific CSP Accuracy. It is known from the literature that the docking accuracy of docking−scoring combinations may not only vary with the protein target, but also with the physicochemistry of the protein−ligand interactions.[4,18] The set of 65 CYP2D6 substrates (32 of which were stereoisomers) included in this work has broad chemical and structural diversity (Table 1). The substrates contain up to four hydrogen bond acceptors, up to five hydrogen bond donors, and molecular weights ranging from 138 to 478 (80% with MW = 250−500). The number of rotatable bonds, reflecting conformational flexibility, ranges from 0 to 14. All compounds contain a positively charged nitrogen atom at physicological pH, a typical pharmacophoric feature of CYP2D6 ligands,[32,33] with the exception of progesterone and spirosulfonamide. The latter substrate was found to be one of the most difficult docking cases in terms of catalytic site prediction.

No clear correlations could be found between CSP accuracy and the molecular weight or the number of rotational bonds. Docking results were in many cases sensitive toward small differences in the chemical structure of the substrate as demonstrated by the fact that all four docking programs (AutoDock, FlexX, GOLD-Goldscore, and GOLD-Chemscore) showed stereoisomer-dependent, regioisomer-dependent, and MDMA- and MAMC-derivative-dependent CSP accuracies. This indicates that very subtle differences may play a crucial role in accurately predicting the site of metabolism. It should be noted that most experimental CYP2D6 metabolism studies have only reported results on racemic mixtures. Ligand descriptors alone will not likely be able to catch these effects, but the inclusion of protein target coordinates (and water molecules) appears crucial for the CSP.

The docking performance of each of the four docking programs showed a significant improvement by re-ranking the ligand poses with the scoring function SCORE. The positive effect of rescoring was most pronounced in the case of AutoDock (Table 1 and Figure 4). Rescoring all pooled AutoDock, FlexX, GOLD-Goldscore, and GOLD-Goldscore docking runs with SCORE yielded CSP accuracies comparable to the most accurate docking−scoring combination, with and without the consideration of active-site water molecules. These findings show that scoring/rescoring is an essential aspect of automated docking, and even predominates docking, a conclusion in agreement with previously published comparisons.[4,67,68] Comparing the overall CSP accuracy of the optimal docking strategies for each docking algorithm (including active-site water and rescoring with SCORE), AutoDock is only slightly better than GOLD-Chemscore and GOLD-Goldscore (CSP accuracies of 82, 80, and 80%, respectively) but superior to FlexX (CSP accuracy of 38%). Previous docking studies comparing the RMSD docking accuracy of FlexX and GOLD-Goldscore showed GOLD-Goldscore to give superior results,[4,9,18] whereas two recent comparative studies demonstrated that the relative performance of AutoDock, FlexX, and GOLD-Goldscore varied with the selected protein target.[5,10] The docking accuracies of

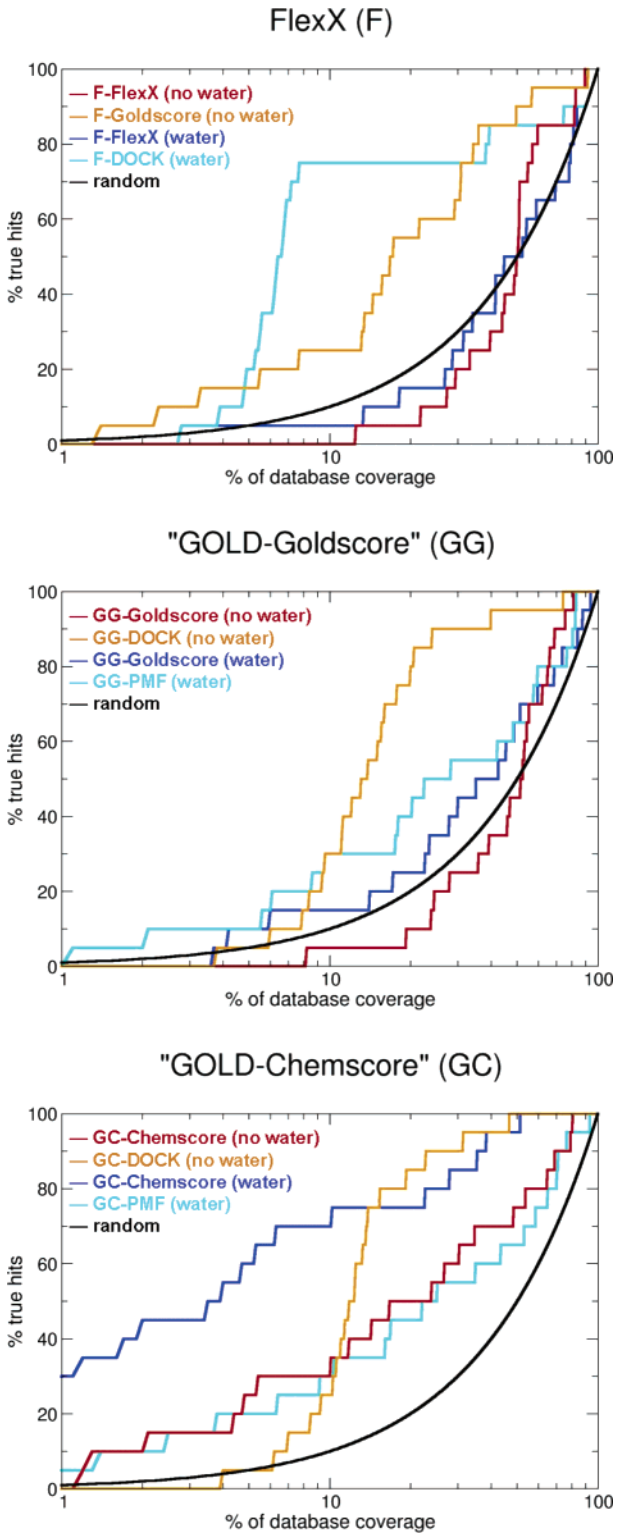


Figure 3. Enrichment in virtual screening of a database of 980 drug-like compounds[4] and 20 known CYP2D6 substrates (true hits) using FlexX, GOLD-Goldscore and GOLD-Chemscore with the program implemented scoring function and the best performing stand-alone scoring function (at top 5%) with and without the consideration of active-site water molecules. The solid black line represents the fraction of actives expected at random.

GOLD-Goldscore and GOLD-Chemscore were found to be comparable in an earlier study.[14]

Database- and Docking-Strategy-Specific Virtual Screening Accuracy. The 980 drug-like compound database used for the evaluation of virtual screening strategies spans the range of

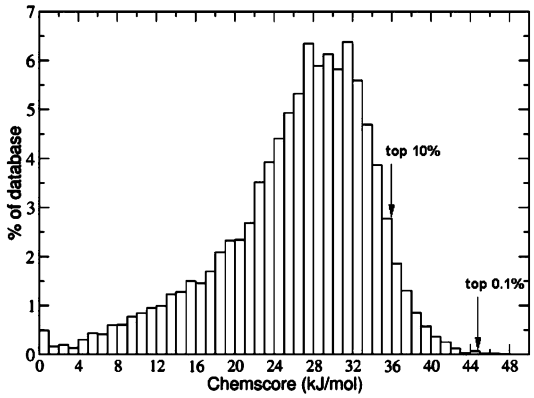


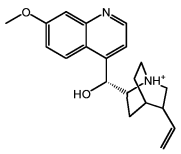
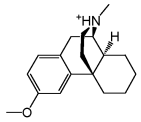
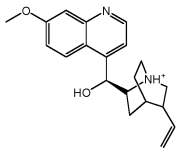
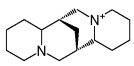
Figure 4. Distribution of Chemscore scores of the proprietary database of 19 619 compounds. The experimentally tested samples (of eight compounds) at the top 0.1% (∼45−48 kJ/mol) and top 10% (∼36 kJ/mol) of the database, shown in Tables 4−5, are indicated with arrows.

chemical properties of the 20 known CYP2D6 substrates added to this database, with the exception of debrisoquine, MDPA, and EMAMC. These compounds have slightly lower molecular weights (MW) than those of the compounds in the database (MW = 250−500). The molecular weights of the proprietary database of 19 619 entries range from 68 to 820 (80% with MW = 250−500). Recently, Verdonk et al. suggested that studies aimed at validating particular protein−ligand docking-based virtual screening methods should use libraries containing compounds with physicochemical properties similar to the actives.[15] For the training of different docking-based virtual screening methods, we used the Rognan database[4] because this database is used in previous comparative docking studies.[4,9] For the validation of our virtual screening strategy, we wanted a database that was pharmaceutically relevant and available for direct experimental verification. Nevertheless, we did not bias the database to have an identical distribution of properties because this is not likely the case in physical screenings carried out in the context of drug discovery. Furthermore, it can be assumed that focused databases have a higher probability of already containing actual high-affinity ligands than random drug-like databases do. This could then result in the underestimation of the virtual screening accuracies of docking strategies.

Virtual screening accuracies were found to be highly docking−scoring-combination dependent, as was found in pervious comparative docking studies.[4−6,8,9,12,15,69] The best docking-based virtual screening strategy, that is, GOLD-Chemscore with the Chemscore scoring function and including active-site water, was found to be superior to all other strategies. The 20 known CYP2D6 substrates seeded into the drug-like database were structurally too diverse to draw conclusions on the terms in the Chemscore scoring function that were responsible for this superiority.

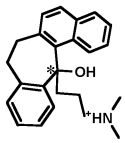
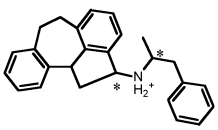
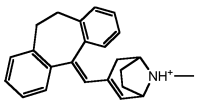
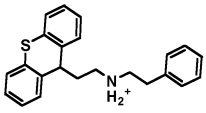
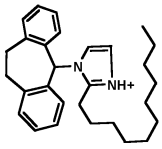
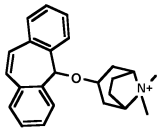
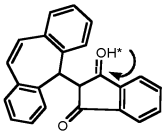
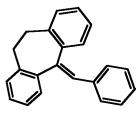
Virtual Screening Hits for CYP2D6. Seven of the eight top 0.1% hits found by virtual screening with GOLD-Chemscore (and considering water) of the proprietary database of 19 619 compounds against CYP2D6 contained a tricyclic dibenzo moiety and six contained a positively charged nitrogen, which are a part of the established CYP2D6 pharmacophore models of inhibitors and substrates.[32,33] Only one compound, BS7129, contained no hydrogen bond donors or hydrogen bond acceptors. This compound was the only false-positively screened hit (IC_{50} > 200 mM) among the top 0.1% of the ranked scorers. It should be noticed that compound BS7129 does not contain a positively charged nitrogen atom. Therefore, a combined CYP2D6 2D/

Table 3. Affinities of Known CYP2D6 Inhibitors, Measured as Their IC_{50} Values in MAMC O-Demethylation[a]

compound	IC_{50} (μM)	compound	IC_{50} (μM)
quinidine 	0.02 ± 0.01	dextromethorphan 	3.4 ± 0.8
quinine 	1.8 ± 0.2	sparteine 	60 ± 10

[a] All values are the means of at least three independent experiments ± standard deviations as described in the Materials and Methods section.

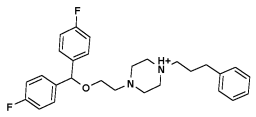
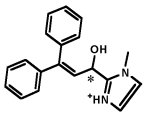
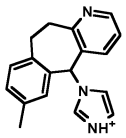
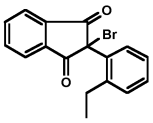
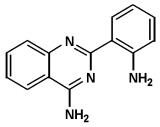
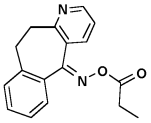
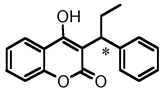
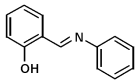
Table 4. CYP2D6 Affinities of Compounds Ranked at the Top 0.1% of Virtual Screening Studies (Figure 4), Measured as Their IC_{50} Values in MAMC O-Demethylation

compound[a]	IC_{50} (μM)[b]	compound[a]	IC_{50} (μM)[b]
BS 7840 	0.06 ± 0.003	BS 7581 	0.15 ± 0.03
GBR 12530 	0.21 ± 0.12	BS 9106 	0.26 ± 0.16
BS 9342 	5.8 ± 1.8	GBR 13036 	60 ± 6
GBR 30111[c] 	60 ± 10	BS 7129 	>200

[a] Chiral centers are indicated with an asterisk. [b] All values are the means of at least three independent experiments ± standard deviations as described in the Materials and Methods section. [c] Indication of tautomeric proton shift.

3D-pharmacophore-search[70] and structure-based virtual screening might exclude this false positive from the top ranked hits. However, compound GBR30111 does not have a positively charged nitrogen atom either, but it does contain a hydrogen bond donor that forms a hydrogen bond with Glu[216] (like *R*-propranolol in Figure 2) and does show affinity for CYP2D6 (IC_{50} value of 60 mM). This illustrates that docking-based virtual screening methods might enable the discovery of leads with chemical features and structures dissimilar to known actives, thus extending the possibilities of drug design. False-positive hits nevertheless may highlight the particular weaknesses of a docking scoring function.[69] All top 0.1% of scorers were predicted to be bound to the CYP2D6 binding pocket in about the same binding mode with their tricyclic dibenzo moiety stacked between the Phe[120] and Phe[483] and (with the exception of BS7129) forming an electrostatic interaction with the negatively charged carboxylic group of Glu[216]. These amino acid residues were previously found to be key determinants in the CYP2D6-mediated metabolism in site-directed mutagenesis studies.[43,44,71−73] BS7129 was found to have a tight steric fit in the CYP2D6 binding pocket, expressed by a high lipophilic score (E_{lip}), a relatively low protein−ligand clash (E_{clash}), and ligand-internal energy (E_{int}) scores in the GOLD-Chemscore docking score.[14]

Table 5. CYP2D6 Affinities of Compounds Ranked at the Top 10% of Virtual Screening Studies (see Figure 4), Measured as Their IC_{50} in MAMC O-demethylation

compound[a]	IC_{50} (μM)[b]	compound[a]	IC_{50} (μM)[b]
GB12819 	1.21 ± 0.07	GBR 12289 	15 ± 5
BS 9086 	40 ± 6	BS 8832 	83 ± 47
BS 9373 	169 ± 56	BS 8558 	>400
BS 7565 	528 ± 58	GBR 11840 	>1000

[a] Chiral centers are indicated with an asterisk. [b] All values are the means of at least three independent experiments ± standard deviations as described in the Materials and Methods section.

Two of the eight hits sampled from the top 10% of scorers, GB12819 (IC_{50} = 1.21 mM) and BS 9086 (IC_{50} = 40 mM), contained a positively charged nitrogen atom via which chemical group these compounds were found to interact with Glu[216] in the docked CYP2D6−ligand complex structure. The first compound is actually the only high-affinity (<10 mM[38]) CYP2D6 inhibitor found in this sample. The top 10% of the ranked scorers are structurally more diverse than the top 0.1% scorers and also more diverse in affinity. Two top 10% scorers, BS9086 and BS8558, contain a tricyclic dibenzo group, as also observed in seven of the eight top 0.1% structures, but only the first compound appeared to have medium-affinity (40 μM) for CYP2D6.

CSP Accuracy versus Virtual Screening Accuracy. Some comparative docking studies show that the ability of docking methods to predict binding modes of protein−ligand complexes is not correlated with their relative virtual screening performances.[4,69] Others, however, did find both properties to be correlated,[9,13] while yet others find this correlation to be target- and docking-method dependent.[6] These discrepancies imply that docking strategies should be separately optimized for the purpose of binding mode/catalytic site prediction and virtual screening. However, if a docking method is good in prioritizing known actives, whereas the generated binding modes do not resemble known binding modes, then it is hard to understand the basis of success and failure.[6] Such a docking method would be useful for lead identification but less valuable with respect to lead optimization. This problem might in principle be solved by the complementary use of pharmacophore restraints and structural knowledge about the way ligands typically bind to a given target to guide database docking experiments.[15,74−76] In the present study, AutoDock, GOLD-Goldscore, and GOLD-Chemscore in combination with the SCORE scoring function and with active-site water, are equally good docking strategies in terms of CSP accuracies (82, 80, and 80%, respectively). AutoDock was not applied to our virtual screening studies because the AutoDock algorithm is not fast enough for this purpose. The GOLD-Chemscore algorithm combined with its native Chemscore scoring function and considering active-site water molecules was superior to other approaches in terms of virtual screening accuracy. Moreover, these CSP and virtual screening strategies were able to select experimentally validated high-affinity inhibitors from another, larger database of 19 169 drug-like compounds.

Protein Target-Specific Training of Docking Strategies. Docking (binding mode or CSP) and scoring accuracies of docking−scoring combinations are often shown to vary considerably with the selected target protein, physicochemical details of target−ligand interactions,[4−6,13,15,16] and even depend on fine details of the protein structure.[77,78] In the current study, it was shown that both docking accuracy and virtual screening accuracy for CYP2D6 is highly docking-strategy dependent. Therefore, a docking−scoring strategy should be tailored to the system of interest and preferably be on the basis of a training set of ligand-bound protein crystal structures. In the absence of such a training set (which is the case for CYP2D6), other experimental data, such as regio-specificity of metabolism (binding mode prediction/CSP accuracy, as applied to CYP2D6 in the present study), binding affinity determinations (scoring/ranking accuracy, as applied to CYP2D6 in the present study), and site-directed mutagenesis studies (binding mode prediction/relevance of specific amino acid residues for binding, as applied to CYP2D6 in other studies[31,43,44]), can be used to validate docking strategies.

Conclusions

We presented an extensive docking study for the catalytic site prediction (CSP) of human CYP2D6 substrates and the first

automated docking-based virtual screening for the high-affinity ligands of this enzyme from a large chemical database. The presence of water molecules at predicted positions in the active site during docking studies was shown to strongly improve the CSP and virtual screening accuracies of various docking−scoring combinations. The CSP accuracy of the AutoDock, FlexX, GOLD-Goldscore, and GOLD-Chemscore programs were more docking-case specific than ligand-specific, and the virtual screening accuracy depended strongly on the combination of the docking program and the scoring function. Rescoring of the poses generated by each of the three different docking algorithms with the scoring function SCORE improved the CSP accuracy of almost all docking−water scenario combinations (docking strategies). GOLD-Chemscore docking in combination with the Chemscore scoring function (and with consideration of active-site water molecules) clearly outperformed the other docking−scoring combinations with respect to virtual screening accuracy and was also one of the most accurate strategies with respect to CSP. This docking strategy was validated experimentally. It was successfully used for the selection of high-affinity inhibitors of CYP2D6 from a large proprietary database. A selection of top-ranked (top 0.1%) compounds included significantly more high-affinity inhibitors than a selection of medium-ranked (top 10%) compounds. The current study shows that protein target specific training of automated docking strategies is essential for accurate predictions of protein−ligand binding modes and affinities.

Materials and Methods

Preparation of CYP2D6 Substrate and Protein Input Structures. A set of 65 known CYP2D6 substrates (including 32 stereoisomers) was selected for this study, on the basis of the availability of experimental information on CYP2D6-catalyzed product formation and the affinity for CYP2D6 (K_m < 200 μM). The protein homology model of CYP2D6 was constructed, refined, and validated as described.[30,31,43] Also, the preparation of protein and substrate input structures, the definition of the binding pocket, and the GRID-based[45] prediction of energetically favorable positions of active-site water molecules was performed as described.[20] The procedure for predicting the locations of active-site water molecules can be summarized as follows. A rectangular grid box of 21.75 × 21.75 × 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) using the GRID21 version of GREATER. An AutoDock 3.0 tool, pdb-volume, was used to calculate the dimensions of a minimal box for dextromethorphan, a typical CYP2D6 substrate with a relatively large size. Predicted water molecules situated within 5.5 Å (half the length of the largest ligand-box dimension) from the active-site center were excluded from docking studies. The positions of the hydrogen atoms of predicted water molecules were optimized using DOWSER.[79] This procedure yielded one single configuration of active-site water molecules that were used for all docking studies described in this article. Reorientation of water−hydrogen positions was not allowed during the docking experiments. The 3D structures of the CYP2D6 homology model and the 65 known substrates are available upon request.

Automated Docking Methodology for Catalytic Site Prediction. Automated docking studies were performed with four different docking algorithms, AutoDock 3.0,[39] FlexX 1 × 10[40] (as implemented in Sybyl 6.8), and GOLD 2.1[41] using the Goldscore and Chemscore fitness functions.[14] Because scoring is a very important second aspect of automated docking methodology, we decided to investigate the effect of re-scoring, that is, the process of re-prioritizing docking solutions (primarily ranked by the native scoring function implemented in the docking program) with an additional stand-alone scoring function. Earlier studies with Cyt P450 and TK crystal structures[20] showed that the docking performance of AutoDock, FlexX, and GOLD was significantly improved in terms of atomic root-mean-square displacement (RMSD) and catalytic-site-prediction (CSP) accuracy by the re-ranking of ligand poses with the scoring function SCORE.[48] In the current study, the docking accuracy of different docking approaches was evaluated with respect to their CSP accuracy.[20] The CSP accuracy is defined as the percentage of docking solutions with binding modes corresponding to experimentally determined major biotransformation products. The ligand atoms were considered to be potential sites of catalysis when they were within 6.0 Å from the CYP2D6 heme Fe-atom.

CSP accuracies are presented as follows. (i) The average percentages of successful catalytic site predictions for all of the solutions (50) of each docking study (illustrating the chance of finding a reliable solution) ($\bar{A}$); (ii) the CSP accuracy for solutions ranked as number one by the program-implemented scoring function (reflecting the ability of the program-implemented scoring functions to properly rank poses after the docking procedure) (I); (iii) the CSP accuracy for solutions ranked as number one by the scoring function SCORE (reflecting the ability of SCORE to properly rank poses after the docking procedure) (S); and (iv) the CSP accuracy for the poses closest to the experimentally determined structure (the propensity of the docking algorithms to find a reliable solution, whatever its ranking) (C).

To compare the overall difference in the performance of the different docking−scoring combinations and water scenarios (referred to as docking strategies), the terms relative increase by rescoring with SCORE (RI_{Sscore}) and relative averaged increase by the consideration of water (RAI_{water}) are defined

$$RI_{score} = \frac{(X_S - X_I)}{X_I}$$

$$RAI_{water} = \frac{(W_{\bar{A}} - N_{\bar{A}})/W_{\bar{A}} + (W_I - N_I)/W_I + (W_S - N_S)/W_S}{3}$$

where $\bar{A}$, I, and S indicate the CSP accuracies defined above for water scenario X (without active-site water (N) and with predicted active-site water molecules (W)).

Preparation of 3D Chemical Databases. For the evaluation of the virtual screening performances of the different docking strategies, a database was prepared by randomly replacing 20 compounds from a 3D database (in mol2 format) of 1000 drug-like molecules[4] with the 20 known CYP2D6 substrates highlighted in Table 1. A second, larger proprietary 3D database was prepared for the identification of new potential high-affinity CYP2D6 ligands and to evaluate the ability of optimal docking strategies to discriminate between high-affinity, medium-affinity, and low-affinity ligands. A 2D database of 5760 compounds was converted to 3D structures (in mol2 format) using the program MOE (version 2004, Chemical Computing Group, Montreal, Canada). Subsequently, structures were washed (filtering counterions and solvents), energy minimized, and MOE was used for the generation of all possible stereoisomers for the database (yielding 9706 entries). The database was tautomerised using AGENT2.0.[80] AGENT generated approximately two tautomers per compound in the database, yielding 19 619 entries in total). For some compounds, AGENT did not generate any tautomers, for others, it generated up to seven tautomers. On average, two stereoisomers were generated for each compound in the original database, and two tautomers were generated for each stereoisomer. After tautomerization, explicit hydrogen atoms were added to the database in Sybyl 6.8.

Automated Docking Methodology for Virtual Screening. FlexX 1.10 and the two GOLD 2.1 automated docking tools (using the Goldscore and Chemscore fitness functions[14]) were used for the virtual screening of the two databases (the combined database of 980 drug-like molecules and 20 known CYP2D6 substrates and the proprietary database of a total of 19 619 molecules, respectively). Standard FlexX settings and GOLD default 4 settings were used. All ligands for which a docking solution had been found were

rescored using the Cscore module of Sybyl6.62 (including the scoring functions Chemscore,[42] DOCK,[46] FlexX,[40] Goldscore,[41]and PMF[47]) and the scoring function SCORE (as part of an spl-script used by Bissantz et al.[4]). It should be noted that FlexX scores calculated either from FlexX or Cscore are very similar,[4] whereas Chemscore and GOLD scores calculated by the GOLD program differ from those calculated by Sybyl and thus cannot be compared.[81] Therefore, the Chemscore, FlexX, and GOLD scores proposed by Sybyl were discarded when the scoring function was coupled to the corresponding docking procedure.

Virtual screening accuracies are calculated in terms of the hit rate, r_H, and yield y

$$r_H = \frac{n_{TF}}{N} \cdot 100$$

$$y = \frac{n_{TF}}{n_T} \cdot 100$$

where N is the total number of compounds on the hit list, n_{TF} the number of true hits found in the hit list, and n_T the total number of true hits in the database.

IC_{50} Determinations. The pSP19T7LT plasmid containing bicistronically human CYP2D6 with a *C*-terminal His_6 tag and human NADPH-cytochrome P450 reductase, was kindly provided by Dr. Ingelman-Sundberg. *E. coli* JM109 was obtained from DSMZ (Braunschweig, Germany). The pSP19T7LT plasmid containing CYP2D6 was transformed into *Escherichia coli* strain JM109. Expression and membrane isolation was carried out as described.[43] The compounds tested (Tables 4 and 5) were taken from a proprietary database. 7-Methoxy-4-(aminomethyl)-coumarin (MAMC) and 7-hydroxy-4-(aminomethyl)-coumarin (HAMC) were synthesized as described.[82] Dextromethorphan hydrobromide, debrisoquine sulfate, and quinidine sulfate dihydrate were obtained from Sigma (St Louis, MD). All other chemicals were of analytical grade and obtained from standard suppliers.

MAMC *O*-demethylation reactions by CYP2D6 were carried out in 96-well plates, in a total volume of 200 μL.[43] The reaction mixture consisted of 5 mM $MgCl_2$ in KPi buffer, 50 μM MAMC, *E. coli* membranes corresponding to 40 nM CYP2D6, and different concentrations of an inhibitor. The inhibitors were dissolved in DMSO and stored as 20−100 mM stocks at −20 °C. The reactions were initiated by the addition of an NADPH regenerating system, resulting in final concentrations of 0.1 mM NADPH, 0.3 mM glucose-6-phosphate, and 0.4 units/mL glucose-6-phosphate dehydrogenase. The reactions were allowed to proceed for 30 min at 37 °C and the fluorescence of the samples was subsequently measured on a Victor^2 1420 multilabel counter (Wallac, Oy, Finland) using λ_{ex} = 405 nm, λ_{em} = 460 nm.

Acknowledgment. We thank Dr. Magnus Ingelman-Sundberg and Dr. Mats Hidestrand for providing the pSP19T7LT plasmid containing CYP2D6 and the human NADPH-CYP reductase. We thank Dr. Didier Rognan for providing the 1000 drug-like compound database. We thank Dr. Iwan de Esch and Andrea van Stolpe for providing the compounds tested from the proprietary database and Atilla Akdemir for collaboration on the preparation of the 3D proprietary database. We thank Patrick Ballmer and Dr. Pavel Pospisil for providing the AGENT program.

Supporting Information Available: Virtual screening results for all docking−scoring combinations. This material is available free of charge via the Internet at http://pubs.acs.org.

Note Added in Proof. Very recently an X-ray crystal structure of substrate-free CYP2D6 was resolved at 3.0 Å resolution (pdb code 2F9Q; Rowland et al. *J. Biol. Chem.* December, 2005). Our homology model was found to agree with most of the details of this crystal structure. Structural differences between the homology model and the crystal structure were the same as those observed between the substrate-free and substrate-bound structures of other CYPs (C. de Graaf, unpublished results).

References

(1) Hou, T.; Xu, X. Recent development and application of virtual screening in drug discovery: an overview. *Curr. Pharm. Des.* **2004**, *10*, 1011−1033.
(2) Congreve, M.; Murray, C. W.; Blundell, T. L. Keynote review: Structural biology and drug discovery. *Drug Discovery Today* **2005**, *10*, 895−907.
(3) 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.
(4) 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.
(5) Bursulaya, B. D.; Totrov, M.; Abagyan, R.; Brooks, C. L., III. Comparative study of several algorithms for flexible ligand docking. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 755−763.
(6) Cummings, M. D.; DesJarlais, R. L.; Gibbs, A. C.; Mohan, V.; Jaeger, E. P. Comparison of automated docking programs as virtual screening tools. *J. Med. Chem.* **2005**, *48*, 962−976.
(7) 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.
(8) 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.
(9) Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* **2004**, *57*, 225−242.
(10) Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of docking performance: comparative data on docking algorithms. *J. Med. Chem.* **2004**, *47*, 558−565.
(11) Kontoyianni, M.; Sokol, G. S.; McClellan, L. M. Evaluation of library ranking efficacy in virtual screening. *J. Comput. Chem.* **2005**, *26*, 11−22.
(12) Mozziconacci, J. C.; Arnoult, E.; Bernard, P.; Do, Q. T.; Marot, C.; et al. Optimization and validation of a docking-scoring protocol; application to virtual screening for COX-2 inhibitors. *J. Med. Chem.* **2005**, *48*, 1055−1068.
(13) Perola, E.; Walters, W. P.; Charifson, P. S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* **2004**, *56*, 235−249.
(14) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **2003**, *52*, 609−623.
(15) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T.; Murray, C. W.; et al. Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793−806.
(16) 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.
(17) Wang, R.; Lu, Y.; Fang, X.; Wang, S. An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2114−2125.
(18) Paul, N.; Rognan, D. ConsDock: A new program for the consensus analysis of protein-ligand interactions. *Proteins* **2002**, *47*, 521−533.
(19) McConkey, B. J.; Sobolev, V.; Edelman, M. The performance of current methods in ligand-protein docking. *Curr. Sci.* **2002**, *83*, 845−856.
(20) de Graaf, C.; Pospisil, P.; Pos, W.; Folkers, G.; Vermeulen, N. P. Binding mode prediction of cytochrome P450 and thymidine kinase protein−ligand complexes by consideration of water and rescoring in automated docking. *J. Med. Chem.* **2005**, *48*, 2308−2318.
(21) Guengerich, F. P. Reactions and significance of cytochrome P-450 enzymes. *J. Biol. Chem.* **1991**, *266*, 10019−10022.
(22) Goeptar, A. R.; Scheerens, H.; Vermeulen, N. P. Oxygen and xenobiotic reductase activities of cytochrome P450. *Crit. Rev. Toxicol.* **1995**, *25*, 25−65.
(23) 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.
(24) Zanger, U. M.; Raimundo, S.; Eichelbaum, M. Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. *Naunyn−Schmiedeberg's Arch. Pharmacol.* **2004**, *369*, 23−37.
(25) Bertilsson, L.; Dahl, M. L.; Dalen, P.; Al-Shurbaji, A. Molecular genetics of CYP2D6: clinical relevance with focus on psychotropic drugs. *Br. J. Clin. Pharmacol.* **2002**, *53*, 111−122.

(26) Oscarson, M. Pharmacogenetics of drug metabolising enzymes: importance for personalised medicine. *Clin. Chem. Lab. Med.* **2003**, *41*, 573−580.
(27) Ingelman-Sundberg, M. Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. *Trends Pharmacol. Sci.* **2004**, *25*, 193−200.
(28) de Graaf, C.; Vermeulen, N. P.; Feenstra, K. A. Cytochrome p450 in silico: an integrative modeling approach. *J. Med. Chem.* **2005**, *48*, 2725−2755.
(29) Kirton, S. B.; Kemp, C. A.; Tomkinson, N. P.; St-Gallay, S.; Sutcliffe, M. J. Impact of incorporating the 2C5 crystal structure into comparative models of cytochrome P450 2D6. *Proteins* **2002**, *49*, 216−231.
(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) Keizers, P. H.; de Graaf, C.; de Kanter, F. J.; Oostenbrink, C.; Feenstra, K. A.; et al. Metabolic regio- and stereoselectivity of cytochrome P450 2D6 towards 3,4-methylenedioxy-*N*-alkylamphetamines: in silico predictions and experimental validation. *J. Med. Chem.* **2005**, *48*, 6117−6127.
(32) Ekins, S.; de Groot, M. J.; Jones, J. P. Pharmacophore and three-dimensional quantitative structure activity relationship methods for modeling cytochrome p450 active sites. *Drug Metab. Dispos.* **2001**, *29*, 936−944.
(33) Vermeulen, N. P. E. Prediction of drug metabolism: the case of cytochrome P450 2D6. *Curr. Top. Med. Chem.* **2003**, *3*, 1227−1239.
(34) Kirton, S. B.; Murray, C. W.; Verdonk, M. L.; Taylor, R. D. Prediction of binding modes for ligands in the cytochromes P450 and other heme-containing proteins. *Proteins* **2005**, *58*, 836−844.
(35) 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.
(36) DeVoss, J. J.; Sibbesen, O.; Zhang, Z. P.; deMontellano, P. R. O. Substrate docking algorithms and prediction of the substrate specificity of cytochrome P450(cam) and its L244A mutant. *J. Am. Chem. Soc.* **1997**, *119*, 5489−5498.
(37) Verras, A.; Kuntz, I. D.; de Montellano, P. R. O. Computer-assisted design of selective imidazole inhibitors for cytochrome p450 enzymes. *J. Med. Chem.* **2004**, *47*, 3572−3579.
(38) Kemp, C. A.; Flanagan, J. U.; van Eldik, A. J.; Marechal, J. D.; Wolf, C. R.; et al. Validation of model of cytochrome P450 2D6: an in silico tool for predicting metabolism and inhibition. *J. Med. Chem.* **2004**, *47*, 5340−5346.
(39) 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.
(40) 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.
(41) 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.
(42) 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*, 425−445.
(43) Keizers, P. H.; Lussenburg, B. M.; de Graaf, C.; Mentink, L. M.; Vermeulen, N. P.; et al. Influence of phenylalanine 120 on cytochrome P450 2D6 catalytic selectivity and regiospecificity: crucial role in 7-methoxy-4-(aminomethyl)-coumarin metabolism. *Biochem. Pharmacol.* **2004**, *68*, 2263−2271.
(44) Lussenburg, B. M.; Keizers, P. H.; de Graaf, C.; Hidestrand, M.; Ingelman-Sundberg, M.; et al. The role of phenylalanine 483 in cytochrome P450 2D6 is strongly substrate dependent. *Biochem. Pharmacol.* **2005**, *70*, 1253−1261.
(45) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(46) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.
(47) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein−ligand interactions: a simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791−804.
(48) 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.
(49) 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.
(50) 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.
(51) 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.
(52) Bohm, H. J. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 593−606.
(53) 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.
(54) Rarey, M.; Kramer, B.; Lengauer, T. The particle concept: placing discrete water molecules during protein-ligand docking predictions. *Proteins* **1999**, *34*, 17−28.
(55) Schnecke, V.; Kuhn, L. A. Virtual screening with solvation and ligand-induced complementarity. *Perspect. Drug Discovery Des.* **2000**, *20*, 171−190.
(56) 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.
(57) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; et al. Modeling water molecules in protein−ligand docking using GOLD. *J. Med. Chem.* **2005**, *48*, 6504−6515.
(58) 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.
(59) Rao, M. S.; Olson, A. J. Modelling of factor Xa-inhibitor complexes: a computational flexible docking approach. *Proteins* **1999**, *34*, 173−183.
(60) 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.
(61) 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.
(62) 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 Transduction* **2002**, *22*, 141−154.
(63) Pickett, S. D.; Sherborne, B. S.; Wilkinson, T.; Bennett, J.; Borkakoti, N.; et al. Discovery of novel low molecular weight inhibitors of IMPDH via virtual needle screening. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1691−1694.
(64) Floriano, W. B.; Vaidehi, N.; Zamanakos, G.; Goddard, W. A., III. HierVLS hierarchical docking protocol for virtual ligand screening of large-molecule databases. *J. Med. Chem.* **2004**, *47*, 56−71.
(65) Bellocchi, D.; Macchiarulo, A.; Costantino, G.; Pellicciari, R. Docking studies on PARP-1 inhibitors: insights into the role of a binding pocket water molecule. *Bioorg. Med. Chem.* **2005**, *13*, 1151−1157.
(66) Aparna, V.; Rambabu, G.; Panigrahi, S. K.; Sarma, J. A.; Desiraju, G. R. Virtual screening of 4-anilinoquinazoline analogues as EGFR kinase inhibitors: importance of hydrogen bonds in the evaluation of poses and scoring functions. *J. Chem. Inf. Model.* **2005**, *45*, 725−738.
(67) Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. Consensus scoring for ligand/protein interactions. *J. Mol. Graphics Modell.* **2002**, *20*, 281−295.
(68) 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.
(69) Graves, A. P.; Brenk, R.; Shoichet, B. K. Decoys for docking. *J. Med. Chem.* **2005**, *48*, 3714−3728.
(70) Susnow, R. G.; Dixon, S. L. Use of robust classification techniques for the prediction of human cytochrome P450 2D6 inhibition. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1308−1315.
(71) Guengerich, F. P.; Hanna, I. H.; Martin, M. V.; Gillam, E. M. Role of glutamic acid 216 in cytochrome P450 2D6 substrate binding and catalysis. *Biochemistry* **2003**, *42*, 1245−1253.

(72) Paine, M. J.; McLaughlin, L. A.; Flanagan, J. U.; Kemp, C. A.; Sutcliffe, M. J.; et al. Residues glutamate 216 and aspartate 301 are key determinants of substrate specificity and product regioselectivity in cytochrome P450 2D6. *J. Biol. Chem.* **2003**, *278*, 4021−4027.
(73) Flanagan, J. U.; Marechal, J. D.; Ward, R.; Kemp, C. A.; McLaughlin, L. A.; et al. Phe120 contributes to the regiospecificity of cytochrome P450 2D6: mutation leads to the formation of a novel dextromethorphan metabolite. *Biochem. J.* **2004**, *380*, 353−360.
(74) Fradera, X.; Knegtel, R. M.; Mestres, J. Similarity-driven flexible ligand docking. *Proteins* **2000**, *40*, 623−636.
(75) Hindle, S. A.; Rarey, M.; Buning, C.; Lengaue, T. Flexible docking under pharmacophore type constraints. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 129−149.
(76) Goto, J.; Kataoka, R.; Hirayama, N. Ph4Dock: pharmacophore-based protein−ligand docking. *J. Med. Chem.* **2004**, *47*, 6804−6811.
(77) Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins* **1999**, *37*, 228−241.
(78) Merlitz, H.; Burghardt, B.; Wenzel, W. Impact of receptor conformation on in silico screening performance. *Chem. Phys. Lett.* **2004**, *390*, 500−505.
(79) Zhang, L.; Hermans, J. Hydrophilicity of cavities in proteins. *Proteins: Struct., Funct., Genet.* **1996**, *24*, 433−438.
(80) Pospisil, P.; Ballmer, P.; Scapozza, L.; Folkers, G. Tautomerism in computer-aided drug design. *J. Recept. Signal Transduction Res.* **2003**, *23*, 361−371.
(81) Cole, J. C.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. Comparing protein-ligand docking programs is difficult. *Proteins* **2005**.
(82) Onderwater, R. C.; Venhorst, J.; Commandeur, J. N. M.; Vermeulen, N. P. E. Design, synthesis, and characterization of 7-methoxy-4-(aminomethyl)coumarin as a novel and selective cytochrome P450 2D6 substrate suitable for high-throughput screening. *Chem. Res. Toxicol.* **1999**, *12*, 555−559.

JM0508538