VU Research Portal

Cytochrome P450-drug interactions: Computational binding mode and affinity predictions in CYP2D6
de Graaf, C.

2006

document version
Publisher's PDF, also known as Version of record

citation for published version (APA)
de Graaf, C. (2006). *Cytochrome P450-drug interactions: Computational binding mode and affinity predictions in CYP2D6*.

Cytochrome P450-Drug Interactions

Computational Binding Mode and Affinity Predictions in CYP2D6

Cytochrome P450-drug interactions: computational binding mode and affinity predictions in CYP2D6

Chris de Graaf

VRIJE UNIVERSITEIT

Cytochrome P450-drug interactions:
Computational binding mode and affinity predictions in CYP2D6

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Exacte Wetenschappen
op maandag 6 november 2006 om 13.45
in de aula van de universiteit,
De Boelelaan 1105

door

Chris de Graaf

geboren te Leiderdorp

"La esperienza non falla, ma sol fallano i nostri giudizi, promettendosi di lei cose che non sono in sua potestà"

"Experience does not fail, only our judgements fail, promising things which are not in her power to achieve"

Leonardo da Vinci - Aforismi, Novelle e Profezie (ca. 1510)

leescommissie: prof.dr. G. Folkers
                   prof.dr. P.D.J. Grootenhuis
                   prof.dr. K. Lammertsma
                   dr. I.J.P. de Esch
                   dr. L. Ridder

Contents

# Introduction

Chapter 1

1.1 General Introduction

The research in this thesis was performed to obtain knowledge on the structure and dynamics of the Cytochrome P450 (CYP) enzyme and CYP2D6-ligand interactions. Computational approaches will be used to predict and rationalize drug metabolism by this enzyme. The aim of the current chapter is to give a general introduction into the background and context of the research presented in the thesis. Paragraph **1.2** deals with the clinical significance of drug metabolism and paragraph **1.3** with the role of ADME prediction in drug development. A general description of the use of computational chemistry in drug metabolism studies is provided in paragraph **1.4**. In the next paragraph **1.5** the clinical significance of Cytochrome P450 enzyme systems are described, with special focus on the human CYP2D6 isoenzyme, which will be the main subject of this thesis. In the final paragraphs **1.6** and **1.7**, the aim, scope, and outline of this thesis is presented.

1.2 Drug metabolism and clinical significance

Drug metabolism determines the fate of a compound entering the body, ultimately controlling whether that compound exerts pharmacological or toxicological effects, or is excreted from the body before doing so. Ideally, drugs and other xenobiotics are degraded to harmless, water soluble metabolites, which are easily excreted via the urine or bile. Occasionally though, an otherwise innocuous compound is converted to a harmful metabolite in a process known as metabolic toxification.[1,2] Other xenobiotic compounds may give rise to drug-drug interactions by inducing or inhibiting the activity of metabolic enzymes, thus affecting the normal detoxification of co-administered drugs. These processes can all lead to clinically significant adverse drug reactions (ADRs).[3,4] Biotransformation of drugs is usually divided into three phases.[5] Phase I reactions involve the introduction of functional groups by hydrolysis, reduction and/or oxidation. In phase II reactions, endogeneous cofactors react directly with substrates or with the functional groups introduced in phase I. The phase II reactions involved are glucoronidation, sulphonation, acetylation, methylation and conjugation with amino acids or glutathione. Phase III metabolism constitutes the further metabolism of the products derived from phase II conjugation reactions. In **Figure 1,** an example is given of possible clinical implications in the case of codeine metabolism. Cytochrome P450 enzymes (CYPs) are the most important phase I drug metabolizing enzymes, implicated in many hundreds of cases of metabolic toxification and clinically significant drug-drug interactions.[6] Understanding these enzymes and their mechanisms of action is therefore invaluable in the process of drug discovery and development.

In many cases, ADRs have been shown to be associated with genetic polymorphisms in specific drug targets, such as GPCRs,[7] or in specific drug metabolizing enzymes, such as CYPs.[3,4] It has for example been estimated that: (i) ADRs cost the US society, US $100 billion[8]; (ii) ADRs cause 100.000 deaths annually in the USA[9]; and that (iii) up to 7% of all hospital admissions in the UK and Sweden are due to ADRs, increasing the length of hospital stays by at least two days at an increased cost of ~$2500 per patient.[10] In addition, the costs of treating patients with polymorphic forms of CYPs are much higher than those required to treat patients who possess non-polymorphic alleles.[11] ADRs are also

one of the most common causes of drug withdrawals, which may have enormous financial implications for the pharmaceutical industry. Furthermore, the number of non-responding patients to drug therapy is high and represents 30–60% of subjects treated with drugs.[12] According to Phillips *et al.*, 56% of drugs that are cited in ADR studies are metabolized by polymorphic phase I enzymes, of which 86% are CYPs.[13] Particularly CYP1A2 and CYP2D6 were found to be relevant in this regard, as they were found to be involved in the metabolism of respectively 75% and 38% of the relevant ADR drugs. Only 20% of the drugs that are substrates for non-polymorphic enzymes are occurring in the ADR reports.
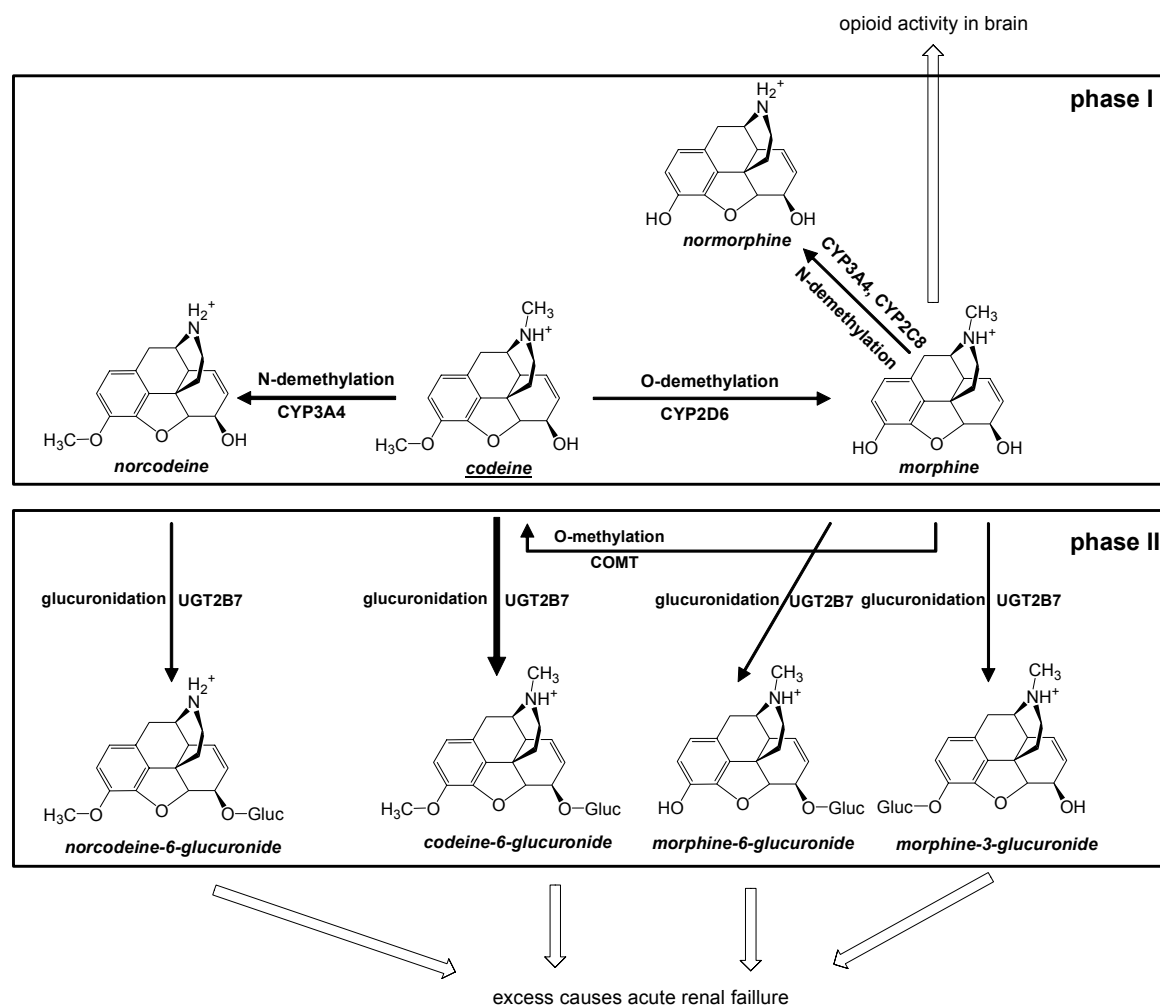
Figure 1: *Metabolic fate of codeine in man, showing phase I O-demethylation and N-demethylation catalysed by CYPs, and phase II O-methylation and glucuronidation catalysed by catechol-O-methyltransferase (COMT) and uridine 5'-triphosphate glucuronosyl-transferase 2B7 (UGT2B7), respectively. The thickness of the arrows indicates the relative importance of the pathways. Notice that formation of morphine, which is responsible for opioid activity in the brain, is normally only observed in relatively small amounts, but that in persons with an ultra-rapid metabolism (UM) CYP2D6 phenotype this pathway becomes more important. Increased formation of morphine and decreased formation of norcodeine (e.g. as a result of CYP3A4 inhibition due to drug-drug interactions) leads to excessive formation of glucuronides, resulting in acute renal failure. On the other hand, for persons with a poor metabolism (PM) CYP2D6 phenotype, codeine therapy will fail because of a decreased formation of the active morphine metabolite. Adapted from Gasche et al.[14]*

Absent or reduced activity of enzymes involved in drug biotransformation can lead to therapeutic failure or adverse drug effects by the following mechanisms (see also **Figure 2** and **Table 4**):[3]

- *Decreased first pass metabolism and drug elimination.* For drugs that are highly dependent on biotransformation to inactive metabolites by a specific (genetically polymorphic) enzyme: i) individuals with a poor metabolizing phenotype (PMs), may experience a larger drug response and/or be at greater risk of toxicity; ii) co-administered inhibitors can diminish enzyme activity, which may result in a larger drug response as well as in toxic effects.
- *Decreased pro-drug metabolism.* When the conversion of a drug into an active metabolite is primarily mediated by a specific enzyme (e.g. conversion of codeine to morphine by CYP2D6[15]), the drug may be ineffective in PMs or, when enzyme inhibitors are co-administered, lead to therapeutic faillure.

Increased activity of enzymes involved in drug biotransformation can lead to therapeutic failure or adverse drug effects as well, e.g. by the following mechanisms:
- *Increased first-pass metabolism and drug elimination.* For drugs that are highly dependent on biotransformation to inactive metabolites by specific (genetically polymorphic) enzymes: i) people with an ultra-rapid metabolizers phenotype (UMs), may have a lower drug response; ii) co-administered inducers can significantly increase enzyme activity and result in lower drug response and therapeutic faillure as well.
- *Increased formation and accumulation of toxic metabolites* formed by drug biotransformation, may result in adverse drug effects.
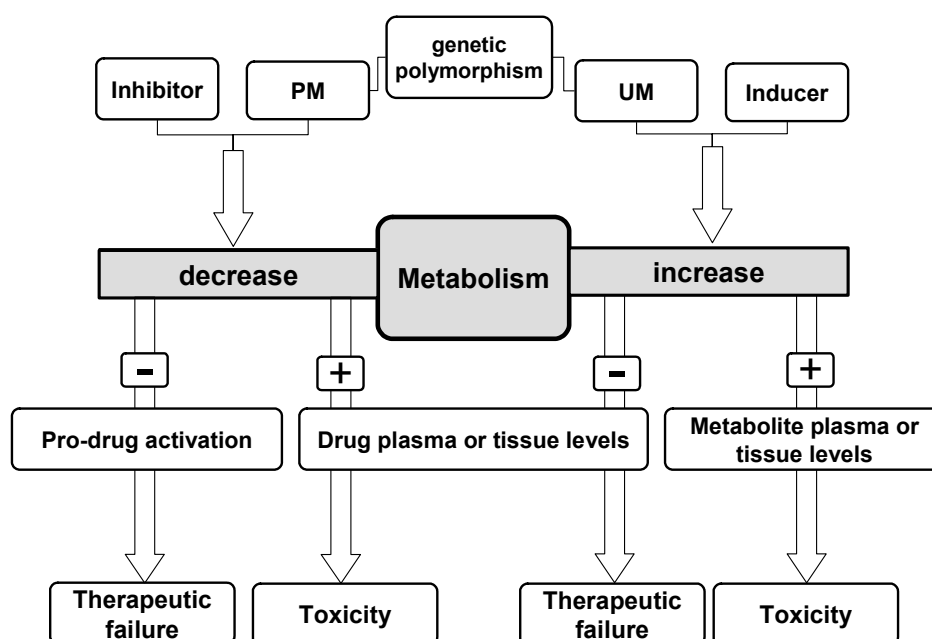
Figure 2: *Effects of drug-drug interactions (inhibitors, inducers) and genetic polymorphisms (poor metabolizing phenotype (PM), ultra-rapid metabolizing phenotype (UM)) on metabolism, affecting the clinical impact of drugs (toxicity and therapeutic failure). "+" and "-" indicate an increase or decrease of the subsequent clinical effect.*

In conclusion, drugs metabolism determines the fate of a compound entering the body, and absent, reduced or increased activity of enzymes involved in drug biotransformation can lead to therapeutic failure or adverse drug effects. Understanding these enzymes and their mechanisms of action is therefore invaluable in the process of drug discovery and development. The next paragraph describes the interplay between *in silico*, *in vitro* and *in vivo* studies to predict processes controlling the fate of drugs: absorption, disposition, metabolism and excretion (ADME).

1.3 The role of ADME(T) in drug discovery: The interplay between *in silico, in vitro* and *in* vivo studies

High throughput synthesis and screening of potential drug candidates yields hundreds of thousands of compounds. High-throughput screening (HTS) tests are performed for various properties, ranging from solubility and metabolic stability to drug potency and toxicity.[16] The number of high-quality lead compounds derived from such HTS tests is rather low, and several studies suggest that poor pharmacokinetics (PK) and toxicity are among the most important causes of late-stage failures of compounds in drug development.[17,18] Early consideration of drug potency, ADME-properties (absorption disposition, metabolism and excretion) and toxicity is therefore increasingly seen as essential for efficient discovery and development of new drugs and drug candidates.[16,19] High-throughput *in vitro* methods are generally used to screen previously synthesized compound libraries while library design now also includes *in silico* predictions to increase the critical hit/lead ratio. **Figure 3** presents two approaches in which ADME/Tox analyses are incorporated in the drug discovery and development process.[20] In the first ("serial") model, compounds are tested for potency first via high-throughput screening (HTS), followed by serial examinations of their absorption, distribution, metabolism, and excretion (ADME) and toxicity (Tox) profiles. Compounds failing appropriate ADME/Tox profiling require new analogues to be synthesized and submitted through the cycle. In the second ("parallel") model, libraries of 'pre-synthesized' analogues are submitted to multiple *in silico* and *in vitro* tests in parallel, building a database of potency, ADME, and toxicity profiles. Best-option molecules are selected for further optimization. This simultaneous optimisation of the pharmacokinetic, toxicological, and pharmacological properties remains one of the greatest challenges in drug design.[21]
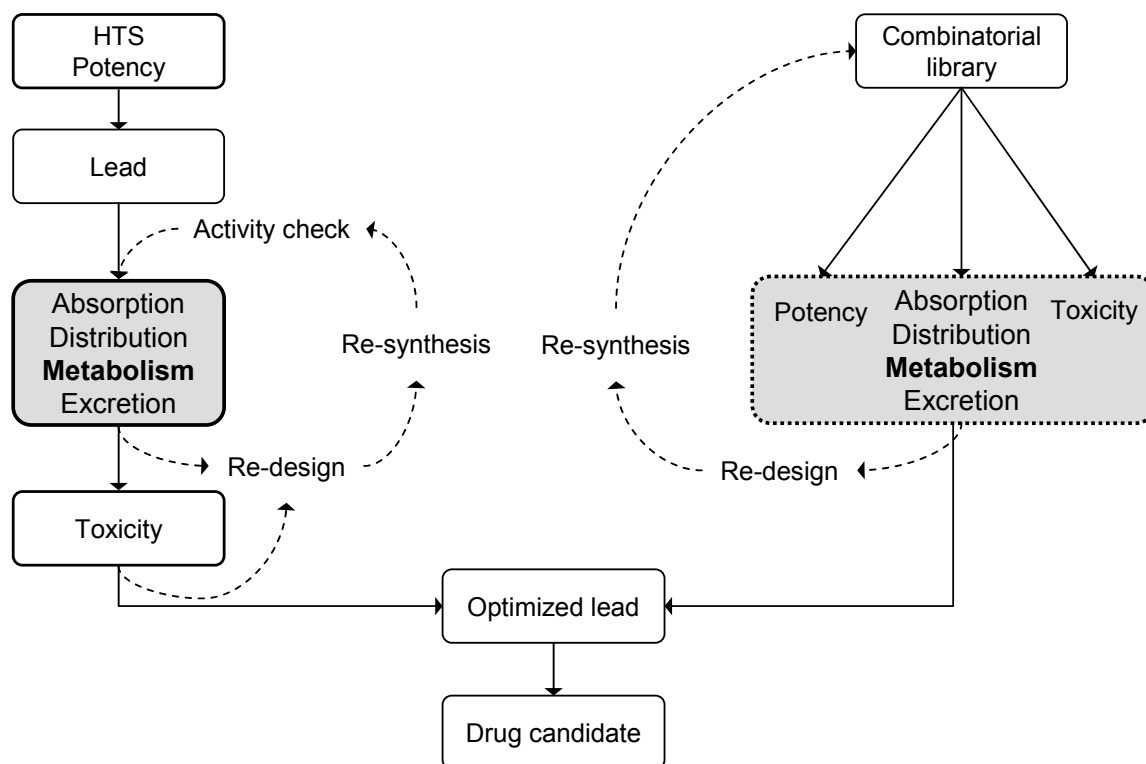
Figure 3: *"Serial" (left scheme) and "parallel" (right scheme) models of lead optimisation (see text). Adapted from Beresford et al.*[20]

Pre-experimental prediction of pharmacological and non-pharmacological drug-properties with *in silico* ADME-Tox screening[21,22] is likely to be one of the most challenging developments with a great scientific and practical impact. **Figure 4** shows the interplay between *in silico* models, *in vitro* screening, and *in vivo* testing in drug lead finding and optimization.
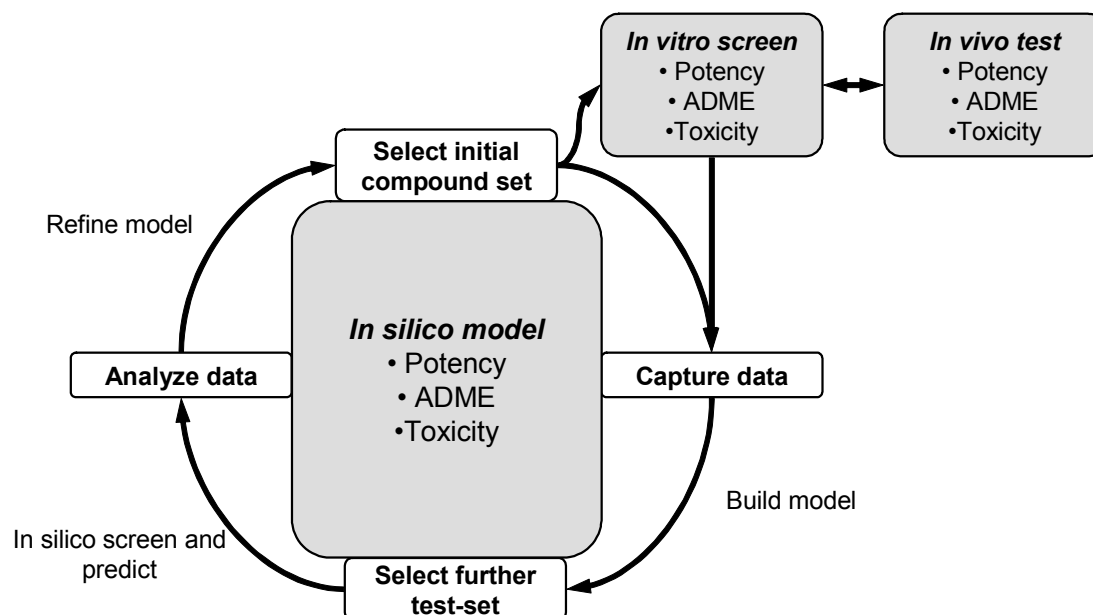
Figure 4: *The interplay between in silico models, in vitro screening, and in vivo testing in drug lead finding and optimization (including the optimization of drug potency and ADME properties, and the minimization of drug toxicity). In iterative cycles, in silico models are built, refined and validated by in vitro screening. Subsequently, optimized leads are tested in vivo. Adapted from Beresford et al.*[20]

Figure 5 gives a schematic overview of predictive *in silico* models necessary for guiding drug lead finding and optimization. *In silico* studies have been applied to predict intestinal absorption and drug distribution including permeation across the blood-brain barrier.[19,23] Based on the assumption that oral bioavailability of metabolically stable compounds are determined by the rate of absorption, some of these methods were also used to predict oral bioavailability. Since the fraction of the oral dose that reaches systemic circulation is influenced by both absorption and first-pass metabolism, prediction techniques for metabolic stability are needed. Lipinski has analyzed large drug databases and identified molecular criteria that should be fulfilled in order to obtain appropriate oral delivery: i) molecular weight, MW < 500 Daltons, ii) octanol/water partition coefficient, log(P) < 5, iii) number of hydrogen bond donors, N_{donor} <5, iv) number of hydrogen bond acceptors, $N_{acc.}$ < 10.[18] *In silico* prediction of drug potency, ADME properties, and toxicity (Tox) can be performed using various approaches, such as database modelling, 1D/2D-quantitative-structure-activity relationships modelling, and ligand and protein target structure-based modelling, reviewed by refs 19,23. A virtual screening cascade typically follows these different approaches in subsequent steps: i) filters like the above mentioned Lipinski rule-of-fives are applied to eleminate compounds possessing non-drug-like properties; ii) topocological 1D/2D searches from known ligands; iii) 3D-similarity or 3D-pharmacophore searches; iv) protein structure-based virtual screening.[24] Computational methods used for prediction of drug metabolism are elaborated in paragraph **1.4**.
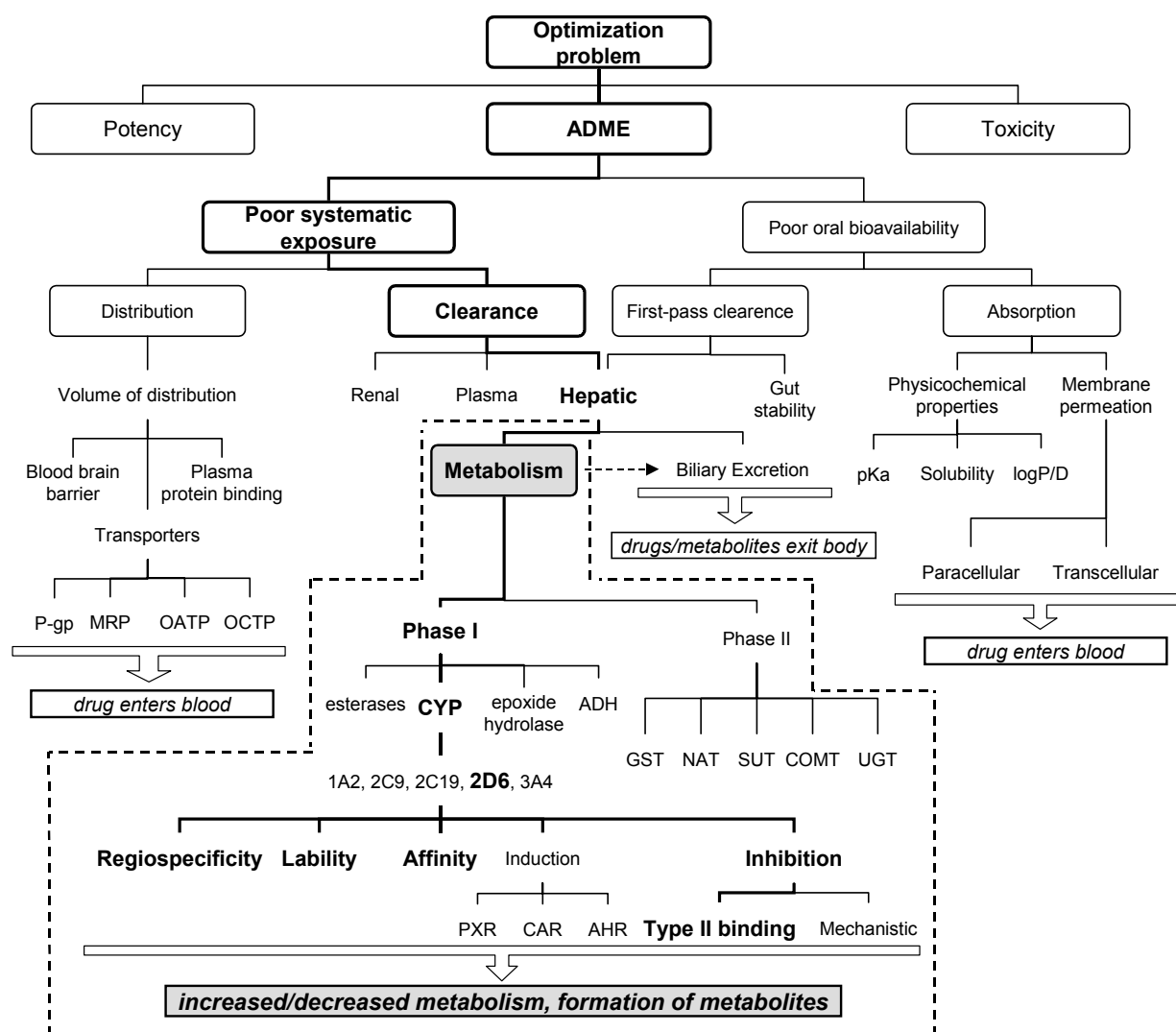
Figure 5: *Schematic overview of predictive in silico models necessary for guiding drug lead optimization. The emphasis is on ADME (Absorption, Distribution, Metabolism, Excretion) processes (apart from the optimization of drug potency and minimization of drug toxicity) and their ultimate effect on drug fate (indicated by arrows). CYP(2D6) metabolism models described in this thesis are highlighthed in bold and enclosed by a dashed box. Abbreviations: AD, alcohol dehydrogenase; NAT, N-acetyl transferase; GST, glutathione S-transferase ; SUT, sulfotransferase, COMT, catechol-O-methyltransferase; UGT, uridine 5'-triphosphate glucuronosyltransferase; AHR, aryl hydrocarbon receptor; CAR; constitutive androsterone receptor, MRP, multidrug resistance protein; OATP, organic anion transporter protein; OCTP, organic cation transport protein; P-gp, P-glycoprotein; PXR, pregnane receptor. Adapted from Ekins et al.[25]*

Early consideration of drug potency, ADME-properties (absorption disposition, metabolism and excretion) and toxicity is increasingly seen as essential for efficient discovery and development of new drugs and drug candidates. Drug lead compounds are discovered and optimized by an integrative process of *in silico* modeling, *in vitro* screening, and *in vivo* testing. The next paragraph focuses on the *in silico* prediction of metabolism.

1.4 Prediction of drug metabolism *in silico*

Since activity, toxicity, bioavailability, distribution, and final elimination all depend on metabolic biotransformations, the prediction of metabolic stability and the identification of potential sites of metabolism in a putative drug molecule can be a significant help in

designing new compounds with a better pharmacokinetic profile.[26-28] When the site of metabolism is known, metabolically labile compounds can be stabilized by removing, replacing or protecting metabolically susceptible groups. Similarly, the formation of toxic metabolites can be avoided by chemically protecting the labile moieties. Finally, knowing the site of metabolism is essential for the design of pro-drugs, where the compound needs to be metabolized in order to become active, and also for drugs showing excessively long half lives. Modifying functional groups in drug lead compounds that are essential for binding to metabolic enzymes may affect the binding affinity for these enzymes as well as their metabolic regiospecificity.[26-28] **Figure 5** gives a schematic overview of predictive *in silico* models currently used for guiding drug lead finding and optimisation and highlights models describing drug metabolism.

Computational chemistry is an especially suitable tool to predict metabolic stability and sites of metabolism in potential drugs. *In silico* approaches to predict metabolism can be divided into i) predictive databases, ii) quantitative structure-relationships (QSARs) derived from physicochemical ('1D') and topological ('2D') properties of biotransformation enzyme substrates, and iii) structure-based ('3D') approaches.[19,29] The first two approaches provide robust but rough classifications of the inhibition potency of ligands without yielding much insight into the possible (molecular) background of the mode of action. The structure-based ('3D') molecular modeling methods, on the other hand, aim to be insightful as well as predictive.[22,30] The latter approaches can be divided into ligand-based (including QM calculations on substrates, pharmacophore models, and 3D-QSAR models), protein-based (including protein crystal structures and homology models) and ligand-protein interaction based models (including issues like binding orientation, ligand dynamics and binding affinity).[29] CYPs are of special clinical importance because of their great impact on phase I drug metabolism.[1,2] These are therefore the most extensively studied drug metabolizing enzymes, *in vivo, in vitro* as well as *in silico*. Examples of computational approaches that use databases to predict drug metabolism are MetabolExpert, META, and Meteor.[31] 1D-QSARs predicting CYP metabolism have been very recently reviewed by Hansch et al.[32] Several successful 2D-QSAR models predicting CYP metabolism and inhibition have been reported as well.[33-36] An extensive overview of structure-based computational investigations used to understand, rationalize and predict CYP structure and function is provided in **Chapter 2**. Molecular modeling studies of other human biotransformation enzymes than CYPs are relatively scarce, although they have been increasingly reported in the past few years: see for example studies on epoxide hydrolase,[37] alcohol dehydrogenase,[38] esterases,[39] sulfotransferase (SUT),[40] catechol-O-methyltransferase (COMT),[41,42] glutathione S-transferase (GST),[43,44] N-acetyl transferase (NAT),[45,46] uridine 5'-triphosphate glucuronosyltransferase (UGT),[47-49] and cysteine cunjugate beta-lyase.[50]

Examples of future and as yet unresolved challenges in computational drug metabolism research are: i) The accurate quantitative prediction of metabolite formation ratios and turnover rates, and ii) the construction of combined models for multiple biotransformation enzymes, enabling the prediction of the relative kinetics of different reaction pathways and prediction of the relative affinities of drugs for different enzymes. This would give insight into the relative importance of different enzymes in the formation of different metabolites and might finally lead to the prediction of (the kinetics of) complete metabolic pathways of drugs.

The next paragraph focuses on the family of Cytochrome P450 enzymes (CYPs), the most important phase I drug metabolizing enzymes, able to catalyse a wide variety of reactions, and implicated in many hundreds of cases of metabolic toxification and clinically significant drug-drug interactions.

1.5 Clinical significance of Cytochromes P450

Cytochromes P450 (CYPs) are hemoproteins which catalyse the oxidation and reduction of a wide variety of endogenous and xenobiotic compounds.[51,52] Originally, they were named Cytochrome because they were a pigment (in Greek: chromos) from within the cells (in Greek: cyto), and P450 because they were the pigment that gave a typical absorbance band at 450 nm.[53,54] CYPs generally detoxify pharmacologically active and potentially hazardous compounds, but in a number of cases non-toxic parent compounds are bioactivated into pharmacologically or toxicologically active metabolites, or procarcinogens into their ultimate carcinogens.[1,2] There are 4500 uniquely named sequences of CYPs known (drnelson.utmem.edu). They are named according to sequence homology. In general, when the amino acid sequences of two CYPs are less than 40% identical, the enzymes are assigned to different gene families, indicated by a number. When sequences are between 40 and 55% identical, they belong to the same family but are assigned to different subfamilies, indicated by a capital letter. When the sequences are more than 55% identical they are classified as members of the same subfamily, but the isoenzyme is indicated by a number.[55] For instance, a CYP from the family 2, subfamily D, isoenzyme 6 is thus named CYP2D6. This paragraph describes some general features of CYPs, in particular their catalytic activity, their role in drug metabolism, and factors influencing their clinical significance, such as genetic polymorphisms and drug-drug interactions. Special attention is given to CYP2D6, the isoenzyme studied in this thesis. An extensive overview of computational investigations used to understand, rationalize and predict CYP structure and function is provided in **Chapter 2**.
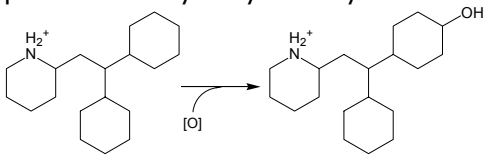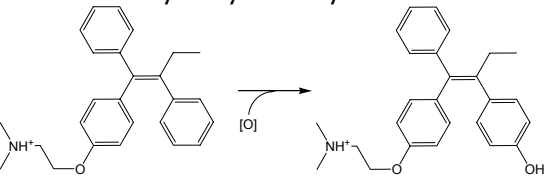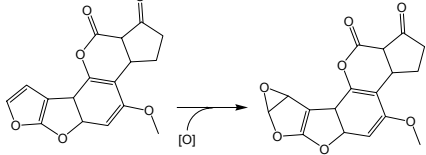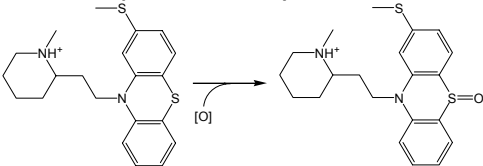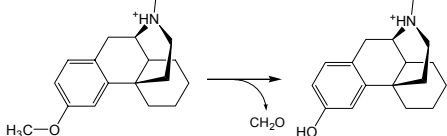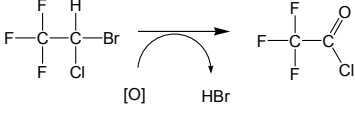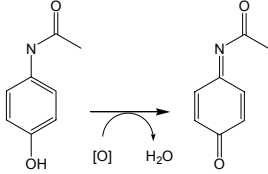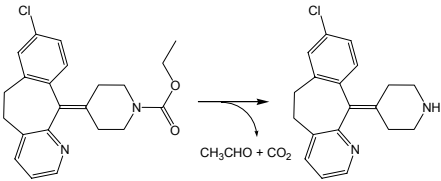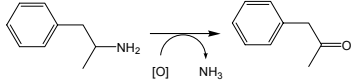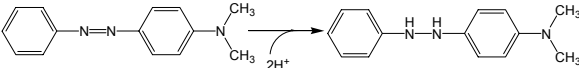
CYP catalytic activities

CYPs catalyse a wide variety of reactions (**Table 1**). The catalytic activities of CYPs can be divided into (i) monooxygenase activity, usually resulting in the incorporation of an oxygen atom into the substrate, (ii) oxidase activity, resulting in the formation of superoxide anion radicals or hydrogen peroxide (uncoupling of the catalytic cycle, see **Figure 6**, $2e^-$ oxidase and $1 e^-$ oxidase activity), and (iii) substrate reductase activity usually producing free radical intermediates under anaerobic conditions.[52,56] Of these three enzymatic functions of CYPs, attention has been almost exclusively on the monooxygenase activity and its proposed mechanisms of action. The two most commonly encountered mechanisms of action, the H-atom abstraction/oxygen rebound[57] and the electron oxidation/recombination[58] have by far been investigated most. **Figure 6** schematically shows the CYP reaction cycle where these three activities and associated reactions are indicated. The CYP catalytic cycle begins with reversible substrate (R) binding, which converts the six-coordinate, low-spin form (species 1) of the enzyme to the five-coordinate, high spin Fe^{III} substrate complex (species 2). Addition of the first electron reduces the enzyme to the five-coordinate Fe^{II} substrate complex (species 3). Binding of molecular oxygen gives the six-coordinate Fe^{II}-O_2 dioxygen intermediate (species 4). Addition of a second electron and two protons followed by cleavage of the oxygen-oxygen bond produces a molecule of water and an oxidizing species, the so-called oxygen intermediate (species 7). Insertion of the iron-bound oxygen into the substrate leads to species 8 and subsequent product release. Species 5 and 6 represent alternative possible species along the reaction pathway. The studies described in this thesis focus on (the transition between) species 1 and 2, the binding of the substrate in the active site.

For the bacterial CYP101, the most extensively studied CYP, only small changes in the protein conformation and in the substrate position and orientation are observed by X-ray crystallography during the catalytic cycle (**Figure 6**). Also the substrate orientation in the bacterial CYP107A was observed to be largely unaffected by binding of dioxygen.[59] For CYP102, on the other hand, it has been suggested that conformational changes of the

protein, driven by electron-transfer, trigger the movement of the substrates into effective orientations for hydroxylation.[60]

Table 1: *Types of reactions catalyzed by Cytochrome P450 isoenzymes (in bold Italics) with an example for each reaction. All types of reactions mediated by CYP2D6 are exemplified for this isoenzyme.* [56,61]

Aliphatic hydroxylation perhexiline 4-hydroxylation by CYP2D6 	***Aromatic hydroxylation*** tamoxifen 4-hydroxylation by CYP2D6
Epoxidation aflatoxine B1 2,3-epoxidation by CYP1A1 	***Heteroatom oxygenation*** thioridazine sulphoxidation by CYP2D6
Heteroatom dealkylation dextromethorphan O-dealkylation by CYP2D6 	***Dehalogenation*** Halothane debromination by CYP2E1
Dehydrogenation Acetaminophen dehydrogenation by CYP2E1, CYP1A2, and CYP3A4 	***Cleavage of esters*** loratadine de-esterification by CYP3A4 and CYP2D6
Oxidative group transfer oxidative deamination of amphetamine 	***Reduction*** dimethylaminoazobenzene reduction by CYP2B

Oxygen reduction
In presence of NAPQI CYP1A reduces oxygen (between species 4 and 2 in **Figure 6**)

$$(Fe^{2+} O_2 RH) \rightarrow (Fe^{3+} RH) + O_2^{\bullet -}$$

Indeed a balance exists between the kinetics of the heme-iron-oxygen reduction and the dynamics of the substrate binding. Changes in the oxygen species may occur faster than substrate binding orientation and relaxation times, leaving the substrate in steps 2-8 of the catalytic cycle in an orientation that was optimal for a previous, but no longer present oxygen species. Nevertheless, small conformational changes in the protein have been hypothesized to facilitate the different steps along the reaction pathway.[62] In the case of CYP101, an important conformational change seems to be the re-orientation of D251 and T251 in the I-helix.[63] This might facilitate stabilization of and proton delivery to the

dioxygen complex (species 4 and 5), eventually leading to the formation of the oxygen intermediate (species 7). For a proper description of the effects of such subtle conformational changes on the reaction cycle, and the types of reactions catalyzed by CYPs (**Table 1**), the use of quantum mechanics (QM) or combined quantum mechanics/molecular mechanics (QM/MM) methods is necessary.
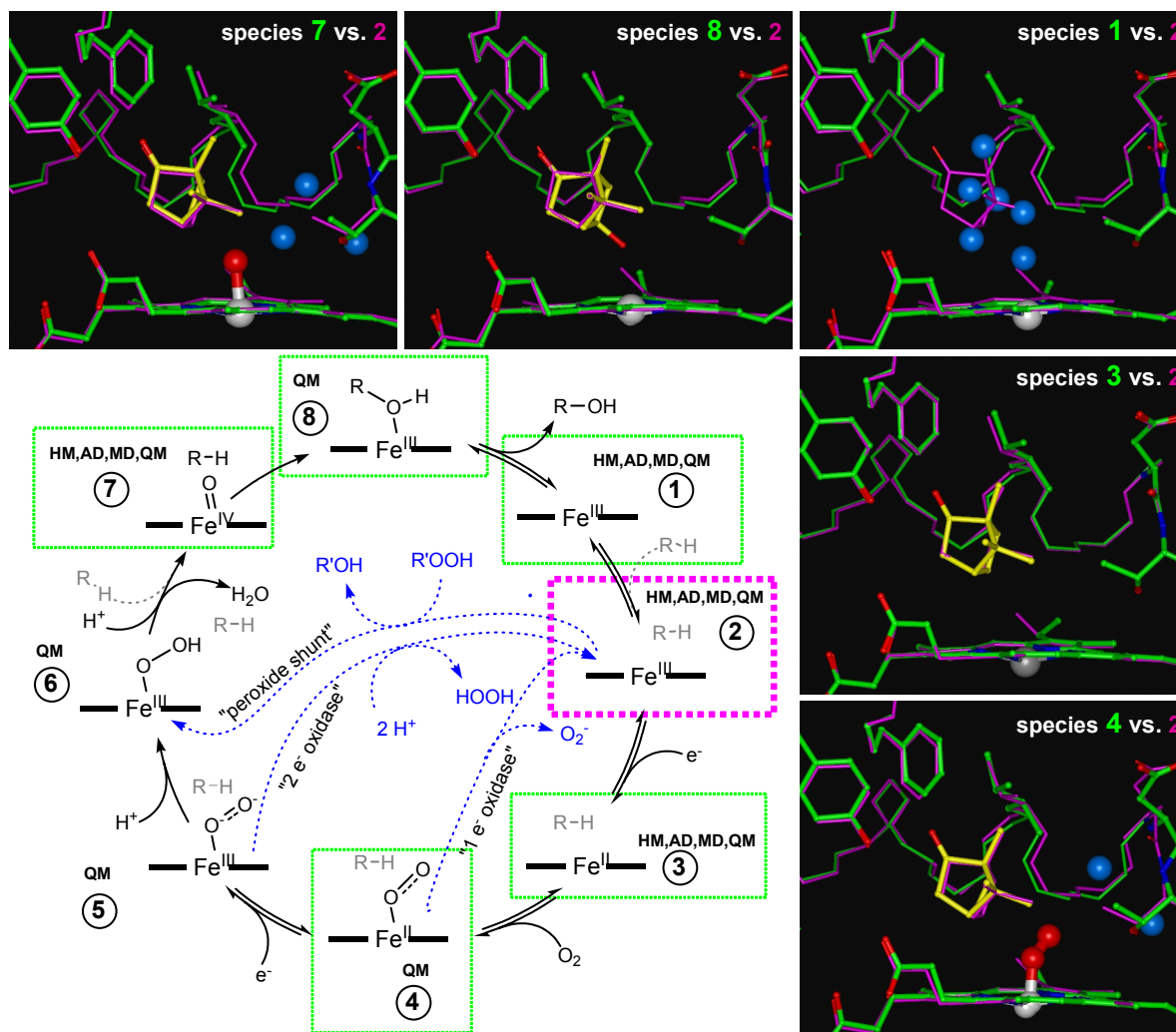
***Figure 6**: Cytochrome P450 reaction cycles, currently known and described in the text, and the molecular modelling methods to describe the different species in these cycles: homology modelling (HM), automated docking (AD), molecular dynamics simulation (MD), and quantum mechanical calculations (QM). This thesis focuses on (the transition between) species 1 and 2. Panels surrounding the catalytic cycle show the active site of crystal structures of species 1,[64] 3,[65] 4, [65] 7,[65] and 8[66] of CYP101 (protein carbon atoms in green, substrate carbon atoms in yellow) compared to the structure of species 2 [65] (protein and substrate carbon atoms in purple). The heme iron atom is depicted in white; water oxygen atoms are colored blue. Only minor changes in the protein conformation and in the position and orientation of the substrate are observed during the catalytic cycle. These small conformational and orientational changes however are hypothesized to facilitate the different steps in the catalytic cycle.[62]*

Indeed, models for species 4, 5, 6, and 8 have only been described by QM or combined QM/MM methods,[67-71] while the other species have also been subjected to (combinations of) other molecular modelling approaches.[29,72] The buried active site of many CYPs, however, implies that the protein must undergo dynamic motions to allow substrates to enter and products to leave the active site (transition between species 1 and 2 and

between species 8 and 1, respectively).[73] Indeed, conformational changes seem to occur during ligand binding, as comparisons between ligand-bound and ligand-free crystal structures of different CYP isoenzymes showed ligand induced changes in size, shape and hydration of the active site.[29] This conformational flexibility of the active site is likely to underlie the capacity of many CYP isoenzymes to metabolize structurally diverse substrates of different sizes and to bind substrates in multiple binding modes (generating different metabolic products). Computational chemistry techniques like MD and automated docking are especially suitable to gain insight into this dynamical process of ligand (un)binding. Consideration of species 2 (and the transition between species 1 and 2) using these techniques is sufficient for proper description and prediction of CYP-ligand interactions and structures (see **Chapter 2**).

Drug metabolism by CYPs
The human genome encodes for 80 CYP proteins, of which approximately fifteen are classified to be primarily involved in the metabolism of xenobiotic compounds (i.e. CYP1A1, 1A2, 1B1, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1, 3A4, 3A5 and 3A7).
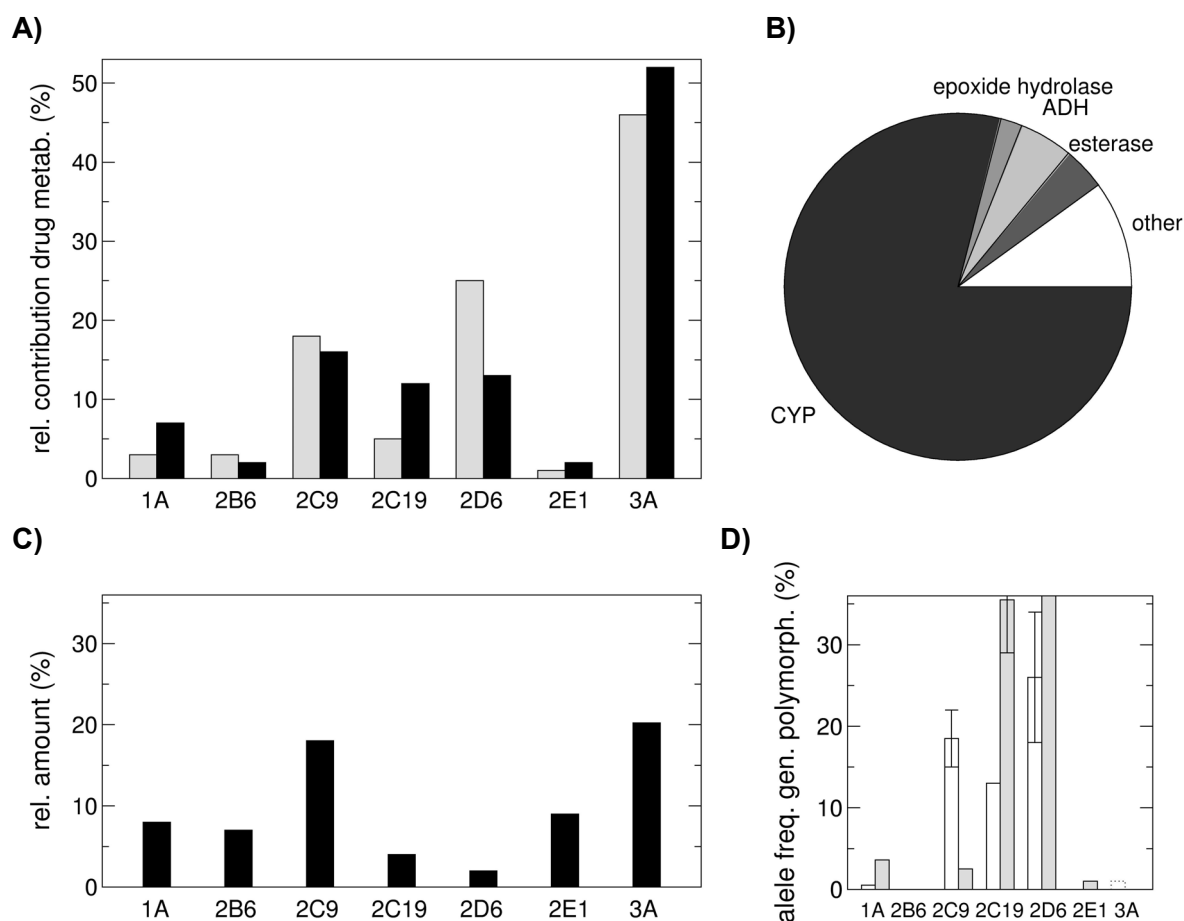
Figure 7: *Panel A shows the relative contributions of CYP isoenzymes to the clearance of 315 drugs in 1997 (grey)[74] and to the metabolism of the top 200 drugs prescribed in 2002 (black).[75] Total contribution of CYPs to the phase I metabolism of all drugs currently marketed[76] are presented in panel B. Relative amounts of major CYPs involved in hepatic metabolism[75] are given in panel C. The frequencies of alleles of clinically relevant genetic polymorphisms are given in panel D (white=Caucasians, grey=Orientals).[77,78] CYP2A6 and CYP2C8 (not included in the graphs) are present in the liver in high amounts (resp. 17% and 4%)[79] and clinically significant genetic polymorphisms for these isoenzymes have been reported,[77,78] but their relative contributions to drug metabolism are small. CYP1A includes CYP1A1 and CYP1A2, CYP3A includes CYP3A4, CYP3A5 and CYP3A7.*

All together these CYPs account for the phase I metabolism of about 70-80% of currently marketed drugs.[74-76] In **Figure 7C** the distribution of the major human liver CYP isoenzymes involved in drug metabolism in a group of 30 Caucasian patients is shown. CYP3A4 is the most abundant isoenzyme, followed by the CYP2C family members, and then CYP1A.[79] When it comes to the amount of drugs for which a CYP isoenzyme is involved in metabolism again CYP3A4 takes the largest share (~50%), followed by the CYP2C family members (~25%) and CYP2D6 (~15%) (**Figure 7A**). So although CYP2D6 only accounts for a small percentage of all hepatic CYPs (~2%), its role in drug metabolism is significantly higher than its hepatic occurence. However, there seems to be a decreasing trend in the relative importance of CYP2D6 versus CYP2C9 and especially CYP2C19 (see **Figure 7A**). Early reviews estimated that CYP2D6 was involved in the clearance of 25%-30% of the drugs metabolized by CYPs, while the relative contributions of CYP2C19 and CYP2C9 were estimated to be ~15%, ~5%, respectively.[74] More recent reviews estimate the relative contributions of CYP2D6, CYP2C9 and CYP2C19 to be ~15%, ~15%, and ~10%, respectively.[75,76] This could be an indication of the fact that CYP2D6 metabolism is more often actively considered in the design of novel drugs.

CYP substrate selectivity and characteristics

Substrates for the different human CYP isoforms can be roughly characterized based on their physicochemical properties (1D),[32] as well as their topology (2D).[80] A summary of the characteristics of the major hepatic CYPs involved in drug metabolism is provided in **Table 2**.

Table 2: *Characteristics of major human hepatic drug metabolizing CYPs. The information represents a summary of several studies reviewed by refs* [78,81,82]

CYP	substrate characteristics				Substrate[e]	Inhibitor[f]	Inducer[g]	GP[h]	X-ray[i]
---	planar[a]	Vol[b]	lipo[c]	ch[d]					
1A1	yes	m	h(m)	0	m-resorufine	furafyline	PAH	-	no
1A2	yes	l	h(m)	0^+	e-resorufine	fluvoxamine	PAH	+	no
2A6	no	l	m(m)	0^+	coumarine	pilocarpine	PB	+	yes[83]
2B6	no	m	m(m)	0^+	phenobarbital	orphenadrine	PB	+	no[c,84]
2C8	no	m	h(m)	0^-	taxol	trimethoprim	PB	+	yes[85]
2C9	no	m	h(h)	0^+	tolbutamide	sulphaphenaz	PB	++	yes[86,87]
2C19	no	m	h(m)	-	*S*-mephenytoin	omeprazole	PB	++	no[c]
2D6	**no**	**m**	**h(l)**	**+**	**d-metorphan**	**quinidine**	**none**	++	**yes[88]**
2E1	yes	l	m(l)	0	chlorzoxazone	isoniazid	ethanol	-	no
3A4	no	h	h(h)	0	nifedipine	ketoconazole	DM	-	yes[89,90]

a) Majority of typical substrates of the specific isoenzyme is considered to be planar or non-planar, as measured by the ratio of the substrate molecular area (a) and depth squared (d^2) (considering $a/d^2 = 3$ as a cut-off); **b**) Majority of typical substrates of the specific isoenzyme have a l(ow) (~<230 $Å^3$), m(edium) (~230-300 $Å^3$) or h(igh) (~>300 $Å^3$) solvent accessible surface volume; **c**) Majority of typical substrates of the specific isoenzyme has a low (l), medium (m) or high (h) lipophilicity as measured by log(P), the logarithm of the octanol/water partition coefficient and or (in brackets) log($D_{7.4}$), the logarithm of the octanol/water partition coefficient at pH 7.4; l(ow): ~<0.5, m(edium): ~0.5-1.5; h(igh): ~>1.5); **d**) Majority of typical substrates of the specific isoenzyme are acidic/negatively charged(-), weakly acidic (0^-), neutral (0), weakly basic (0^+), or basic/positively charged (+). **e**) Marker substrate. It should be noted that the substrate can also (too minor extent) be metabolized by other CYP isoforms. Abbreviations: methoxyresorufin (m-resorufine), ethoxyresorufin (e-resorufine), dextromethorphan (d-methorphan). **f**) Marker inhibitor. It should be noted that the compound can also inhibit other CYP isoforms (too minor extent). **g**) Isoenzyme inducible by: polyclic aromatic hydrocarbons (PAH), Phenobarbital (PB), ethanol, or dexamethasone (DM); CYP2D6 is considered to be non-inducible. **h**) Clinical significance of genetic polymorphisms: significant (+); very significant (++); not significant (-) **i**) Availability of an X-ray structure for the isoenzyme. **j**) Crystal structure of (an) isoenzyme(s) of the same subfamily has been resolved.

1D/2D-QSARs models for CYPs have provided robust but rough classifications of the inhibition potency of ligands, but do not give much insight into the possible (molecular) background of the mode of action. The structure-based ('3D') molecular modeling methods on the other hand, aim to be insightful as well as predictive. The latter approaches are for example needed in the many cases in which the same substrate is (to different extents) metabolized by different CYP isoforms (yielding different metabolites), as for example is the case for clozapine (CYP2D6, 1A2, 2C9, 2C19, and 3A4),[91] imipramine (CYP2D6, 1A2, 2C19 and 3A4),[92] and MDMA (CYP2D6, 2B6 and 3A4).[93]

Genetic polymorphism of CYPs

The efficacy and pharmacokinetics of a drug can be determined genetically. Genes encoding for proteins involved in the pharmacological effects, transport, or metabolism of the drug for instance, can be present in the population in two or more allelic forms which may therefore result in varying therapeutic activities and safety margins of the drug. When at least two alleles have frequencies greater than 1% this is called a genetic polymorphism. In some cases these polymorphisms are caused by a mutation, deletion or insertion of just a single nucleotide, called single nucleotide polymorphisms (SNPs). Many drug-metabolizing enzymes are known to be polymorphic: aldehyde dehydrogenase, UDP-glucoronyltransferases (UGT2 family members), sulfotransferase (SUT1A3), catechol O-methyl transferase, N-acetyl transferases (NAT1 and NAT2), glutathion S-transferase (STM1 and GSTT1), and many CYPs.[94] There are as many as 22 human CYPs expressed from different alleles (www.imm.ki.se/CYPalleles).[95] Nevertheless, of these 22 enzymes, only few were found with significant clinical effects caused by the genetic polymorphisms (see **Figure 7D**). Genetic polymorphisms in enzymes may lead to abolished, reduced, altered, or increased enzyme activity.[3] Minor clinical effects were reported for CYPs 1A2, 2A6, 2B6 and 2C8, and major clinical effects of polymorphism were reported for CYPs of the 2 family, namely CYP2C9, CYP2C19 and CYP2D6 (see **Tables 2**-**4**).[11]

The Cytochrome P450 2D6 isoenzyme

Although the expression levels of CYP2D6 represent only 2% of all hepatic CYPs (**Figure 7C**), it is the second most important drug metabolizing enzyme after CYP3A4, and involved in the metabolism of about 15-30% of the currently marketed drugs (**Figure 7A**).[96,97] Large interindividual differences exist in CYP2D6 activity, due to gene multiplicity and polymorphisms further emphasizing its clinical importance (**Figure 7D**, **8**, **Tables 2**-**4**).[11,94,98] The early identification of potential CYP2D6 substrates and prediction of their metabolism is therefore advantageous in the discovery and development of new drugs.

CYP2D6 catalyzes a broad variety of reactions (**Table 6**) and is responsible for the metabolism of many broadly prescribed pharmaceuticals, including anti-arrhythmics, antidepressants, anti-psychotics, beta-blockers, and analgesics (**Table 4**).[61,96,99] The preference for drugs acting on the central nervous system is interesting, suggesting overlap between CYP2D6 specificity and the specificity of proteins involved in neurotransmission processes in the brain, such as serotonine transporters and GPCRs. In general, these drugs are lipophilic, basic, medium sized, positively charged at neutral pH (which decreases their lipophilicity) (**Table 2**).[81] In addition, it has been shown that 5-methoxytrypamine, a metabolite and precursor of melatonin, is a specific and high turnover endogeneous substrate of CYP2D6 that is being converted to the neurotransmitter serotonin.[100] The broad substrate specificity of CYP2D6 makes this CYP-isoform prone to inhibition by numerous drugs from these and other therapeutically important classes, setting the stage for clinically significant drug-drug interactions. Quinidine, clemastine, fluoxetine, paroxetine, terbinafine, and ritonavir are just a few inhibitors of CYP2D6, which have the potential to block the metabolism of other drugs prescribed simultaneously (**Table 4**).[6]

The enzyme expression levels are, in contrast to other hepatic xenobiotic metabolising cytochrome P450s, not regulated by any known environmental agents and are not inducible by known hormones. CYP2D6 is primarily expressed in the liver, but low expression levels of CYP2D6 are also found in the skin,[101] the gastrointestinal tract,[102] and in the brain.[103] A frameshift mutation in the pseudogene CYP2D7 led to a brain specific expression of a splice variant of this enzyme in Indian people.[104] This enzyme was shown to be even more active than wild-type CYP2D6 in morphine formation from codeine. In addition to the earlier mentioned preference of CYP2D6 for the metabolism of psychoactive drugs, and its association with serotonin metabolism, the latter study further supports the important role of CYP2D6 in influencing mood and mental health.

In the 1970s it was shown that debrisoquine, an antihypertensive drug, and sparteine, an anti-arrhythmic drug, were found to be inefficiently metabolized in some individuals, thus causing unexpected adverse reactions.[105,106] Later studies suggested that these and other drug metabolism defects were caused by the same genetic deficiency;[107-109] it was for the first time that genetic polymorphisms were related to drug metabolizing enzymes. In the 1980s, first rat debrisoquine hydroxylase[110] and later the responsible human enzyme[111] were purified and biochemically characterized and the genetic basis behind the debrisoquine polymorphism was elucidated. The CYPIID genetic locus was isolated[112] and sequenced[113,114] and defective alleles could be isolated and sequenced to characterize the molecular defect at the level of genomic DNA.[115] After establishment of the official cytochrome P450 nomenclature in 1993, the unique name CYP2D6 was given to the human debrisoquine hydroxylase.[55]

To date more than 80 mutant alleles of the CYP2D6 gene have been identified (see *http://www.imm.ki.se/CYPalleles/CYP2D6.htm*). The consequence of these different alleles is varying activity of the enzyme between different individuals (**Figure 7D, Table 3**). Phenotypes of the CYP2D6 activity, are divided into the extensive, poor, and ultrarapid metabolizers (EM, PM, and UM, respectively) according to the enzymatic activities (**Table 3**, **Figure 8**).[97]
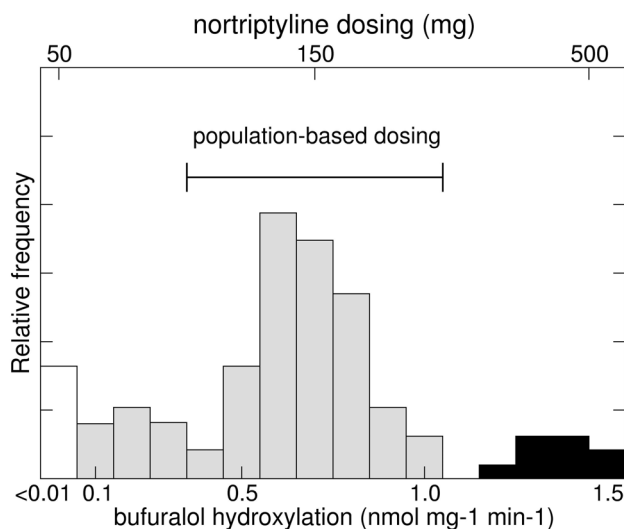
Figure 8: *Variation in drug metabolism and nortriptyline dosing in the European population, based on CYP2D6 activity (1′-hydroxylation of R-bufuralol). Within the population four phenotypes can be identified: poor metabolizers (white, PMs), extensive metabolizers (grey, EMs), and ultrarapid metabolizers (black, UMs), and optionally, in between PMs and EMs, intermediary metabolizers (grey, IMs) (see text). The doses of nortriptyline that are required to achieve therapeutic levels in all phenotypes are given. Despite this variation in metabolizing capability, population-based dosing is used today, and is based on the average plasma levels obtained in a given population for a given dose. Figure adapted from Ingelman-Sundberg.[11]*

variant	mutation	effect	PT	Allele frequency			
				Caucasians	Asians	Black Africans	Ethiopians + Saudi Arabians
CYP2D6*2xN	Gene duplication	Increased activity	UM	1-5	0-2	2	10-16
CYP2D6*4	Defective splicing	Inactive enzyme	PM	12-21	1	2	1-4
CYP2D6*5	Gene deletion	No enzyme	PM	4-6	6	4	1-3
CYP2D6*10	P34S, S486T	Unstable enzyme	PM/IM	1-2	50	6	3-9
CYP2D6*17	T107I, R296C, S486T	Reduced affinity for substrates	PM/IM	0	ND	34	3-9

Therapeutic class	Drug	Catalyzed reaction[a]	Phenotype[b]	Interacting drug[c]	Effect	mechanism[d]
analgesica/ antitussies	codeine	O-demethylation	PM		therapeutic failure	no active M
	ethyl-morphine	O-dealkylation	UM		side effects metabolite	M↑
	hydrocodone	O-dealkylation	UM	RTN	side effects metabolite	M↑
	oxycodone	O-demethylation	PM		therapeutic failure	no active M
			UM	RTN	side effects metabolite	M↑
	tramodol	O-demethylation	PM	RTN	therapeutic failure	no active M
antiarrythmics	encainide	O-demethylation	UM		arrhytmias	M↑
	flecainide	O-dealkylation	PM	FLX		CL↓ / accum.
	mexiletine	aromatic hydroxylation	PM	PRX,RTN		CL↓ / accum.
	propafenone	aromatic hydroxylation	PM	amiodarone,SRT,QND	vomiting	CL↓ / accum.
	sparteine	aliphatic hydroxylation	PM	QND		CL↓ / accum.
tricyclic antidepressants	amitriptyline	benzylic hydroxylation	PM	RTN	cardiotoxicity	CL↓ / accum.
	desipramine	aromatic hydroxylation	PM	CIM,FLX,PRX,SRT,QND,RTN	confusion	CL↓ / accum.
	imipramine	aliphatic hydroxylation	PM	CIM,FLX,PRX,SRT,QND,RTN	confusion	CL↓ / accum.
	nortriptiline	benzylic hydroxylation	PM	CIM,FLX,PRX,SRT,QND,terb.	confusion	CL↓ / accum.
			UM		therapeutic failure	CL↑

Therapeutic class	Drug	Catalyzed reaction[a]	Phenotype[b]	Interacting drug[c]	Effect	mechanism[d]
other	citalopram	N-dealkylation	PM		confusion	CL ↓ / accum.
antidepressants			UM		therapeutic failure	CL ↑
	fluoxetine	N-demethylation	PM	clarithromycin,RTN	confusion	CL ↓ / accum.
	fluvoxamine	unclear	PM		confusion	CL ↓ / accum.
	mianserin	aromatic hydroxylation	PM		confusion	CL ↓ / accum.
			UM		therapeutic failure	CL ↑
	mirtazapine	aromatic hydroxylation	UM		therapeutic failure	CL ↑
	paroxetine	demethylenation	UM		serotonine syndrome	CL ↑
	venlafaxine	O-demethylation	PM	dextromethorphan	confusion	CL ↓ / accum.
			UM	RTN	therapeutic failure	CL ↑
antidiabetics	phenformin	aromatic hydroxylation	PM		lactic acidosis	CL ↓ / accum.
antiemetic drugs	ondansetron	aromatic hydroxylation	UM		therapeutic failure	CL ↑
antiestrogens	tamoxifen	aromatic hydroxylation	PM	SSRIs	therapeutic failure	no active M
antihypertensive	debrisoquine	benzylic hydroxylation	PM	cimetidine,QND	enh. pharmacol. effect	CL ↓
anti-muscarinics	tolterodine	aliphatic hydroxylation	PM		therapeutic failure	CL ↓ / accum.
antipsychotics	haloperidol	N-dealkylation	PM	FLX,RTN	Parkinsonism	CL ↓ / accum.
			UM		therapeutic failure	CL ↑
	perphenazine	N-dealkylation	PM		oversedation	CL ↓ / accum.
	zuclopnethixol	N-dealkylation	PM		oversedation	CL ↓ / accum.
	thioridazine	sulfoxidation	PM		oversedation	CL ↓ / accum.
apetite suppress.	dexfenluramine	N-dealkylation	PM	FLX	nausea	CL ↓ / accum.
calcium antagon.	perhexilene	aliphatic hydroxylation	PM		hepatoxicity, neuropath	CL ↓ / accum.
beta adrenergic	metoprolol	aliphatic hydroxylation	PM	propafenone,RTN,SSRI,QND	bradicarida	CL ↓ / accum.
blocking agents	propranolol	aromatic hydroxylation	PM	QND		CL ↓ / accum.
	bufuralol	benzylic hydroxylation	PM			CL ↓ / accum.
	tropisetron	aromatic hydroxylation	UM		therapeutic failure	CL ↓ / accum.
recreat. drugs	MDMA	O- demethylenation	PM	PRX		CL ↓ / accum.

a) Examples of reactions schemes for CYP2D6 and other CYPs are shown in **Table 1**

b) Poor metabolisers (PM) who lack the functional enzyme, ultrarapid metabolisers (UM) who carry duplicated or multiduplicated functional CYP2D6 genes

c) Abbreviations for CYP2D6 inhibitors: cimetidine (CIM), fluoxetine (FLX), paroxetine (PRX), sertraline (SRT), quididine (QND), ritononavir (RTN), selective serotonin reuptake inhibitor (SSRI)

d) no active metabolite formed (no active M), increase (toxic) metabolite formation (M ↑), decreased metabolism/clearance causing drug accumulation (CL ↓/accum.), increased metabolism/clearance (CL ↑).

The PM phenotype, with no CYP2D6 being expressed, is caused by a gene deletion or single nucleotide polymorphism in the gene. In PM (~20% of Caucasian population and 5% of Asian population), CYP2D6 substrates are metabolised at a reduced rate and, therefore, higher plasma concentrations than expected are observed at conventional doses. In contrast, the UM phenotype, carrying additional copies of a functional CYP2D6 gene (1*xN or 2*xN allele, up to ~5% of the Caucasian and up to 15% of the Ethiopian population), shows an increased rate of drug metabolism and requires a higher dose than normally recommended to achieve a therapeutic plasma drug concentration.[99] In between PM and UM are the normal extensive metabolizing (EM) phenotype, with one or two functional genes and the intermediate metabolizing (IM) phenotype, characterized by at least one partially deficient allele.[96] Absent or reduced CYP2D6 activity can lead to ADRs by the same mechanisms described earlier in paragraph **1.2** and **Figure 2**. **Table 4** gives an overview of the clinical effect of genetic polymorphisms and drug-drug interactions on CYP2D6-mediated metabolism of selected drugs, belonging to specific classes.

1.6 Scope & Objectives

As outlined in this introduction, CYPs and particularly CYP2D6 play an important role in drug metabolism. CYP2D6 is the second most important drug metabolizing enzyme after CYP3A4. Large interindividual differences exist in CYP2D6 activity due to polymorphisms, further emphasizing its clinical importance. The early identification of potential CYP2D6 substrates and prediction of their metabolism is therefore advantageous in the discovery and development of new drugs. Computational chemistry is a very suitable tool to predict metabolic stability and sites of metabolism in potential drugs. *In silico* predictive database approaches and quantitative structure-activity relationships (QSARs) derived from physicochemical (1D) and topological (2D) properties of CYP ligands provide robust but rough classifications of the inhibition potency of ligands, without yielding much insight into the possible (molecular) background of the mode of action. The structure-based ('3D') molecular modeling methods, on the other hand, aim to be insightful as well as predictive[30] and were applied in this thesis to obtain novel knowledge on the structure and dynamics of CYP2D6-ligand interactions. Of all human CYPs the human CYP2D6 isoenzyme is a particularly suited model system to investigate the structure and dynamics of protein-ligand interactions. Apart from the intrinsic relevance of this isoenzyme, considerable computational research has been performed and offers a starting point for new modelling investigations. At the beginning of this research no experimentally determined structure was available, but several important ligand-protein interactions could be described by pharmacophore models.[115,117-124] From these models it became clear that CYP2D6 specifically binds compounds containing a basic nitrogen and an aromatic moiety. The structural specificity is not as pronounced as for e.g. CYP101, and not as diverse as for e.g. CYP3A4. Moreover, CYP2D6 has been shown to metabolise drug substrates at different sites, suggesting the existence of multiple binding modes. Therefore, at the start of these in investigations in 2002, CYP2D6 represented a manageable challenge to study ligand-protein interactions, from which general insights could be derived.

High-quality protein models

At the start of this study, several CYP2D6 protein homology models had already been constructed and successfully used to predict substrate and inhibitor selectivity and regiospecificity of metabolism.[50,60,117,122,125-129] The availability of crystal structures of stronger homologous CYP isoforms, and improvements of computational modelling techniques, create new possibilities to improve the predictive quality of these models. In the past few years, crystal structures of many mammalian CYPs have become available: CYP2A6,[83] CYP2B4,[84,130] CYP2C5,[131-133] CYP2C8,[85] CYP2C9,[86,87] and CYP3A4.[89,90] And very recently, during the final writing process of this thesis, a substrate-free crystal structure of CYP2D6 was solved.[88]

However, as was mentioned above, comparisons between ligand-bound and ligand-free crystal structures of many CYPs showed ligand induced changes in size, shape and hydration of the active site. The construction, refinement and validation of an accurate substrate-bound three-dimensional protein model of CYP2D6 is therefore still a crucial first stage in modeling CYP2D6-ligand interactions.

Protein structure-based virtual screening methods

At the start of this study, several classical ligand-based pharmacophore models of CYP2D6 had already been developed and optimized, facilitating the prediction of possible involvement of CYP2D6 in the metabolism of substrates as well as likely sites of catalysis.[115,117-124] Pharmacophore and 3D-QSAR models derive information on the active site of proteins from the shape, electronic properties and conformations of ligands (substrates, inhibitors or metabolic products) and are constructed by superimposing abstracted descriptors of essential functional groups and topologies of the ligands. In addition to 3D-QSAR models, successful QSAR models derived from physicochemical (1D)[32] and topological (2D)[35] ligand properties of CYP2D6 ligands had been reported as well. These *in silico* physico-chemical, topological and ligand structure-based models are suitable to make predictions at a very early stage of the drug-development process, in a high-throughput fashion and at low costs.[21] As was already mentioned, structure-based approaches offer more insight into the molecular background of the mode of action of drug candidates. Among structure-based approaches, ligand-based methods have two major weaknesses compared to protein-based methods: i) at least a partial solution of the molecular alignment of ligands is needed; ii) it is difficult to build accurate models based on ligands (and aimed at finding ligands) having different underlying chemotypes.[134]

Automated ligand-protein docking is the most commonly used protein-based virtual screening approach.[135,136] Automated docking algorithms generate energetically favorable conformations and orientations of ligands in the binding pocket of a protein and scoring functions are used to predict the tightness of the protein-ligand interactions. The accuracies of these predictions, however, are found to be highly dependent of the docking-scoring combination used, suggesting that a docking-scoring strategy should be specifically optimised for the system under study.[137-150] 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.[151,152] When a structural model of a CYP isoenzyme is available, docking methods can be used to predict the substrate catalytic site and binding affinity for CYP ligands (substrates and inhibitors), to identify potential CYP substrates in a chemical database (protein-based virtual screening), and to identify amino acid residues involved in substrate binding. The heme prosthetic group, the presence of water in the active site, the relatively large size and large conformational flexibility of the active site, the broad range of substrates and abundance of catalytic sites in most substrates, make CYPs rather challenging docking targets.

Although ligand-based and protein-based virtual screening can be equally successful and can be used in a complementary way,[24,134] protein-based virtual screening still has as an advantage over ligand-based virtual screening as it: i) visualises the actual binding mode of a ligand in the protein; ii) enables the identification of ligands with different binding modes and different binding modes for the same ligand, by addressing novel interaction sites that not have yet been used by known ligands.[24]

Molecular dynamics of protein-ligand interactions

As was discussed above, the few studies offering experimental information on the structures of CYP catalytic cycle intermediates suggest that once the substrate is bound, the conformation of CYPs does not change to a very large extent with the different stages.[63,65] The buried active site of many CYPs, however, implies that the protein must undergo motion

to allow substrates to enter the active site and products to leave the active site.[73] Indeed, conformational changes seem to occur during ligand binding, as comparisons between ligand-bound and ligand-free crystal structures of different CYP isoenzymes show ligand induced changes in size, shape and hydration of the active site.[29] This conformational flexibility of the active site is likely to underlie the capacity of many CYP isoenzymes to metabolize structurally diverse ligands of different sizes and to bind substrates in multiple binding modes (resulting in different metabolic products). Although docking methods that allow partial treatment of protein flexibility have recently been reported,[153-157] no automated ligand-protein docking methods that attempt to explicitly consider the full flexibililty of both the ligand and protein structure, have been validated extensively.[136,151] Furthermore, automated docking alone is not suitable for accurate determination of relative probabilities of different substrate binding modes and for discrimination between substrates and enzymes of high similarity, such as regio/stereoisomers and mutants of enzymes.

Molecular dynamics (MD) simulations describe the movements of atoms according to Newton's classical equations of motions and can give detailed insight in the behavior of molecular systems, in both space and time. To describe the dynamics of CYP2D6-ligand interactions and to take full protein and ligand flexibility into account, molecular dynamics simulations of CYP2D6-ligand complexes have to be performed. Performing multiple MD simulations starting from different ligand docking poses as starting structures might even further increase conformational sampling of the protein-ligand complex.

Aims of this thesis

The primary aim of the investigations described in this thesis was to obtain knowledge on the structure and dynamics of CYP2D6-ligand interactions, enabling the prediction and rationalization of drug metabolism by this specific Cytochrome P450 enzyme, and offering new approaches for the study of CYP-ligand interactions in general.

Our work particularly focused on three challenging aspects of CYP research:

1) The construction of a high-quality CYP2D6 homology model suitable for designing and rationalizing experimental studies.
2) The development of protein structure-based virtual screening methods for the prediction of CYP2D6-ligand interactions in a high-throughput fashion.
3) The description of the dynamics of CYP2D6-ligand interactions by using multiple experimental and in silico techniques in an integrative way.

To summarize, CYP2D6 homology models constructed from the recently reported more homologous and higher resolution crystal structures, finally are suitable enough to be subjected to more detailed analysis using molecular modelling methods like 3D-QSAR modelling, automated ligand-protein docking, and (long time) molecular dynamics simulations. In addition, the molecular modelling techniques mentioned have rapidly evolved over the past few years, and are considered valuable *in silico* tools for the optimization and evaluation of structural models and for the description of (the dynamics of) ligand-protein interactions. Another clear trend observed in many recent computational studies is the combination of several modelling techniques, to arrive at meaningful and successful rationalizations of experimental data and at interpretations of CYP function. Finally, computational chemistry is especially suitable to gain insight into the dynamics of ligand binding, but *in silico* models need to be validated using experimental techniques. In this thesis a basis is laid to study the dynamics of human CYP2D6 using novel computational approaches in conjunction with experimental research. The present studies have been performed within the framework of the Chemistry of Complex Molecules (CCM) research program at the Vrije Universiteit Amsterdam. The CCM program brings a number of chemical disciplines together in order to gain more insights in how complex molecular systems, like proteins, function in space and time. With a combined approach of quantum chemical

calculations, computer modeling, time resolved and static spectroscopy, enzymology, and mutagenesis studies this goal could be reached. This offered oppurtunities to use new experimental data to validate and optimize our *in silico* models and, vice versa, using our models to design and rationalize experimental studies, as well as to develop new integrative approaches for the molecular modelling of CYP2D6-ligand interactions.

1.7 General outline of this thesis

The research presented in this thesis describes the structure and dynamics of CYP2D6-ligand interactions using various *in silico* modelling techniques, in conjunction with *in vitro* experimental data, enabling the prediction and rationalization of drug metabolism mediated by this Cytochrome P450 enzyme. Three different lines of research correspond to the aims described in the previous paragraph: **1**) the construction and validation of a high-quality CYP2D6 homology model; **2**) The development of methods for the prediction of CYP-ligand interactions in a high-throughput fashion; **3**) The description of the dynamics of CYP-ligand interactions by using multiple different experimental and in silico techniques in an integrative way. These three lines are addressed in the different chapters of this thesis.

Chapter 1 gives a general introduction into the background and context of the research presented in the thesis and into the model enzyme system investigated: human Cytochrome P450 2D6. It deals with the clinical impact of drug metabolism, the role of the prediction of metabolism in drug discovery and development, the use of computational chemistry in drug metabolism studies and the catalytic activity and clinical significance of the human CYP2D6 isoenzyme. **Chapter 2** reviews in an integrative manner the computational approaches used to understand, rationalize, and predict the activity and substrate selectivity of CYPs, as well as the possibilities and limitations of these approaches now and in the future.

Chapter 3 focuses on research line **1** and describes the construction and optimization of the CYP2D6 homology model, its comparison to the very recently solved CYP2D6 crystal structure, and its use to design experimental site-directed mutation studies to probe the enzyme active site and to validate the molecular modelling work.

The next two chapters cover research line **2**. Automated docking strategies, considering active-site water molecules and rescoring, are evaluated and optimized with respect to their capacity to reproduce the binding modes of X-ray crystal structures of Cytochrome P450 and thymidine kinase protein-ligand complexes and to predict the sites of catalysis of the different ligands (**Chapter 4**). The same docking strategies are subsequently used for the catalytic site prediction and virtual screening of substrates in our CYP2D6 homology model (**Chapter 5**).

In the next three chapters, focusing on research line **3**, a combination of automated docking and molecular dynamics is used to predict the regio- and stereo-specificity of catalysis of MDMA and MDMA-analogs (**Chapter 6**) and the relative binding affinity of propranolol stereo-isomers (**Chapter 7**) towards wild-type and mutant CYP2D6, and to explore substrate access and product exit channels in CYP2D6 (**Chapter 8**).

The final **Chapter 9** summarizes the overall conclusions of the investigations desribed in this thesis and provides an outlook for future computational research to understand, rationalize and predict Cytochrome P450 structure, function, and dynamics.

References

(1) 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.

(2) Sheweita, S. A. Drug-metabolizing enzymes: Mechanisms and functions. *Current Drug Metabolism* **2000**, *1*, 107-132.

(3) Pirmohamed, M.; Park, B. K. Cytochrome P450 enzyme polymorphisms and adverse drug reactions. *Toxicology* **2003**, *192*, 23-32.

(4) Pirmohamed, M.; Park, B. K. Genetic susceptibility to adverse drug reactions. *Trends Pharmacol Sci* **2001**, *22*, 298-305.

(5)   Ariens, E. J.; Simonis, A. M. *In: Towards a better safety of drugs and pharmaceutical products*; Elsevier Biomedical Press, Amsterdam, 1980.

(6)   Michalets, E. L. Update: clinically significant cytochrome P-450 drug interactions. *Pharmacotherapy* **1998**, *18*, 84-112.

(7)   Thompson, M. D.; Burnham, W. M.; Cole, D. E. The G protein-coupled receptors: pharmacogenetics and disease. *Crit Rev Clin Lab Sci* **2005**, *42*, 311-392.

(8)   Marshall, A. Getting the right drug into the right patient. *Nat Biotechnol* **1997**, *15*, 1249-1252.

(9)   Lazarou, J.; Pomeranz, B. H.; Corey, P. N. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *Jama* **1998**, *279*, 1200-1205.

(10)   Bates, D. W.; Spell, N.; Cullen, D. J.; Burdick, E.; Laird, N. et al. The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group. *Jama* **1997**, *277*, 307-311.

(11)   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.

(12)   Spear, B. B.; Heath-Chiozzi, M.; Huff, J. Clinical application of pharmacogenetics. *Trends Mol Med* **2001**, *7*, 201-204.

(13)   Phillips, K. A.; Veenstra, D. L.; Oren, E.; Lee, J. K.; Sadee, W. Potential role of pharmacogenomics in reducing adverse drug reactions: a systematic review. *Jama* **2001**, *286*, 2270-2279.

(14)   Gasche, Y.; Daali, Y.; Fathi, M.; Chiappe, A.; Cottini, S. et al. Codeine intoxication associated with ultrarapid CYP2D6 metabolism. *N Engl J Med* **2004**, *351*, 2827-2831.

(15)   Dayer, P.; Desmeules, J.; Leemann, T.; Striberni, R. Bioactivation of the narcotic drug codeine in human liver is mediated by the polymorphic monooxygenase catalyzing debrisoquine 4-hydroxylation (cytochrome P-450 dbl/bufI). *Biochem Biophys Res Commun* **1988**, *152*, 411-416.

(16)   Oprea, T. L. Virtual screening in lead discovery: A viewpoint. *Molecules* **2002**, *7*, 51-62.

(17)   Dimasi, J. A. Risks in new drug development: approval success rates for investigational drugs. *Clin Pharmacol Ther* **2001**, *69*, 297-307.

(18)   Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **2001**, *46*, 3-26.

(19)   van de Waterbeemd, H.; Gifford, E. ADMET in silico modelling: towards prediction paradise? *Nat Rev Drug Discov* **2003**, *2*, 192-204.

(20)   Selick, H. E.; Beresford, A. P.; Tarbit, M. H. The emerging importance of predictive ADME simulation in drug discovery. *Drug Discov Today* **2002**, *7*, 109-116.

(21)   Oprea, T. I.; Matter, H. Integrating virtual screening in lead discovery. *Curr Opin Chem Biol* **2004**, *8*, 349-358.

(22)   Hou, T. J.; Xu, X. J. Recent development and application of virtual screening in drug discovery: An overview. *Current Pharmaceutical Design* **2004**, *10*, 1011-1033.

(23)   Ekins, S.; Waller, C. L.; Swaan, P. W.; Cruciani, G.; Wrighton, S. A. et al. Progress in predicting human ADME parameters in silico. *J Pharmacol Toxicol Methods* **2000**, *44*, 251-272.

(24)   Evers, A.; Hessler, G.; Matter, H.; Klabunde, T. Virtual screening of biogenic amine-binding G-protein coupled receptors: comparative evaluation of protein- and ligand-based virtual screening protocols. *J Med Chem* **2005**, *48*, 5448-5465.

(25)   Ekins, S.; Nikolsky, Y.; Nikolskaya, T. Techniques: application of systems biology to absorption, distribution, metabolism, excretion and toxicity. *Trends Pharmacol Sci* **2005**, *26*, 202-209.

(26)   Rowley, M.; Hallett, D. J.; Goodacre, S.; Moyes, C.; Crawforth, J. et al. 3-(4-Fluoropiperidin-3-yl)-2-phenylindoles as high affinity, selective, and orally bioavailable h5-HT(2A) receptor antagonists. *J Med Chem* **2001**, *44*, 1603-1614.

(27)   Cruciani, G.; Carosati, E.; De Boeck, B.; Ethirajulu, K.; Mackie, C. et al. MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. *J Med Chem* **2005**, *48*, 6970-6979.

(28)   Deng, B. L.; Hartman, T. L.; Buckheit, R. W., Jr.; Pannecouque, C.; De Clercq, E. et al. Synthesis, anti-HIV activity, and metabolic stability of new alkenyldiarylmethane HIV-1 non-nucleoside reverse transcriptase inhibitors. *J Med Chem* **2005**, *48*, 6140-6155.

(29)   de Graaf, C.; Vermeulen, N. P.; Feenstra, K. A. Cytochrome p450 in silico: an integrative modeling approach. *J Med Chem* **2005**, *48*, 2725-2755.

(30)   Congreve, M.; Murray, C. W.; Blundell, T. L. Structural biology and drug discovery. *Drug Discov Today* **2005**, *10*, 895-907.

(31)   Langowski, J.; Long, A. Computer systems for the prediction of xenobiotic metabolism. *Adv Drug Deliv Rev* **2002**, *54*, 407-415.

(32)   Hansch, C.; Mekapati, S. B.; Kurup, A.; Verma, R. P. QSAR of cytochrome P450. *Drug Metab Rev* **2004**, *36*, 105-156.

(33)   Molnar, L.; Keseru, G. M. A neural network based virtual screening of cytochrome P450 3A4 inhibitors. *Bioorg Med Chem Lett* **2002**, *12*, 419-421.

(34)   Zuegge, J.; Fechner, U.; Roche, O.; Parrott, N. J.; Engkvist, O. et al. A fast virtual screening filter for cytochrome P450 3A4 inhibition liability of compound libraries. *Quantitative Structure-Activity Relationships* **2002**, *21*, 249-256.

(35)   Susnow, R. G.; Dixon, S. L. Use of robust classification techniques for the prediction of human cytochrome P450 2D6 inhibition. *Journal of Chemical Information and Computer Sciences* **2003**, *43*, 1308-1315.

(36)   O'Brien, S. E.; de Groot, M. J. Greater than the sum of its parts: combining models for useful ADMET prediction. *J Med Chem* **2005**, *48*, 1287-1291.

(37)   Lewis, D. F.; Lake, B. G.; Bird, M. G. Molecular modelling of human microsomal epoxide hydrolase (EH) by homology with a fungal (Aspergillus niger) EH crystal structure of 1.8 A resolution: structure-activity relationships in epoxides inhibiting EH activity. *Toxicol In Vitro* **2005**, *19*, 517-522.

(38)   Yang, Z. N.; Bosron, W. F.; Hurley, T. D. Structure of human chi chi alcohol dehydrogenase: a glutathione-dependent formaldehyde dehydrogenase. *J Mol Biol* **1997**, *265*, 330-343.

(39)   Wierdl, M.; Morton, C. L.; Nguyen, N. K.; Redinbo, M. R.; Potter, P. M. Molecular modeling of CPT-11 metabolism by carboxylesterases (CEs): use of pnb CE as a model. *Biochemistry* **2004**, *43*, 1874-1882.

(40)   Gamage, N. U.; Duggleby, R. G.; Barnett, A. C.; Tresillian, M.; Latham, C. F. et al. Structure of a human carcinogen-converting enzyme, SULT1A1. Structural and kinetic implications of substrate inhibition. *J Biol Chem* **2003**, *278*, 7655-7662.

(41)   Chen, D.; Wang, C. Y.; Lambert, J. D.; Ai, N.; Welsh, W. J. et al. Inhibition of human liver catechol-O-methyltransferase by tea catechins and their metabolites: structure-activity relationship and molecular-modeling studies. *Biochem Pharmacol* **2005**, *69*, 1523-1531.

(42)   Palma, P. N.; Bonifacio, M. J.; Loureiro, A. I.; Wright, L. C.; Learmonth, D. A. et al. Molecular modeling and metabolic studies of the interaction of catechol-O-methyltransferase and a new nitrocatechol inhibitor. *Drug Metab Dispos* **2003**, *31*, 250-258.

(43)   Ridder, L.; Rietjens, I. M.; Vervoort, J.; Mulholland, A. J. Quantum mechanical/molecular mechanical free energy simulations of the glutathione S-transferase (M1-1) reaction with phenanthrene 9,10-oxide. *J Am Chem Soc* **2002**, *124*, 9926-9936.

(44)   Hu, X.; O'Donnell, R.; Srivastava, S. K.; Xia, H.; Zimniak, P. et al. Active site architecture of polymorphic forms of human glutathione S-transferase P1-1 accounts for their enantioselectivity and disparate activity in the glutathione conjugation of 7beta,8alpha-dihydroxy-9alpha,10alpha-ox y-7,8,9,10-tetrahydrobenzo(a)pyrene. *Biochem Biophys Res Commun* **1997**, *235*, 424-428.

(45)   Ouzzine, M.; Gulberti, S.; Levoin, N.; Netter, P.; Magdalou, J. et al. The donor substrate specificity of the human beta 1,3-glucuronosyltransferase I toward UDP-glucuronic acid is determined by two crucial histidine and arginine residues. *J Biol Chem* **2002**, *277*, 25439-25445.

(46)   Rodrigues-Lima, F.; Delomenie, C.; Goodfellow, G. H.; Grant, D. M.; Dupret, J. M. Homology modelling and structural analysis of human arylamine N-acetyltransferase NAT1: evidence for the conservation of a cysteine protease catalytic domain and an active-site loop. *Biochem J* **2001**, *356*, 327-334.

(47)   Battaglia, E.; Elass, A.; Drake, R. R.; Paul, P.; Treat, S. et al. Characterization of a new class of inhibitors of the recombinant human liver UDP-glucuronosyltransferase, UGT1*6. *Biochim Biophys Acta* **1995**, *1243*, 9-14.

(48)   Miners, J. O.; Smith, P. A.; Sorich, M. J.; McKinnon, R. A.; Mackenzie, P. I. Predicting human drug glucuronidation parameters: application of in vitro and in silico modeling approaches. *Annu Rev Pharmacol Toxicol* **2004**, *44*, 1-25.

(49)   Takenaga, N.; Ishii, M.; Kamei, T.; Yasumori, T. Structure-activity relationship in O-glucuronidation of indolocarbazole analogs. *Drug Metab Dispos* **2002**, *30*, 494-497.

(50)   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.

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

(52)   Goeptar, A. R.; Scheerens, H.; Vermeulen, N. P. Oxygen and xenobiotic reductase activities of cytochrome P450. *Crit Rev Toxicol* **1995**, *25*, 25-65.

(53)   Estabrook, R. W. A passion for P450s (rememberances of the early history of research on cytochrome P450). *Drug Metab Dispos* **2003**, *31*, 1461-1473.

(54)   Coon, M. J. Cytochrome P450: nature's most versatile biological catalyst. *Annu Rev Pharmacol Toxicol* **2005**, *45*, 1-25.

(55)   Nelson, D. R.; Kamataki, T.; Waxman, D. J.; Guengerich, F. P.; Estabrook, R. W. et al. The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA Cell Biol* **1993**, *12*, 1-51.

(56)   Hanna, I. H.; Kim, M. S.; Guengerich, F. P. Heterologous expression of cytochrome P450 2D6 mutants, electron transfer, and catalysis of bufuralol hydroxylation: the role of aspartate 301 in structural integrity. *Arch Biochem Biophys* **2001**, *393*, 255-261.

(57)   Groves, J. T. Key Elements of the Chemistry of Cytochrome-P-450 - the Oxygen Rebound Mechanism. *Journal of Chemical Education* **1985**, *62*, 928-931.

32

(58)   Newcomb, M.; Shen, R.; Choi, S. Y.; Toy, P. H.; Hollenberg, P. F. et al. Cytochrome P450-catalyzed hydroxylation of mechanistic probes that distinguish between radicals and cations. Evidence for cationic but not for radical intermediates. *Journal of the American Chemical Society* **2000**, *122*, 2677-2686.

(59)   Nagano, S.; Cupp-Vickery, J. R.; Poulos, T. L. Crystal structures of the ferrous dioxygen complex of wild-type cytochrome P450eryF and its mutants, A245S and A245T: investigation of the proton transfer system in P450eryF. *J Biol Chem* **2005**, *280*, 22102-22107.

(60)   Modi, S.; Paine, M. J.; Sutcliffe, M. J.; Lian, L. Y.; Primrose, W. U. et al. A model for human cytochrome P450 2D6 based on homology modeling and NMR studies of substrate binding. *Biochemistry* **1996**, *35*, 4540-4550.

(61)   Rendic, S. Summary of information on human CYP enzymes: human P450 metabolism data. *Drug Metab Rev* **2002**, *34*, 83-448.

(62)   Poulos, T. L. Intermediates in P450 catalysis. *Philos Transact A Math Phys Eng Sci* **2005**, *363*, 793-806; discussion 1035-1040.

(63)   Nagano, S.; Cupp-Vickery, J. R.; Poulos, T. L. Crystal structures of the ferrous dioxygen complex of wild-type cytochrome P450eryF and its mutants, A245S and A245T: Investigation of the proton transfer system in P450ERYF. *J Biol Chem* **2005**.

(64)   Poulos, T. L.; Finzel, B. C.; Howard, A. J. Crystal structure of substrate-free Pseudomonas putida cytochrome P-450. *Biochemistry* **1986**, *25*, 5314-5322.

(65)   Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A. M.; Maves, S. A. et al. The catalytic pathway of cytochrome p450cam at atomic resolution. *Science* **2000**, *287*, 1615-1622.

(66)   Li, H. Y.; Narasimhulu, S.; Havran, L. M.; Winkler, J. D.; Poulos, T. L. Crystal-Structure of Cytochrome P450(Cam) Complexed with Its Catalytic Product, 5-Exo-Hydroxycamphor. *Journal of the American Chemical Society* **1995**, *117*, 6297-6299.

(67)   Harris, D. L.; Loew, G. H. Investigation of the proton-assisted pathway to formation of the catalytically active, ferryl species of P450s by molecular dynamics studies of P450eryF. *J Am Chem Soc* **1996**, *118*, 6377-6387.

(68)   Bathelt, C. M.; Zurek, J.; Mulholland, A. J.; Harvey, J. N. Electronic structure of compound I in human isoforms of cytochrome P450 from QM/MM modeling. *J Am Chem Soc* **2005**, *127*, 12900-12908.

(69)   Harris, D. L.; Park, J. Y.; Gruenke, L.; Waskell, L. Theoretical study of the ligand-CYP2B4 complexes: effect of structure on binding free energies and heme spin state. *Proteins* **2004**, *55*, 895-914.

(70)   Swart, M.; Groenhof, A. R.; Ehlers, A. W.; Lammertsma, K. QM/MM study on the catalytic cycle of cytochrome P450: The importance of selecting an appropriate density functional theory functional. *Abstracts of Papers of the American Chemical Society* **2004**, *227*, U1449-U1449.

(71)   Friesner, R. A.; Guallar, V. Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Annu Rev Phys Chem* **2005**, *56*, 389-427.

(72)   de Groot, M. J.; Kirton, S. B.; Sutcliffe, M. J. In silico methods for predicting ligand binding determinants of cytochromes P450. *Curr Top Med Chem* **2004**, *4*, 1803-1824.

(73)   Wade, R. C.; Winn, P. J.; Schlichting, I.; Sudarko A survey of active site access channels in cytochromes P450. *J Inorg Biochem* **2004**, *98*, 1175-1182.

(74)   Bertz, R. J.; Granneman, G. R. Use of in vitro and in vivo data to estimate the likelihood of metabolic pharmacokinetic interactions. *Clin Pharmacokinet* **1997**, *32*, 210-258.

(75)   Williams, J. A.; Hyland, R.; Jones, B. C.; Smith, D. A.; Hurst, S. et al. Drug-drug interactions for UDP-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios. *Drug Metab Dispos* **2004**, *32*, 1201-1208.

(76)   Evans, W. E.; Relling, M. V. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* **1999**, *286*, 487-491.

(77)   Ingelman-Sundberg, M.; Oscarson, M.; McLellan, R. A. Polymorphic human cytochrome P450 enzymes: an opportunity for individualized drug treatment. *Trends Pharmacol Sci* **1999**, *20*, 342-349.

(78)   Ingelman-Sundberg, M. The human genome project and novel aspects of cytochrome P450 research. *Toxicol Appl Pharmacol* **2005**, *207*, 52-56.

(79)   Guengerich, F. P.; Wheeler, J. B.; Chun, Y. J.; Kim, D.; Shimada, T. et al. Use of heterologously-expressed cytochrome P450 and glutathione transferase enzymes in toxicity assays. *Toxicology* **2002**, *181-182*, 261-264.

(80)   Kriegl, J. M.; Eriksson, L.; Arnhold, T.; Beck, B.; Johansson, E. et al. Multivariate modeling of cytochrome P450 3A4 inhibition. *Eur J Pharm Sci* **2005**, *24*, 451-463.

(81)   Lewis, D. F. On the recognition of mammalian microsomal cytochrome P450 substrates and their characteristics: towards the prediction of human p450 substrate specificity and metabolism. *Biochem Pharmacol* **2000**, *60*, 293-306.

(82)   Smith, D. A.; Ackland, M. J.; Jones, B. C. Properties of cytochrome P450 isoenzymes and their substrates .2. properties of cytochrome P450 substrates. *Drug Discovery Today* **1997**, *2*, 479-486.

(83)   Yano, J. K.; Hsu, M. H.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen. *Nat Struct Mol Biol* **2005**, *12*, 822-823.

(84)   Scott, E. E.; White, M. A.; He, Y. A.; Johnson, E. F.; Stout, C. D. et al. Structure of mammalian cytochrome P450 2B4 complexed with 4-(4-chlorophenyl)imidazole at 1.9-A resolution: insight into the range of P450 conformations and the coordination of redox partner binding. *J Biol Chem* **2004**, *279*, 27294-27301.

(85)   Schoch, G. A.; Yano, J. K.; Wester, M. R.; Griffin, K. J.; Stout, C. D. et al. Structure of human microsomal cytochrome P450 2C8. Evidence for a peripheral fatty acid binding site. *J Biol Chem* **2004**, *279*, 9497-9503.

(86)   Williams, P. A.; Cosme, J.; Ward, A.; Angove, H. C.; Matak Vinkovic, D. et al. Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* **2003**, *424*, 464-468.

(87)   Wester, M. R.; Yano, J. K.; Schoch, G. A.; Yang, C.; Griffin, K. J. et al. The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-A resolution. *J Biol Chem* **2004**, *279*, 35630-35637.

(88)   Rowland, P.; Blaney, F. E.; Smyth, M. G.; Jones, J. J.; Leydon, V. R. et al. Crystal structure of human cytochrome P450 2D6. *J Biol Chem* **2006**, *281*, 7614-7622.

(89)   Williams, P. A.; Cosme, J.; Vinkovic, D. M.; Ward, A.; Angove, H. C. et al. Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* **2004**, *305*, 683-686.

(90)   Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D. et al. The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-A resolution. *J Biol Chem* **2004**, *279*, 38091-38094.

(91)   Olesen, O. V.; Linnet, K. Contributions of five human cytochrome P450 isoforms to the N-demethylation of clozapine in vitro at low and high concentrations. *J Clin Pharmacol* **2001**, *41*, 823-832.

(92)   Yang, T. J.; Krausz, K. W.; Sai, Y.; Gonzalez, F. J.; Gelboin, H. V. Eight inhibitory monoclonal antibodies define the role of individual P-450s in human liver microsomal diazepam, 7-ethoxycoumarin, and imipramine metabolism. *Drug Metab Dispos* **1999**, *27*, 102-109.

(93)   Kreth, K.; Kovar, K.; Schwab, M.; Zanger, U. M. Identification of the human cytochromes P450 involved in the oxidative metabolism of "Ecstasy"-related designer drugs. *Biochem Pharmacol* **2000**, *59*, 1563-1571.

(94)   Wormhoudt, L. W.; Commandeur, J. N.; Vermeulen, N. P. Genetic polymorphisms of human N-acetyltransferase, cytochrome P450, glutathione-S-transferase, and epoxide hydrolase enzymes: relevance to xenobiotic metabolism and toxicity. *Crit Rev Toxicol* **1999**, *29*, 59-124.

(95)   Oscarson, M.; Ingelman-Sundberg, M. CYPalleles: a web page for nomenclature of human cytochrome P450 alleles. *Drug Metab Pharmacokinet* **2002**, *17*, 491-495.

(96)   Zanger, U. M.; Raimundo, S.; Eichelbaum, M. Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. *Naunyn Schmiedebergs Arch Pharmacol* **2004**, *369*, 23-37.

(97)   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.

(98)   Oscarson, M. Pharmacogenetics of drug metabolising enzymes: importance for personalised medicine. *Clin Chem Lab Med* **2003**, *41*, 573-580.

(99)   Ingelman-Sundberg, M. Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity. *Pharmacogenomics J* **2005**, *5*, 6-13.

(100)   Yue, Q. Y.; Alm, C.; Svensson, J. O.; Sawe, J. Quantification of the O- and N-demethylated and the glucuronidated metabolites of codeine relative to the debrisoquine metabolic ratio in urine in ultrarapid, rapid, and poor debrisoquine hydroxylators. *Ther Drug Monit* **1997**, *19*, 539-542.

(101)   Board, P. G.; Coggan, M.; Watson, S.; Gage, P. W.; Dulhunty, A. F. CLIC-2 modulates cardiac ryanodine receptor Ca2+ release channels. *Int J Biochem Cell Biol* **2004**, *36*, 1599-1612.

(102)   Prueksaritanont, T.; Dwyer, L. M.; Cribb, A. E. (+)-bufuralol 1'-hydroxylation activity in human and rhesus monkey intestine and liver. *Biochem Pharmacol* **1995**, *50*, 1521-1525.

(103)   Siegle, I.; Fritz, P.; Eckhardt, K.; Zanger, U. M.; Eichelbaum, M. Cellular localization and regional distribution of CYP2D6 mRNA and protein expression in human brain. *Pharmacogenetics* **2001**, *11*, 237-245.

(104)   Pai, H. V.; Kommaddi, R. P.; Chinta, S. J.; Mori, T.; Boyd, M. R. et al. A frameshift mutation and alternate splicing in human brain generate a functional form of the pseudogene cytochrome P4502D7 that demethylates codeine to morphine. *J Biol Chem* **2004**, *279*, 27383-27389.

(105)   Mahgoub, A.; Idle, J. R.; Smith, R. L. Genetically determined variability in drug metabolism: dual slow acetylation and drug oxidation traits. *Lancet* **1979**, *2*, 154.

(106)   Eichelbaum, M.; Spannbrucker, N.; Dengler, H. J. Influence of the defective metabolism of sparteine on its pharmacokinetics. *Eur J Clin Pharmacol* **1979**, *16*, 189-194.

(107)   Bertilsson, L.; Eichelbaum, M.; Mellstrom, B.; Sawe, J.; Schulz, H. U. et al. Nortriptyline and antipyrine clearance in relation to debrisoquine hydroxylation in man. *Life Sci* **1980**, *27*, 1673-1677.

(108)   Eichelbaum, M.; Bertilsson, L.; Sawe, J.; Zekorn, C. Polymorphic oxidation of sparteine and debrisoquine: related pharmacogenetic entities. *Clin Pharmacol Ther* **1982**, *31*, 184-186.

(109)   Kleinbloesem, C. H.; van Brummelen, P.; Faber, H.; Danhof, M.; Vermeulen, N. P. et al. Variability in nifedipine pharmacokinetics and dynamics: a new oxidation polymorphism in man. *Biochem Pharmacol* **1984**, *33*, 3721-3724.

(110) Larrey, D.; Distlerath, L. M.; Dannan, G. A.; Wilkinson, G. R.; Guengerich, F. P. Purification and characterization of the rat liver microsomal cytochrome P-450 involved in the 4-hydroxylation of debrisoquine, a prototype for genetic variation in oxidative drug metabolism. *Biochemistry* **1984**, *23*, 2787-2795.

(111) Distlerath, L. M.; Reilly, P. E.; Martin, M. V.; Davis, G. G.; Wilkinson, G. R. et al. Purification and characterization of the human liver cytochromes P-450 involved in debrisoquine 4-hydroxylation and phenacetin O-deethylation, two prototypes for genetic polymorphism in oxidative drug metabolism. *J Biol Chem* **1985**, *260*, 9057-9067.

(112) Eichelbaum, M.; Baur, M. P.; Dengler, H. J.; Osikowska-Evers, B. O.; Tieves, G. et al. Chromosomal assignment of human cytochrome P-450 (debrisoquine/sparteine type) to chromosome 22. *Br J Clin Pharmacol* **1987**, *23*, 455-458.

(113) Gonzalez, F. J.; Vilbois, F.; Hardwick, J. P.; McBride, O. W.; Nebert, D. W. et al. Human debrisoquine 4-hydroxylase (P450IID1): cDNA and deduced amino acid sequence and assignment of the CYP2D locus to chromosome 22. *Genomics* **1988**, *2*, 174-179.

(114) Kimura, S.; Umeno, M.; Skoda, R. C.; Meyer, U. A.; Gonzalez, F. J. The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and identification of the polymorphic CYP2D6 gene, a related gene, and a pseudogene. *Am J Hum Genet* **1989**, *45*, 889-904.

(115) Meyer, U. A.; Gut, J.; Kronbach, T.; Skoda, C.; Meier, U. T. et al. The molecular mechanisms of two common polymorphisms of drug oxidation--evidence for functional changes in cytochrome P-450 isozymes catalysing bufuralol and mephenytoin oxidation. *Xenobiotica* **1986**, *16*, 449-464.

(116) Cascorbi, I. Genetic basis of toxic reactions to drugs and chemicals. *Toxicol Lett* **2005**.

(117) de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A.; Jones, B. C. Novel approach to predicting P450-mediated drug metabolism: Development of a combined protein and pharmacophore model for CYP2D6. *Journal of Medicinal Chemistry* **1999**, *42*, 1515-1524.

(118) Strobl, G. R.; von Kruedener, S.; Stockigt, J.; Guengerich, F. P.; Wolff, T. Development of a pharmacophore for inhibition of human liver cytochrome P-450 2D6: molecular modeling and inhibition studies. *J Med Chem* **1993**, *36*, 1136-1145.

(119) Islam, S. A.; Wolf, C. R.; Lennard, M. S.; Sternberg, M. J. A three-dimensional molecular template for substrates of human cytochrome P450 involved in debrisoquine 4-hydroxylation. *Carcinogenesis* **1991**, *12*, 2211-2219.

(120) Koymans, L.; Vermeulen, N. P.; van Acker, S. A.; te Koppele, J. M.; Heykants, J. J. et al. A predictive model for substrates of cytochrome P450-debrisoquine (2D6). *Chem Res Toxicol* **1992**, *5*, 211-219.

(121) Wolff, T.; Distelrath, L. M.; Worthington, M. T.; Groopman, J. D.; Hammons, G. J. et al. Substrate specificity of human liver cytochrome P-450 debrisoquine 4-hydroxylase probed using immunochemical inhibition and chemical modeling. *Cancer Res* **1985**, *45*, 2116-2122.

(122) de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A.; Jones, B. C. A novel approach to predicting P450 mediated drug metabolism. CYP2D6 catalyzed N-dealkylation reactions and qualitative metabolite predictions using a combined protein and pharmacophore model for CYP2D6. *Journal of Medicinal Chemistry* **1999**, *42*, 4062-4070.

(123) de Groot, M. J.; Bijloo, G. J.; Martens, B. J.; van Acker, F. A.; Vermeulen, N. P. A refined substrate model for human cytochrome P450 2D6. *Chem Res Toxicol* **1997**, *10*, 41-48.

(124) de Groot, M. J.; Bijloo, G. J.; van Acker, F. A.; Fonseca Guerra, C.; Snijders, J. G. et al. Extension of a predictive substrate model for human cytochrome P4502D6. *Xenobiotica* **1997**, *27*, 357-368.

(125) Koymans, L. M.; Vermeulen, N. P.; Baarslag, A.; Donne-Op den Kelder, G. M. A preliminary 3D model for cytochrome P450 2D6 constructed by homology model building. *J Comput Aided Mol Des* **1993**, *7*, 281-289.

(126) De Rienzo, F.; Fanelli, F.; Menziani, M. C.; De Benedetti, P. G. Theoretical investigation of substrate specificity for cytochromes P450 IA2, P450 IID6 and P450 IIIA4. *J Comput Aided Mol Des* **2000**, *14*, 93-116.

(127) 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.

(128) Snyder, R.; Sangar, R.; Wang, J. B.; Ekins, S. Three-dimensional quantitative structure activity relationship for CYP2D6 substrates. *Quantitative Structure-Activity Relationships* **2002**, *21*, 357-368.

(129) de Groot, M. J.; Vermeulen, N. P.; Kramer, J. D.; van Acker, F. A.; Donne-Op den Kelder, G. M. A three-dimensional protein model for human cytochrome P450 2D6 based on the crystal structures of P450 101, P450 102, and P450 108. *Chem Res Toxicol* **1996**, *9*, 1079-1091.

(130) Scott, E. E.; He, Y. A.; Wester, M. R.; White, M. A.; Chin, C. C. et al. An open conformation of mammalian cytochrome P450 2B4 at 1.6-A resolution. *Proc Natl Acad Sci U S A* **2003**, *100*, 13196-13201.

(131) Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dansette, P. M.; Mansuy, D. et al. Structure of a substrate complex of mammalian cytochrome P450 2C5 at 2.3 A resolution: evidence for multiple substrate binding modes. *Biochemistry* **2003**, *42*, 6370-6379.

(132) Williams, P. A.; Cosme, J.; Sridhar, V.; Johnson, E. F.; McRee, D. E. Microsomal cytochrome P450 2C5: comparison to microbial P450s and unique features. *J Inorg Biochem* **2000**, *81*, 183-190.

(133) Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dijols, S.; Dansette, P. M. et al. Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 A resolution: evidence for an induced fit model of substrate binding. *Biochemistry* **2003**, *42*, 9335-9345.

(134) Jain, A. N. Virtual screening in lead discovery and optimization. *Curr Opin Drug Discov Devel* **2004**, *7*, 396-403.

(135) 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.

(136) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* **2002**, *47*, 409-443.

(137) 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.

(138) Bursulaya, B. D.; Totrov, M.; Abagyan, R.; Brooks, C. L., 3rd Comparative study of several algorithms for flexible ligand docking. *J Comput Aided Mol Des* **2003**, *17*, 755-763.

(139) 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.

(140) 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.

(141) 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.

(142) 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.

(143) Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of docking performance: comparative data on docking algorithms. *J Med Chem* **2004**, *47*, 558-565.

(144) Kontoyianni, M.; Sokol, G. S.; McClellan, L. M. Evaluation of library ranking efficacy in virtual screening. *J Comput Chem* **2005**, *26*, 11-22.

(145) 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.

(146) 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.

(147) 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.

(148) 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.

(149) 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.

(150) 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.

(151) McConkey, B. J.; Sobolev, V.; Edelman, M. The performance of current methods in ligand-protein docking. *Current Science* **2002**, *83*, 845-856.

(152) 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.

(153) Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: Efficient molecular docking considering protein structure variations. *Journal of Molecular Biology* **2001**, *308*, 377-395.

(154) Kairys, V.; Gilson, M. K. Enhanced docking with the mining minima optimizer: Acceleration and side-chain flexibility. *Journal of Computational Chemistry* **2002**, *23*, 1656-1670.

(155) 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.

(156) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. FDS: flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. *J Comput Chem* **2003**, *24*, 1637-1656.

(157) Vigers, G. P.; Rizzi, J. P. Multiple active site corrections for docking and virtual screening. *J Med Chem* **2004**, *47*, 80-89.

CYP in silico

Cytochrome P450 in silico: An integrative molecular modeling approach

Chris de Graaf, Nico P.E. Vermeulen, and K. Anton Feenstra**

Leiden Amsterdam Center for Drug Research (LACDR)/Division of Molecular Toxicology, Department of Chemistry and Pharmacochemistry, Vrije Universiteit, The Netherlands
* Both authors contributed equally

The current chapter presents in an integrative manner structure based computational approaches used to understand, rationalize and predict the activity and substrate selectivity of CYPs, as well as the possibilities and limitations of these approaches now and in the future. One can divide structure-based models and methods into: i) ligand-based models (QM calculations on substrates, pharmacophore, and QSAR models), ii) protein-based models (crystal structures, homology models and various ways in which this structural information is analyzed and used), and ligand-protein interaction based models (e.g. automated docking and molecular dynamics simulations to describe ligand binding orientation, protein-ligand dynamics and ligand binding affinity). Apart from the earlier mechanism-based *ab initio* calculations on substrates, also pharmacophore modeling of ligands (*i.e.* both substrates and inhibitors) and protein homology modeling in combination with automated docking and molecular dynamics simulations have been used successfully for the rationalization and prediction of metabolite formation by several CYP isoenzymes. In the past six years, the number of available crystal structures, especially for mammalian and human isoforms, has increased significantly. In addition, the lack of detailed structural knowledge from crystallography for many other isoforms is being overcome in part by the prediction of CYP structures using computer-aided homology modeling techniques in combination with pharmacophore or small molecule models and mechanism-based *ab initio* and semi-empirical calculations. It has been shown that a great variety of *in silico* modeling approaches have been applied to CYP enzymes, thereby successfully adding to our understanding of CYP structure and function in a way that is complementary to experimental studies. Another clear trend in many recent computational studies is the combination of multiple modeling techniques to arrive at meaningful and successful rationalization of experimental data, and interpretations and predictions of CYP function.

2.1 Introduction

Cytochromes P450 (CYPs) constitute the most important family of biotransformation enzymes involved in drug metabolism, playing an important role in the disposition of drugs, and their pharmacological and toxicological effects. Early consideration of ADME-properties (absorption disposition, metabolism and excretion) is increasingly seen as essential for efficient discovery and development of new drugs and drug candidates[1]. Apart from *in vitro* tools, this necessitates the application of novel *in silico* tools that can accurately predict ADME properties of drug candidates already in early stages of the lead finding and optimization process.[2,3]

CYPs generally detoxify potentially harmful xenobiotic compounds and in a number of cases non-toxic compounds are bioactivated to toxic reactive intermediates, and pro-carcinogens into ultimate carcinogens. Furthermore, CYPs catalyze key reactions in the formation of endogenous compounds, such as hormones and steroids[4]. The catalytic activities of CYPs can be divided into *i)* monooxygenase activity, usually resulting in the incorporation of an oxygen atom into the substrate, *ii)* oxidase activity, resulting in the formation of superoxide anion radicals or hydrogen peroxide (uncoupling of the catalytic cycle), and *iii)* substrate reductase activity usually producing free radical intermediates under anaerobic conditions.[5] **Figure 1** schematically shows the CYP reaction cycle where these three activities and associated reactions are indicated, for more details we refer to a comprehensive review by Guengerich.[6] Apart from the classical incorporation of oxygen into substrates, other reactions can be performed by CYPs, *e.g.* de-saturation, dehydrogenation, ring formation and dehalogenation. CYPs can be classified according to the electron transfer chain that delivers the electrons for the one-electron reductions from NAD(P)H: class I CYPs are found in bacteria and in eukaryotic mitochondrial membranes and require a flavin adenine dinucleotide (FAD) containing reductase and an iron-sulfur protein (putidaredoxin); class II CYPs are bound to the endoplasmic reticulum and interact directly with a cytochrome P450 reductase containing FAD and flavin mononucleotide (FMN).[7]

In all known CYP crystal structures available today, the same general three-dimensional fold is present with a conserved core region containing the C-terminal half of helix I, helices L, E, K, and K', and the heme coordination region.[8] The three-dimensional structure of these regions is well conserved despite a low sequence homology. Other regions (*e.g.*, the active site region containing the B' helix, the loops between helices C and D, the region spanning helices F and G, and most of the β-strands) are more variable. The largest structural variations are found in the B' helix and helices F and G and their connecting loop, which makes sense since these regions are known to be involved in substrate access and recognition.[7] Six substrate recognition sites (SRS's) have been assigned based on mutagenesis and sequence alignment studies with representative members of the CYP2 family and CYP101 (*cam*).[9] Selected recent example CYP structures are given in **Figure 2** that shows CYP2C9 with two different substrates bound, CYP3A4 without ligand, and a homology model of CYP2D6. **Figure 3-B** illustrates the organization of SRS's and secondary structure elements in CYP2D6, which shares all major features of CYPs.

In general terms, the prediction of enzyme activity and selectivity includes everything from substrate binding through the complete catalytic process, including the input of electrons and protons, to product exit. Many of the steps involved are reversible, so the combined kinetics of all steps determines final products, or the ratio between possible products, formed. A comprehensive approach that provides a complete *in silico* description of this process would allow a complete *in silico* prediction of enzyme reactivity, and would include the prediction of such diverse components as enzyme/active site structure, substrate binding orientation and conformation and binding affinity including multiple binding modes and/or water mediated binding, intrinsic (regio-specific) substrate

reactivity, orientation and/or conformation dependent reactivity, and product affinity.[10] Needless to say, no such combined method is available today.
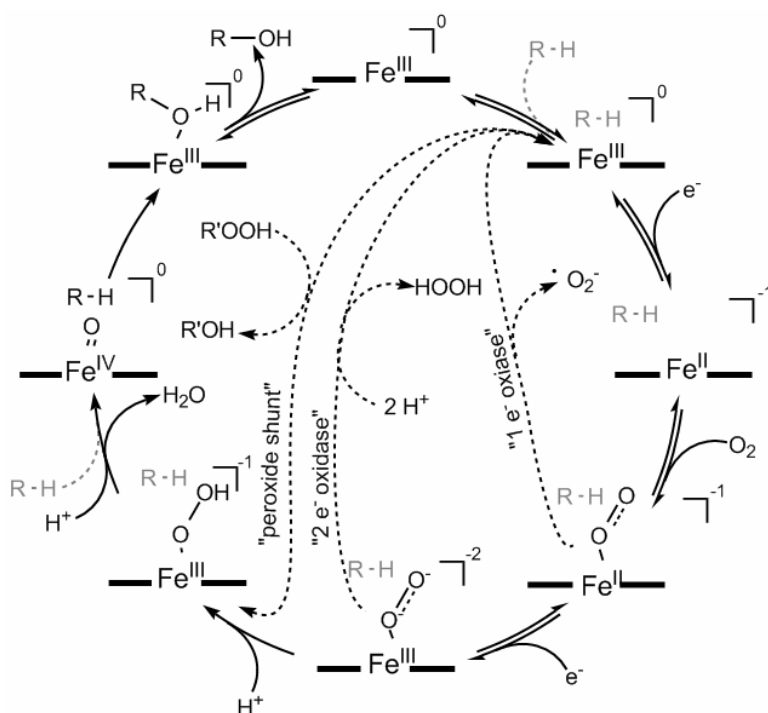
Figure 1: *The generalized CYP reaction cycle; the main reaction cycle corresponds to the general mono-oxygenase function that can produce several reactions (carbon hydroxylation, heteroatom release or oxygenation, epoxydation and group migration). Not for all CYPs is the order of substrate and oxygen binding established, therefore the substrate ('R-H') enters the cycle in two possible places and is drawn in gray for part of the cycle. Three alternative reaction pathways are indicated with dashed arrows, and correspond to the 1- and 2-electron oxidase uncoupling reactions and the peroxide shunt.[6] The 1-electron oxidase produces a superoxide radical which in turn can produce a radical derivative of the substrate. The peroxide shunt is an alternative way to produce an activated iron-oxygen species. Indicated are the heme-iron species (Fe[N]) and oxidation state (N), the substrate and an alternate reducing peroxide agent (R'OOH). In the resting state (at the top of the cycle) a water molecule (or hydroxy anion[11]) is assumed to be bound to the Fe[III].*

The state of the art of computational methods used in CYP modeling varies considerably, as does the reliability of the results obtained. The quality of a homology model is often not sufficient for reliable docking of a substrate, unless the sequence homology with the crystal structure template is very high. Binding affinities are notoriously hard to predict quantitatively, certainly not without reliable knowledge of ligand binding orientations[12], and likewise details of product formation are hard to predict without knowing the binding orientations.

For accurate prediction of reactivities the dependence on kinetic factors and the influence of the active site electrostatic potential should be accounted for, and, in turn, this can depend on details of the protein structure and the binding conformation and orientation of the substrate. Finally, the absence of a certain predicted metabolite may be due to slow conversion compared to other metabolic reactions, from competing metabolic pathways or from other enzymes. In many cases it is possible to identify rate-limiting steps (which may vary between CYP isoforms, between substrates in the same isoform, and between different environments), implying that product formation will be dominated by the characteristics of that particular step and thereby greatly simplifying the complexity of the prediction.[13]
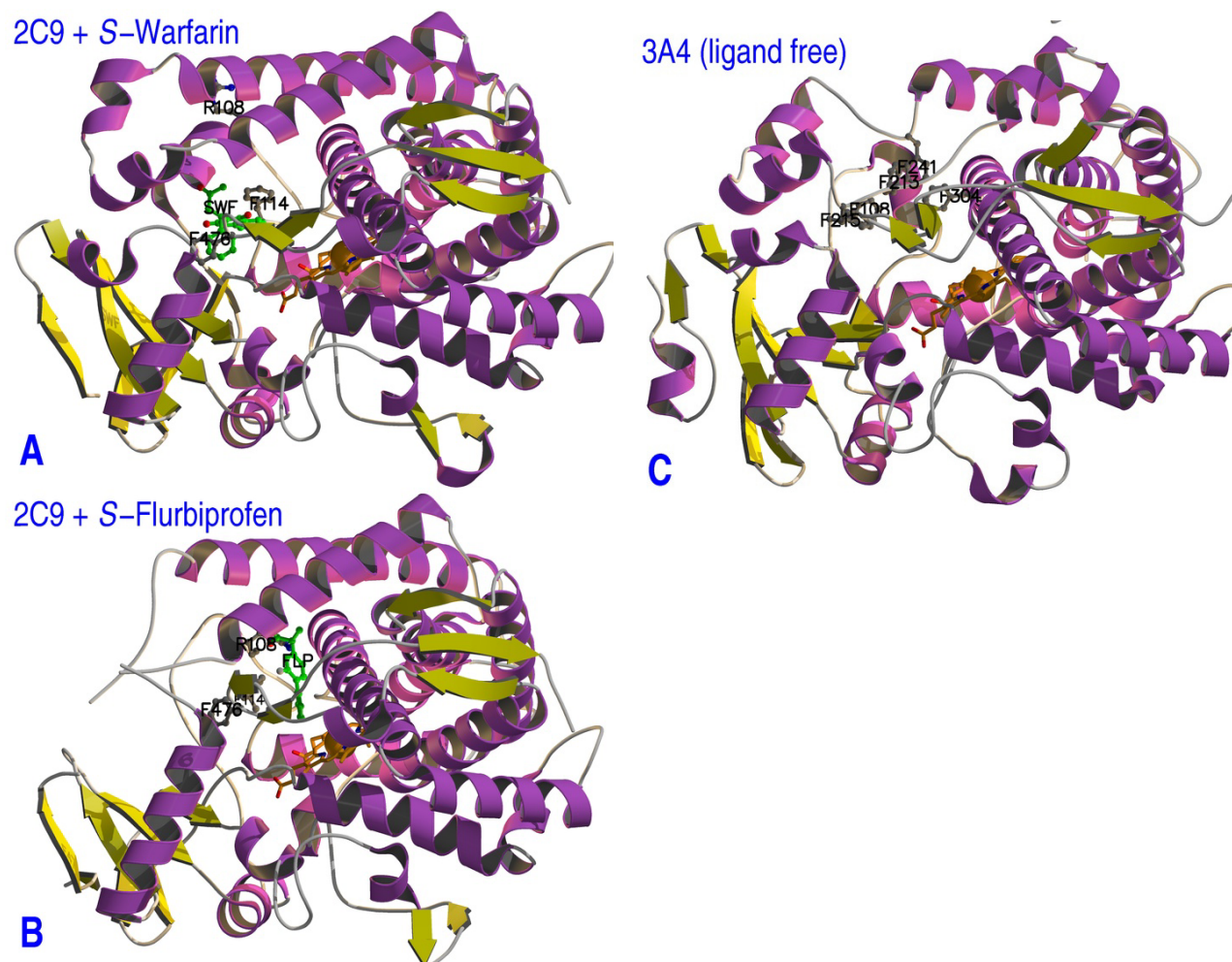
Figure 2: *Representative CYP crystal structures. Central in the structures is the heme group in orange with the iron as a space-filling sphere. The ligands (in green) and key interacting residues in the binding pockets are shown in ball-and-stick. Panels* **(A)** *and* **(B)** *show the 2C9 crystal structures with S-Warfarin (SWF)[14] and S-Flurbiprofen (FLP)[15] bound, respectively. Panel* **(C)** *shows the ligand-free 3A4 crystal structure[16]. The similarity in fold and topology between the isoforms is evident. Most notable differences which determine the overall size and shape of the binding cavity are in SRS2 and 3 (helices F and G, respectively), the connecting FG-loop, SRS1 (helix B') and SRS6 (sheet 4). Secondary structure elements are colored analogous to* **Figure 3**. *The CYP topology is schematically shown in* **Figure 3B**. *The figures were generated using Molscript and Raster3D (see* **Table 4**).

A remaining challenge in predicting product formation in CYPs is the elucidation of the relevant reactive heme-iron-oxygen species that is ultimately responsible for the chemistry of the enzyme reaction, and the coupling between the kinetics of the reaction cycle of the heme-iron-oxygen moiety and the dynamics of substrate and oxygen and release of the product. Elucidation of the reaction cycle of CYPs is complicated by the differences in influence of substrate, sequence of events and rate-limiting step between different isoforms.[13] Modeling approaches dealing with these details of the reaction cycle, mainly QM studies, have been described recently in the review of Shaik *et al.*[17].
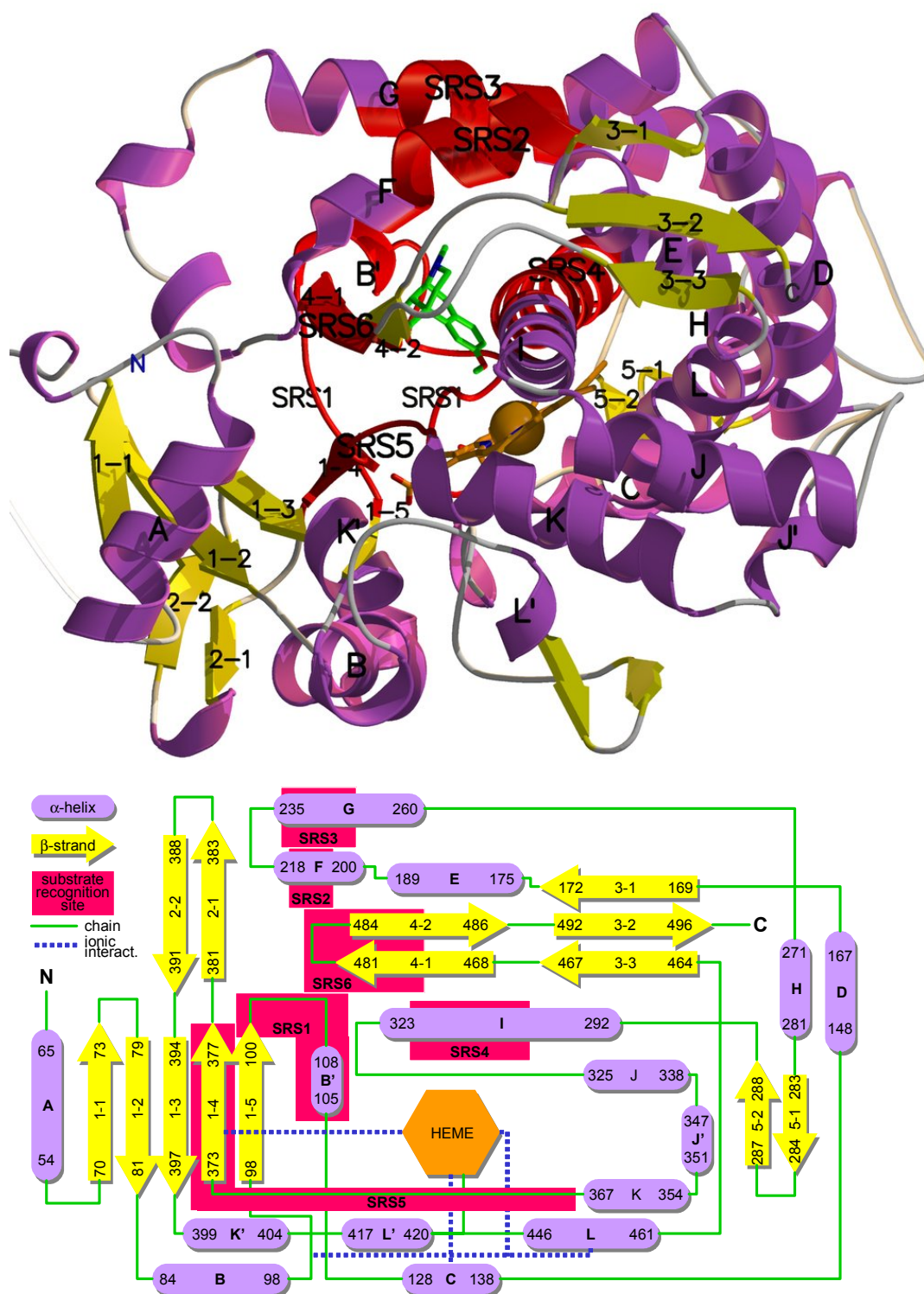
Figure 3: *Protein homology model, topology and tertiary organization of CYP2D6. α-Helices are colored purple, and β-strands are yellow. In red the substrate recognition sites (SRS's, after Gotoh[9]) are indicated in their relative orientation to each other and to the heme catalytic center. Panel* **(A)** *shows the CYP2D6 homology model 3D-structure with dextromethorphan bound[18] (in green) and key interacting residues in the binding pockets in ball-and-stick. A corresponding depiction of a CYP2D6 pharmacophore model (with bufuralol) is shown in* ***Figure 4*** *in the same orientation. Panel* **(B)** *shows the CYP topology and tertiary organization illustrated with details of CYP2D6.*

The primary aim of this paper is to present in an integrative manner computational approaches used to understand, rationalize and predict the activity and substrate selectivity of CYPs, as well as the possibilities and limitations of these approaches now and in the future. Apart from the earlier mechanism-based *ab initio* calculations on substrates, also pharmacophore modeling of ligands (*i.e.* both substrates and inhibitors)[19] and protein homology modeling in combination with automated docking and molecular dynamics simulations have been used successfully for the rationalization and prediction of metabolite formation by several CYP isoenzymes. In recent years, the number of available crystal structures, especially for mammalian and human isoforms, has increased significantly. In addition, the lack of detailed structural knowledge from crystallography for many other isoforms is being overcome in part by the prediction of CYP structures using computer-aided homology modeling techniques in combination with pharmacophore or small molecule models and mechanism-based *ab initio* and semiempirical calculations. Quantitative structure-activity relationship (QSAR) approaches, derived from physicochemical ('1D') and topological ('2D') properties[20] or the 3D structure[21] of CYP ligands, have been used as well for the prediction of drug metabolism by CYPs. A pragmatic, rapid and rather successful approach to this are topological screening methods recently developed for CYP2D6[22] and CYP3A4[23], which provide robust but rough classifications of the inhibition potency of ligands without much insight into possible (molecular) background of the mode of action. The structure-based ('3D') molecular modeling methods on the other hand, aim to be insightful as well as predictive. It has been concluded that not one computational approach is capable of rationalizing and reliably predicting metabolite formation by CYPs, but that it is rather the combination of the various complimentary approaches.[10]

In this paper, we divide structure-based models and methods into ligand based, protein based, and ligand-protein interaction based. In the corresponding sections each of the categories is described, giving an overview of the methodology applied, a general background on each of these, some highlight examples of applications, and conclusions. Finally, a general summary and conclusions are presented, and an outlook of expected developments in the field. The first **section 2** ´*Ligand Based Models*´, deals with QM calculations on substrates, pharmacophore, and QSAR models. **Section 3** ´*Protein Structure and Analysis*´ deals with crystal structures, homology models and various ways in which this structural information is analyzed and used. Finally, in **section 4,** ´*Ligand-Protein Interactions*´**,** issues on binding orientation, ligand dynamics and binding affinity are covered. Four comprehensive tables are supplied that list most important pharmacophore models (**Table 1**), all available CYP crystal structures (**Table 2**), most important homology models (**Table 3**), and most commonly used software for CYP molecular modeling (**Table 4**).

2.2 Ligand Based Models

2.2.1 Introduction

Models for the active sites of CYP isoenzymes and CYP-ligand interactions can be derived from ligand based models, based on a set of known CYP isoenzyme-specific substrates or inhibitors. Ligand based models can be divided into two broad categories, namely quantum mechanical (QM) and pharmacophore models. QM calculations are based on certain levels of *ab initio* or semiempirical QM theory and produce numbers on activation barriers, charge distributions and minimum energy conformations. Pharmacophore modeling concerns empirical techniques to (indirectly) derive information on the active site from the shape, electronic properties, and conformations of substrates, inhibitors, or metabolic products. An example of a pharmacophore model is shown in **Figure 4**.

The so-called "classical" pharmacophore models are constructed by superimposing substrate and/or inhibitor molecules onto each other by matching chemically similar groups. Current approaches, however, utilize the three-dimensional quantitative structure-activity relationship (3D-QSAR) approaches in which abstracted descriptors of essential functional groups and topologies of the molecules are superimposed, possibly supplemented by details of the surrounding active site. (This distinction is not fundamental; a molecule consisting of atoms and bonds in a "classical" pharmacophore model can be viewed as a type of "abstracted descriptor", although rather detailed, in a 3D-QSAR model.) Recent reviews on CYP pharmacophore and 3D-QSAR methods are provided by Ekins and de Groot.[19,21]

In this section, methods used to construct CYP pharmacophore and 3D-QSAR models will be described. A summary of most recently published CYP pharmacophore models is provided in **Table 1** and selected successful applications will be highlighted using examples from this table. Commonly used software for CYP pharmacophore model building is listed in **Table 4**.
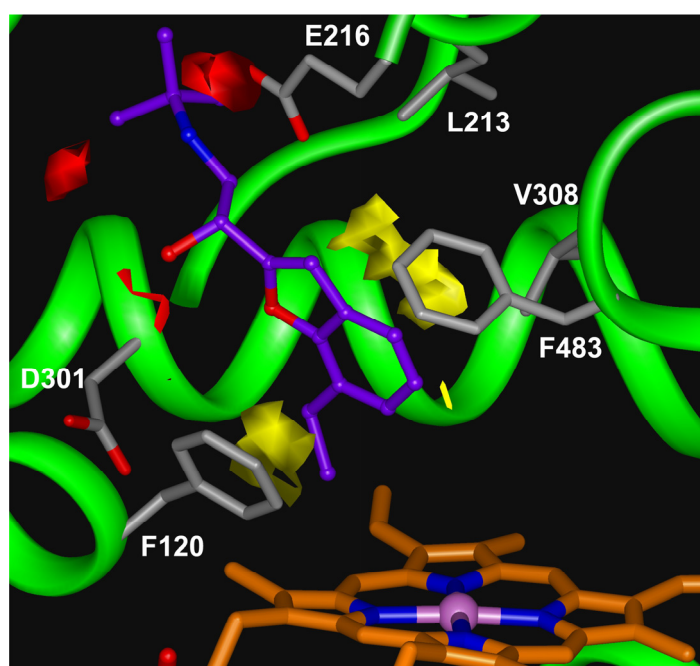
Figure 4: *Examples of the pharmacophore model for CYP2D6 illustrated by a molecular interaction field (MIF)[24] of a carboxylic probe (in red at E216) and a hydrophobic probe (yellow at F120 and between L213, V308 and F483) for S-bufuralol (in purple). Selected side chains are shown as stick models, the rest of the active site is shown as green ribbon. The complete structure of the homology model is shown in* **Figure 2**.

2.2.2 Quantum-chemistry
Quantum-chemistry Methods Background

In CYP enzymes, quantum-chemical calculations can yield valuable information on activation barriers for substrates or heme-oxygen species during the catalytic cycle, and relative energies between species and relevant geometries. Gas phase calculations, commonly at the Hartree-Fock (HF) level, are mainly applied to isolated substrates in order to assess relative reactivities for a given reaction at different sites in the substrate. Using density-functional theory (DFT) instead of HF calculations allows extension of the quantum-system to include the heme-oxygen species and the associated fifth ligand cysteine. Recently, hybrid quantum-mechanical/molecular-mechanical (QM/MM) methods have been applied that include the complete enzyme environment in a molecular mechanics (MM) description as well.[25-27]

Of the three enzymatic functions of CYPs (monooxygenase, oxidase and reductase as summarized by Goeptar *et al.*[5] and Guengerich[6], and indicated the reaction cycle in **Figure 1**) attention has been almost exclusively on the monooxygenase activity and its proposed mechanisms of action. The two most commonly encountered mechanisms of action, the H-atom abstraction/oxygen rebound[28] and the electron oxidation/recombination[29] have been investigated the most by far.

Quantum-chemistry Applications
Gas phase *ab initio* calculations with two sequential one-election oxidation steps, radical intermediates, statistical spin-distribution and hydroxyl-radical recombination have been used to rationalize metabolite formation for CYPs in general[30,31] and in more depth for CYP2D6.[32] Very recently, Sing *et al.*[33] developed a rapid semi quantitative model based on a trend vector model (derived from AM1 semi-empirical calculations) for estimating H-abstraction energy and surface area exposure of the H-atom to predict the relative susceptibilities of different sites on 50 drug molecules to metabolism by CYP3A4. The regiospecific hydroxylation of *S*- and *R*-nicotine[34], 5 camphor analogs[35], and carbofuran[36] by CYP101 (*cam*) and 6-deoxyerythronolide B by CYP107A (*EryF*)[37] was successfully predicted employing geometric criteria obtained from molecular dynamic simulations of enzyme-substrate interactions in conjunction with thermodynamic criteria of relative radical energetics derived from *ab initio* and semiempirical quantum mechanics. Park and Harris[38] employed DFT calculations on CYP2E1 homology models and substrate conformations from automated docking and molecular dynamics simulations to predict metabolism for seven substrates. Relative radical energies were in good agreement with differential H-abstraction activation energies by oxyferryl heme, and optimized oxyferryl-heme-ligand geometries were in good agreement with MD-sampled conformations. DFT calculations were performed to estimate stabilization energies on cooperative steroid-substrate interaction in the active site of a CYP3A4 homology model[39]. Recently, for CYP2B4 the effects of ligand binding on the high/low spin equilibrium were explored with DFT calculations, showing that strong interactions with the heme favor the low iron spin-state[40]. The two-state abstraction/oxygen rebound mechanism of CYP monooxygenase activity, as proposed by Groves *et al.*[28] was supported by recent studies on CYP101 (*cam*) and of the mechanisms of C-H hydroxylation.[27]

Gas phase calculations on the activation energy barrier for breaking a C-H bond in alkanes and fatty acids, as described by Jones *et al.*[41], could not explain observed product formation for CYP102 (*BM3*) wild-type and several mutants (Feenstra *et al.*, unpublished results). In this approach a hydroxyl radical is used to activate the C-H bond, thereby mimicking the catalytic effect of the heme-iron-oxygen species of CYP enzymes, and it is based on unrestricted HF at the semi-empirical AM1 level. Significant modulation of activation energy barrier heights by addition of the electrostatic environment of the CYP protein were observed, underlining the importance of incorporation of the active site properties of the enzyme in determining accurate regiospecific reactivities. Combination of these reactivities with binding statistics derived from MD simulations, give trends in predictions of product formation that correspond to experimental observations, as will be described in more detail in **section 2.4.3** '*Molecular Dynamics*'.

Quantum-chemistry Conclusions
While early applications of quantum chemical calculations on isolated substrates in the gas phase have yielded valuable information about the intrinsic regio-specific susceptibility of molecules to metabolism by CYP enzymes in general, it has become clear that gas phase QM calculations are insufficient to explain product formations for all CYP isoforms and substrates. In some cases, a combination of the QM derived substrate susceptibilities with predicted binding orientations, *e.g.* from a pharmacophore model or from docking and MD

simulations in a protein structure, yields a more realistic estimate of product formation, as has been done for the CYPs 101 (*cam*), 107A (*EryF*), 2E1, 2B4 and in ongoing work for CYP102 (*BM3*). Although many QM studies have been devoted to the identification of heme-iron compounds in the enzyme redox cycle (which is outside the scope of this paper), so far few studies have addressed details of the actual reaction with substrates. Several QM studies on CYP101 (*cam*) have found the two-state abstraction/oxygen rebound mechanism to be the most important for C-H hydroxylation.

Finally, the steady increase in computational power brings some sophisticated novel methods into the range of feasibility, that make it possible to include substantial parts of the active site in the QM calculation, *e.g.* the heme-iron moiety and nearest charged amino acid residues. Promising examples are methods based on the fragment molecular orbital (FMO) approach that in principle allow the incorporation of the whole protein into the quantum HF or DFT calculations. These methods eliminate the need for arbitrary selections of a restricted quantum system, as used in *e.g.* a QM/MM approach, and the many artifacts and uncertainties that arise from that restriction. Although successful application of a full-protein DFT calculation to cytochrome C has been reported[42], to date these have not been applied to cytochromes P450.

2.2.3 Classical Pharmacophore Models

Classical Pharmacophore Models Methods Background

Pharmacophore models derive information on the active site from the shape, electronic properties and conformations of substrates, inhibitors or metabolic products. The so-called "classical" pharmacophore models are constructed by superimposing these molecules onto each other by matching chemically similar groups, possibly supplemented by details of the surrounding active site. This approach can be used to perform catalytic site prediction (CSP) and (qualitative) predictions with respect to CYP activity and (binding) affinity (K_s, K_i, IC$_{50}$, K_m). Primary assumptions for this methodology are that all ligands share the same binding site and binding mode in the active site (both electronically and sterically) and that the minimum energy conformation of the ligands in vacuum corresponds to the biologically active conformation. Moreover, it must be stressed that CYP active sites allow some degree of flexibility and therefore may allow multiple pharmacophores to exist[43], as is described in more detail in **section 2.3.2** *'Crystallographic Protein Models Active Site'*. During the past two decades, pharmacophore models have not only represented useful tools to rationalize properties of CYP active sites in the absence of properly validated crystallographic or homology model structures for these enzymes but have also been combined with these structural models to improve capabilities of predicting CYP substrate selectivity and metabolism, as demonstrated in **Table 1** and **Table 3**.

Building a pharmacophore model is an iterative process of repeated cycles of selection of relevant substrates and/or inhibitors and their functional groups, structure minimization and superposition, and validation, as is outlined in a flowchart in **Figure 5**. Commonly used software for CYP pharmacophore model building is listed in **Table 4**. When building a pharmacophore model for substrates or inhibitors of a specific enzyme by the "classical" pharmacophore method, *i.e.* manual or automated superposition of 3D-structures, a primary requisite is an appropriate template molecule. This molecule ideally is a ligand that: *i)* is specifically metabolized by the (iso-)enzyme under investigation, *ii)* is large, *iii)* is relatively rigid, in order to limit conformational freedom, *iv)* contains essential functional groups, and *v)* is regio- and/or stereoselectively metabolized. After selection of a template molecule, additional ligands are superimposed onto the template ligand, matching some essential functional groups, such as H-bond acceptors/donors or aromatic rings, or by using conformations generated by automated docking in a CYP (homology) model. A check whether the superimposed structures are in an energetically favorable conformation

can be done before, during or after the fitting, using rigid pre-energy-minimized (EM) structures, distance restrained EM, or EM or energy calculation of the fitted structures, respectively.[44] Models for CYP inhibitors are generally more difficult to construct compared to substrate models as the site of catalysis is lacking in inhibitors and can therefore not be used as an easily identifiable site to be superimposed. The development in classical pharmacophore modeling methodology is exemplified below by the cases of CYP2C9 and CYP2D6.
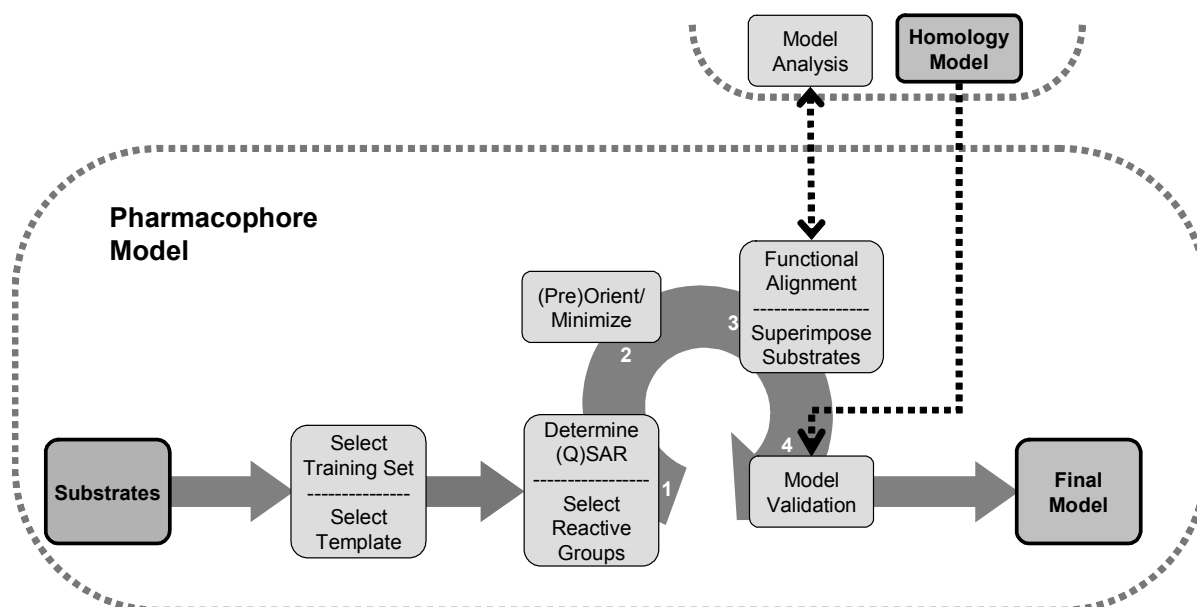
Figure 5: *Flowchart for pharmacophore and QSAR modeling. The cycle is iterated to improve the model; the improved model can yield better choices for pharmacophoric groups or QSAR descriptors. Functional alignments in QSAR, and superposition of substrate in the pharmacophore model can also be improved by knowledge from a homology model and analysis of structural details of a homology model can help in the final pharmacophore or QSAR model validation.*

Classical Pharmacophore Models Applications

Early pharmacophore models of CYP2C9 suggested a H-bond donor or acceptor within the active site.[45] A second pharmacophore model, supported by data of 10 tienilic-acid analogues, superimposed substrates using their sites of oxidation and brought all anionic heteroatoms in the substrate within 3.5-4.8 Å of a (hypothetical) cationic site within the CYP2C9 protein. This model was later updated, leading to the inclusion of a hydrophobic zone between the hydroxylation site and the cationic site of the protein.[46] The first combined pharmacophore-homology model of de Groot *et al.*[41] includes an aromatic region for which F476 is suggested, and an anionic ligand binding site that may involve R108. Evidence for the importance of an anionic site in the substrates was strengthened by site-directed mutagenesis experiments highlighting the importance of arginine residues within the CYP2C9 active site[47]. In the recent CYP2C9 crystal structure[14] F476 indeed interacts with *S*-warfarin, however, no Arg is present in the binding pocket, while in a more recent substrate bound crystal structure F114 and R108 interact with *S*-flurbiprofen[15], as can be seen in **Figure 2** and described in more detail in **section 2.3.2**.

   Early pharmacophore models for CYP2D6 had a basic nitrogen at distances of either 5 Å[48] or 7 Å[49] from the site of oxidation as well as an aromatic ring system, but none could explain the other group of substrates. An extended model[50] combined these models by indicating a distance between 5 and 7 Å and in addition contained the oxygen bound heme moiety. 15 compounds could be fitted onto the template debrisoquine, and one was correctly predicted not to be a substrate for 2D6.[50] A hypothetical carboxylate sidechain

binding the basic nitrogen atom of the substrate was first included by Koymans *et al.*[51], thereby also explaining both the 5 Å and 7 Å distances, and allowed the prediction of four compounds giving 14 possible 2D6-dependent metabolites, of which 13 were experimentally verified initially. Several additional metabolites were rationalized using a refined version of the model[52], and the model was extended with more substrates[53]. Adding the heme moiety and I-helix as predicted from homology modeling identified D301 as the source of the carboxylate group[4] which was later experimentally confirmed[54], while further studies included E216 as key determinant for binding and the source of the carboxylate group, and a possible indirect role of D301.[55]

   A combined pharmacophore and homology model for CYP2D6, consisting of distinct pharmacophores for O-dealkylation and N-dealkylation, was able to predict the metabolism of a variety of compounds.[56] Finally, an inhibitor model of 2D6[57] appeared to share many features with the previous substrate models[4,50,51] suggesting they may be combined. The main features of the currently known CYP2D6 pharmacophore models are highlighted in **Figure 4** where molecular interaction fields (MIFs, explained further **section 2.2.4**) calculated for bufuralol and overlaid on the substrate docked into the CYP2D6 active site clearly correlate with the specific residues responsible for substrate interaction.

Classical Pharmacophore Models Conclusions
The earlier CYP substrate/inhibitor pharmacophore models were based on comparison of ligand properties only, and in that respect were relatively crude. The more recent pharmacophore models are constructed using also structural information of protein models, as can be seen in **Table 1** for CYP2C9 and CYP2D6 and most other CYP isoforms as well. Conversely, most recent homology models have been built using knowledge from an existing pharmacophore model or in parallel with building a new pharmacophore model, as is described in **section 2.3.3** '*Homology Based Protein Models*'. The CYP2C9 pharmacophore model involves an anionic and a hydrophobic/aromatic site; the aromatic site could be attributed to F476 and several other residues in the crystal structure. In CYP2D6, the pharmacophore model involves a basic nitrogen and an aromatic site; the nitrogen is thought to bind to E216 and/or D301 based on homology models and site-directed mutagenesis studies. Recent classical pharmacophore models predict possible involvement of CYP2D6 or CYP2C9 (and some other isoforms) and likely sites of catalysis in the metabolism of substrates, but for the quantitative prediction of CYP binding affinity parameters 3D-QSAR models are necessary (see **section 2.2.4**).

2.2.4 3D-QSAR Pharmacophore Models
3D-QSAR Pharmacophore Models Methods Background
In 3D-QSAR methods[58] spatial atomic descriptors, *e.g.* molecular interaction fields (MIFs), functional groups and their (relative) orientations, electronic properties, or shape, of a training set of molecules are characterized in three-dimensional space, and condensed into a matrix. Subsequently, this matrix is analyzed using multivariate analysis like Principal Component Analysis (PCA) to extract and condense the information contained in the matrix by discovering patterns, trends and clusters and further analyzed using regression techniques like Partial Least Squares (PLS) analysis to explain the chemical properties of these molecules, *e.g.* (binding) affinities (K_d, K_i, IC_{50}, K_m). The constructed 3D-QSAR model can be used to perform quantitative predictions of properties for molecules that were not used to build the QSAR model and to predict sites of catalysis of CYP substrates. It is generally advantageous if the molecules used in 3D-QSAR studies both cover wide ranges in molecular structure and affinity to maximize conformational/chemical space and to gain as much information as possible regarding the binding site in order to increase the statistical significance of the model. Several parameters directly or indirectly related to CYP

binding affinity have been used to construct and validate 3D-QSAR models, like Michaelis-Menten constants (K_m), inhibition constants (K_i), concentration required for 50% inhibition of the metabolism of a specific substrate (IC_{50}), and actual binding constants (K_s). It is important to note that for CYPs K_s binding constants are spectroscopically determined parameters, whereas K_m, K_i and IC_{50} are activity measurements. This can lead to differences due to binding modes or experimental conditions that do not reflect different binding properties[59]. When an inhibitor is not metabolized by the CYP studied, K_i can be similar to K_s, unlike the K_m of a substrate, which is usually not equal to the K_s. IC_{50} and K_i values of an inhibitor which is metabolized depend on the affinity of the substrate(s) competing with the inhibitor.[60] In most cases K_s will provide the most direct measure of the binding site properties of the CYP isoenzyme.

Like classical pharmacophore models, 3D-QSAR approaches have not only represented useful tools in understanding CYP active sites in the absence of proper crystallographic or homology model structures for these enzymes, but they have also been combined with structural protein models to improve the prediction of CYP substrate specificity and metabolism. **Figure 6** shows an example of a correlation between predicted and experimentally determined K_i values for CYP2C9, **Figure 4** shows an example of MIFs of a substrate docked into a CYP2D6 homology model, and **Figure 3-A** the corresponding CYP2D6 homology model.

Traditionally, methods for performing 3D-QSAR rely upon alignment of the training set molecules. Energetically favorable conformations of substrates and/or inhibitors can be generated using energy minimization and/or automated docking of ligands into a CYP (homology) model. This often is time-consuming and can introduce user bias, the resulting model being dependent upon and sensitive to the alignment used. Commonly used software for CYP pharmacophore/3D-QSAR model building is listed in **Table 4**, examples being CoMFA, Catalyst, PLS MS-WHIM, and ALMOND**.** ALMOND is a recent 3D-QSAR program specifically developed for generating and handling alignment independent descriptors, *e.g.* grid independent descriptors (GRIND). However, even when using this approach, models usually depent on the ligand conformations. The possibility of constructing conformer independent 3D-QSAR models for CYPs was explored very recently by the use of flexible GRIND MIFs.[24,61] A strategy for the incorporation of water molecules present in a ligand binding site into 3D-QSARs has also been described.[62] In addition to ligand-based 3D-QSAR pharmacophore models, protein-based inverse pharmacophore models can also be derived from the analysis of MIFs in active sites of several protein targets using consensus principal component analysis (CPCA). Very recently, methods for homology and inverse 3D-QSAR pharmacophore model building have been evaluated for CYP2C5 and 2C9. They were concluded to depend critically on the alignment strategy, on the quality of the template crystal structure, as well as on the optimization procedures applied.[63]

Alignment-dependent 3D-QSAR models for CYP inhibitors are generally more difficult to construct than substrate models, as the site of catalysis is lacking in inhibitor models and can therefore not be used as an easily identifiable site to be superimposed. However, until now, more inhibitor models than substrate models have been developed (see **Table 2**), probably because K_s data for CYP ligands (*i.e.* substrates as well as inhibitors) are relatively scarce. The possibilities and limitations in 3D-QSAR pharmacophore modeling will be demonstrated by models described below for the cases of CYP2C9 and CYP2D6.

Table 1: *Overview of CYP pharmacophore and 3D-QSAR models.*

Method[a]	Sets Training[b]	Test[c]	Predicted property	Key Features of Pharmacophore	Ref.
CYP1A2					
COMBINE and GRID/ GOLPE + homology model	12	n.a.	Mutagenicity	Inhibitor model with H-bonding and hydrophobic binding sites (on the protein). Several residues in the homology model are found to interact with the pharmacophore model.	64
CYP2A5					
CoMFA	16	n.a.	IC_{50}	Substrates and inhibitors have a negative electrostatic potential close to a lactone moiety, steric effects around a methoxy group on methoxasalen and positive electrostatic potential para to a methoxy group.	65
PLS MS-WHIM	16	n.a.	K_i	Potent inhibitors have a positive molecular electrostatic potential and a H-bond acceptor.	66
CYP2A6					
CoMFA and GRID/GOLPE	23	5 (s.s.)	IC_{50}	Potent CYP2A6 inhibitors do not include a lactone moiety.	67
CYP2B6					
Catalyst and PLS MS-WHIM	16	n.a. (s.s.)	K_m, CSP	Substrate model includes at least three hydrophobic regions 3.1, 4.6 and 5.3 Å from a H-bond acceptor.	68
Catalyst + homology model	16	5 (4)	K_m, CSP	Substrate model includes two hydrophobic regions and one H-bond acceptor. Substrate catalytic site 4.0 and 3.4 Å from two hydrophobic regions and 4.6 Å from a H-bond acceptor in model A, and 4.9 and 4.1, and 7.8 Å, in model B.	69
CYP2C8/9/18/19					
GRID/CPCA + homology model + docking	n.a.	n.a.	CSP	Combined protein-based inverse substrate model includes two hydrophobic and two electropositive binding sites (on the protein).	70
CYP2C8					
Catalyst + homology model + docking	7	n.a.	CSP	Substrate model with H-bond acceptor, a hydrophobic and a cationic region.	71
CYP2C9					
Manual superposition[c]	8	n.a.	CSP	Substrate model protein H-bond donor 7 Å from substrate catalytic site.	72
Manual superposition[c]	20	n.a.	CSP	Substrate model includes anionic site 7.8 Å from catalytic site, in between a hydrophobic region.	46,73,74
CoMFA + homology model	27	14 (13)	K_i, CSP	Inhibitor model includes two cationic binding sites, along with an aromatic binding region, and a steric regions (on the protein). Substrates posses a partial negative charge at 10 Å and an anionic site at 6 Å from the catalytic site.	45 75
Catalyst and PLS-WHIM	9 29 13	14 (10) 14 (12)	K_i K_i	Inhibitor models include at least one hydrophobic and one H-bond acceptor at 3 to 5.8 Å from each other. H-bond acceptor and H-bond donor/acceptor are 3.4 to 5.7 Å apart.	43
GRID/GOLPE + homology model + docking	21	8 (8)	K_i	Inhibitor model in agreement with site directed-mutagenesis data indicating important roles Leu102, Val113, F114, and Leu362.	76

Method[a]	Sets Training[b]	Test[c]	Predicted property	Key Features of Pharmacophore	Ref.
GRID/GOLPE/ ALMOND/ + homology model + docking	42 21	39 (29) 12 (11)	K_i K_i	Alignment independent inhibitor model includes two H-bond donors at 6 Å from each other, a hydrophobic region at 13 Å from a H-bond donor, and two H-bond acceptors at 12 and 15 Å from two a H-bond donors (on the protein).	77
Manual superposition[c] + homology model	27	4 (4)	CSP	Substrate model includes a H-bond acceptor site and an aromatic region, suggests R108 for electrostatic and F476 for aromatic interactions.	41
Catalyst	36	n.a. (s.s.)	HA	Heteroactivator model includes two hydrophobic regions at 11.6-11.8 Å from a H-bond acceptor and at 10.3-10.9 Å from an aromatic ring, and a H-bond acceptor 4.8 Å from an aromatic ring.	78
GRID/GOLPE/ ALMOND + homology model	43	n.a. (22-39)	CSP	Alignment independent protein-based inverse model, represents substrates as set of fingerprints for each H-atom present.	79
GRID/GOLPE/ ALMOND + homology model + docking	22	12 (11)	K_i	Alignment and conformer independent inhibitor model includes two H-bond donors, and three hydrophobic binding sites (on the protein).	61
CYP2D6					
Manual superposition[c]	15	n.a.	CSP	Substrates possess a basic nitrogen atom at 5 or 7 Å from site of catalysis, and coplanar aromatic rings.	48-50
Manual superposition[c]	156	n.a.	CSP	Carboxylate group in protein responsible for the 5 or 7 Å distances between basic nitrogen and site of catalysis.	51
Manual superposition[c] + homology model	10	n.a.	CSP	D301 identified as residue binding to basic nitrogen atom in ligands	80
Manual superposition[c] + homology model + molecular orbital calc.	40+14	7 (6)	CSP	Substrates have a basic nitrogen atom at 5, 7 or 10 Å from site of catalysis. Separate models for O-demethylation and N-dealkylation. E216, D301 and F481 identified as interaction sites.	56,81
Manual superposition[c]	6	n.a.		Inhibitors contain a tertiary nitrogen atom (protonated at physiological pH), a flat hydrophobic region, and two regions in which functional groups with lone pairs are allowed.	57
Catalyst	20 31	15 (9) 15 (10)	K_i K_i	Inhibitor models include a H-bond acceptor and a H-bond donor and two to three hydrophobic regions.	82
Catalyst and PLS MS-WHIM + homology model	24 52	28 (s.s.) n.a.	K_m	Substrate model includes one cationic site at 5.5 Å from a H-bond acceptor, and two hydrophobic regions, respectively at 8.7 Å and 8.3 Å from the H-bond acceptor and at 19° 'counterclockwise' from the cationic site, and 28° from the first hydrophobic region. D301 and F481 are involved in electrostatic and aromatic interactions, respectively.	69
ComFA	24	15 (s.s.)	K_m	Substrates possess a positive electrostatic site 5, 7 or 10 Å away from the site of catalysis.	83
CYP2E1					
CoMFA	12	n.a.	Cl_{intr}	Substrate model with long hydrophobic access channel.	43

Method[a]	Training[b]	Test[c]	Predicted property	Key Features of Pharmacophore	Ref.
GRID/GOLPE/ ALMOND/ + homology model + docking	42	39 (29)	K_i	Alignment independent inhibitor model includes two H-bond donors at 6 Å from each other, a hydrophobic region at 13 Å from a H-bond donor, and two H-bond acceptors at 12 and 15 Å from two a H-bond donors (on the protein).	77
	21	12 (11)	K_i		
Manual superposition[c] + homology model	27	4 (4)	CSP	Substrate model includes a H-bond acceptor site and an aromatic region, suggests R108 for electrostatic and F476 for aromatic interactions.	41
Catalyst	36	n.a. (s.s.)	HA	Heteroactivator model includes two hydrophobic regions at 11.6-11.8 Å from a H-bond acceptor and at 10.3-10.9 Å from an aromatic ring, and a H-bond acceptor 4.8 Å from an aromatic ring.	78
GRID/GOLPE/ ALMOND + homology model	43	n.a. (22-39)	CSP	Alignment independent protein-based inverse model, represents substrates as set of fingerprints for each H-atom present.	79
GRID/GOLPE/ ALMOND + homology model + docking	22	12 (11)	K_i	Alignment and conformer independent inhibitor model includes two H-bond donors, and three hydrophobic binding sites (on the protein).	61
CYP2D6					
Manual superposition[c]	15	n.a.	CSP	Substrates possess a basic nitrogen atom at 5 or 7 Å from site of catalysis, and coplanar aromatic rings.	48-50
Manual superposition[c]	156	n.a.	CSP	Carboxylate group in protein responsible for the 5 or 7 Å distances between basic nitrogen and site of catalysis.	51
Manual superposition[c] + homology model	10	n.a.	CSP	D301 identified as residue binding to basic nitrogen atom in ligands	80
Manual superposition[c] + homology model + molecular orbital calc.	40+14	7 (6)	CSP	Substrates have a basic nitrogen atom at 5, 7 or 10 Å from site of catalysis. Separate models for O-demethylation and N-dealkylation. E216, D301 and F481 identified as interaction sites.	56,81
Manual superposition[c]	6	n.a.		Inhibitors contain a tertiary nitrogen atom (protonated at physiological pH), a flat hydrophobic region, and two regions in which functional groups with lone pairs are allowed.	57
Catalyst	20	15 (9)	K_i	Inhibitor models include a H-bond acceptor and a H-bond donor and two to three hydrophobic regions.	82
	31	15 (10)	K_i		
Catalyst and PLS MS-WHIM + homology model	24	28 (s.s.)	K_m	Substrate model includes one cationic site at 5.5 Å from a H-bond acceptor, and two hydrophobic regions, respectively at 8.7 Å and 8.3 Å from the H-bond acceptor and at 19° 'counterclockwise' from the cationic site, and 28° from the first hydrophobic region. D301 and F481 are involved in electrostatic and aromatic interactions, respectively.	69
	52	n.a.			
ComFA	24	15 (s.s.)	K_m	Substrates possess a positive electrostatic site 5, 7 or 10 Å away from the site of catalysis.	83
CYP2E1					
CoMFA	12	n.a.	Cl_{intr}	Substrate model with long hydrophobic access channel.	43

3D-QSAR Pharmacophore Models Applications
One of the first COMFA models for CYP2C9 predicted two anionic sites, an aromatic region and a steric region near the site of metabolism as important features.[45] This model predicted 13 inhibitors correctly out of 14 and inclusion of additional substrates did not change the model appreciably.[75] Based on a homology model of CYP2C9, F110 and F114 were implicated in aromatic, and Arg105 and Asp293 in electronic interactions, respectively[75]. Using Catalyst and PLS-WHIM with three different sets of inhibition constants (K_i's), models for 2C9 have been constructed that included at least one hydrophobic region and at least one H-bond acceptor[43], similar to the previous COMFA model. A combined 3D-QSAR-homology model for 2C9 inhibitors, based on a training set of 21 inhibitors and using automated docking for conformer selection, was able to predict the K_i of eight other compounds to within 0.5 log units[76], as is shown in **Figure 6**. An inverse, protein-based, pharmacophore model of 2C9 was generated by the analysis of molecular interaction fields (MIFs) in active sites of the CYP2C isoforms 8, 9, 18, and 19 using consensus PCA (CPCA).[70] CYP2C8 was the most different, with changes at positions 114 (Ser in 2C8, Phe in the others), 205 (Phe in 2C8, Ile in 2C9/19, and Leu in 2C18), and 476 (Ile in 2C8, Phe in the others) causing different size and hydrophobicity of the binding pocket.
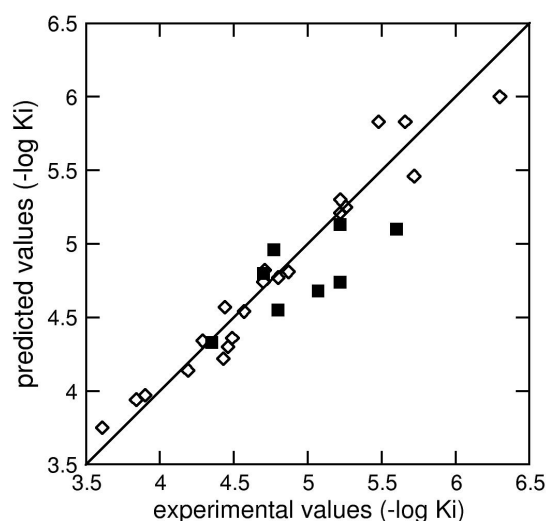
Figure 6: *Correlation for a combined CYP2C9 3D-QSAR-homology inhibitor model of Afzelius et al. [Mol. Pharmacol. 2001 **59** 909][76] between measured and predicted pK_i's, with $r^2 = 0.947$ and $q^2 = 0.730$. (triangles) training set (n=21); (squares) test set (n=8).*

Using alignment-independent GRIND descriptors in ALMOND, a CYP2C9 inhibitor model was constructed with two H-bond donors at 6 Å distance from each other, a hydrophobic region and a H-bond donor 13 Å from each other, and a nitrogen interaction point and a H-bond donor 12 to 15 Å from each other[77]. For construction and validation of this model, automated docking of template molecules in a CYP2C9 homology model was used for conformer selection of 35 inhibitors (21 for the training set, 14 for the test set) and 46 non-inhibitors (21 for the training set, 25 for the test set). 74% of the compounds were correctly predicted to be either inhibitor or non-inhibitor, and 11 out of 12 K_i's of inhibitors not used to derive the model were predicted within 0.5 log-units. An inverse alignment-independent protein-based CYP2C9 3D-QSAR pharmacophore model was combined with a homology model to represent substrates as a set of fingerprints for each H-atom present in the molecules.[79] This model was able to predict sites of catalysis of 39 out of 43 substrates. A Catalyst model based on 36 CYP2C9 heteroactivators contained one H-bond acceptor, one aromatic ring and two hydrophobic regions.[78] 94% of the training set of heteroactivators was predicted within 1 log unit. The model also correctly identified close

to 70% of inhibitors as high-affinity CYP2C9 binders, suggesting that heteroactivators and inhibitors share some common structural 2C9 binding features. Using ALMOND, a conformer- and alignment-independent CYP2C9 model based on 22 flexible inhibitors was built very recently by the use of flexible, alignment independent MIFs, and was able to predict the K_i's of 11 out of 12 other inhibitors to within 0.5 log units.[61]

For CYP2D6, a set of two Catalyst 3D-QSAR models has been created containing a H-bond acceptor and a H-bond donor and two to three hydrophobic regions[82]. Both models, based on a training set of respectively 20 and 31 inhibitors, were only capable of predicting 7 out of 15 K_i's of inhibitors not used to derive the model within 0.5 log units. A CoMFA model for CYP2D6, with an additional distance component included was based on a training set of 24 substrates with a basic nitrogen at either 5, 7, or 10Å from the site of catalysis predicted K_m's of 13 out of 15 other substrates to within 0.5 log unit.[83] A combined Catalyst 3D-QSAR/homology model for CYP2D6 has been constructed based on a training set of 24 substrates and an alignment from fluoxetine docked into the homology model using GOLD, contained one H-bond acceptor, two hydrophobic regions, and one cationic site.[69] The model was found statistically significant, however it could predict K_m's of only 14 out of 28 other substrates within 0.5 log units.

3D-QSAR Pharmacophore Models Conclusions

The recent 3D-QSAR models for several CYP isoforms show generalized features and molecular properties of CYP substrates/inhibitors. These features have been correlated with and used for the prediction of qualitative properties like site of catalysis, and they have been used for prediction of quantitative properties like (binding) affinity, as is illustrated in the QSAR correlation plot in **Figure 6**. **Table 1** lists for each of the current models reported, the key features as well as the relevant predicted substrate/inhibitor properties. Especially for CYP2C9, and also CYP2D6 and CYP3A4, several models with quantitative predictive value for K_i are available and for other isoforms often IC_{50} values and sometimes K_m values are predicted. In many cases the QSAR model was not validated against a test-set. Almost all recent classical and 3D-QSAR pharmacophore models have been improved by including details of the active site structure, and in turn they have been used to improve details of the alignment and structure of homology models for CYPs 17, 1A2, 2B6, 2C8/9/18/19, 2D6 and 3A4.

Alignment-independent 3D-QSAR models, as applied for CYP2C9, are not limited by a suitable template compound for superposition, so they can be constructed in a fairly automated fashion. In contrast, classical pharmacophore models must be built manually, and can only be used for qualitative predictions. Other recent developments in 3D-QSAR methods consider active site water molecules, not yet applied to CYP's, and the construction of conformer-independent models by the use of flexible molecular interaction fields (MIFs), very recently applied to CYP2C9. These developments will make it possible to build models without an alignment rule for chemically diverse ligands and without using information about the protein structure.

2.2.5 Conclusions Ligand-based Models

Intrinsic regio-specific susceptibilities of molecules to CYP metabolism have been predicted using available QM methodologies successfully for CYPs in general. Significant improvements in the prediction of site of catalysis have been reported for CYPs 101 (*cam*), 107A (*eryF*), 2E1, 2B4 and in ongoing work for CYP102 (*BM3*) by inclusion of known or predicted substrate binding orientations in the active site or of relative orientations in a pharmacophore model. Further improvements have been seen from including a description of the relevant heme-iron-oxygen activated reactive species at a higher level in the QM calculations, and of inclusion of more of the active site and protein surroundings that can have significant influence on electronic properties of the substrate mainly through

electrostatic interactions. Classical CYP pharmacophore models have been successful for the prediction of qualitative properties like substrate selectivity and site of catalysis, especially in the cases of CYP2C9 and CYP2D6, see **Table 1**. More recent 3D-QSAR pharmacophore models have in addition been used for the quantitative prediction (binding) affinity, especially for CYP2C9, as is listed in **Table 1**.

   The future role of pharmacophore/QSAR models seems to be to consolidate available experimental data on substrates and inhibitors and to combine these with structural details included in CYP crystallographic and homology model structures (see also the next **section 2.3.2** '*Protein Structure and Analysis*'). On the other hand, the use of alignment and conformer independent models make it possible to build models without alignment rules for chemically diverse ligands and without using information about the protein structure. Predictions based on QM calculations on the substrate will likely be used as a complementary source of information obtained from pharmacophore/QSAR models.

2.3 Protein Structure and Analysis

2.3.1 Introduction

A wealth of detailed information on CYP enzyme action and substrate or inhibitor interactions with the binding cavity becomes accessible when a structure of the CYP protein is available. The main source for these structures is X-ray crystallography, which are summarized in **Table 2**, but an even larger number of CYP structures have been built on the basis of these crystal structures using homology modeling techniques, which are summarized in **Table 3**. The progress in fields of X-ray crystallography and homology model building, as well as their mutual complementarity, will be illustrated and discussed in the following sections. Commonly used software for CYP homology model building and CYP protein structure analysis is listed in **Table 4**.

2.3.2 Crystallographic Protein Models

Crystallographic Protein Models Structure Determination

Crystal structures have been solved for several cytosolic bacterial (n = 77), cytosolic eukaryotic (n = 24), and, recently, also of membrane-bound mammalian (n = 13) CYP oxygenases (by the end of 2004). In addition to these 114 CYP oxygenase structures, including 3 CYP102 (*BM3*) oxygenase-reductase complexes, 6 individual CYP reductase structures have been solved. **Table 2** presents an overview of all currently available CYP oxygenase crystal structures, including information on X-ray resolution, bound ligands, mutants and heme oxidation state. In addition, several new structures have recently become available on the Protein Databank, most notably the human CYP3A4.[16] The progress in CYP crystallography has recently culminated in a fair number of a human CYPs, of which CYP2C9[14] was the first. **Figure 2** shows two ligand-bound structures of CYP2C9 (panels **A** and **B**), and a ligand-free structure of CYP3A4 (panel **C**).

   In several cases, different protein conformations have been found in crystal structures for the same isoform, and the question arises which of those is (the most) relevant for describing the catalytic properties of the enzyme. For example, in the first CYP2C9 crystal structure[14], the distance of the bound *S*-warfarin substrate to the heme iron is large (~10Å), see **Figure 2-A**, which makes it unlikely that this crystal structure corresponds to a catalytically active state. In line with this suggestion, Arg 105 and Arg 108, implicated in the formation of putative anionic-binding sites by mutagenesis, pharmacophore and homology modeling studies (see **sections 2.2.3, 2.2.4** and **2.3.3**) both point away from the active site. In a more recent substrate bound CYP2C9 crystal structure, however, the substrate *S*-flurbiprofen is within a catalitically reactive distance from the heme iron (~5Å), see **Figure 2-B**, and interacts with F114 and R108.[15] In the substrate-bound CYP102 (*BM3*) structures, unreasonably long substrate-heme distances have been found previously.[96,97] It has been suggested that protein conformational changes driven by electron-transfer trigger the movement of these substrates into effective

positions for hydroxylation.[98] In crystal structures of CYP101 (*cam*)[99,100] and CYP2C5[101], multiple ligand binding modes were observed. In CYP107A (*EryF*)[102] homotropic cooperativity of multiple ligands was observed. These effects probably result from ligand-induced conformational changes of the active site and the relatively large volume of the active site as compared to the volume occupied by the ligand, as described in more detail below in **section** *'Crystallographic Models Active Site'*.

Although crystallographic protein structures are firmly based on experimental data, it must be borne in mind that for resolutions of around 2Å and worse, the electron density maps are not sufficiently detailed to resolve individual atoms. In order to circumvent this problem, molecular modeling techniques are often applied and usually yield reliable structural models that better represent the measured diffraction patterns[103]. However, the likelihood of errors in the structure, like wrong rotamer states, mis-threading, or misplacement of secondary structure elements, increases rapidly with diminishing resolution. Furthermore, non-active site mutations introduced to facilitate crystallization of mammalian membrane-bound CYP crystal structures might cause significant conformational changes compared to crystal structures of less extensively modified constructs, as was recently observed for CYP2C9[15,16] and is clearly visible in **Figure 2-A** and **B**. In some cases, it is more appropriate to characterize these structures as crystallographic protein *models* to emphasize the distinction with atomic-resolution crystallographic *structures* – hence the title for this section: ''Crystallographic Protein Models. Like the homology models discussed in **section 2.3.3** *'Homology Based Protein Models'*, these crystallographic models warrant additional validation.

Crystallographic Protein Models Access & Exit Channels
A thermal motion pathway analysis of B-factors in the crystal structure of CYP101 (*cam*) with camphor bound revealed the existence of three distinct pathway families. Exit channels were explored by molecular dynamics (MD) simulations by imposing an artificial expulsion force on the substrate in addition to the standard MD force field.[104] This random expulsion molecular dynamics (REMD) method was later also tested in simulations of the palmitoleic acid bound structure of CYP102 (*BM3*), a system for which a substrate-access channel is closed in the substrate-bound and open in the substrate-free crystal structure[105] and the closing motion has been reproduced by MD simulations.[106,107] It was applied to complexes of CYP101 (*cam*) with camphor, endo-borneol allyl ether and 5-hydroxy camphor (product of camphor). The three ligand expulsion pathways of CYP101 observed by REMD matched well with three putative ligand channels found by thermal motion pathway analysis.[108] In a complementary study, estimates of the relative probabilities of ligand passage through the three channels were computed using steered molecular dynamics (SMD) and adiabatic mapping.[109] The REMD and SMD approaches combined with essential dynamics (ED) analysis were used for 6-deoxyerythronolide B bound CYP107A (*EryF*), and the results obtained for this complex were compared to those for CYP101 (*cam*) and 102 (*BM3*).[110] Several possible exit/entrance routes in each enzyme were identified and one route, located between the F/G loop, the β1 sheet and either the B' helix (CYP101 and 102) or B'/B loop (*EryF*) (see **Figure 4**), was common to all three proteins. The mechanisms of substrate entrance and/or product exit along this pathway, however, were different for each of the three isoforms and apparently adapted to the properties of the substrate. Based on this observation, it is suggested that the characteristics of the active site itself may not be the only determinant of substrate specificity. The access channel properties may thus play a key role by kinetic modulation of substrate recognition.[110] Very recently, open channels leading to the active site of currently available CYP crystal structures (see **Table 2**) were analyzed, providing experimental evidence for the existence of four substrate exit/entrance routes that were identified by MD simulation.[111]

Table 2: *Overview of all currently available CYP oxygenase crystal structures (by the end of 2004).*

CYP Isoform	Species[a]	Resolution (Å) [b]	Ligands[c]	Mutants[d]	Fe state[e]	PDB-code [f]
Soluble bacterial crystal structures						
101 (cam)	*Pseudomonas putida*	1.6 (1.55- 2.3)	R-camphor, S-camphor (mb), 5-exo-hydroxycamphor, adamantane, adamantanone, camphane, norcamphor, thiocamphor (mb), 1-, 2- & 4-phenylimidazole (cv), metyrapone (cv), nicotine (mb,cv), phenyl radical (cv), N-butyl-isocyanide (cv), sulphyhydryl reagent (cv), 1,3,5-trichlorobenzene, chiral inhibitor (cv), Ru-diimine photosensitizer, D-8-Ad, (+)-α-pinene (mb)	F87W/Y96F/ V247L/C334A, F87W/Y96F, R109K, R112K, D251N, T252I, T252A, C334A L358P/C334A	(II), (III), O_2, (IV)-O, CN, CO	2CPP[112]
102 (BM-3)	*Bacilus megaterium*	1.65 (1.65– 2.7)	Palmitoleic acid, N-palmitoylglycine	A264E, T268A, F393H	(III)	1BU7[113]
107A1 (EryF)	*Saccharopolyspora erythrea*	2.10 (2.10-2.35)	9-aminophenanthrene (hc), androstendione (hc), ketoconazole (cv), 6-deoxyerythronolide B	A245S	(III)	1JIO[114]
108 (terp)	*Pseudomonas*	2.3	-	-	(III)	1CPT[115]
119	*Sulfolobus solfactaricus*	1.50 (1.93-2.69)	Imidazole (cv), 4-phenylimidazole (cv)	-	(III)	1IO7[116]
	Sulfolobus Tokodaii	3.00	-		(III)	1UE8[117]
152A1	*Bacilus subtilis*	2.10	Palmitoleic acid		(III)	1IZO[118]
154A1	*Streptomyces coelicolor A3(2)*	1.85	4-phenylimidazole (cv)		(III)	1ODO[119]
154C1	*Streptomyces coelicolor A3(2)*	1.92	-		(III)	1GWI[120]
158A2	*Streptomyces coelicolor A3(2)*	1.50	4-phenylimidazole (cv)		(III)	1S1F[121]
175A1	*Thermus thermophilus*	1.80	-		(III)	1N97[122]
	Thermus Thermophilus Hb8	2.00	-		(III)	1WIY
EpoK	*Sorangium cellulosum*	2.65 (1.93-2.10)	Epothilone B, Epothilone D		(III)	1Q5E[123]
OxyB	*Amycolatopsis orientalis*	1.70 (1.70-2.20)	-	-	(III)	1LFK[124]
OxyC	*Amycolatopsis orientalis*	1.90	-	-	(III)	1UED[125]

CYP Isoform	Species[a]	Resolution (Å)[b]	Ligands[c]	Mutants[d]	Fe state[e]	PDB-code[f]
Soluble eukariotic crystal structure						
51 (14α)	*Mycobacterium tuberculosis*	2.05 (*1.55-2.21*)	4-phenylimidazole (cv), fluconazole (cv), estriol	C37L/C151T /C442A	(II), (III)	1H5Z[126]
55A1 (nor)	*Fusarium oxysporum*	1.00 (*1.50-1.60*)	N-butyl-isocyanide (cv)	T243A, T243N, T243V, S286V, S286T, S73G/S75G	(II), (III), CN, CO	1JFB[127]
121	*Mycobacterium tuberculosis*	1.06 (*1.80*)	Iodopyrazole	-	(III)	1N4O[128]
Soluble mammalian crystal structures						
2A6	*Homo sapiens*	*1.90 (2.05)*	Coumarin, methoxsalen			1Z10[129]
2B4	*Oryctolagus cuniculus (rabbit)*	*1.60 (1.90)*	4-(4-chlorophenyl)imidazole	(d)	(III)	1PO5[130]
2C5/3	*Oryctolagus cuniculus (rabbit)*	3.00 (*2.10-2.30*)	Diclofenac, 4-methyl-N-methyl-N-(2-phenyl-2H-pyrazol-3-yl)benzenesulfonamide (mb)	(d)	(III)	1DT6[131]
2C8	*Homo sapiens*	*2.70*	Palmitic acid[g]	(d)	(III)	1PQ2[132]
2C9	*Homo sapiens*	*2.60 (2.0-2.55)*	S-warfarin, S-flurbiprofen	(d)	(III)	1R9O[15]
2D6	*Homo sapiens*	3.0	-	(d)	(III)	2F9Q[133]
3A4	*Homo sapiens*	2.05 (*2.65-2.74*)	Progesterone[g], metyrapone	(d)	(III)	1TQN[134]

a) Generally, CYPs were cloned and expressed in *E. coli*. CYPs 101, 108, and 55A1 were initially produced from their natural organisms.
b) Highest resolution of ligand-free enzyme. For CYPs 107A1, 51, 55A1 and 152A only ligand-bound structures have been solved, here the highest resolution is in *italics*. The range of resolutions for all ligand-bound structures is shown in *italics* in brackets.
c) Binding in multiple binding modes (mb), covalently bound (cv), or via homotropic cooperativity (hc) is indicated.
d) Of all CYP isoenzymes mentioned, at least the wildtype was crystallized. Listed are active site mutants. For mammalian CYP 2A6, 2B4, 2C5/3, 2C8, 2C9, 2D6 and 3A4 crystal structures, non-active site mutations (in the N-terminus, and in the case of CYP 2C5/3 and some CYP 2C9, 2D6 and 3A4 structures, also in the F'-G' region) are present to improve crystallization.
e) Unless stated otherwise, the heme iron atom is assumed to be in a ferric state (Fe(III)).
f) Reference of ligand-free enzyme structure with highest resolution.
g) Two ligand molecules bound in the dimer interface.

Crystallographic Protein Models Active Site

Helms and Wade calculated the free energies of hydration of two cavities in the active site of two liganded complexes of CYP101 (*cam*) by multiconfigurational thermodynamic integration (TI) calculations.[135] Based on computed free energy differences, they concluded that a crystallographically well-defined water molecule mediating H-bonds between 2-phenylimidazole and CYP101 is energetically favorable, while solvation of camphor bound CYP101 is energetically unfavorable. Free energy calculations for having five to eight water molecules in the active site cavity of the unliganded enzyme showed that occupation by five to six water molecules is thermodynamically most favorable, which is in agreement with estimates from crystal structures and studies employing hydrostatic pressure. Potential of Mean Force (PMF) calculations were used to identify functional water channels in CYPs 101 (*cam*), 102 (*BM3*), 108 (*terp*), and 107A (*EryF*)[136] and proton-transfer pathways in CYP101.[137]

Recently, ligand induced changes in the active sites of several proteins, among which CYP101 (*cam*), were analyzed using classical molecular interaction potentials, in conjunction with protein geometry and protein cavity calculations.[138] It was found that the binding of CYP101 ligands introduces small perturbations at the binding site of the proteins, and in some structures of CYP101 these perturbations result in important alterations in the recognition pattern for ligand binding. Comparisons between ligand-bound and ligand-free crystal structures of other isoenzymes, *i.e.* CYPs 102 (*BM3*)[105], 107A (*EryF*)[114], 119[139], 51 (*14α*)[140], and 2C5[101,141], showed ligand induced changes in size, shape and hydration of the active site. This conformational flexibility of the active site is likely to underlie the capacity of many CYP isoenzymes to metabolize structurally diverse ligands of different sizes and to bind ligands in multiple binding modes (generating different metabolic products).
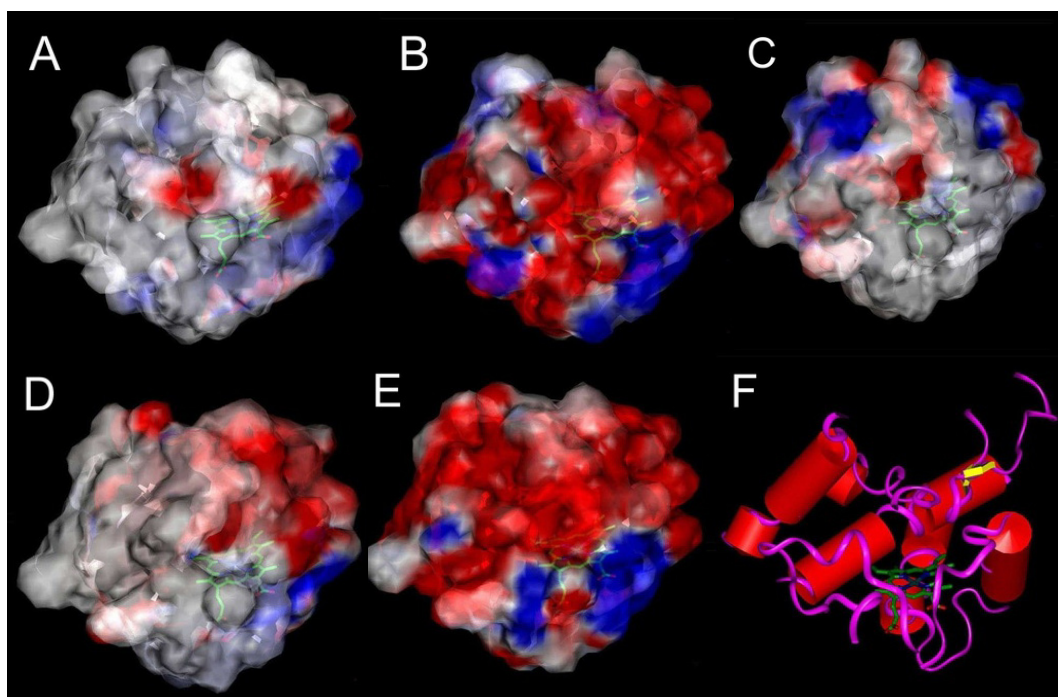
Figure 7: *Electrostatic potential surfaces of the active site of (A) CYP2D1; (B) CYP2D2; (C) CYP2D3; (D) CYP2D4; and (E) CYP2D6. Red, blue and white indicate negative, positive and neutral electrostatic potentials, respectively. The electrostatic potential surfaces are shown in the same orientation as the protein secondary structure shown in (F), with the heme-group at the bottom of the active site. Modified after Venhorst et al. [J Med Chem 2003 **46** 74][10] with permission.*

In human and rat CYP2D isoenzymes, electrostatic potential calculations on the binding pocket residues showed large differences in the negative charge of active sites of different isoenzymes, as is shown in **Figure 6**. These differences correlated with observed IC_{50}'s[10]. This observation provided novel insights into differences in active site topology of human and rat 2D-isoforms and on the validity of the rat as a model for human CYP2D6 activity.

Crystallographic Protein Models Conclusions
The first human CYP crystal structure (*i.e.* 2C9) together with some other crystal structures of mammalian CYPs (rabbit 2B4 and 2C5 and human 2A6, 2C8, 2D6 and 3A4) provide a still growing basis for protein homology modeling work. Examples of crystal structures are shown in **Figure 2**, for CYPs 2C9 and 3A4. **Table 2** presents an overview of all currently available CYP oxygenase crystal structures (by the end of 2004) and the growth in available CYP crystal structures is shown in **Figure 8**. Especially in the cases of CYP101, 102, and 107A, these structures have successfully been used as the basis for studies on structural and dynamic properties of active site hydration, substrate access and binding, and product exit. Structural differences between substrate-free and substrate-bound CYP isoenzymes, as seen for the crystal structures of CYPs 101, 102, 107A, 119*,* 51, 2B4, 2C5, and 2C9, indicate that conformational flexibility of the active site is likely to undelie the capacity of many CYP isoenzymes to metabolize structurally diverse substrates of different sizes and to generate different metabolic products from these substrates. Furthermore, CYP crystal structures are used as templates for CYP homology modeling. The most recent homology models are based on the rabbit CYP2C5 structure, as shown in **Table 3** and described in the **section 2.3.3**, and it is expected that new homology models will be based on the new mammalian structures (*e.g.* CYPs 2A6, 2B4, 2C8, 2C9, 2D6 and 3A4).
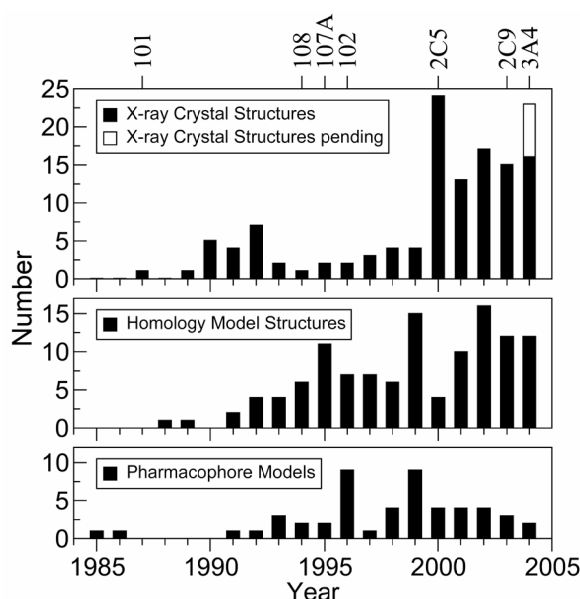
Figure 8: *The number of new crystal structures, homology models and pharmacophore models published over the years. The overall steady increase is clearly visible, as is the increase in homology and pharmacophore models during (temporary) declines in number of crystal structures, suggesting the complementary nature of experimental structure determination and model building. The structures of isoforms that have been or will be extensively used for homology model building are indicated at the top.*

When using crystal structures for CYP-modeling studies, special attention should be given to the state in which the CYP was crystallized, as this may not correspond to a

catalytically active state. It is possible to combine such a structure with a relevant active site loop structure from a related CYP crystal structure in which the loop is known to be in an active state. Although the resulting protein model would often be classified as 'only' a homology model, when rigorously validated it could be more relevant for predicting enzyme activity than the crystal structure.

2.3.3 Homology-based Protein Models

Homology-based Protein Models Methods Background
In addition to X-ray crystal structures, CYP protein homology model structures can be a valuable source of detailed information on binding cavity characteristics and other details of the ligand-protein interactions. In addition, a variety of biochemical, spectroscopical and mutational data may be incorporated during the model building process, providing a powerful way to consolidate and rationalize the different experimental insights into the structure and function of the CYP enzyme. Commonly used software for CYP homology model building is listed in **Table 4** and a selection of the many homology models built so far based on available CYP crystal structures (by the end of 2004) is presented in **Table 3**.

Building a homology model of a protein is a highly iterative process involving (often many) cycles of sequence or structural alignments, and model building, and analysis and validation, as is depicted in the flowchart in **Figure 9.**[142] Starting with the sequence of the target CYP protein, homologous sequences are found by scanning against sequences of proteins in the Protein Databank using one of the many search algorithms available, such as PSI-BLAST, more are listed in **Table 4**. This is step 1 in the top cycle in **Figure 9**. Often first a multiple sequence alignment is performed against homologous sequences of the CYP subfamily, as available from sequence databases, which is step 2 in this cycle. The sequence alignment of templates must then be validated, *e.g.* using mutational studies and isoenzyme specificity data to verify if the location of essential amino acids is correct, which is step 3 in this cycle, and if necessary it has to be further refined by structural alignments, step 4 in the top cycle in **Figure 9.** Although all CYPs share the same fold, variations in length and position of secondary structure elements between isoforms make structural alignments less than straightforward. Alignments can be carried out in two different ways[7]. The first method is to align the sequence according to a superposition of available protein crystal structures. Using this approach, Hasemann *et al.* made the first CYP structural sequence alignment using crystal structures of CYPs 101 (*cam*), 102 (*BM3*), and 108 (*ter*).[8] The second method focuses on local structural similarities between equivalent structure elements. Jean *et al.* developed an algorithm to define the so-called 'common structure building blocks' (CBS's) by comparing torsion angles of protein backbones[143]. From the same three CYP crystal structures 15 CBS's were derived, consisting of most CYP secondary structure elements except helices A and B' (see also **Figures 2** and **3**). Very recently, other methods for homology model building have been evaluated for CYP2C5 and 2C9 and concluded to depend critically on the alignment strategy, on the quality of the template crystal structure, as well as on the optimization procedures applied.[63] Use of manually optimized multiple sequence alignments corroborated with experimental data (most notably mutagenesis data) generally improves the models, as does careful optimization using molecular dynamics simulations.

Once the templates have been aligned, step 1 in the central cycle and step 4 in the top cycle in **Figure 9**, the amino acid sequence of the target protein is aligned with them, this is step 2 in the central cycle**,** and this alignment must be validated using mutational studies to verify that the location of essential amino acids is correct, which is step 3 in the central cycle. For the parts of the target protein that have the best alignment with the template structure, or one of the template structures, the coordinates of the backbone of the target are taken from the homologous parts of the backbone of the template. Side

chain coordinates are transferred as well using the 'maximum overlap' principle to keep the coordinates of all atoms from the template residue sidechain that have topologically corresponding atoms in the target. Missing sidechain atoms are added, and conformations are generated for inserted residues and loops and, finally, the structure can be optimized using, *e.g.,* energy minimization, simulated annealing, or 'room temperature' MD simulations. Based on basic protein structure quality parameters (*e.g.* Ramachandran distributions, packing quality) one of several generated model structures is selected. Together this constitutes the 'model building' step, number 4 in the central cycle in **Figure 9**.
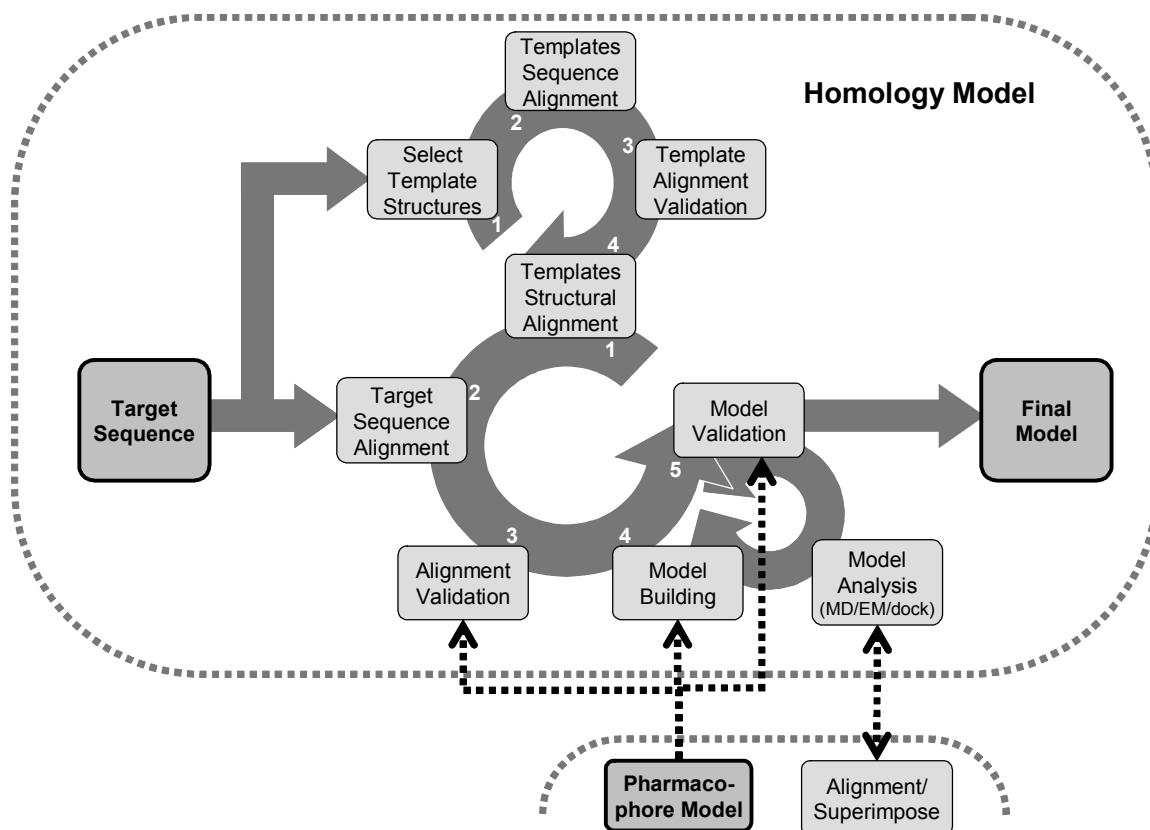
Figure 9: *Flowchart for protein homology modeling modified after Kirton et al. [Adv Drug Deliv Rev 2002 **54** 385].[142] The highly iterative processes in alignment, model building, model analysis and validation, involving three connected cycles, are visible. Details from a pharmacophore model can be used to optimize sequence alignments, to choose arrangements of certain important groups during model building and for final validation.*

The 3D homology model of the protein structure is validated as much as possible with experimental data, like mutagenesis data, NMR spin-relaxation measurements[10,98], active site chemical probe studies[144] and predictions of CYP substrate and/or inhibitor binding and regioselective metabolism and isoenzyme specificity, using for example automated docking (**section 2.4.2**) and MD simulations (**section 2.4.3**), which is step 5 in the central cycle in **Figure 9**. Depending on the quality of the model as indicated by the validations used, it will be refined by additional cycles of structure optimization and repeated model building, which is the bottom-right cycle in **Figure 9** and also involves steps 3 and 4 of the central cycle, as well as by improving the alignment with the template sequence(s) and/or structure(s) (top cycle) and repeating all steps in the central cycle.

Table 3 shows a selection of the many homology models built so far based on available CYP crystal structures (by the end of 2004). Especially in the last few years the use of

homology modeling has increased considerably, see **Figure 8**. For each CYP isoform only the most recent and most properly refined and validated homology models are included in this review. The dependency of homology model building on the availability of (quality) crystal structures is obvious, the correlated growth in published crystal structures and homology models is clearly visible in **Figure 8**. In total we counted about 144 homology models published of which 62 are included in **Table 3**, and nearly all of the earlier models, until 2000, were discussed by Ter Laak *et al.*[44]. The progress in homology modeling will be demonstrated below by the case of CYP2D6.

Homology-based Protein Models Applications
The first preliminary homology model for the human CYP2D6, containing only active site regions of the protein (11 segments) indicated D301 as an important amino acid residue for catalytic activity.[145] For one of the first complete CYP2D6 homology models, a structural sequence alignment of the crystal structures of bacterial CYP101, 102, and 108 isoenzymes[8] was aligned with a multiple sequence alignment for members of the CYP2-family using site-directed mutagenesis data.[4] Three known substrates and one inhibitor were docked into the active site of the CYP2D6 model[4] indicating that the protein model was able to accommodate large substrates and inhibitors. The docked conformations indicated a crucial H-bond with the amino acid D301, which was in turn used to improve the previously described pharmacophore model for CYP2D6 substrates (see **sections 2.2.3** and **2.2.3** on 'classical' and 3D-QSAR pharmacophore models, respectively). Two amino acids for which site-directed mutagenesis data were available, namely D301[54] and Val374 were part of the active site of the protein model[4]. Later, a cluster of 51 manually superimposed and energy minimized substrates was docked and evaluated in a refined model explaining 72 metabolic pathways catalyzed by CYP2D6[56]. It appeared that this model could predict correctly six out of eight metabolites observed in a test set of compounds. A CYP2D6 model was published together with models of CYP1A2 and CYP3A4 based on four bacterial crystal structures[146]. In total, 14 CYP2D6 substrates and 4 nonspecific substrates known to be metabolized by 2D6 were successfully automatically docked into the active site. Almost all substrates had important Van der Waals interactions with Val307, F483, and Leu484, whereas D301 was always involved in charge-reinforced H-bonds with the protonated nitrogen atom of the substrates. A novel approach for the construction of CYP2D6 protein models was introduced by Modi *et al.*[98], who used NMR spin-relaxation rate derived distance restraints for codeine and MPTP to guide the homology modeling process.

Recently, 4 different sets of comparative models of CYP2D6 (including the first 2D6 homology models based on rabbit CYP2C5, the first mammalian CYP crystal structure[131]), were constructed using different structural alignments of combinations of 4 bacterial and/or CYP2C5 crystal structure templates[142]. Selection of the best model was based on stereochemical quality and side chain environment. The resulting model was further refined and validated by automated docking in combination with NMR-derived distance restraints mentioned above[98,147] and GRID/GOLPE-analysis of the active site. This latter analysis suggested that both D301 and E216 are required for metabolism of basic substrates, as E216 was identified as a key determinant in the binding of the basic moiety of substrates while D301 might exert indirect effects on protein-ligand binding by movement of the B'-C loop and creating a net negative charge together with E216. This hypothesis was later confirmed by site-direct mutagenesis studies which showed that neutralizing both E216 and D301 alters CYP2D6 substrate recognition.[55]

More recently, a new homology model of CYP2D6 based on rabbit CYP2C5 was reported[10]. The model was validated on its ability to *i*) accommodate codeine in the binding orientation determined by NMR by Modi *et al.*[98], *ii*) reproduce the binding orientations corresponding to metabolic routes of 10 substrates with rigid docking studies,

iii) remain stable during unrestrained molecular dynamics at 300 K and *iv*) reproduce inhibition characteristics of 11 CYP2D6 model-ligands. Analogous homology models were generated for the rat CYP2D-isoforms, 2D1/2/3/4.[10] Electrostatic potential calculations on the binding pocket residues showed large differences in the negative charge of active sites of the different isoenzymes, as is shown in **Figure 7.** These differences correlated well with measured IC50's.[10] Interestingly, in addition to D301 and E216, this study also indicated F483 and F120 as key interaction residues. The importance of F120 in CYP2D6 metabolism was recently confirmed by site-directed mutagenesis studies.[18,148] Very recently, a substrate-free CYP2D6 crystal has been solved.[133] **Chapter 3** describes the construction, refinement and validation of the CYP2D6 homology model used in this thesis, and its comparison with the recently solved CYP2D6 crystal structure.[133]

In summary, protein homology models of CYP2D6 have provided a detailed insight in the overall protein structure as well as in the active site topology. The availability of the first crystal structure of a membrane bound CYP, *i.e.* the rabbit CYP2C5, as well as more advanced computational modeling techniques, *e.g.* molecular dynamics simulations, have significantly improved the predictive quality of the resulting protein models. The very recent human CYP2C9 crystal structure has so far only been used in the construction of a new CYP2B4 model.[40] Homology models can be already very suitable to predict specificity and sites of metabolism for CYP substrates, but the quantitative prediction of (binding) affinities and turnover rates remains a challenge for the near future.

Homology-based Protein Models Conclusions
As shown in **Table 3,** the trend in CYP homology model building is clearly towards a multi-integrated approach using multiple sequence alignments, (multiple) structural alignments, in combination with mutational, spectroscopic and enzyme kinetic experimental data and new or available pharmacophore models, automated docking, and molecular dynamics (MD) simulations. Mutational data and experimental knowledge of SRS's are very frequently used to optimize sequence and structural alignments and to validate final homology models. Recent homology models show good agreement with site-directed mutagenesis studies, especially in the case of CYP2D6 where F120, F483, E216, and D301 were indicated as key determinants of selectivity and regiospecificity of substrate binding and catalysis by this isoenzyme. Several cases are known in which CYP protein homology models are combined with pharmacophore models (*i.e.* CYPs 1A2, 2B6, 2C8/9/18/19, and 2D6) or QM calculations (*i.e.* CYPs 2E1 and 3A4). In recent studies, for nearly all CYP isoforms, predictions of site(s) of catalysis in substrates, substrate-selectivity, isoenzyme specificity, *e.g.*, CYPs 1A2, 2D6 and 3A4[146], 2B1/4/5[149], and 2C8/9/18/19[70], and ligand (binding) affinity by automated docking and MD simulations have been used during the model building process to evaluate intermediate stages, as well as for a first validation of the final model. An upcoming and potentially powerful approach is the use of MD simulation techniques for optimization of the final model, *e.g.* for CYP2B6.[150] Although CYP homology models are rarely used for quantitative binding affinity predictions (*e.g.* CYP1A1, as described in **section 2.4.4**), CYP homology models for many isoforms have been used successfully for the prediction of substrate selectivity and site of metabolism by automated docking and MD (see **sections 2.4.2** and **2.4.3**, respectively).

2.3.4 Conclusions Protein Structure and Analysis
The growth in available CYP crystal structures, as presented in **Table 2** and **Figure 8**, has inspired many studies on detailed analysis of CYP structures and of CYP-ligand interactions, and it has been the basis for much homology modeling work (**Table 3** and **Figure 8**). In that respect, homology modeling has been complementary to X-ray crystallography, especially for mammalian CYPs where crystal structures have become available very recently, *i.e.* rabbit CYPs 2B4 and 2C5, and human CYPs 2C8, 2C9 and 3A4. Examples of crystal structures of CYPs 2C9 and 3A4 are shown in **Figure 2**.

Table 3: Selection of 62 of the 144 CYP homology models built so far based on available CYP crystal structures (by the end of 2004). For each CYP isoform the most recent and most properly refined and validated homology models are included. For CYP2D6, all homology models are included.

CYP Isoform	CYP Template(s)	Predicted property[a]	Refinement and validation[b]	Remarks	Ref.[c]
1A1	101				151
	2C5	$CSP^{(c)}$, K_m	MD, SM		152
1A2	102	CSP, mutagenicity	AD, P^{64}, SM	Evaluation of alignment strategies	153
	101/102/107A/108	CSP, IS	AD	Comparison with 2D6 and 3A4. Suggestion for membrane attachment	146
	2C5	$CSP^{(c)}$		Rationalization of inhibition mechanisms	154
1B1	2C5	$CSP^{(c)}$, $K_m^{(d)}$	SM		155
2A1/2A4/2A5	102	$CSP^{(c)}$, $K_m^{(d)}$, IS	SM		156
2A6/2A13	2C5	$CSP^{(c)}$, IS	SM		157
2B1/2B2	101/102/108	$CSP^{(c)}$	SM^{158}	2B2 model built from 2B1 template	159
2B1/2B4/2B5	2C5	IS	SM		149
2B4	101/102/107A1/108		SM, MD	Suggestion for binding regions with redox partners	160
	2C5/2C9	K_s	SM, MD, QM	Comparison of models derived from 101, 2C5 or 2C9	40
2B6	101/102/107A1/108	CSP, IS	SM		161
	102/2C5	CSP	AD, MD	Bound and unbound forms of 102 as additional templates for the F-G-loop	150
	2C5	CSP, K_m	AD, P		69
2C8/2C9/2C19	2C5	CSP, IS	AD, P	Comparison of 2C active sites	71
2C8/2C9/2C18/2C19	2C5	CSP, IS	AD, P	Comparison of 2C active sites	70
2C9	102	$CSP^{(c)}$, IS	MD, SM		162
	102/2C5	$CSP^{(c)}$	P	B'-region from 102	41
2C18/2C19	102	$CSP^{(c)}$, IS	MD		163
2D6	101	$CSP^{(c)}$		Preliminar model of active site regions (11 segments)	145
	101/102/108	$CSP^{(c)}$	P		4,56,81
	101/102/108	CSP		Refinement and validation using NMR-derived ligand-heme distance restraints	98,147
	101/102/107A1/108	CSP, IS	AD	Comparison with 1A2 and 3A4. Includes suggestion for membrane attachment	146
	101/102/2C5	CSP, K_m	AD, P		69
	101/102/107A1/108/2C5	CSP, $IC_{50}^{(e)}$	AD^{148}, SM^{55}	Evaluation of alignment strategies	142
	2C5	CSP, IS	AD, MD	Comparison with rat 2D1/2/3/4 models built on 2D6. Refinement and validation using NMR-derived ligand-heme distance restraints	10

CYP Isoform	CYP Template(s)	Predicted property[a]	Refinement and validation[b]	Remarks	Ref.[c]
2E1	102	CSP[c], IS	SM		164
	2C5/2C9	CSP			165
3A4	101/102/107A1/108	CSP	AD, MD, QM[39,166,167]	Active site chemical probe studies[168]	169
	101/102/107A1/108	CSP, IS	AD	Comparison with 1A2 and 2D6. Suggestion for membrane attachment	146
	2C5	CSP, K_m[e]	AD, P		71
4A1/4A4/4A11	102	CSP[e]	SM		170
4A11	102	CSP[e]	MD		107
4F11/4F3A	101/102/107A1/108/2C5	IS			171
6B1	102	CSP	AD, SM[172]		173
11A (SCC)	102		AD, MD, SM	Protein-protein docking with adrenodoxin redox-partner	174
17 (17α)	101	CSP[e]	MD, SM		175
	108	CSP[e]	MD		176
19 (arom)	102	CSP[e]	SM		177
	2C5	CSP[e]	SM		178
51 (14α)	101		SM		179
	101/102/107A1	CSP[e]	MD		180
27A1	102		SM	Suggestion for membrane attachment. F-G-loop from Arc Repressor.	181
27B1	2C5		SM		71
73A1	101/102/107A1/108	CSP[e]	SM	Refinement and validation using NMR-derived ligand-heme distance restraints	182
105C1A1/B1	101	CSP[e]		Models with different alignments.	183
105C1	101/102/107A1/108		MD		184
119	101/102/107A1/108/55A1		MD		185

a) Predicted property: 'Affinity' (Michaelis-Menten constant K_m or spectral dissociation constant K_s); Catalytic Site Prediction (CSP); Isoenzyme Specificity (IS)
b) Method of prediction and validation: Automated docking (AD) of substrates in homology model to predict site of catalysis; Molecular dynamics (MD) simulations to predict site of catalysis or affinity (LIE method); Combined with pharmacophore model (P); Site-directed mutagenesis (SM) studies used for refinement and validation of sequence alignment and/or homology model; Quantum mechanics (QM) calculations to predict site of catalysis.
c) Catalytic site prediction using manual docking in combination with energy minimizations and/or molecular dynamics (MD) simulations
d) Affinity prediction using manual docking in combination with energy minimizations and 1D-QSARs
e) Affinity prediction using automated docking

Several CYP isoforms have only been crystallized in a non-active ligand-bound state, and for some it is not known whether the crystal structure corresponds to an active state or not. Ligand-induced changes can be observed in the size, shape and hydration of the active sites of several CYPs, *i.e.* 101, 102, 107A, 119*,* 51, 2C5, and 2C9 (see **Figure 2-A** and **B**) and these are likely to undelie the capacity of many CYP isoenzymes to metabolize structurally diverse substrates of different sizes and to generate different metabolic products from these substrates.

The trend in CYP homology model building is clearly towards an integrated approach of multiple sequence and structural alignments, combination with new or available pharmacophore models, automated docking, molecular dynamics (MD) simulations and validation with mutational, spectroscopic, structural and enzyme kinetic experimental data. Notably, recent models of CYP2D6 show very good agreement with site-directed mutation studies. Predictions of substrate selectivity, site of metabolism and isoenzyme specificity have been successful using homology models of many CYP isoforms.

2.4 Ligand-Protein Interactions
2.4.1 Introduction
One of the critical issues in predicting CYP metabolism is the combination of statistical, orientational and energetic factors that each contributes to various degrees in different CYP enzymes. A first step is made by automated docking programs that aim to predict energetically favorable binding conformations of ligands in the active site cavity. Further exploration of binding conformations and relative probabilities can be done using molecular dynamics simulations, which can also be used for predictions of binding affinities. Commonly used software for study of CYP-ligand interaction predictions is listed in **Table 4**.

2.4.2 Automated Ligand-Protein Docking
Automated Ligand-Protein Docking Methods Background
Automated ligand-protein docking methods predict energetically favorable conformations and orientations of ligands in the binding pocket of a protein. Algorithms to generate different poses (docking) are combined with scoring functions that consider the tightness of the protein-ligand interactions[186]. When a structural model of a CYP isoenzyme is available, from X-ray crystallography or homology model building, docking methods can be used for predictions of the catalytic site and binding affinity for CYP ligands (substrates and inhibitors), and to identify potential CYP substrates in a chemical database (protein-based virtual screening). In addition, automated docking can be used to generate input structures for pharmacophore modeling (see **sections 2.2.3** and **2.2.4**) and MD-simulations (see **section 2.4.3**). The heme prosthetic group, the presence of water in the active site, the relatively large size and large conformational flexibility of the active site, the broad range of substrates and many catalytic sites in most substrates, as described in **section 2.3.2**, make CYPs rather challenging docking targets.

Several docking algorithms and scoring functions have been described in the past few years, the most commonly used for CYPs are GOLD, FlexX, DOCK and AutoDock, and the scoring functions from C-Score (TRIPOS Inc., St. Louis); more are listed in **Table 4**. Docking accuracy, *i.e.* structure prediction, and scoring accuracy, *i.e.* prediction of binding free energy, of docking-scoring combinations vary considerably with the selected target protein and physicochemical details of target-ligand interactions[187], and even depend on fine details of the CYP protein structure, as shown for CYP101 (*cam*)[188]. Therefore, a docking-scoring strategy must be tailored to the system of interest, based on a test-set of ligand-bound protein crystal structures. For CYP crystal structures, such a methodological evaluation was performed very recently[18] and so far has not been explored for CYP homology models. Water molecules can play an essential role in ligand-protein binding[189],

and water-mediated ligand-protein interactions have been observed in crystal structures of most CYP isoforms, *e.g.* rabbit CYP2C5[101,141] and human CYP2C9.[14,15] Most docking tools, however, ignore the possibility of water-mediated interactions between protein and ligand.[190,191] Ligand-induced changes in the active sites of CYPs can lead to important alterations in ligand recognition patterns (see **2.3.2** '*Crystallographic Protein Models Active Site*' and are accounted for in some new methods[190,192] but the inclusion of full protein flexibility in automated ligand-protein docking gives an enormous increase in complexity[186,190], and until now no such method has been validated extensively.[191]

Automated Ligand-Protein Docking Applications
Automated docking has primarily been applied to refine and validate CYP pharmacophore models and protein homology models (see **Table 3**), and manual docking has been extensively used for this purpose as well in the past. Validation is achieved by comparing predicted binding modes of substrates with reported metabolic product(s), and was used during the homology modeling process of CYPs 1A2[146], 2B6[69,150], the 2C subfamily[70,71], 2D6[10,142,146], 2E1[165], 3A4[71,146,169], and 6B1.[173] In **Figures 4** and **3-A** dextromethorphan is shown in its docked binding orientation in the binding cavity of the CYP2D6 homology model.

Experimentally determined binding affinities of 11 different CYP101(*cam*)-ligand complexes showed no clear correlation with values from different C-Score scoring functions on CYP101(*cam*)-ligand crystal structures nor on complexes produced by the automated docking program FlexX.[193] Nevertheless, the docking and scoring could still be used to prioritize *in silico* screening hits of a chemical database. For wildtype CYP101 (*cam*) DOCK was used for *in silico* screening to identify potential substrates[194], and imidazole inhibitors.[59] The automated docking program GOLD was recently used to predict binding of octane, octanoic, and lauric acid to CYP102 (*BM3*) (Feenstra *et al.*, unpublished results). Predicted binding modes were further analyzed with MD simulations and QM calculations on substrate reactivity, as described in more detail in **sections 2.4.3** and **2.2.2**, respectively. Marked differences were seen in the probabilities of different binding modes in wild type CYP102 and three mutants, and these trends were comparable to experimental product formation.

Recently, three popular automated docking programs AutoDock, FlexX and GOLD were evaluated on their ability to predict the binding mode of ligands from 19 crystallized complexes of CYPs and ligands, either without water, with crystal waters or with water in positions based on predictions by GRID. Although the individual programs showed variable performance, pooling and rescoring of all solutions improved the prediction accuracy significantly. Binding mode predictions were strongly improved by including either crystal or predicted water, and even with higher deviation from the reference crystal structure, correct predictions of catalytic sites (CSPs) were still obtained.[18]

Automated Ligand-Protein Docking Conclusions
Manual docking has been extensively used in the past and automated docking approaches are nowadays successfully applied for the prediction of the site of catalysis in substrates, the refinement and validation of CYP homology models (*i.e.* CYPs 1A2, 2B6, the 2C subfamily, 2D6, 2E1, 3A4, and 6B1, see **Table 3** and **Figure 9**), and the construction of pharmacophore models (*i.e.*, CYPs of the 2C subfamily, see **Table 1** and **Figure 6**). Docking algorithms are successful in determining binding conformations and orientations of CYP-ligand complexes, but docking scoring functions are still not suitable for the accurate prediction of ligand binding affinity, as will be described in more detail in **section 2.4.4**. Despite the fact that database docking methods have been successfully applied to crystal structures[187] as well as protein homology models[195], structure-based *in silico* screening studies on CYPs have so far only been reported for CYP101 (*cam*).

New docking strategies, considering explicit water molecules, partial protein flexibility, and (consensus) re-scoring of docking poses have already been found to improve binding mode prediction of CYP ligands. It is expected that these improved docking strategies will also be used more and more for *in silico* CYP screening studies in the future. However, inclusion of explicit water molecules and especially the inclusion of flexibility for the whole enzyme, leads to significantly higher complexity. Without proper algorithmical treatment of this complexity, this can lead to a final lower reliability of the docking results. The combination of docking with MD simulations (next section, **2.4.3**) to improve docking predictions and to explore conformational flexibility of substrates and CYP enzymes to a greater extent, as well as the explicit inclusion of water, is seen in an increasing number of studies and shows much promise.

2.4.3 Molecular Dynamics

Molecular Dynamics Methods Background

The role of CYPs as biotransformation enzymes and the broad substrate selectivity of many CYPs is reflected by the size of the active site, which is generally large compared to the size of substrates. Furthermore, many substrates and inhibitors exhibit high degrees of internal flexibility, and large conformational changes of the CYP protein structure upon ligand binding are known to play an important role in the CYP reaction cycle. Using molecular dynamics (MD) simulations, these processes associated with CYP-ligand interaction can be described and understood in atomic detail[107,196].

Molecular dynamics (MD) simulations provide detailed insight in the behavior of molecular systems, in both space and time, with ranges of up to nanometers and nanoseconds attainable for a system of the size of a CYP enzyme in solution. However, MD-simulations are based on empirical molecular mechanics force field descriptions of interactions in the system, and depend therefore directly on the quality of the force field parameters[197]. Commonly used MD programs for CYPs are AMBER, CHARMm, GROMOS, and GROMACS, and results seem to be comparable between methods. These programs are also listed in **Table 4**. For validation, direct comparisons between measured parameters and parameters calculated from MD simulations are possible, *e.g.* for fluorescence[198] and NMR (cross-)relaxation.[199] In many applications where previously only energy minimization would be applied, it is now common to perform one or several MD simulations as well, as is for example performed in studies of substrate entrance and product exit by Wade and co-workers.[110]

Starting conformations of MD simulations are generally obtained from automated docking methods (see also the previous **section**, **2.4.2**). Conversely, structural snapshots from MD runs can be used, with or without energy minimization, to include protein flexibility into docking simulations. Due to limited time scales of simulations, results can remain dependent on the starting conformations. The state-of-the-art is to solvate the simulated system in explicit water in a periodic box which, although computationally more expensive than continuum solvent or limited solvation schemes, avoids many of the possible pitfalls associated with them. Furthermore, several methodological developments to increase the efficiency of simulations of proteins (and ligand) in water are coming available, *e.g.*, by using methods that allow increased time steps[200], or by placing ligands in a pre-simulated environment.[201] Analysis methods are available to estimate statistics of ligand-binding and -orientation, and energies of ligand-protein interactions or free energies of binding.[202]

Molecular Dynamics Applications

Many applications of MD can be found in refinement and optimization of homology models, as highlighted in **section 2.3.3** and **Table 3**. This use of MD simulations is rapidly expanding. In some cases, the dynamics of the ligand is included and, *e.g.*,

distance of oxidation site in the substrate to the heme iron during the simulation is taken as a quality measure for the homology model being constructed, *e.g.* CYP2B6[150] and CYP2D6.[10]

Extensive MD simulations were used to identify key conformational changes upon ligand binding for CYP102 (*BM3*).[106,107] MD simulations were also used to study substrate access and product exit, as described in **section 2.3.2** *'Crystallographic Models Acces and Exit Channels'*, by random expulsion simulations and related methods for CYPs 101 (*cam*), 102 (*BM3*), and 107A (*EryF*).[104,108-110] Free energy of hydration of the active site of CYP101 has been studied using thermodynamic integration (TI).[135,203] Putative protein domain boundaries of CYP102-*BM3* were identified using dynamical cross-correlation (DCMM) analysis[204]. Based on MD simulations of substrates in the bound state in the enzyme and free state in solution, substrate binding affinities have been predicted with the LIE method for CYP101[34,74] and CYP1A1[152], free energy perturbation (FEP) for CYP101[203] and normal modes and Poisson-Boltzmann for CYP2B4[40], as is described further in **section 2.4.4**.

For the prediction of regiospecific substrate oxidation, QM calculations and MD simulations have been combined to account for differences in reactivity in the substrate as well as distributions of multiple binding conformations, as applied for CYPs 101 (*cam*)[34-36], 107A (*EryF*)[37] and 2E1[165], as presented in more detail in **section 2.2.2**.

In ongoing work (Feenstra *et al*, unpublished results) on CYP102 (*BM3*), probabilities for substrate orientation based on statistics of substrate binding from MD simulations were combined with probabilities for substrate reaction based on QM calculations of H-atom abstraction by a hydroxyl radical.[41] The experimental known preference for sub-terminal, distal hydroxylation of octane, octanoic acid, and lauric acid are reproduced.

Molecular Dynamics Conclusions
MD is used in various ways to study CYP-ligand interactions. As shown in **Table 3**, applications for homology model optimization and validation of model stability and the prediction of sites of catalysis in substrates are becoming common practice. Prediction of substrate and inhibitor binding affinity and orientation have been reliable in the cases of CYPs 101 (*cam*), 2B4 and 1A1, and combined with QM calculations on the substrate for predictions of product formation for CYPs 101 (*cam*), 107A (*EryF*), and 2E1 and in ongoing work for CYP102 (*BM3*).

Currently, most MD studies have been limited to short (~100 picoseconds, ps) or even very short (~10 ps) simulation times and only a single simulation. A notable exception is Winn *et al.*[110] who performed seventeen simulations of several hundreds of pico- to a few nanoseconds to describe substrate entrance to and product exit from CYPs 101 (*cam*), 102 (*BM3*) and 107A (*EryF*). For a proper description of ligand dynamics, and conformational changes in the enzyme, simulation times exceeding tens or hundreds of nanoseconds may well be needed.

2.4.4 Binding Affinity
Binding Affinity Methods Background
Binding affinity calculations of compounds to CYPs are a crucial ingredient to successful *in-silico* ADME(T) approaches, *e.g.* in database screening applications and during early stages of drug discovery. Methods for calculating binding affinity fall into two major categories, one in which the structure of the protein is used, and one in which it is not. In the first category, the assumption is made that similarity in biological response is reflected by chemical similarity of the ligands, and the second considers geometrical and chemical complementarities between the ligand and protein.[12] 3D-QSAR methods belong to the first category and are discussed in **2.4.4**. In the second category, scoring functions from automated ligand-protein docking programs (as described in **section 2.4.2**) can predict binding affinity based on the binding orientation and conformation of the ligand.

Table 4: *Commercially or freely available computational molecular modeling software used in studies described in this paper.*

Program	Description	Reference
	Homology Modeling	
Expasy	Proteomics web-server for the analysis of protein sequences and structures	www.expasy.org[205]
(PSI-)BLAST	Algorithm for searching of nucleotide and protein databases	www.ncbi.nlm.nih.gov/BLAST/[206]
ClustalW	Sequence alignment program/web server	www.ebi.ac.uk/clustalw/[206]
DALI	Automatic structure comparison program for structural sequence alignments	www.ebi.ac.uk/dali/[207]
GCG	Sequence alignment program	www.accelrys.com/products/gcg_wisconsin_package/[208]
JPRED	Secondary structure prediction and sequence alignment web server	www.compbio.dundee.ac.uk/~www-jpred/[209]
T-COFFEE	Sequence alignment program	www.ch.embnet.org/software/TCoffee.html[210]
NnPredict	Secondary structure prediction web server	www.cmpharm.ucsf.edu/~nomi/nnpredict.html[211]
Prosa	Algorithm for searching of nucleotide and protein databases	biologia.iztacala.unam.mx/tools/prosa/
PSIPRED	Automatic sequence alignment, protein fold recognition, and secondary structure prediction web server	bioinf.cs.ucl.ac.uk/psipred/[212]
Consensus	Comparative homology modeling program	www.accelrys.com/insight/consensus.html[213]
Modeller	Comparative homology modeling program	salilab.org/modeller/modeller.html[214]
Swissmodel	Automated comparative protein modeling server	expasy.hcuge.ch/swissmod/SWISS-MODEL.html[215]
Predict Protein	Automatic sequence alignment, protein fold recognition, and secondary structure prediction web server	www.embl-heidelberg.de/predictprotein/[216]
	Protein Structure Analysis	
Errat	Program to evaluate the amino acid side chain environment (packing quality) of protein structures	www.doe-mbi.ucla.edu/Services/ERRAT/[217]
PROCHECK	Program to evaluate the stereochemical quality of protein structures	www.biochem.ucl.ac.uk/~roman/procheck/procheck.html[218]
PROSA II	Program to evaluate the fold of protein structures	www.came.sbg.ac.at/Services/prosa.html[219]
PROVE	Program for protein volume evaluation	www.ucmb.ulb.ac.be/SCMBB/PROVE/[220]
Verify3D	Program to evaluate the amino acid side chain environment (solvent accessibility) of protein structures	www.doe-mbi.ucla.edu/Services/Verify_3D/[221]
	Pharmacophore/3D-QSAR Modeling	
Apex3d	3D-QSAR program implemented into Sybyl into InsightII	www.netsci.org/Science/Compchem/feature09.html
Catalyst	Software environment for several pharmacophore/3D-QSAR applications	www.accelrys.com/catalyst/
CoMFA	3D-QSAR program implemented into Sybyl	www.tripos.com/sciTech/inSilicoDisc/strActRelationship/QSAR.html[222]
GOLPE	Chemometric toolbox for 3D-QSAR	www.miasrl.com/GOLPE.htm[223]
GRID	Program to determine energetically favorable binding sites on proteins or ligands	www.moldiscovery.com/soft_grid.php[24]
EL3DMD/MS-WHIM	3D-QSAR program using descriptors derived from molecular surface properties	[66]

Program	Description	Reference
MetaSite	Program to predict the site of catalysis for substrates starting from the 3D structure of a compound	www.moldiscovery.com/soft_metasite.php[224]
Molecular Dynamics (MD) Simulations		
AMBER	MD simulation engine and force field	amber.ch.ic.ac.uk/[225]
CHARMm	MD simulation engine and force field	www.charmm.org[226]
GROMACS	MD simulation engine and force field	www.gromacs.org[227]
GROMOS	MD simulation engine and force field	www.igc.ethz.ch/gromos/[228]
Automated Docking (AD)		
AutoDock	AD program with genetic algorithm and regression-based scoring function	www.scripps.edu/pub/olson-web/doc/autodock/[229]
Chemscore	Regression-based scoring function	www.tripos.com/sciTech/inSilicoDisc/virtualScreening/cscore.html[230]
Cscore	Suite of scoring functions (ChemScore, DOCK, FlexX, GOLD, PMF) and consensus scoring tool	www.tripos.com/sciTech/inSilicoDisc/virtualScreening/cscore.html[231]
DOCK	AD program with incremental construction and random search algorithm and force-field based scoring function	dock.compbio.ucsf.edu[232]
FlexX	AD program with incremental construction algorithm and regression-based scoring function	www.biosolveit.de/FlexX/[233]
GLIDE	AD program with exhaustive search algorithm and regression-based scoring function	www.schrodinger.com/Products/glide.html[234]
GOLD	AD program with genetic algorithm and force-field-based scoring function	www.ccdc.cam.ac.uk/products/life_sciences/GOLD/[235]
PMF	Knowledge-based scoring function	www.tripos.com/sciTech/inSilicoDisc/virtualScreening/cscore.html[236]
SCORE	Regression-based scoring function	www.es.embnet.org/Doc/score/intro.html[237]
SLIDE	AD program with multi-level and mean field algorithm and regression-based scoring function	www.bch.msu.edu/labs/kuhn/web/projects/slide/home.html[238]
X-SCORE	Suite of three regression-based scoring functions and consensus scoring tool	sw16.im.med.umich.edu/software/xtool/[239]
General modeling programs		
InsightII	Program with modules for homology modeling, protein structure analysis, pharmacophore/3D-QSAR modeling and molecular dynamics	www.accelrys.com/insight/Insight2.html
QUANTA	Program with modules for homology modeling, protein structure analysis and MD	www.accelrys.com/quanta/
Sybyl	Program with modules for homology modeling, protein structure analysis, pharmacophore/3D-QSAR modeling, molecular dynamics and automated docking	http://www.tripos.com/data/SYBYL/SYBYLBH.pdf
WHATIF	Program with modules for homology modeling, protein structure analysis and MD	www.cmbi.kun.nl/whatif/[240]
Visualization programs		
MOLSCRIPT	Script-based 3D visualization program	www.avatar.se/molscript[241]
RASTER3D	3D rendering and ray-tracing engine	www.bmsc.washington.edu/raster3d/raster3d.html[242]

Molecular dynamics (MD) and Monte Carlo (MC) methods, as described in the previous section, **2.4.3**, explicitly include the statistics of protein-ligand interactions and can therefore be considered as a more appropriate tool to include flexibility on a large scale[243]. Conformations derived from MD/MC are used in traditional free energy calculations, *i.e.* free energy perturbation (FEP) or thermodynamic integration (TI)[244], and in the linear interaction energy (LIE) approximation[152] to determine or predict the free energy difference between two states (*e.g.*, one ligand in bound and unbound state, two ligands bound, or two mutant proteins).

Binding Affinity Applications

Scoring functions available in docking programs show little or no correlation with (experimental) binding affinity of CYP101 (*cam*)[193], 2D6[148], and 3A4[71]. However, using LIE, the absolute and relative binding energies of 11 CYP101 complexes were generally well produced (mean errors of 0.55 and 0.77 kcal/mol respectively)[74]. Binding free energies of two out of three CYP1A1 substrates could be computed within 0.9 kcal/mol of experimentally observed values by LIE[152] using coefficients from the previous study[74]. The binding free energy of camphor to CYP101 (*cam*) could be computed within 0.8 kcal/mol from experimental values by FEP with full treatment of solvent degrees of freedom[203], while the binding free energy difference of *S*- and *R*-nicotine to CYP101 could be computed within 0.3 kcal/mol.[34] Free energy of binding of methoxyflurane, benzphetamine and 4-phenylimidazole to CYP2B4 was calculated based on a homology model using a combination of MD simulations, normal modes analysis for estimating entropic contribution and Poisson-Boltzmann solvation energy (MMPB-SA), to within 0.5 to 2 kcal/mol.[40]

Binding Affinity Conclusions

Predictions of binding affinities based on the scoring functions of docking algorithms are in general not accurate, in contrast to 3D-QSAR affinity predictions for CYPs (see **section 2.4.4**). The applications of FEP, LIE and MMPB-SA methods to CYPs 101 (*cam*), 1A1 and 2B4 are significantly more accurate than the QSAR results, although they are limited to a more narrow range of chemical diversity and computationally much more costly. For *e.g.* the estrogen receptor for which ligands are much less diverse, successful application of LIE for a large set of ligands and a wide range of affinities is reported[245]. The challenges specific to CYP enzymes arise from specific characteristics of the CYP enzyme family, namely a very broad range of substrates, of which many are highly charged and/or highly flexible. It is known that the MD simulations and forcefields used in the FEP, LIE and MMPB-SA methods perform best for substrates with little or only moderate charge and flexibility.

2.4.5 Conclusions Ligand-Protein Interactions

Automated docking approaches are nowadays successfully applied for prediction of the site of catalysis in substrates for many CYP isoforms, for the refinement and optimization of CYP homology models (*i.e.* CYPs 1A2, 2B6, the 2C subfamily, 2D6, 2E1, 3A4, and 6B1) and for the construction of pharmacophore models (*i.e.* CYPs of the 2C subfamily). Docking is still not suitable for the quantitative prediction of ligand binding affinity, and structure-based *in silico* screening studies on CYPs have so far only been reported for CYP101 (*cam*), which may be attributed at least in part to lack of proper consideration of active site water and protein flexibility. MD simulations allow the inclusion of explicit water and full protein flexibility but at the expense of greatly increased computational demands. MD based methods have shown to be reliable for prediction of binding affinities for CYPs 101(*cam*), 1A1, and 2B4, for finding a more comprehensive description of multiple sites of catalysis for several CYP substrates, and for exploring ligand binding in CYP homology

models and the protein flexibility of the homology model in more detail, *e.g.* for CYP2B6 and CYP2D6.

A remaining hurdle in the accurate calculation of CYP ligand binding affinity is the diversity of binding modes of substrates, reflected in a flexible binding cavity that is large in relation to the size of the substrates, and, subsequently, for many substrates the binding mode in CYPs is not as strictly defined as is the case for many other more substrate specific enzymes. This restricts the application of MD based methods that can provide the most accurate free-energy calculations.

2.5 Summary & Conclusions

2.5.1 Introduction

Early consideration of properties of the cytochrome P450 (CYP) family of biotransformation enzymes involved in drug metabolism is increasingly seen as essential for efficient discovery and development of new drugs and drug candidates. Apart from the application of *in vitro* tools, this necessitates the application of *in silico* tools that can accurately predict metabolic properties of drug candidates already in early stages of the lead finding and optimization process.

The state of the art of computational methods used in CYP modeling varies considerably, as does the reliability of the results obtained. This arises from limitations of the methods applied as well as from difficulties related to the complexity of the CYP enzymatic action itself. Mechanism-based quantum-chemical calculations on substrates and the enzyme, pharmacophore modeling of ligands and protein homology modeling in combination with automated docking and molecular dynamics simulations have been used for the rationalization and prediction of ligand binding and metabolite formation. A remaining challenge in predicting product formation by CYPs is the elucidation of the relevant reactive heme-iron-oxygen species that is ultimately responsible for the chemistry of the enzyme reaction, and the coupling between the kinetics of the reaction cycle of the heme-iron-oxygen moiety and the dynamics of substrate and oxygen and the release of product. It may be concluded that not one computational approach is capable of rationalizing and reliably predicting metabolite formation by CYPs, but that it is rather the combination of the various complimentary approaches.

The primary aim of this paper was to present an integrative modeling approach to understand, rationalize and predict the activity and substrate selectivity of CYPs, as well as the possibilities and limitations of these approaches now and in the future. Software that implements these computational methods and has been applied to CYPs is listed in **Table 4**. We divided the models and methods into ligand based, protein based, and ligand-protein interaction based. Results for (ligand based) pharmacophore models are listed in **Table 1**, for (protein based) crystal structures in **Table 2** and homology model structures in **Table 3**. Results for ligand-protein interaction based models were presented in **section 2.4**, and are mentioned in some instances for pharmacophore models in **Table 1** and homology models in **Table 3**. In the following sections, a summary and general conclusions will first be presented for the different modeling methods applied to CYPs, then for each of the most important CYP isoforms, and finally for CYPs in general.

2.5.2 Methods

Quantitative structure-activity relationship (QSAR) approaches, derived from physicochemical and topological properties or the 3D structure of CYP ligands, have been used often for the prediction of drug metabolism by CYPs.[19-21] A pragmatic, rapid and rather successful approach to this are topological screening methods recently developed for some CYPs, which provide robust but rough classifications of inhibitory potency of ligands without much insight in the possible (molecular) background of the mode of

action.[23] The structure-based molecular modeling methods on the other hand, aim to be insightful as well as predictive.

Intrinsic regio-selective susceptibilities of molecules to CYP metabolism have been predicted using available quantum mechanical (QM) methodologies for CYPs, and significant improvements in the prediction of site of catalysis have been achieved for several CYPs by inclusion of known or predicted substrate binding orientations in the active site or of relative substrate orientations in pharmacophore models.[34,35,37,165] Further improvements have been seen from including a description of the relevant heme-iron-oxygen activated reactive species at a higher level in the QM calculations.[25,27] Additional improvements were seen from inclusion of more of the active site and protein surroundings that can have a significant influence on electronic properties and activation energies of the substrate mainly through electrostatic interactions (Feenstra *et al.* unpublished results). Classical CYP pharmacophore models have been successful for the prediction of qualitative properties like substrate selectivity and site of catalysis, and the more recent 3D-QSAR pharmacophore models have in addition been used for the quantitative prediction of binding affinities[19,21], as is listed in **Table 1**.

The growth in available CYP protein structures from X-ray crystallography, presented in **Figure 8**, has inspired several investigators to detailed analysis of CYP structure and of CYP-ligand interaction[8,111,135-138], and it has been the basis for much homology modeling work[44,142], see **Table 2** for crystal and **Table 3** for homology model structures. In this respect, homology structures have been complementary to crystal structures, especially for the mammalian CYPs now that mammalian crystal structures have very recently become available.[14,16,131,132,246] Several CYP isoforms have only been crystallized in a non-active ligand-bound state, and for some it is not known whether the crystal structure corresponds to an active state or not.[14,96,97] It is important to realize that this may make the crystallized structure of a certain CYP isoenzyme not directly applicable to *e.g.* explaining observed substrate selectivity and product formation of the enzyme. Ligand-induced changes in the size, shape and hydration of the active sites of several CYPs are also observed[105,114,139-141], and these are likely to undelie the capacity of many CYPs to metabolize structurally diverse substrates of different sizes and to generate different metabolic products from these substrates.

Automated docking approaches[186] are successfully applied for prediction of the binding mode of ligands and for prediction of sites of catalysis in substrates for several CYP isoforms, for the refinement and optimization of homology models[10,69-71,146,150,165,169,173] and for the construction of pharmacophore models.[41,61,70,71,76,77] Docking is still not suitable for quantitative prediction of ligand binding affinity[71,148,193], which may be attributed at least in part to the lack of proper consideration of active site water and protein flexibility.[18,191] Structure-based *in silico* screening studies on CYPs have so far been reported only for CYP101 (*cam*).[59,193,194] Molecular dynamics based methods for prediction of binding affinities[34,40,74,152,203], for finding a more comprehensive description of multiple sites of catalysis in substrates[34-37,165], for exploring ligand binding and for exploring protein flexibility of CYP crystal structures and homology models in more detail[104,106-110,204], have shown to be more reliable in several cases.

2.5.2 CYP Isoforms

Results and conclusions for the most studied, and most interesting, CYP isoforms are summarized for each isoform, and additional details of the relevant data and references can be found in the method **sections 2-4** and in the **Tables 1-3**. Bacterial and eukaryotic isoforms, and general conclusions are treated in three separate sections.

Bacterial CYPs

CYP101 (*cam*) was the first CYP to be crystallized[112], and remains the most extensively studied CYP by X-ray crystallography. Based on this first structure and the very many ligand-bound and ligand-free structures solved later, extensive QM, docking, and MD studies have been performed, as well as homology model building for other isoforms. Only minor ligand-induced conformational changes in the protein were observed in the crystal structures. Calculations on the protein structures revealed water channels and proton-transfer pathways[136,137], but did not clearly show possible substrate access and product exit channels.[111] Analysis of B-factors and MD-simulations revealed the existence of several distinct access/exit channels.[104,108,109] Differences in entrance and exit mechanisms in CYPs 102 (*BM3*) and 107A (*EryF*) are thought to contribute to substrate recognition for these isoforms.[110] Substrate binding has been studied by automated docking to accurately predict substrate catalytic sites and for the selection of substrates and inhibitors from databases.[59,193,194] Substrate binding affinities and free energy of hydration of the active site of CYP101 were calculated based on MD simulations.[34,35,74,135,203] QM calculations of the rate-limiting H-atom abstraction step of C-H hydroxylation reactions of the enzyme, based on conformations from MD simulations, could describe additional details of product formation, and supported a two-state abstraction/oxygen rebound mechanism of the CYP mono-oxygenase activity.[34-36]

CYP102 (*BM3*), the second CYP to be crystallized, spurred much MD and homology modeling work. The ligand-bound and ligand-free conformations differ considerably and these differences revealed the existence of a substrate access and product exit channel[105]. MD simulations have given plausible descriptions of the ligand-induced changes and could confirm the access and exit channels as well.[106,107] Calculations on the structures revealed water channels and proton-transfer pathways.[136] Even in the ligand-bound structures the distances between ligand and the heme-iron catalytic center are too large for catalytic activity, and additional conformational changes driven by the electron-transfer process are thought to precede substrate metabolism.[96-98] Differences in entrance and exit mechanisms with CYPs 101 (*cam*) and 107A (*EryF*) are thought to contribute to substrate recognition for these isoforms.[108-110] Analysis of correlated dynamics in MD simulations revealed three domains in the protein, while analysis of the protein structure only identified two.[204] Catalytic site prediction in substrates based on docking, MD simulations and QM calculations is the subject of ongoing work (Feenstra *et al*, unpublished results).

For CYP107A (*EryF*) several ligand-bound crystal structures have been solved and QM, docking, and MD studies have been performed, and in combination with some other bacterial CYPs used for homology model building for other isoforms. Ligand-induced conformational changes were observed in different ligand-bound crystal structures.[114] Calculations on the structures revealed water channels and proton-transfer pathways.[136] MD simulations revealed the existence of several distinct access/exit channels.[110] Differences in entrance and exit mechanisms with CYPs 101 (*cam*) and 102 (*BM3*) are thought to contribute to substrate recognition for these isoforms.[110] QM calculations of the formation of the twice-reduced activated oxygen species, based on conformations from MD simulations, could describe additional details of product formation, and identified Glu360 as a plausible source for the proton necessary in this reaction.[37]

For CYP108 (*terp*) one ligand-free crystal structure has been solved, and in combination with some other bacterial CYPs it has been used for homology model building for other isoforms. Calculations on the protein structure revealed water channels and proton-transfer pathways.[136]

Eukaryotic CYPs

Early eukaryotic homology models, as reviewed by Ter Laak *et al*[44], have been based on bacterial template crystal structures, *e.g.* CYPs 101, 102, 107A and 108. Most have since

been superseded by newer models based, at least in part, on more appropriate eukaryotic template crystal structures, mainly CYP2C5. It stands to reason that in the near future new homology models will be built using the new CYP3A4 and CYP2C9 crystal structures as well.

For the CYP1A1 and CYP1A2, several pharmacophore models have been constructed[64], and homology models[64,152,154] have been built based on the CYP101 and CYP102 crystal structures and later CYP2C5 crystal structures. Details of the alignment and structure of the homology models have been improved based on CYP1A1/2 pharmacophore models, and validated using a combination of 3D-QSAR and automated docking. Substrate binding affinities have been calculated for CYP1A1 with linear interaction energy calculations, based on the homology model structure.

For several members of the CYP2A subfamily, *i.e.* 2A1/4/5/6, pharmacophore models[67] and homology models[156,157], based on the CYP102 and CYP2C5 crystal structures, have been constructed separately, and recently X-ray crystal structures of CYP2A6 have been solved.[129]

For several members of the CYP2B subfamily, *i.e.* 2B1/2/4/5/6, homology models[40,69,149,159-161] have been built based on the CYPs 101, 102, 107A, 108 and 2C5 crystal structures, and QM, docking, and MD studies have been performed. For CYP2B4 recently an open, unbound[246] and a closed, inhibitor bound[130] X-ray crystal structure have been solved and extensive conformational changes in the regions between helices B and C (where helix B' can be present) and between helices F and G are observed. Two CYP2B6 homology models[69,150] based on the CYP2C5 crystal structure have been the most extensively refined and validated using pharmacophore models, automated docking, MD simulations and CYP2B subfamily isoenzyme specificity. Ligand binding and spin-state equilibria were found to be connected in QM studies on CYP2B4, and in combination with conformations from MD simulations additional details of product formation could be described.[40]

For CYP2C5 ligand-bound[101,141] and ligand-free[131] crystal structures have been solved, which have been used extensively for homology model building work for other isoforms. Ligand-induced conformational changes, multiple substrate binding modes and water-mediated ligand binding were observed in different ligand-bound crystal structures.

CYP2C9 was the first human CYP for which a crystal structure was solved[14], see **Figure 2-A**. Many QM, 3D-QSAR, docking, and MD studies have been performed. Pharmacophore (mostly 3D-QSAR)[41,43,45,46,61,75-79] and homology models[41,70,162], based on crystal structures of CYPs 101, 102 and 2C5, have been constructed separately as well as in conjunction with each other. The homology model structures have been extensively refined and validated using pharmacophore models, automated docking and MD simulations. Pharmacophore models for CYP2C9 inhibitors and substrates predict binding affinities reasonably well, and suggest substrates bind an aromatic and a cationic site for which respectively F476 and R108 were suggested based on homology modeling work. Involvement of these residues was confirmed by site-directed mutagenesis experiments. The first ligand-bound CYP2C9 crystal structure shows interactions with F476[14], see **Figure 2-A,** while in contrast to this, a more recent substrate bound crystal structure shows interactions with F114 and R108[15], see **Figure 2-B**.

For several members of the CYP2C subfamily, *i.e.* 2C8/9/18/19, homology models have been built[70,71,163] in conjunction with a pharmacophore model, based on the CYPs 101, 102, 107A, 108 and 2C5 crystal structures, and many QM, 3D-QSAR, docking, and MD studies have been performed. The very recent CYP2C8 crystal structure[132] adds to the growing experimental structural basis for homology modeling work of human CYPs. Predictions of CYP2C subfamily isoenzyme specificity were based on 3D-QSAR, docking, and MD studies.

For CYP2D6, homology models based on crystal structures of CYPs 101, 102, 107A, 108 and 2C5 have been constructed[4,10,56,69,81,98,142,145-147] separately as well as in conjunction with each other. The homology model structures have been extensively refined and validated using automated docking, MD simulations and comparison with site-directed mutagenesis studies and NMR spin-relaxation rate measurements. The latest CYP2D6 homology model is shown in **Figure 3-A**. Many QM, pharmacophore, docking and MD studies have been performed. Gas phase QM calculations involving radical formation, redistribution and recombination could rationalize CYP2D6 metabolite formation[30,32]. Pharmacophore models (notably 3D-QSAR) [48-51,56,57,69,80-83] for CYP2D6 inhibitors and substrates predict binding affinities well, and suggest substrates bind two or three aromatic/hydrophobic sites and a carboxylic site for which F120 and F483, respectively E216 or D301 were suggested in combination with homology modeling work and supported by mutagenesis studies. **Figure 4** shows molecular interaction fields for dextromethorphan overlaid on the corresponding sidechains in the CYP2D6 active site. Differences between rat CYP2D isoforms (2D1/2/3/4) and human CYP2D6, and details of the 2D6 active site topology, could be described by the electrostatic properties of the active site and hydrophobic substitutions at the 120 position.[10] Very recently a substrate-free CYP2D6 crystal has been solved[133], which is described in **Chapter 3** in more detail.

For CYP2E1, pharmacophore[43] and homology models[164,165] based on the CYPs 102, 2C5 and 2C9 crystal structures, have been built, and QM, docking, and MD studies have been performed[165]. The homology models were refined by comparison of species differences between rat, mouse and human, and validated by catalytic site prediction based on QM, docking and MD calculations for several substrates. QM calculations on substrate radical species, based on conformations produced by docking and MD simulations, could predict binding affinities and could describe additional details of product formation. These calculations supported a hydrogen-abstraction mechanism of the CYP mono-oxygenase activity involving radical intermediates.

For CYP3A4, homology models have been built[71,146,169] based on the CYP 101, 102, 107A, 108 and 2C5 crystal structures, and QM, 3D-QSAR, docking, and MD studies have been performed. Very recently ligand-bound[16] and ligand-free[16,134] crystal structures have been solved. The highest resolution ligand-free structure is shown in **Figure 2-C**. The homology models were refined and verified using automated docking, MD simulations, QM calculations and comparison with pharmacophore models.[39,166,167] 3D-QSAR pharmacophore models could predict ligand binding affinities accurately. QM calculations of the rate-limiting H-atom abstraction step of a C-H hydroxylation reaction of the enzyme, based on conformations from MD simulations, could predict plausible catalytic sites in a number of substrates and, in combination with the homology model protein structure, could describe stabilizing and cooperative effects of multiple ligand binding.

For CYP19 (*arom*), pharmacophore models have been developed[90-92] and homology models have been built[177,178] based on the CYP 101 and 102 crystal structures. The homology models were refined and verified using predictions of catalytic sites and comparison with pharmacophore models. 3D-QSAR pharmacophore models could predict inhibitor binding reasonably well.

For CYP51 (*14α*), pharmacophore models have been developed[93-95] and homology models have been built[180] based on the CYP 101, 102 and 107A crystal structures. Several ligand-bound and ligand-free crystal structures have been solved and MD studies have been performed. Ligand-induced conformational changes were observed in different crystal structures.[140] The homology models were refined and verified using MD simulations and comparison with pharmacophore models. 3D-QSAR pharmacophore models could predict inhibitor binding reasonably well.

CYPs in General

The bacterial CYPs 101 (*cam*), 102 (*BM3*), and 108 (*terp*) were used for building the first structural alignment[8] and later a new alignment was built with the addition of 107A (*EryF*), from which also an identification of common structural building blocks (CSBs) was made.[143] This alignment and the CSBs have been the basis of many homology models for other CYP isoforms. Identification of substrate recognition sites (SRS's) was based on additional alignment of the bacterial CYPs with members of the CYP2 subfamily[9]. The SRS's are shown structurally and topologically in **Figure 3**.

CYPs generally have broad substrate selectivities and often substrates are found in multiple binding modes[99-101], instead of in a shape-complementary fit that is common for other enzymes with high specificities (and some CYPs as well). For several CYP isoforms it is known that water-mediated ligand binding plays an important role.[15,101,141]

Another important factor is the possible effect of atomic or molecular oxygen bound to the heme iron, and to know which oxygen species, if any, is present during ligand binding. Moreover, a balance exists between the kinetics of the heme-iron-oxygen reduction and the dynamics of the substrate binding. Changes in the oxygen species may occur faster than substrate binding orientation relaxation times, leaving the substrate during the actual reaction in an orientation that was optimal for a previous, but no longer present oxygen species. Additionally, the existence of substrate-bound crystal structures where the substrate apparently is not in a catalytically active position with respect to the heme-iron[14,96,97] may indicate the importance of the kinetics and dynamics of substrate binding and oxygen binding and reduction.[98]

2.5.4 Future Prospects

The previous sections have described the development of several modeling methodologies, and successful applications to understanding CYP structure and enzyme function, and predicting CYP metabolic activity. This field of computational medicinal chemistry and molecular toxicology is rapidly evolving. Current applications of several techniques can be expected to become more widespread, and (more) combinations of methods can be expected to be successful as well. This will be covered in the next section. Expected future developments of novel methodologies and of the application of existing methods not yet applied to CYPs, is covered in the final **section 2.5.4**.

Current Applications

A number of pharmacophore models have been derived for CYPs, either based on suitable template molecules or on a variety of substrates or inhibitors when a single compound was inappropriate as a template molecule.[19,21] Several of these pharmacophore models, among which are those of CYP2D6[56,81] and CYP2C9[61], have been shown to have a good predictive value concerning qualitative metabolism and substrate/inhibitor selectivity, a property especially relevant for isoenzymes which are subject to genetic polymorphisms (*e.g.*, CYP2D6 or the CYP2C subfamily). Despite potential benefits (especially for the pharmaceutical industry), the development of pharmacophore models for biotransformation enzymes has received relatively little attention yet, in contrast to pharmacophore models for receptor ligands.

For the prediction of catalytically susceptible sites in the substrate, further improvements have been achieved by including a description of the relevant heme-iron-oxygen activated reactive species at a higher level in the QM calculations.[25,27] Additional improvements were seen from inclusion of more of the active site and protein surroundings[25] that can have a significant influence on electronic properties and activation energies of the substrate mainly through electrostatic interactions (Feenstra *et al.*, unpublished results).

An emerging role of pharmacophore/QSAR models is to consolidate available experimental data on substrates and inhibitors and to combine these with structural details

included in CYP X-ray crystallographic or homology model structures.[56,71,76,81] On the other hand, the use of alignment and conformer independent models make it possible to build models without an alignment rule for chemically diverse ligands and without the use of information about the protein structure[61]. Predictions based on QM calculations on substrates will likely be used as a complementary source of information obtained from pharmacophore/QSAR models.

It has been shown for CYP2D6[10,142] and CYP2C9[41,70] that good protein homology models can be built for CYPs, and used to obtain useful information concerning amino acids important for substrate and/or inhibitor binding. Due to relatively low sequence and structure homology in the ligand-binding site region of some CYPs, these predictions, however, have to be experimentally verified and carefully considered. For elucidation of the role of specific amino acids in the catalytic activity of CYPs, homology models are used as well to rationalize observed differences, however, the activities cannot yet be predicted accurately. Indications of substrate selectivity can be obtained qualitatively on the basis of homology models in specific cases, although these predictions should always be experimentally verified and carefully considered as well.

The trend in CYP homology model building is clearly towards an integrated approach of multiple sequence and structural alignments, combination with new or available pharmacophore models, automated docking, molecular dynamics simulations and validation with mutational, spectroscopic, structural and enzyme kinetic data.[10] Predictions of selectivity and site of metabolism have been successful for homology models of many CYP isoforms, and, for example, recent models of CYP2D6 show very good agreement with site-directed mutation studies as well[18,148]. Furthermore, the use of homology modeling building methods to enrich X-ray crystallographic structures with other experimental data, *e.g.* biochemical or spectroscopical, on activities and binding orientations of substrates, is a topic of ongoing research and expected to become important in the near future.

CYPs generally have broad substrate selectivities and often substrates are found in multiple binding modes[101], instead of in a shape-complementary fit that is common for other enzymes with high specificities (and some CYPs as well). This makes statistical considerations like relative weights of different binding modes crucial and in addition increases the flexibility of ligands within each binding mode[109](Feenstra *et al.* unpublished results). Energetic consideration of detailed ligand-protein interactions gives insight in the (relative) importance of different sites of interaction between the ligand and the protein. Finally, the inclusion of explicit water molecules in the active site allows the elucidation of the (mediating) role of water in ligand binding.[18,203] A remaining hurdle in the accurate calculation of CYP-ligand binding affinity is the prediction of multiple binding modes for substrates, reflected in a flexible binding cavity that is large compared to the size of the substrates, and, subsequently, for many substrates the binding mode is not as strictly defined as is the case for many other more substrate selective enzymes. This restricts the application of MD based methods that can provide the most accurate free-energy calculations.

Evaluation of different scoring functions, docking algorithms and pooling of docking results from different programs, has been shown to be critical in obtaining good results for ligand docking in CYP crystal structures[18] and also make docking into homology model structures feasible.[247] Much additional gain is expected from these methodological developments. Moreover, addition of water bound at specific (predicted) locations in the active sites is the topic of current investigations and is expected to significantly improve the quality of docking results[18]. Another important factor is the possible effect of atomic or molecular oxygen bound to the heme iron, and the oxygen species, if any, present during ligand binding. Obviously, a balance exists between the kinetics of the heme-iron-oxygen reduction and the dynamics of the substrate binding. Changes in the oxygen species may

occur faster than substrate binding-orientation relaxation times. In that case, during the actual reaction the substrate may be left in an orientation that was optimal for a previous, but no longer present, oxygen species. This important issue has only occasionally received attention in literature, *e.g.*, some studies were performed using MD simulations that investigate the formation of the active, bound complex with subsequent QM calculations of the reactivity.[165] Additionally, the existence of substrate-bound crystal structures where the substrate apparently is not in a catalytically active position with respect to the heme-iron may indicate the importance of the kinetics and dynamics of substrate binding and oxygen binding and reduction.[14,97]
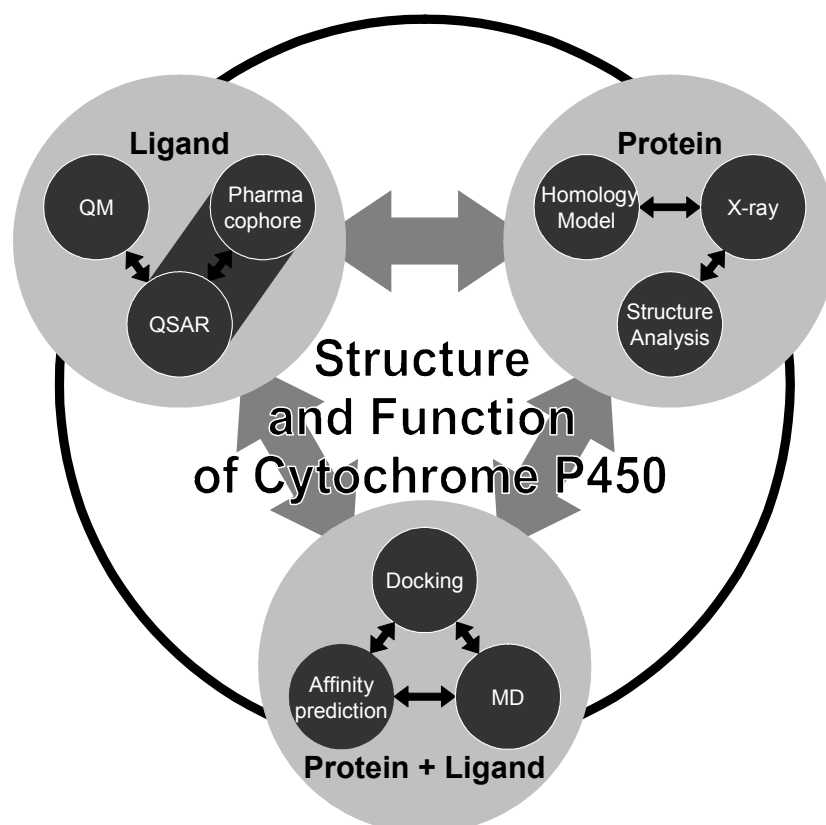
Figure 10: *The integrative modeling approach of cytochromes P450; schematic arrangement of modeling methods and the division into ligand-based, protein-based and ligand-protein interaction-based methods, with relationships between methods and between method groups indicated with arrows. Ligand-based methods are quantum-mechanical (QM) calculations and classical and QSAR pharmacophore models, Protein-based methods are X-ray crystallography, protein homology model building and protein structure analysis; and Protein + Ligand-based methods are automated ligand docking, molecular dynamics (MD) simulations and affinity prediction calculations.*

Expected New Developments
Generally speaking, it is clear that computational approaches in parallel with high(er) throughput experimental technologies are among the newer and fast developing approaches in drug discovery, drug metabolism, and toxicology. New links with other current developments, such as in neural network computing, genomics, proteomics and bioinformatics are within reach and successful exploitation of these links will allow significant progress towards *in silico* prediction of drug activity, drug metabolism and toxicity[10]. Future developments towards a more complete *in silico* prediction of ADME(T) will need more accurate models for the prediction of drug absorption (A), disposition (D) and elimination (E) in addition to the models for prediction of metabolism (M) and toxicology

(T). Furthermore, in general a better understanding of the key issues in drug activity should be developed.[2,248] This type of pre-experimental prediction of non-pharmacological drug-properties with *in silico* ADME(T) screening[1] is likely to be one of the challenging developments with a great scientific and practical impact.

It has been shown that a great variety of *in silico* modeling approaches have been applied to CYP enzymes, thereby successfully adding to our understanding of CYP structure and function in a way that is complementary to experimental studies. Another clear trend in many recent computational studies is the combination of several modeling techniques to arrive at meaningful and successful rationalization of experimental data, and interpretations and predictions of CYP function. Indeed, much of the recent progress in this area stems from integration of different computational methods, as well as improvements of individual methods. In **Figure 10**, the integration of computational approaches is shown as connections between the modeling methods, and it is clear that not only between methods *within* the 'ligand-based', 'protein-based' and 'ligand-protein interaction based' groups of methods, but also *between* the three groups, extensive methodological connections are present. It is our conclusion that these connections, in fact, are due to a necessity of combining existing methodology into an integrative *in silico* approach for modeling of CYP function, which is the key for efficient, reliable and successful application of *in silico* approaches. Finally, it is concluded that the combination of new computational methods and solid experimental data will prove to be critical and essential for the elucidation of as yet unknown functional and structural properties and mechanisms of CYPs.

References

(1)   Hou, T. J.; Xu, X. J. Recent development and application of virtual screening in drug discovery: An overview. *Current Pharmaceutical Design* **2004**, *10*, 1011-1033.

(2)   van de Waterbeemd, H.; Gifford, E. ADMET in silico modelling: towards prediction paradise? *Nat Rev Drug Discov* **2003**, *2*, 192-204.

(3)   Yue, Q. Y.; Alm, C.; Svensson, J. O.; Sawe, J. Quantification of the O- and N-demethylated and the glucuronidated metabolites of codeine relative to the debrisoquine metabolic ratio in urine in ultrarapid, rapid, and poor debrisoquine hydroxylators. *Ther Drug Monit* **1997**, *19*, 539-542.

(4)   de Groot, M. J.; Vermeulen, N. P.; Kramer, J. D.; van Acker, F. A.; Donne-Op den Kelder, G. M. A three-dimensional protein model for human cytochrome P450 2D6 based on the crystal structures of P450 101, P450 102, and P450 108. *Chem Res Toxicol* **1996**, *9*, 1079-1091.

(5)   Goeptar, A. R.; Scheerens, H.; Vermeulen, N. P. Oxygen and xenobiotic reductase activities of cytochrome P450. *Crit Rev Toxicol* **1995**, *25*, 25-65.

(6)   Guengerich, F. P. Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem Res Toxicol* **2001**, *14*, 611-650.

(7)   Li, H. Cytochrome P450. *Handbook of metalloproteins*; Wiley: Chichester, 2001; pp 267-282.

(8)   Hasemann, C. A.; Kurumbail, R. G.; Boddupalli, S. S.; Peterson, J. A.; Deisenhofer, J. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure* **1995**, *3*, 41-62.

(9)   Gotoh, O. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem* **1992**, *267*, 83-90.

(10)   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.

(11)   Groenhof, A. R.; Swart, M.; Ehlers, A. W.; Lammertsma, K. Electronic ground states of iron porphyrin and of the first species in the catalytic reaction cycle of cytochrome P450s. *Journal of Physical Chemistry A* **2005**, *109*, 3411-3417.

(12)   Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angewandte Chemie-International Edition* **2002**, *41*, 2645-2676.

(13)   Guengerich, F. P. Rate-limiting steps in cytochrome P450 catalysis. *Biol Chem* **2002**, *383*, 1553-1564.

(14)   Williams, P. A.; Cosme, J.; Ward, A.; Angove, H. C.; Matak Vinkovic, D. et al. Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* **2003**, *424*, 464-468.

(15)   Wester, M. R.; Yano, J. K.; Schoch, G. A.; Yang, C.; Griffin, K. J. et al. The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-A resolution. *J Biol Chem* **2004**, *279*, 35630-35637.

(16)   Williams, P. A.; Cosme, J.; Vinkovic, D. M.; Ward, A.; Angove, H. C. et al. Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* **2004**, *305*, 683-686.

(17)   Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. *Chem Rev* **2005**, *105*, 2279-2328.

(18)   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.

(19)   de Groot, M. J.; Ekins, S. Pharmacophore modeling of cytochromes P450. *Advanced Drug Delivery Reviews* **2002**, *54*, 367-383.

(20)   Hansch, C.; Mekapati, S. B.; Kurup, A.; Verma, R. P. QSAR of cytochrome P450. *Drug Metab Rev* **2004**, *36*, 105-156.

(21)   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 Metabolism and Disposition* **2001**, *29*, 936-944.

(22)   Susnow, R. G.; Dixon, S. L. Use of robust classification techniques for the prediction of human cytochrome P450 2D6 inhibition. *Journal of Chemical Information and Computer Sciences* **2003**, *43*, 1308-1315.

(23)   Zuegge, J.; Fechner, U.; Roche, O.; Parrott, N. J.; Engkvist, O. et al. A fast virtual screening filter for cytochrome P450 3A4 inhibition liability of compound libraries. *Quantitative Structure-Activity Relationships* **2002**, *21*, 249-256.

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

(25)   Schoneboom, J. C.; Cohen, S.; Lin, H.; Shaik, S.; Thiel, W. Quantum mechanical/molecular mechanical investigation of the mechanism of C-H hydroxylation of camphor by cytochrome P450cam: theory supports a two-state rebound mechanism. *J Am Chem Soc* **2004**, *126*, 4017-4034.

(26)   Swart, M.; Groenhof, A. R.; Ehlers, A. W.; Lammertsma, K. Substrate binding in the active site of cytochrome P450cam. *Chemical Physics Letters* **2005**, *403*, 35-41.

(27)   Shaik, S.; Cohen, S.; de Visser, S. P.; Sharma, P. K.; Kumar, D. et al. The "Rebound controversy": An overview and theoretical modeling of the rebound step in C-H hydroxylation by cytochrome P450. *European Journal of Inorganic Chemistry* **2004**, 207-226.

(28)   Groves, J. T. Key Elements of the Chemistry of Cytochrome-P-450 - the Oxygen Rebound Mechanism. *Journal of Chemical Education* **1985**, *62*, 928-931.

(29)   Newcomb, M.; Shen, R.; Choi, S. Y.; Toy, P. H.; Hollenberg, P. F. et al. Cytochrome P450-catalyzed hydroxylation of mechanistic probes that distinguish between radicals and cations. Evidence for cationic but not for radical intermediates. *Journal of the American Chemical Society* **2000**, *122*, 2677-2686.

(30)   de Groot, M. J.; Donne-Op den Kelder, G. M.; Commandeur, J. N.; van Lenthe, J. H.; Vermeulen, N. P. Metabolite predictions for para-substituted anisoles based on ab initio complete active space self-consistent field calculations. *Chem Res Toxicol* **1995**, *8*, 437-443.

(31)   Bathelt, C. M.; Ridder, L.; Mulholland, A. J.; Harvey, J. N. Aromatic hydroxylation by cytochrome P450: model calculations of mechanism and substituent effects. *J Am Chem Soc* **2003**, *125*, 15004-15005.

(32)   Lightfoot, T.; Ellis, S. W.; Mahling, J.; Ackland, M. J.; Blaney, F. E. et al. Regioselective hydroxylation of debrisoquine by cytochrome P4502D6: implications for active site modelling. *Xenobiotica* **2000**, *30*, 219-233.

(33)   Singh, S. B.; Shen, L. Q.; Walker, M. J.; Sheridan, R. P. A model for predicting likely sites of CYP3A4-mediated metabolism on drug-like molecules. *Journal of Medicinal Chemistry* **2003**, *46*, 1330-1336.

(34)   Jones, J. P.; Trager, W. F.; Carlson, T. J. The Binding and Regioselectivity of Reaction of (R)-Nicotine and (S)-Nicotine with Cytochrome-P-450cam - Parallel Experimental and Theoretical-Studies. *Journal of the American Chemical Society* **1993**, *115*, 381-387.

(35)   Harris, D.; Loew, G. Prediction of Regiospecific Hydroxylation of Camphor Analogs by Cytochrome-P450(Cam). *Journal of the American Chemical Society* **1995**, *117*, 2738-2746.

(36)   Keseru, G. M.; Kolossvary, I.; Bertok, B. Cytochrome P-450 catalyzed insecticide metabolism. Prediction of regio- and stereoselectivity in the primer metabolism of carbofuran: A theoretical study. *Journal of the American Chemical Society* **1997**, *119*, 5126-5131.

(37)   Harris, D. L.; Loew, G. H. Investigation of the proton-assisted pathway to formation of the catalytically active, ferryl species of P450s by molecular dynamics studies of P450eryF. *J Am Chem Soc* **1996**, *118*, 6377-6387.

(38)   Park, J. Y.; Harris, D. Construction and assessment of models of CYP2E1: Predictions of metabolism from docking, molecular dynamics, and density functional theoretical calculations. *J Med Chem* **2003**, *46*, 1645-1660.

(39)   Torimoto, N.; Ishii, I.; Hata, M.; Nakamura, H.; Imada, H. et al. Direct interaction between substrates and endogenous steroids in the active site may change the activity of cytochrome P450 3A4. *Biochemistry* **2003**, *42*, 15068-15077.

(40)   Harris, D. L.; Park, J. Y.; Gruenke, L.; Waskell, L. Theoretical study of the ligand-CYP2B4 complexes: effect of structure on binding free energies and heme spin state. *Proteins* **2004**, *55*, 895-914.

(41)   de Groot, M. J.; Alex, A. A.; Jones, B. C. Development of a combined protein and pharmacophore model for cytochrome P4502C9. *Journal of Medicinal Chemistry* **2002**, *45*, 1983-1993.

(42)   Sato, F.; Yoshihiro, T.; Era, M.; Kashiwagi, H. Calculation of all-electron wavefunction of hemoprotein cytochrome c by density functional theory. *Chemical Physics Letters* **2001**, *341*, 645-651.

(43)   Ekins, S.; Waller, C. L.; Swaan, P. W.; Cruciani, G.; Wrighton, S. A. et al. Progress in predicting human ADME parameters in silico. *J Pharmacol Toxicol Methods* **2000**, *44*, 251-272.

(44)   Ter Laak, A. M.; Vermeulen, N. P.; de Groot, M. J. Molecular Modeling Approaches to Predicting Drug Metabolism and Toxicity. *Drug-Drug Interactions*; Marcel Dekker, Inc: New York, 2002; pp 505-548.

(45)   Jones, J. P.; He, M. X.; Trager, W. F.; Rettie, A. E. Three-dimensional quantitative structure-activity relationship for inhibitors of cytochrome P4502C9. *Drug Metabolism and Disposition* **1996**, *24*, 1-6.

(46)   Mancy, A.; Dijols, S.; Poli, S.; Guengerich, P.; Mansuy, D. Interaction of sulfaphenazole derivatives with human liver cytochromes P450 2C: molecular origin of the specific inhibitory effects of sulfaphenazole on CYP 2C9 and consequences for the substrate binding site topology of CYP 2C9. *Biochemistry* **1996**, *35*, 16205-16212.

(47)   Ridderstrom, M.; Masimirembwa, C.; Trump-Kallmeyer, S.; Ahlefelt, M.; Otter, C. et al. Arginines 97 and 108 in CYP2C9 are important determinants of the catalytic function. *Biochem Biophys Res Commun* **2000**, *270*, 983-987.

(48)   Wolff, T.; Distelrath, L. M.; Worthington, M. T.; Groopman, J. D.; Hammons, G. J. et al. Substrate specificity of human liver cytochrome P-450 debrisoquine 4-hydroxylase probed using immunochemical inhibition and chemical modeling. *Cancer Res* **1985**, *45*, 2116-2122.

(49)   Meyer, U. A.; Gut, J.; Kronbach, T.; Skoda, C.; Meier, U. T. et al. The molecular mechanisms of two common polymorphisms of drug oxidation--evidence for functional changes in cytochrome P-450 isozymes catalysing bufuralol and mephenytoin oxidation. *Xenobiotica* **1986**, *16*, 449-464.

(50)   Islam, S. A.; Wolf, C. R.; Lennard, M. S.; Sternberg, M. J. A three-dimensional molecular template for substrates of human cytochrome P450 involved in debrisoquine 4-hydroxylation. *Carcinogenesis* **1991**, *12*, 2211-2219.

(51)   Koymans, L.; Vermeulen, N. P.; van Acker, S. A.; te Koppele, J. M.; Heykants, J. J. et al. A predictive model for substrates of cytochrome P450-debrisoquine (2D6). *Chem Res Toxicol* **1992**, *5*, 211-219.

(52)   de Groot, M. J.; Bijloo, G. J.; Hansen, K. T.; Vermeulen, N. P. Computer prediction and experimental validation of cytochrome P4502D6-dependent oxidation of GBR 12909. *Drug Metab Dispos* **1995**, *23*, 667-669.

(53)   de Groot, M. J.; Bijloo, G. J.; van Acker, F. A.; Fonseca Guerra, C.; Snijders, J. G. et al. Extension of a predictive substrate model for human cytochrome P4502D6. *Xenobiotica* **1997**, *27*, 357-368.

(54)   Ellis, S. W.; Hayhurst, G. P.; Smith, G.; Lightfoot, T.; Wong, M. M. et al. Evidence that aspartic acid 301 is a critical substrate-contact residue in the active site of cytochrome P450 2D6. *J Biol Chem* **1995**, *270*, 29055-29058.

(55)   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.

(56)   de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A.; Jones, B. C. A novel approach to predicting P450 mediated drug metabolism. CYP2D6 catalyzed N-dealkylation reactions and qualitative metabolite predictions using a combined protein and pharmacophore model for CYP2D6. *Journal of Medicinal Chemistry* **1999**, *42*, 4062-4070.

(57)   Strobl, G. R.; von Kruedener, S.; Stockigt, J.; Guengerich, F. P.; Wolff, T. Development of a pharmacophore for inhibition of human liver cytochrome P-450 2D6: molecular modeling and inhibition studies. *J Med Chem* **1993**, *36*, 1136-1145.

(58)   Kubinyi, H. QSAR and 3D QSAR in drug design .1. methodology. *Drug Discovery Today* **1997**, *2*, 457-467.

(59)   Verras, A.; Kuntz, I. D.; de Montellano, P. R. O. Computer-assisted design of selective imidazole inhibitors for cytochrome p450 enzymes. *Journal of Medicinal Chemistry* **2004**, *47*, 3572-3579.

(60)   Hollenberg, P. F. Characteristics and common properties of inhibitors, inducers, and activators of CYP enzymes. *Drug Metab Rev* **2002**, *34*, 17-35.

(61)   Afzelius, L.; Zamora, I.; Masimirembwa, C. M.; Karlen, A.; Andersson, T. B. et al. Conformer- and alignment-independent model for predicting structurally diverse competitive CYP2C9 inhibitors. *J Med Chem* **2004**, *47*, 907-914.

(62)   Pastor, M.; Cruciani, G.; Watson, K. A. A strategy for the incorporation of water molecules present in a ligand binding site into a three-dimensional quantitative structure--activity relationship analysis. *J Med Chem* **1997**, *40*, 4089-4102.

(63) Afzelius, L.; Raubacher, F.; Karlen, A.; Jorgensen, F. S.; Andersson, T. B. et al. Structural analysis of CYP2C9 and CYP2C5 and an evaluation of commonly used molecular modeling techniques. *Drug Metabolism and Disposition* **2004**, *32*, 1218-1229.

(64) Lozano, J. J.; Pastor, M.; Cruciani, G.; Gaedt, K.; Centeno, N. B. et al. 3D-QSAR methods on the basis of ligand-receptor complexes. Application of COMBINE and GRID/GOLPE methodologies to a series of CYP1A2 ligands. *J Comput Aided Mol Des* **2000**, *14*, 341-353.

(65) Poso, A.; Juvonen, R.; Gynther, J. Comparative molecular field analysis of compounds with CYP2A5 binding affinity. *Quantitative Structure-Activity Relationships* **1995**, *14*, 507-511.

(66) Bravi, G.; Wikel, J. H. Application of MS-WHIM descriptors: 1. Introduction of new molecular surface properties and 2. Prediction of binding affinity data. *Quantitative Structure-Activity Relationships* **2000**, *19*, 29-38.

(67) Poso, A.; Gynther, J.; Juvonen, R. A comparative molecular field analysis of cytochrome P450 2A5 and 2A6 inhibitors. *J Comput Aided Mol Des* **2001**, *15*, 195-202.

(68) Ekins, S.; Bravi, G.; Ring, B. J.; Gillespie, T. A.; Gillespie, J. S. et al. Three-dimensional quantitative structure activity relationship analyses of substrates for CYP2B6. *Journal of Pharmacology and Experimental Therapeutics* **1999**, *288*, 21-29.

(69) Snyder, R.; Sangar, R.; Wang, J. B.; Ekins, S. Three-dimensional quantitative structure activity relationship for CYP2D6 substrates. *Quantitative Structure-Activity Relationships* **2002**, *21*, 357-368.

(70) Bruns, C. M.; Hubatsch, I.; Ridderstrom, M.; Mannervik, B.; Tainer, J. A. Human glutathione transferase A4-4 crystal structures and mutagenesis reveal the basis of high catalytic efficiency with toxic lipid peroxidation products. *J Mol Biol* **1999**, *288*, 427-439.

(71) Yamamoto, J.; Ikeda, Y.; Iguchi, H.; Fujino, T.; Tanaka, T. et al. A Kruppel-like factor KLF15 contributes fasting-induced transcriptional activation of mitochondrial acetyl-CoA synthetase gene AceCS2. *J Biol Chem* **2004**, *279*, 16954-16962.

(72) Jones, B. C.; Hawksworth, G.; Horne, V. A.; Newlands, A.; Morsman, J. et al. Putative active site template model for cytochrome P4502C9 (tolbutamide hydroxylase). *Drug Metab Dispos* **1996**, *24*, 260-266.

(73) Mancy, A.; Broto, P.; Dijols, S.; Dansette, P. M.; Mansuy, D. The substrate binding site of human liver cytochrome P450 2C9: an approach using designed tienilic acid derivatives and molecular modeling. *Biochemistry* **1995**, *34*, 10365-10375.

(74) Paulsen, M. D.; Ornstein, R. L. Binding free energy calculations for P450cam-substrate complexes. *Protein Engineering* **1996**, *9*, 567-571.

(75) Rao, S.; Aoyama, R.; Schrag, M.; Trager, W. F.; Rettie, A. et al. A refined 3-dimensional QSAR of cytochrome P4502C9: Computational predictions of drug interactions. *Journal of Medicinal Chemistry* **2000**, *43*, 2789-2796.

(76) Afzelius, L.; Zamora, I.; Ridderstrom, M.; Andersson, T. B.; Karlen, A. et al. Competitive CYP2C9 inhibitors: Enzyme inhibition studies, protein homology modeling, and three-dimensional quantitative structure-activity relationship analysis. *Molecular Pharmacology* **2001**, *59*, 909-919.

(77) Afzelius, L.; Masimirembwa, C. M.; Karlen, A.; Andersson, T. B.; Zamora, I. Discriminant and quantitative PLS analysis of competitive CYP2C9 inhibitors versus non-inhibitors using alignment independent GRIND descriptors. *J Comput Aided Mol Des* **2002**, *16*, 443-458.

(78) Egnell, A. C.; Eriksson, C.; Albertson, N.; Houston, B.; Boyer, S. Generation and evaluation of a CYP2C9 heteroactivation pharmacophore. *Journal of Pharmacology and Experimental Therapeutics* **2003**, *307*, 878-887.

(79) 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.

(80) de Groot, M. J.; Bijloo, G. J.; Martens, B. J.; van Acker, F. A.; Vermeulen, N. P. A refined substrate model for human cytochrome P450 2D6. *Chem Res Toxicol* **1997**, *10*, 41-48.

(81) de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A.; Jones, B. C. Novel approach to predicting P450-mediated drug metabolism: Development of a combined protein and pharmacophore model for CYP2D6. *Journal of Medicinal Chemistry* **1999**, *42*, 1515-1524.

(82) Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, J. S.; Ring, B. J. et al. Three and four dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2D6 inhibitors. *Pharmacogenetics* **1999**, *9*, 477-489.

(83) Haji-Momenian, S.; Rieger, J. M.; Macdonald, T. L.; Brown, M. L. Comparative molecular field analysis and QSAR on substrates binding to cytochrome p450 2D6. *Bioorg Med Chem* **2003**, *11*, 5545-5554.

(84) Lewis, D. F.; Eddershaw, P. J.; Goldfarb, P. S.; Tarbit, M. H. Molecular modelling of CYP3A4 from an alignment with CYP102: Identification of key interactions between putative active site residues and CYP3A-specific chemicals. *Xenobiotica* **1996**, *26*, 1067-1086.

(85) Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, J. S.; Ring, B. J. et al. Three- and four-dimensional quantitative structure activity relationship analyses of cytochrome P-450 3A4 inhibitors. *Journal of Pharmacology and Experimental Therapeutics* **1999**, *290*, 429-438.

(86)   Ekins, S.; Bravi, G.; Wikel, J. H.; Wrighton, S. A. Three-dimensional-quantitative structure activity relationship analysis of cytochrome P-450 3A4 substrates. *Journal of Pharmacology and Experimental Therapeutics* **1999**, *291*, 424-433.

(87)   Ekins, S.; Stresser, D. M.; Williams, J. A. In vitro and pharmacophore insights into CYP3A enzymes. *Trends in Pharmacological Sciences* **2003**, *24*, 161-166.

(88)   Schappach, A.; Holtje, H. D. Investigations on inhibitors of human 17 alpha-hydroxylase-17,20-lyase and their interactions with the enzyme. Molecular modelling of 17 alpha-hydroxylase-17,20-lyase, Part II. *Pharmazie* **2001**, *56*, 835-842.

(89)   Clement, O. O.; Freeman, C. M.; Hartmann, R. W.; Handratta, V. D.; Vasaitis, T. S. et al. Three dimensional pharmacophore modeling of human CYP17 inhibitors. Potential agents for prostate cancer therapy. *J Med Chem* **2003**, *46*, 2345-2351.

(90)   Nagy, P. I.; Tokarski, J.; Hopfinger, A. J. Molecular shape and QSAR analyses of a family of substituted dichlorodiphenyl aromatase inhibitors. *J Chem Inf Comput Sci* **1994**, *34*, 1190-1197.

(91)   Oprea, T. I.; Garcia, A. E. Three-dimensional quantitative structure-activity relationships of steroid aromatase inhibitors. *J Comput Aided Mol Des* **1996**, *10*, 186-200.

(92)   Recanatini, M.; Cavalli, A. Comparative molecular field analysis of non-steroidal aromatase inhibitors: an extended model for two different structural classes. *Bioorg Med Chem* **1998**, *6*, 377-388.

(93)   Tafi, A.; Anastassopoulou, J.; Theophanides, T.; Botta, M.; Corelli, F. et al. Molecular modeling of azole antifungal agents active against Candida albicans .1. A comparative molecular field analysis study. *Journal of Medicinal Chemistry* **1996**, *39*, 1227-1235.

(94)   Talele, T. T.; Kulkarni, S. S.; Kulkarni, V. M. Development of pharmacophore alignment models as input for comparative molecular field analysis of a diverse set of azole antifungal agents. *Journal of Chemical Information and Computer Sciences* **1999**, *39*, 958-966.

(95)   Bargar, T. M.; Secor, J.; Markley, L. D.; Shaw, B. A.; Erickson, J. A. A comparative molecular field analysis study of obtusifoliol 14 alpha-methyl demethylase inhibitors. *Pesticide Science* **1999**, *55*, 1059-1069.

(96)   Li, H. Y.; Poulos, T. L. The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. *Nature Structural Biology* **1997**, *4*, 140-146.

(97)   Haines, D. C.; Tomchick, D. R.; Machius, M.; Peterson, J. A. Pivotal role of water in the mechanism of P450BM-3. *Biochemistry* **2001**, *40*, 13456-13465.

(98)   Modi, S.; Paine, M. J.; Sutcliffe, M. J.; Lian, L. Y.; Primrose, W. U. et al. A model for human cytochrome P450 2D6 based on homology modeling and NMR studies of substrate binding. *Biochemistry* **1996**, *35*, 4540-4550.

(99)   Raag, R.; Poulos, T. L. Crystal-Structures of Cytochrome-P-450cam Complexed with Camphane, Thiocamphor, and Adamantane - Factors Controlling P-450 Substrate Hydroxylation. *Biochemistry* **1991**, *30*, 2674-2684.

(100)   Strickler, M.; Goldstein, B. M.; Maxfield, K.; Shireman, L.; Kim, G. et al. Crystallographic studies on the complex behavior of nicotine binding to P450cam (CYP101). *Biochemistry* **2003**, *42*, 11943-11950.

(101)   Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dansette, P. M.; Mansuy, D. et al. Structure of a substrate complex of mammalian cytochrome P450 2C5 at 2.3 A resolution: evidence for multiple substrate binding modes. *Biochemistry* **2003**, *42*, 6370-6379.

(102)   Cupp-Vickery, J.; Anderson, R.; Hatziris, Z. Crystal structures of ligand complexes of P450eryF exhibiting homotropic cooperativity. *Proc Natl Acad Sci U S A* **2000**, *97*, 3050-3055.

(103)   Brunger, A. T.; Adams, P. D. Molecular dynamics applied to X-ray structure refinement. *Acc Chem Res* **2002**, *35*, 404-412.

(104)   Lüdemann, S. K.; Carugo, O.; Wade, R. C. Substrate access to Cytochrome P450cam: A comparison of a thermal motion pathway analysis with molecular dynamics simulation data. *Journal of Molecular Modeling* **1997**, *3*, 369-374.

(105)   Li, H.; Poulos, T. L. Conformational dynamics in cytochrome P450-substrate interactions. *Biochimie* **1996**, *78*, 695-699.

(106)   Paulsen, M. D.; Ornstein, R. L. Dramatic Differences in the Motions of the Mouth of Open and Closed Cytochrome P450bm-3 by Molecular-Dynamics Simulations. *Proteins-Structure Function and Genetics* **1995**, *21*, 237-243.

(107)   Chang, Q.; Harter, T. M.; Rikimaru, L. T.; Petrash, J. M. Aldo-keto reductases as modulators of stress response. *Chem Biol Interact* **2003**, *143-144*, 325-332.

(108)   Ludemann, S. K.; Lounnas, V.; Wade, R. C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. *Journal of Molecular Biology* **2000**, *303*, 797-811.

(109)   Ludemann, S. K.; Lounnas, V.; Wade, R. C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways. *Journal of Molecular Biology* **2000**, *303*, 813-830.

(110)   Winn, P. J.; Ludemann, S. K.; Gauges, R.; Lounnas, V.; Wade, R. C. Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a

novel functional role for a buried arginine. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 5361-5366.

(111)   Wade, R. C.; Winn, P. J.; Schlichting, I.; Sudarko A survey of active site access channels in cytochromes P450. *J Inorg Biochem* **2004**, *98*, 1175-1182.

(112)   Poulos, T. L.; Finzel, B. C.; Howard, A. J. Crystal structure of substrate-free Pseudomonas putida cytochrome P-450. *Biochemistry* **1986**, *25*, 5314-5322.

(113)   Sevrioukova, I. F.; Li, H.; Zhang, H.; Peterson, J. A.; Poulos, T. L. Structure of a cytochrome P450-redox partner electron-transfer complex. *Proc Natl Acad Sci U S A* **1999**, *96*, 1863-1868.

(114)   Cupp-Vickery, J. R.; Garcia, C.; Hofacre, A.; McGee-Estrada, K. Ketoconazole-induced conformational changes in the active site of cytochrome P450eryF. *J Mol Biol* **2001**, *311*, 101-110.

(115)   Hasemann, C. A.; Ravichandran, K. G.; Peterson, J. A.; Deisenhofer, J. Crystal structure and refinement of cytochrome P450terp at 2.3 A resolution. *J Mol Biol* **1994**, *236*, 1169-1185.

(116)   Park, S. Y.; Yamane, K.; Adachi, S.; Shiro, Y.; Weiss, K. E. et al. Crystallization and preliminary X-ray diffraction analysis of a cytochrome P450 (CYP119) from Sulfolobus solfataricus. *Acta Crystallogr D Biol Crystallogr* **2000**, *56 (Pt 9)*, 1173-1175.

(117)   Oku, Y.; Ohtaki, A.; Kamitori, S.; Nakamura, N.; Yohda, M. et al. Structure and direct electrochemistry of cytochrome P450 from the thermoacidophilic crenarchaeon, Sulfolobus tokodaii strain 7. *J Inorg Biochem* **2004**, *98*, 1194-1199.

(118)   Lee, D. S.; Yamada, A.; Sugimoto, H.; Matsunaga, I.; Ogura, H. et al. Substrate recognition and molecular mechanism of fatty acid hydroxylation by cytochrome P450 from Bacillus subtilis - Crystallographic, spectroscopic, and mutational studies. *Journal of Biological Chemistry* **2003**, *278*, 9761-9767.

(119)   Podust, L. M.; Bach, H.; Kim, Y.; Lamb, D. C.; Arase, M. et al. Comparison of the 1.85 A structure of CYP154A1 from Streptomyces coelicolor A3(2) with the closely related CYP154C1 and CYPs from antibiotic biosynthetic pathways. *Protein Sci* **2004**, *13*, 255-268.

(120)   Podust, L. M.; Kim, Y.; Arase, M.; Neely, B. A.; Beck, B. J. et al. The 1.92-angstrom structure of Streptomyces coelicolor A3(2) CYP154C1 - A new monooxygenase that functionalizes macrolide ring systems. *Journal of Biological Chemistry* **2003**, *278*, 12214-12221.

(121)   Zhao, B.; Guengerich, F. P.; Bellamine, A.; Lamb, D. C.; Izumikawa, M. et al. Binding of two flaviolin substrate molecules, oxidative coupling, and crystal structure of Streptomyces coelicolor A3(2) cytochrome P450 158A2. *J Biol Chem* **2005**, *280*, 11599-11607.

(122)   Yano, J. K.; Blasco, F.; Li, H. Y.; Schmid, R. D.; Henne, A. et al. Preliminary characterization and crystal structure of a thermostable cytochrome P450 from Thermus thermophilus. *Journal of Biological Chemistry* **2003**, *278*, 608-616.

(123)   Nagano, S.; Li, H.; Shimizu, H.; Nishida, C.; Ogura, H. et al. Crystal structures of epothilone D-bound, epothilone B-bound, and substrate-free forms of cytochrome P450epoK. *J Biol Chem* **2003**, *278*, 44886-44893.

(124)   Zerbe, K.; Pylypenko, O.; Vitali, F.; Zhang, W. W.; Rouse, S. et al. Crystal structure of OxyB, a cytochrome P450 implicated in an oxidative phenol coupling reaction during vancomycin biosynthesis. *Journal of Biological Chemistry* **2002**, *277*, 47476-47485.

(125)   Pylypenko, O.; Vitali, F.; Zerbe, K.; Robinson, J. A.; Schlichting, I. Crystal structure of OxyC, a cytochrome P450 implicated in an oxidative C-C coupling reaction during vancomycin biosynthesis. *J Biol Chem* **2003**, *278*, 46727-46733.

(126)   Podust, L. M.; Yermalitskaya, L. V.; Lepesheva, G. I.; Podust, V. N.; Dalmasso, E. A. et al. Estriol bound and ligand-free structures of sterol 14alpha-demethylase. *Structure (Camb)* **2004**, *12*, 1937-1945.

(127)   Shimizu, H.; Park, S. Y.; Shiro, Y.; Adachi, S. X-ray structure of nitric oxide reductase (cytochrome P450nor) at atomic resolution. *Acta Crystallographica Section D-Biological Crystallography* **2002**, *58*, 81-89.

(128)   Leys, D.; Mowat, C. G.; McLean, K. J.; Richmond, A.; Chapman, S. K. et al. Atomic structure of Mycobacterium tuberculosis CYP121 to 1.06 A reveals novel features of cytochrome P450. *J Biol Chem* **2003**, *278*, 5141-5147.

(129)   Yano, J. K.; Hsu, M. H.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen. *Nat Struct Mol Biol* **2005**, *12*, 822-823.

(130)   Scott, E. E.; White, M. A.; He, Y. A.; Johnson, E. F.; Stout, C. D. et al. Structure of mammalian cytochrome P450 2B4 complexed with 4-(4-chlorophenyl)imidazole at 1.9-A resolution: insight into the range of P450 conformations and the coordination of redox partner binding. *J Biol Chem* **2004**, *279*, 27294-27301.

(131)   Williams, P. A.; Cosme, J.; Sridhar, V.; Johnson, E. F.; McRee, D. E. Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol Cell* **2000**, *5*, 121-131.

(132) Schoch, G. A.; Yano, J. K.; Wester, M. R.; Griffin, K. J.; Stout, C. D. et al. Structure of human microsomal cytochrome P450 2C8. Evidence for a peripheral fatty acid binding site. *J Biol Chem* **2004***, 279*, 9497-9503.

(133) Rowland, P.; Blaney, F. E.; Smyth, M. G.; Jones, J. J.; Leydon, V. R. et al. Crystal structure of human cytochrome P450 2D6. *J Biol Chem* **2006***, 281*, 7614-7622.

(134) Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D. et al. The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-A resolution. *J Biol Chem* **2004***, 279*, 38091-38094.

(135) Helms, V.; Wade, R. C. Thermodynamics of water mediating protein-ligand interactions in cytochrome P450cam: a molecular dynamics study. *Biophys J* **1995***, 69*, 810-824.

(136) Oprea, T. I.; Hummer, G.; Garcia, A. E. Identification of a functional water channel in cytochrome P450 enzymes. *Proc Natl Acad Sci U S A* **1997***, 94*, 2133-2138.

(137) Taraphder, S.; Hummer, G. Protein side-chain motion and hydration in proton-transfer pathways. Results for cytochrome p450cam. *J Am Chem Soc* **2003***, 125*, 3931-3940.

(138) Fradera, X.; De La Cruz, X.; Silva, C. H.; Gelpi, J. L.; Luque, F. J. et al. Ligand-induced changes in the binding sites of proteins. *Bioinformatics* **2002***, 18*, 939-948.

(139) Park, S. Y.; Yamane, K.; Adachi, S.; Shiro, Y.; Weiss, K. E. et al. Thermophilic cytochrome P450 (CYP119) from Sulfolobus solfataricus: high resolution structure and functional properties. *J Inorg Biochem* **2002***, 91*, 491-501.

(140) Podust, L. M.; Poulos, T. L.; Waterman, M. R. Crystal structure of cytochrome P450 14alpha -sterol demethylase (CYP51) from Mycobacterium tuberculosis in complex with azole inhibitors. *Proc Natl Acad Sci U S A* **2001***, 98*, 3068-3073.

(141) Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dijols, S.; Dansette, P. M. et al. Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 A resolution: evidence for an induced fit model of substrate binding. *Biochemistry* **2003***, 42*, 9335-9345.

(142) 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.

(143) Jean, P.; Pothier, J.; Dansette, P. M.; Mansuy, D.; Viari, A. Automated multiple analysis of protein structures: application to homology modeling of cytochromes P450. *Proteins* **1997***, 28*, 388-404.

(144) Devoss, J. J.; Demontellano, P. R. O. Computer-Assisted, Structure-Based Prediction of Substrates for Cytochrome P450(Cam). *Journal of the American Chemical Society* **1995***, 117*, 4185-4186.

(145) Koymans, L. M.; Vermeulen, N. P.; Baarslag, A.; Donne-Op den Kelder, G. M. A preliminary 3D model for cytochrome P450 2D6 constructed by homology model building. *J Comput Aided Mol Des* **1993***, 7*, 281-289.

(146) De Rienzo, F.; Fanelli, F.; Menziani, M. C.; De Benedetti, P. G. Theoretical investigation of substrate specificity for cytochromes P450 IA2, P450 IID6 and P450 IIIA4. *J Comput Aided Mol Des* **2000***, 14*, 93-116.

(147) Modi, S.; Gilham, D. E.; Sutcliffe, M. J.; Lian, L. Y.; Primrose, W. U. et al. 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine as a substrate of cytochrome P450 2D6: allosteric effects of NADPH-cytochrome P450 reductase. *Biochemistry* **1997***, 36*, 4461-4470.

(148) 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.

(149) Spatzenegger, M.; Wang, Q. M.; He, Y. Q.; Wester, M. R.; Johnson, E. F. et al. Amino acid residues critical for differential inhibition of CYP2B4, CYP2B5, and CYP2B1 by phenylimidazoles. *Molecular Pharmacology* **2001***, 59*, 475-484.

(150) Bathelt, C.; Schmid, R. D.; Pleiss, J. Regioselectivity of CYP2B6: homology modeling, molecular dynamics simulation, docking. *Journal of Molecular Modeling* **2002***, 8*, 327-335.

(151) Zvelebil, M. J.; Wolf, C. R.; Sternberg, M. J. A predicted three-dimensional structure of human cytochrome P450: implications for substrate specificity. *Protein Eng* **1991***, 4*, 271-282.

(152) Szklarz, G. D.; Paulsen, M. D. Molecular modeling of cytochrome P450 1A1: enzyme-substrate interactions and substrate binding affinities. *J Biomol Struct Dyn* **2002***, 20*, 155-162.

(153) 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.

(154) Chou, W. H.; Yan, F. X.; Robbins-Weilert, D. K.; Ryder, T. B.; Liu, W. W. et al. Comparison of two CYP2D6 genotyping methods and assessment of genotype-phenotype relationships. *Clin Chem* **2003***, 49*, 542-551.

(155) Lewis, D. F.; Gillam, E. M. J.; Everett, S. A.; Shimada, T. Molecular modelling of human CYP1B1 substrate interactions and investigation of allelic variant effects on metabolism. *Chemico-Biological Interactions* **2003***, 145*, 281-295.

(156) Lewis, D. F.; Lake, B. G. Molecular modelling of members of the P4502A subfamily: application to studies of enzyme specificity. *Xenobiotica* **1995***, 25*, 585-598.

(157) He, X. Y.; Shen, J.; Hu, W. Y.; Ding, X.; Lu, A. Y. et al. Identification of Val117 and Arg372 as critical amino acid residues for the activity difference between human CYP2A6 and CYP2A13 in coumarin 7-hydroxylation. *Arch Biochem Biophys* **2004**, *427*, 143-153.

(158) Kobayashi, Y.; Fang, X. J.; Szklarz, G. D.; Halpert, J. R. Probing the active site of cytochrome P450 2B1: Metabolism of 7-alkoxycoumarins by the wild type and five site-directed mutants. *Biochemistry* **1998**, *37*, 6679-6688.

(159) Szklarz, G. D.; He, Y. A.; Halpert, J. R. Site-directed mutagenesis as a tool for molecular modeling of cytochrome P450 2B1. *Biochemistry* **1995**, *34*, 14312-14322.

(160) Chang, Y. T.; Stiffelman, O. B.; Vakser, I. A.; Loew, G. H.; Bridges, A. et al. Construction of a 3D model of cytochrome P450 2B4. *Protein Eng* **1997**, *10*, 119-129.

(161) Domanski, T. L.; Schultz, K. M.; Roussel, F.; Stevens, J. C.; Halpert, J. R. Structure-function analysis of human cytochrome P-4502B6 using a novel substrate, site-directed mutagenesis, and molecular modeling. *Journal of Pharmacology and Experimental Therapeutics* **1999**, *290*, 1141-1147.

(162) Payne, V. A.; Chang, Y. T.; Loew, G. H. Homology modeling and substrate binding study of human CYP2C9 enzyme. *Proteins* **1999**, *37*, 176-190.

(163) Payne, V. A.; Chang, Y. T.; Loew, G. H. Homology modeling and substrate binding study of human CYP2C18 and CYP2C19 enzymes. *Proteins* **1999**, *37*, 204-217.

(164) Lewis, D. F.; Bird, M. G.; Parke, D. V. Molecular modelling of CYP2E1 enzymes from rat, mouse and man: An explanation for species differences in butadiene metabolism and potential carcinogenicity, and rationalization of CYP2E substrate specificity. *Toxicology* **1997**, *118*, 93-113.

(165) Pirmohamed, M.; Park, B. K. Cytochrome P450 enzyme polymorphisms and adverse drug reactions. *Toxicology* **2003**, *192*, 23-32.

(166) Xue, L.; Wang, H. F.; Wang, Q.; Szklarz, G. D.; Domanski, T. L. et al. Influence of P450 3A4 SRS-2 residues on cooperativity and/or regioselectivity of aflatoxin B(1) oxidation. *Chem Res Toxicol* **2001**, *14*, 483-491.

(167) Kuhn, B.; Jacobsen, W.; Christians, U.; Benet, L. Z.; Kollman, P. A. Metabolism of sirolimus and its derivative everolimus by cytochrome P450 3A4: insights from docking, molecular dynamics, and quantum chemical calculations. *J Med Chem* **2001**, *44*, 2027-2034.

(168) Yamaguchi, Y.; Khan, K. K.; He, Y. A.; He, Y. Q.; Halpert, J. R. Topological changes in the CYP3A4 active site probed with phenyldiazene: Effect of interaction with NADPH-cytochrome P450 reductase and cytochrome B5 and of site-directed mutagenesis. *Drug Metabolism and Disposition* **2004**, *32*, 155-161.

(169) Szklarz, G. D.; Halpert, J. R. Molecular modeling of cytochrome P450 3A4. *J Comput Aided Mol Des* **1997**, *11*, 265-272.

(170) Lewis, D. F.; Lake, B. G. Molecular modelling of CYP4A subfamily members based on sequence homology with CYP102. *Xenobiotica* **1999**, *29*, 763-781.

(171) Kalsotra, A.; Turman, C. M.; Kikuta, Y.; Strobel, H. W. Expression and characterization of human cytochrome P450 4F11: Putative role in the metabolism of therapeutic drugs and eicosanoids. *Toxicol Appl Pharmacol* **2004**, *199*, 295-304.

(172) Pan, L. P.; Wen, Z. M.; Baudry, J.; Berenbaum, M. R.; Schuler, M. A. Identification of variable amino acids in the SRS1 region of CYP6B1 modulating furanocoumarin metabolism. *Archives of Biochemistry and Biophysics* **2004**, *422*, 31-41.

(173) Baudry, J.; Li, W.; Pan, L.; Berenbaum, M. R.; Schuler, M. A. Molecular docking of substrates and inhibitors in the catalytic site of CYP6B1, an insect cytochrome p450 monooxygenase. *Protein Eng* **2003**, *16*, 577-587.

(174) Usanov, S. A.; Graham, S. E.; Lepesheva, G. I.; Azeva, T. N.; Strushkevich, N. V. et al. Probing the interaction of bovine cytochrome P450scc (CYP11A1) with adrenodoxin: evaluating site-directed mutations by molecular modeling. *Biochemistry* **2002**, *41*, 8310-8320.

(175) Laughton, C. A.; Zvelebil, M. J.; Neidle, S. A detailed molecular model for human aromatase. *J Steroid Biochem Mol Biol* **1993**, *44*, 399-407.

(176) Schappach, A.; Holtje, H. D. Molecular modelling of 17 alpha-hydroxylase-17,20-lyase. *Pharmazie* **2001**, *56*, 435-442.

(177) Auvray, P.; Nativelle, C.; Bureau, R.; Dallemagne, P.; Seralini, G. E. et al. Study of substrate specificity of human aromatase by site directed mutagenesis. *Eur J Biochem* **2002**, *269*, 1393-1405.

(178) Chen, S.; Zhang, F.; Sherman, M. A.; Kijima, I.; Cho, M. et al. Structure-function studies of aromatase and its inhibitors: a progress report. *J Steroid Biochem Mol Biol* **2003**, *86*, 231-237.

(179) Ishida, N.; Aoyama, Y.; Hatanaka, R.; Oyama, Y.; Imajo, S. et al. A single amino acid substitution converts cytochrome P450(14DM) to an inactive form, cytochrome P450SG1: complete primary structures deduced from cloned DNAS. *Biochem Biophys Res Commun* **1988**, *155*, 317-323.

(180) Holtje, H. D.; Fattorusso, C. Construction of a model of the Candida albicans lanosterol 14-alpha-demethylase active site using the homology modelling technique. *Pharm Acta Helv* **1998**, *72*, 271-277.

(181) Murtazina, D.; Puchkaev, A. V.; Schein, C. H.; Oezguen, N.; Braun, W. et al. Membrane-protein interactions contribute to efficient 27-hydroxylation of cholesterol by mitochondrial cytochrome P450 27A1. *J Biol Chem* **2002**, *277*, 37582-37589.

(182) Schoch, G. A.; Attias, R.; Le Ret, M.; Werck-Reichhart, D. Key substrate recognition residues in the active site of a plant cytochrome P450, CYP73A1. Homology guided site-directed mutagenesis. *Eur J Biochem* **2003**, *270*, 3684-3695.

(183) Braatz, J. A.; Bass, M. B.; Ornstein, R. L. An evaluation of molecular models of the cytochrome P450 Streptomyces griseolus enzymes P450SU1 and P450SU2. *J Comput Aided Mol Des* **1994**, *8*, 607-622.

(184) Chang, Y. T.; Loew, G. H. Construction and evaluation of a three-dimensional structure of cytochrome P450choP enzyme (CYP105C1). *Protein Eng* **1996**, *9*, 755-766.

(185) Chang, Y. T.; Loew, G. Homology modeling, molecular dynamics simulations, and analysis of CYP119, a P450 enzyme from extreme acidothermophilic archaeon Sulfolobus solfataricus. *Biochemistry* **2000**, *39*, 2484-2498.

(186) 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.

(187) 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.

(188) Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins* **1999**, *37*, 228-241.

(189) 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.

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

(191) McConkey, B. J.; Sobolev, V.; Edelman, M. The performance of current methods in ligand-protein docking. *Current Science* **2002**, *83*, 845-856.

(192) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* **2002**, *47*, 409-443.

(193) 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.

(194) 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. *Journal of the American Chemical Society* **1997**, *119*, 5489-5498.

(195) Bissantz, C.; Bernard, P.; Hibert, M.; Rognan, D. Protein-based virtual screening of chemical databases. II. Are homology models of G-Protein Coupled Receptors suitable targets? *Proteins* **2003**, *50*, 5-25.

(196) Wade, R. C.; Gabdoulline, R. R.; Ludemann, S. K.; Lounnas, V. Electrostatic steering and ionic tethering in enzyme-ligand binding: insights from simulations. *Proc Natl Acad Sci U S A* **1998**, *95*, 5942-5949.

(197) van Gunsteren, W. F.; Berendsen, H. J. C. Computer simulation of molecular dynamics: Methodology, applications and perspectives in chemistry. *Angew Chem Int Ed English* **1990**, *29*, 992-1023.

(198) van den Berg, P. A. W.; Feenstra, K. A.; Mark, A. E.; Berendsen, H. J. C.; Visser, A. Dynamic conformations of flavin adenine dinucleotide: Simulated molecular dynamics of the flavin cofactor related to the time- resolved fluorescence characteristics. *Journal of Physical Chemistry B* **2002**, *106*, 8858-8869.

(199) Feenstra, K. A.; Peter, C.; Scheek, R. M.; van Gunsteren, W. F.; Mark, A. E. A comparison of methods for calculating NMR cross-relaxation rates (NOESY and ROESY intensities) in small peptides. *Journal of Biomolecular Nmr* **2002**, *23*, 181-194.

(200) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. Improving Efficiency of Large Time-Scale Molecular Dynamics Simulations of Hydrogen-Rich Systems. *J. Comput. Chem.* **1999**, *20*, 786-798.

(201) Oostenbrink, C.; van Gunsteren, W. F. Single-step perturbations to calculate free energy differences from unphysical reference states: limits on size, flexibility, and character. *J Comput Chem* **2003**, *24*, 1730-1739.

(202) Hansson, T.; Oostenbrink, C.; van Gunsteren, W. Molecular dynamics simulations. *Curr Opin Struct Biol* **2002**, *12*, 190-196.

(203) Deprez, E.; Gill, E.; Helms, V.; Wade, R. C.; Hui Bon Hoa, G. Specific and non-specific effects of potassium cations on substrate-protein interactions in cytochromes P450cam and P450lin. *J Inorg Biochem* **2002**, *91*, 597-606.

(204) Arnold, G. E.; Ornstein, R. L. Molecular dynamics study of time-correlated protein domain motions and molecular flexibility: cytochrome P450BM-3. *Biophys J* **1997**, *73*, 1147-1159.

(205) Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D. et al. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **2003**, *31*, 3784-3788.

(206) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **1997**, *25*, 3389-3402.

(207) Holm, L.; Sander, C. Dali/FSSP classification of three-dimensional protein folds. *Nucleic Acids Res* **1997**, *25*, 231-234.

(208) Butler, B. A. Sequence analysis using GCG. *Methods Biochem Anal* **1998**, *39*, 74-97.

(209) Cuff, J. A.; Clamp, M. E.; Siddiqui, A. S.; Finlay, M.; Barton, G. J. JPred: a consensus secondary structure prediction server. *Bioinformatics* **1998**, *14*, 892-893.

(210) Notredame, C.; Higgins, D. G.; Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **2000**, *302*, 205-217.

(211) Kneller, D. G.; Cohen, F. E.; Langridge, R. Improvements in protein secondary structure prediction by an enhanced neural network. *J Mol Biol* **1990**, *214*, 171-182.

(212) McGuffin, L. J.; Bryson, K.; Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **2000**, *16*, 404-405.

(213) Havel, T. F.; Snow, M. E. A new method for building protein conformations from sequence alignments with homologues of known structure. *J Mol Biol* **1991**, *217*, 1-7.

(214) Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **1993**, *234*, 779-815.

(215) Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* **2003**, *31*, 3381-3385.

(216) Rost, B.; Liu, J. The PredictProtein server. *Nucleic Acids Res* **2003**, *31*, 3300-3304.

(217) Colovos, C.; Yeates, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* **1993**, *2*, 1511-1519.

(218) Laskowski, R. A.; Macarthur, M. W.; Moss, D. S.; Thornton, J. M. Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography* **1993**, *26*, 283-291.

(219) Sippl, M. J. Recognition of errors in three-dimensional structures of proteins. *Proteins* **1993**, *17*, 355-362.

(220) Pontius, J.; Richelle, J.; Wodak, S. J. Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J Mol Biol* **1996**, *264*, 121-136.

(221) Luthy, R.; Bowie, J. U.; Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* **1992**, *356*, 83-85.

(222) Cramer, R. D., 3rd; Patterson, D. E.; Bunce, J. D. Recent advances in comparative molecular field analysis (CoMFA). *Prog Clin Biol Res* **1989**, *291*, 161-165.

(223) Cruciani, G.; Watson, K. A. Comparative molecular field analysis using GRID force-field and GOLPE variable selection methods in a study of inhibitors of glycogen phosphorylase b. *J Med Chem* **1994**, *37*, 2589-2601.

(224) Cruciani, G.; Carosati, E.; De Boeck, B.; Ethirajulu, K.; Mackie, C. et al. MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. *J Med Chem* **2005**, *48*, 6970-6979.

(225) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E. et al. Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Computer Physics Communications* **1995**, *91*, 1-41.

(226) MacKerell, A. D., Jr. Atomistic Models and Force Fields. *In: Computational Biochemistry and Biophysics, O.M. Becker, A.D. MacKerell, Jr., B. Roux and M.Watanabe, Eds. Marcel Dekker, Inc. New York* **2001**, p. 7-38.

(227) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* **2001**, *7*, 306-317.

(228) Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R. et al. The GROMOS biomolecular simulation program package. *Journal of Physical Chemistry A* **1999**, *103*, 3596-3607.

(229) 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. *Journal of Computational Chemistry* **1998**, *19*, 1639-1662.

(230) 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.

(231) 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.

(232) 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.

(233) 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.

(234) 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.

(235) 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.

(236)   Muegge, I.; Martin, Y. C. A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *Journal of Medicinal Chemistry* **1999**, *42*, 791-804.

(237)   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. *Journal of Molecular Modeling* **1998**, *4*, 379-394.

(238)   Schnecke, V.; Kuhn, L. A. Virtual screening with solvation and ligand-induced complementarity. *Perspectives in Drug Discovery and Design* **2000**, *20*, 171-190.

(239)   Wang, R. X.; Lai, L. H.; Wang, S. M. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of Computer-Aided Molecular Design* **2002**, *16*, 11-26.

(240)   Vriend, G. What If - a Molecular Modeling and Drug Design Program. *Journal of Molecular Graphics* **1990**, *8*, 52-&.

(241)   Kraulis, P. J. Molscript - a Program to Produce Both Detailed and Schematic Plots of Protein Structures. *Journal of Applied Crystallography* **1991**, *24*, 946-950.

(242)   Merritt, E. A.; Murphy, M. E. P. Raster3d Version-2.0 - a Program for Photorealistic Molecular Graphics. *Acta Crystallographica Section D-Biological Crystallography* **1994**, *50*, 869-873.

(243)   Åqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. Ligand binding affinities from MD simulations. *Accounts of Chemical Research* **2002**, *35*, 358-365.

(244)   Oostenbrink, B. C.; Pitera, J. W.; van Lipzig, M. M.; Meerman, J. H.; van Gunsteren, W. F. Simulations of the estrogen receptor ligand-binding domain: affinity of natural ligands and xenoestrogens. *J Med Chem* **2000**, *43*, 4594-4605.

(245)   van Lipzig, M. M. H.; ter Laak, A. M.; Jongejan, A.; Vermeulen, N. P. E.; Wamelink, M. et al. Prediction of ligand binding affinity and orientation of xenoestrogens to the estrogen receptor by molecular dynamics simulations and the linear interaction energy method. *Journal of Medicinal Chemistry* **2004**, *47*, 1018-1030.

(246)   Scott, E. E.; He, Y. A.; Wester, M. R.; White, M. A.; Chin, C. C. et al. An open conformation of mammalian cytochrome P450 2B4 at 1.6-A resolution. *Proc Natl Acad Sci U S A* **2003**, *100*, 13196-13201.

(247)   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.

(248)   Clark, D. E.; Grootenhuis, P. D. J. Progress in computational methods for the prediction of ADMET properties. *Current Opinion in Drug Discovery & Development* **2002**, *5*, 382-390.

Construction of a CYP2D6 structural model

Carving out key positions

Molecular modeling-guided site-directed mutagenesis of cytochrome P450 2D6

Chris de Graaf,[‡] Chris Oostenbrink,[‡] Peter H.J. Keizers,[‡] Barbara M.A. van Vugt-Lussenburg,[‡] Robert A.B. van Waterschoot,[‡] Richard A. Tschirret-Guth,[§] Jan N.M. Commandeur,[‡] and Nico P.E. Vermeulen[‡]

[‡]Leiden Amsterdam Center for Drug Research (LACDR)/Division of Molecular Toxicology, Department of Chemistry and Pharmacochemistry, Vrije Universiteit, The Netherlands
[§]Department of Drug Metabolism, Merck Research Laboratories, Rahway, USA

Experimentally, the active site of CYP2D6 can be probed by site directed mutagenesis studies. Such studies can be designed from structural models of enzyme-substrate complexes and the observed effect of mutations on metabolism and inhibition can subsequently be rationalized at an atomic resolution by modeling approaches. The current chapter will present the construction, refinement and validation of the CYP2D6 homology model used in our laboratory for the prediction and rationalisation of CYP2D6 substrate metabolism and CYP2D6-ligand interactions. The protein model could explain previously reported site-directed mutagenesis data (for example, mutations of E216 and D301). Furthermore, based on the model, new CYP2D6 mutants were constructed and studied in our laboratory, and also for these mutants a rationalization of experimentally observed characteristics could be achieved (I106E, F120A, T309V, F483A). CYP2D6-substrate interaction fingerprint analysis of docked substrates in our homology model suggests that several other active site residues are probably interacting with ligands as well, opening the way for further mutagenesis studies. Importantly, our homology model was found to agree with most of the details of the recently solved substrate-free CYP2D6 crystal structure [Rowland et al. *J Biol Chem* **2006**, *281*, 7614-7622]. Structural differences between our homology model and the crystal structure were the same as the differences observed between substrate-free and substrate-bound structures of other CYPs, suggesting that these conformational changes are required upon substrate binding. The CYP2D6 crystal structure further validates our homology modeling approach and clearly demonstrates on one hand the success and accuracy of state of the art CYP homology modeling. On the other hand, we show that computational chemistry is a useful and valuable tool to provide models for substrate-bound complexes of CYPs which give insight into CYP-ligand interactions.

3.1 Introduction

Cytochromes P450 (CYPs) are heme containing enzymes which can be found in virtually all living organisms. This large family of enzymes is capable of oxidizing and reducing a broad range of endogenous and exogenous substrates, such as steroids, carcinogens and drugs.[1,2] In humans one of the most relevant drug metabolizing CYPs is CYP2D6. Although the expression levels of CYP2D6 are only 2% of all hepatic CYPs, following CYP3A4, it is the second most important drug metabolizing enzyme, involved in the metabolism of about 15-30% of the currently marketed drugs, including β-blockers, neuroleptics, antidepressants and antiarythmics.[3-5] Large interindividual differences exist in CYP2D6 activity, due to gene multiplicity and polymorphisms, thus further increasing its clinical importance.[6,7] The rationalization and prediction of potential CYP2D6 substrates is therefore advantageous in the discovery and development of new drugs.

A wealth of detailed information on CYP enzyme action and substrate or inhibitor interactions with the binding cavity becomes accessible when a structure of the CYP protein is available. In the past few years, crystal structures of many mammalian CYPs have become available: CYP2A6[8], CYP2B4[9,10], CYP2C5[11-13], CYP2C8[14], CYP2C9[15,16], and CYP3A4.[17,18] Very recently, a substrate-free crystal structure of CYP2D6 was solved.[19] Comparisons between ligand-bound and ligand-free crystal structures of many CYPs however, showed ligand induced changes in size, shape and hydration of the active site. Therefore the construction of properly validated three-dimensional structural models describing substrate-bound complexes of CYP2D6 are still necessary to obtain knowledge on CYP2D6 structure and function, leading to a better understanding, rationalization and prediction of the catalytic activity, ligand specificity, and regiospecificity of metabolism of this CYP isoform. Such a computational structural model can be used in combination with other modeling techniques, like automated docking and molecular dynamics simulations, to design site-directed mutagenesis experiments to explore the active site for amino acid residues determining ligand binding and regiospecificity of metabolism by CYP2D6. The primary aim of this chapter is to present the building, refinement and validation of a CYP2D6 homology model and its usage to explain and design experimental site-directed mutagenesis studies. The evolution of this CYP2D6 model can be considered, as a case study demonstrating the possibilities and limitations in the development, validation and application of enzyme homology models.

The increased number of available CYP protein structures from X-ray crystallography has been the basis for much homology modeling work (see **Chapter 2**).[20,21] In this respect, homology models have been complementary to crystal structures, especially for the mammalian CYPs now that mammalian crystal structures have very recently become available. Several CYP isoforms have only been crystallized in a non-active ligand-bound state, and for some it is not known whether the crystal structure corresponds to an active state or not.[15,22,23] It is important to realize that this may render the crystal structure of a certain CYP isoenzyme not directly applicable to *e.g.* explaining observed substrate selectivity and product formation by the enzyme. Ligand-induced changes in the size, shape and hydration of the active sites of several CYPs are also observed[13,24-27], and these are likely to underlie the capacity of many CYPs to metabolize substrates that are diverse in size and structure and to generate different metabolic products from these substrates. In all known CYP crystal structures available today, the same general three-dimensional fold is observed. A recent structural alignment of CYP crystal structures showed that over 28% of structurally aligned residues have their C-alpha atoms within a root-mean-square deviation of 2Å. This conserved core region contains the C-terminal half of helix I, helices L, E, K, and K', and the heme coordination region.[28,29] The three-dimensional structure of these regions is well conserved despite a low sequence homology (as low as ~15%).[29] Other regions (*e.g.*, the active site region containing the B' helix, the loops between helices C and D, the region spanning helices F and G, and most of the β-strands) are more

variable. The largest structural variations are found in the B' helix and helices F and G and their connecting loop. These regions are suggested to be involved in substrate access and recognition.[29,30] Six substrate recognition sites (SRS's) have been assigned based on mutagenesis and sequence alignment studies with representative members of the CYP2 family and CYP101 (*cam*).[31] However, the modeling studies on CYP2D6 presented in this chapter suggest that only specific amino acid residues located in these SRSs are actually directly involved in CYP2D6-ligand binding (e.g. F120[32] (SRS 1) and F483[33] (SRS 6)), while other residues have an indirect effect on ligand binding and catalysis by i) influencing the protein conformation and shape of the active site (e.g. I106[34] (SRS 1)); ii) stabilizing catalytically active heme-oxygen intermediates (e.g. T309[35] (SRS 4)); iii) controlling substrate entrance and/or product exit from CYP2D6 (see **Chapter 8**).

The current chapter will present the construction, refinement and validation of the CYP2D6 homology model used in this thesis and other work[32-37] for the prediction and rationalisation of CYP2D6 substrate metabolism and CYP2D6-ligand interactions. The structure has been used to explain reported site directed mutagenesis data. Furthermore, the homology model is compared with the recently solved substrate-free CYP2D6 crystal structure[19], discussing and demonstrating the value of computational chemistry in describing substrate-induced conformational changes in protein structures. This chapter will finally describe the design of site-directed mutagenesis studies from molecular modelling, exploring the active site of CYP2D6 for key determinants in CYP2D6-mediated ligand binding and metabolism. The experimental results of these mutagenesis studies, performed in our laboratory, are described in more detail elsewhere.[32-36]

3.2. Homology modeling of CYP2D6

A protein homology model of human CYP2D6 was constructed based on the crystal structures of rabbit CYP2C5 in complex with dimethylsulphophenazole derivative and a second structure of rabbit CYP2C5 in complex with diclofenac (PDB codes 1N6B[11] and 1NR6[13], respectively). In general, homology modeling, model refinement and model validation was performed according to the approach described earlier by Venhorst *et al.* [38]. A flowchart scheme of the homology modeling, refinement, and validation process is shown in **Figure 1**.

Using the amino acid alignment shown in **Figure 2** (step 1), a hundred CYP2D6 models were generated with the restraint-based comparative modeling program Modeller[39] (step 2). The primary structures of CYP2C5 and CYP2D6 share 43% sequence identity and 68% sequence homology. Subsequently, three models were selected with best protein conformation and amino acid environment (step 3, **Table 1**). Finally, one single model was selected which could accommodate codeine best in the experimentally determined binding orientation (step 4). The final model is optimized and refined using the known ligand-based CYP2D6 pharmacophore model and geometry optimization (step 5). This optimized model was validated and used in several applications that are described in this thesis (step 6). First, its ability to reproduce substrate binding orientations that lead to the formation of experimentally observed metabolites and its ability to perform virtual screening (VS) for high affinity CYP2D6 substrates in automated docking (AD) studies was explored (**Chapter 5**). Second, the substrate binding poses predicted by automated docking could be used to identify important active site residues influencing CYP2D6-mediated metabolism (this chapter and refs 32-37). Third, a combination of docking and molecular dynamics (MD) was used to predict the regio- and stereo-specificity of catalysis of MDMA (methylenedioxymethamphetamine) and MDMA-analogs (**Chapter 6**). Fourth, the relative binding affinity of propranolol stereo-isomers towards wild-type and mutant CYP2D6 was rationalised (**Chapter 7**). Finally, the substrate access and product exit channels in the model were explored using random acceleration molecular dynamics (RAMD, **Chapter 8**).
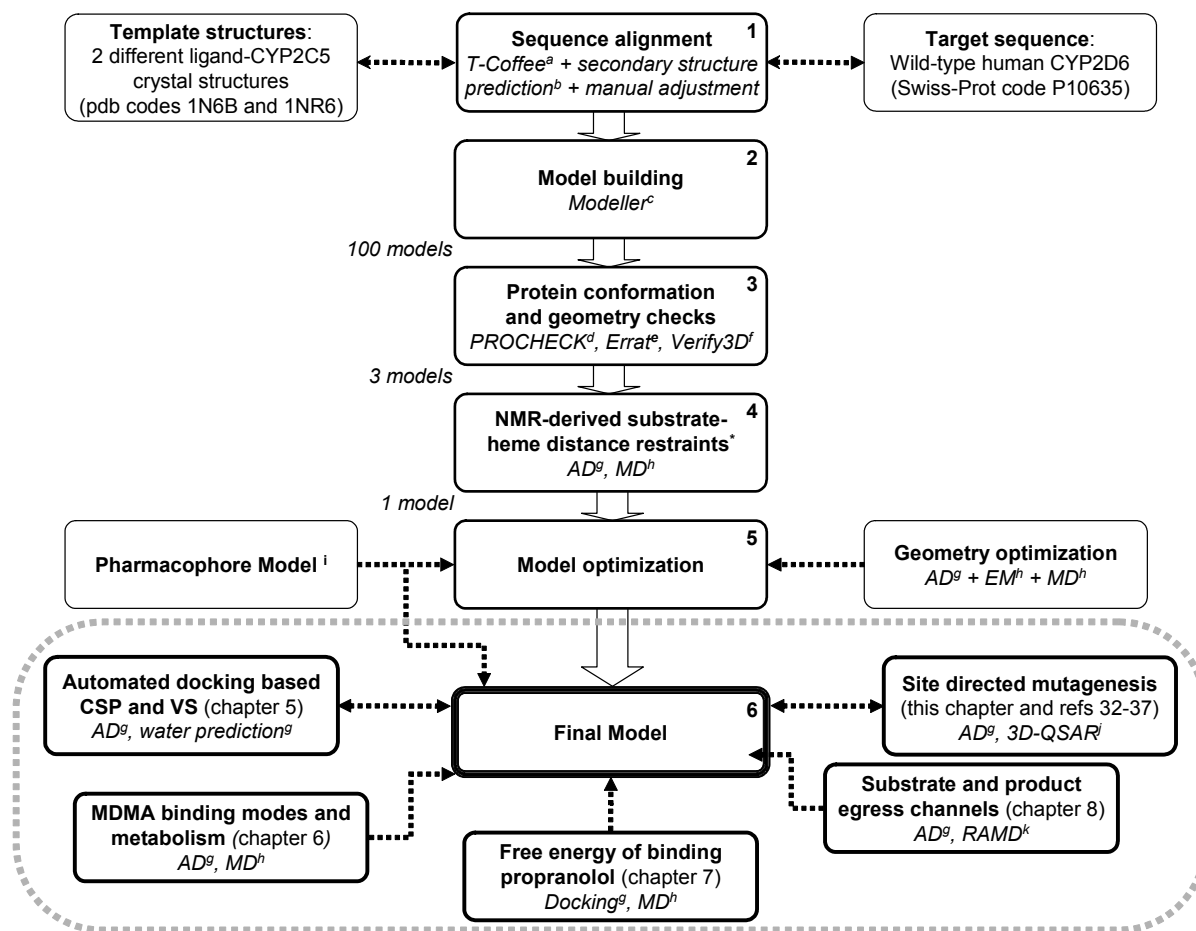
Figure 1: *Flow chart summarizing the construction, optimization, and validation of the protein homology model of CYP2D6 described in this chapter. The construction phase is covered in the first four steps: amino acid alignment to template X-ray crystal structures, model building and two subsequent steps of model selection. The model is optimized and refined using the known ligand-based CYP2D6 pharmacophore model and geometry optimization. Validation studies of the final model, described in chapters 3, 5, 6, 7, 8 and refs 32-37 (in brackets), are enclosed by a dashed box. Molecular modeling methods and software programs used in the different steps are shown in Italics.*

Captions: **a**) Sequence alignment program[40]; **b**) Secondary structure prediction performed with Predict Protein[41], JPred[42], and NnPredict[43]; **c**) Comparative homology modeling program[39]; **d**) Program to evaluate the stereochemical quality of protein structures[44]; **e**) Program to evaluate the amino acid side chain environment (packing quality) of protein structures[45]; **f**) Program to measure the compatibility of a protein model with its sequence[46]; **g**) Automated docking (AD) performed with AutoDock[47], FlexX[48], GOLD-Goldscore[49], GOLD-Chemscore[50], CScore, SCORE[51], and GRID[52]; **h**) Molecular dynamics simulations (MD) and energy minimization (EM) performed with Gromacs[53] and the GROMOS force field[54]; **i**) Following the features of CYP2D6 pharmacophore models desrcribed before (e.g. refs 55-63); **j**) 3D-QSAR performed with GRID/Almond[64]; **k**) (Random Accelleration) MD[65] performed with AMBER[66]

Template selection and amino acid sequence alignment

The high degree of structure conservation among CYP crystal structures[28,29], especially among crystal structures of members of the CYP2 family[29], suggests that it should be possible to derive models of overall good quality for CYP2D6 using recently published substrate-bound CYP2C5 crystal structures (pdb codes 1N6B and 1NR6).[11,13] These structures have a significantly higher resolution than the first resolved mammalian CYP crystal structure, the ligand-free CYP2C5 structure (1DT6)[12], in which the F-G-loop is not

resolved at all. The substrate-bound CYP2C5 structures show high overall structural similarity (superimposing all C alpha atoms yields an RMSD of 0.7 Å), but are structurally different in portions of substrate binding pocket (B', F and G helices). Inclusion of multiple templates is shown to improve the final quality of CYP homology models. In addition to the CYP2C5 structures, many more mammalian CYP crystal structures have been solved in the past few years. These structures became available during or after the performance of the modeling studies described and referred to in this chapter. It is not expected that the inclusion of these templates in the CYP2D6 modeling process would yield significantly better homology models as the latest crystallized CYPs do not have significantly higher sequence similarity (CYP2A6 (35%), CYP2B4 (40%), CYP2C8 (41%), CYP2C9 (40%), CYP3A4 (18%)). Secondly, for many of the latest mammalian CYP crystal structures regions which are part of or close to the active site are badly resolved or not resolved at all (e.g. F-G loop in CYP2C9 (1R9O) and CYP2D6 (2F9Q)). Thirdly, many of the new structures are believed to be in a catalytically non-active state, in which i) one CYP monomer is docked into the active site of the other monomer (1PO5 (CYP2B4); ii) the substrate is in the active site, but too far from the heme catalytic center (e.g. 1OG5 (2C9)); iii) the substrate is binding to a peripheral binding site (e.g. CYP2C8 (1PQ2), CYP3A4 (1W0F)). Finally, the CYP2A6 structure is not believed to be more suitable for modeling CYP2D6 than CYP2C5, as the size of typical CYP2A6 substrates is much smaller than the size of typical CYP2D6 substrates (and therefore probably also the sizes of CYP2A6 and CYP2D6 active sites will be significantly different).

```
CYP2C5  GKLPPGPTPFPIIGNILQIDAKDISKSLTKFSECYGPVFTVYLGMKPTVVLHGYEAVKEALVDLGEEFAGRGSVPILEKVSKG---LGIAFSN-AKTWKEM
CYP2D6  ---PPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGVFLARYGPAWREQ
                          A'        A        1-1      1-2       B          1-5      B'           C

CYP2C5  RRFSLMTLRNFGMGKRSIEDRIQEEARCLVEELRKTNASPCDPTFILGCAPCNVICSVIFHNRFDYKDEEFLKLMESLHENVELLGTPWLQ-VYNNFPALL
CYP2D6  RRFSVSTLRNLGLGKKSLEQWTEEAACLCAAFANHSGRPFRPNGLLDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEES-GFLREVLNAVPVLL
                              D                 E                        F

CYP2C5  DYFPGIHKTLLKNADYIKNFIMEKVKEHQKLLDVNN-PRDFIDCFLIKMEQEN---NLEFTLESLVIAVSDLFGAGTETTSTTLRYSLLLLLKHPEVAARV
CYP2D6  H-IPALAGKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRV
                  G                      H                            I

CYP2C5  QEEIERVIGRHRSPCMQDRSRMPYTDAVIHEIQRFIDLLPTNLPHAVTRDVRFRNYFIPKGTDIITSLTSVLHDEKAFPNPKVFDPGHFLDESGNFKKSDY
CYP2D6  QQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFVKPEA
            J      J'        K          1-4   2-1  2-2  1-3     K'              L'

CYP2C5  FMPFSAGKRMCVGEGLARMELFLFLTSILQNFKLQSLVEPKDLDITAVVNGFVSVPPSYQLCFIPIH
CYP2D6  FLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQP-RPSHHGVFAFLVSPSPYELCAVPR-
                    L        3-3     4-1  4-2    3-2
```

SRS 1, SRS 2, SRS 3, SRS 4, SRS 5, SRS 6

Figure 2: *Amino acid sequence alignment between CYP2D6 and the CYP2C5 X-ray crystal structure template (sharing 43% sequence identity and 68% sequence homology). Secondary structure elements are assigned to the final CYP2D6 model using the DSSP program[67] (CYP2D6 and CYP2C5 share fairly the same fold and secondary structure organization). Six Structure Recognition Sites (SRSs) as defined by Gotoh[31] are enclosed by boxes. Residues interacting with ligands, based on automated docking studies (see **Table 2**) are marked black. Triangles are assigned to residues which are mutated in our laboratory (filled) or by others (empty).*

In order to improve crystallization of CYP2C5 structures, the first 26 amino acid residues are truncated from the polypeptide chain. In addition, five amino acid positions in the F-G loop of CYP2C5 have been changed for the corresponding residues in CYP2C3. Although the N-terminus is believed to be anchored into the membrane, and parts of the F-G loop are believed to be embedded in the membrane as well, the crystal structures were shown to remain catalytically active.[11-13] The first 33 amino acids of CYP2D6 are omitted from the

model as this part of the structure was also not present in the template crystal structures. The first amino acid sequence alignment was generated using the T-Coffee programme [40]. This alignment was manually adjusted to avoid gaps and insertions in secondary structure elements, to concentrate gaps and insertions in short segments and to obtain a proper alignment of turns. The amino acid sequence alignment exhibits only minor gaps in the F/G loop and in the coil between the 3-3 and 4-1 β strands, and only minor insertions in the B´/C loop, F/G loop, and H/I loop. In the final alignment, CYP2C5 and CYP2D6 share 43% sequence identity and 68% sequence homology.

Table 1: *Sequence homology between CYP2D6 and CYP2C5 and comparison of their protein structures in terms of mainchain stereochemistry and side-chain environment.*

|	% seq. sim. (*homol.*)[c]	RMSD (Å) (*% RMSD<2 Å*)[d]	mainchain angles[e]	Rama-chandran[f]	Errat[g]	Verify3D[h]
CYP2D6 model	43 (*68*)	-	77	96 (*78*)	92	194
CYP2D6 (2F9Q)	43 (*68*)	2.9/2.3 (*88/90*)	85	98 (*84*)	95	198
CYP2C5 (1N6B)[a]	-	1.4 (*97*)	89	98 (*86*)	96	193
CYP2C5 (1NR6)[a]	-	1.4 (*97*)	88	99 (*87*)	98	195
CYP2C5 (1DT6)[b]	-	2.2 (*91*)	84	95 (*71*)	94	182

a) Crystal structure templates used for the construction of the CYP2D6 model; **b**) Crystal structure not used for the construction of the CYP2D6 model; **c**) Sequence similarity and sequence homology (between brackets) following the alignment between CYP2D6 and CYP2C5 in **Figure 2** and homology assignment according to ref 68; **d**) C_α root-mean-square deviation (RMSD) from the CYP2D6 homology model; between brackets the percentage of C_α atoms within 2 Å distance from equivalent C_α atoms in the CYP2D6 homology model (according to the alignment in **Figure 2**). For the CYP2D6 (2F9Q) entry the second value refers to the same measure excluding residues 229-239 of the F-G loop.; **e**) Percentage of residues with mainchain angle conformation 'within limits' as defined in refs 44,69; **f**) Percentage of residues with φ, ψ conformation in the 'allowed' and 'most favoured' (between brackets) regions of the Ramachandran plot[44,70]; **g**) Percentage of residues with patterns of non-bonded atomic interactions that satisfy the Errat definitions[45]; **h**) The total Verify3D[46] score summed over all residues, measuring the compatibility of a protein model with its sequence. For a protein the same size as CYP2C5 and CYP2D6 a score of approximately 200 is expected and a score less than 100 would indicate an incorrect structure.

Protein conformation and quality checks

From the hundred models generated by Modeller[39], three models were selected with the best loop conformations as determined by visual inspection, stereochemical parameters using PROCHECK[44], and side-chain environment using Errat[45] and Verify3D.[46] In **Table 1**, the protein quality check scores of the final model are compared with the substrate-free CYP2D6 crystal structure (2F9Q[19]), the two substrate bound CYP2C5 crystal structures used as homology modeling templates (1N6B[11] and 1NR6[13]), and the substrate-free CYP2C5 crystal structure (1DT6[12]), which was used for the modeling of earlier CYP2D6 models.[38,71] The final CYP2D6 model has approximately the same protein quality scores as the substrate-free CYP2D6 and substrate-bound CYP2C5 crystal structures, only the mainchain angles deviate more from the parameters defined by Engh and Huber.[69] It should be stated however, that these deviations can be the result of slight discrepancies between the Engh and Huber bond-angle parameter set and the parameter set of the GROMOS forcefield[54], used to optimize the final CYP2D6 model. Furthermore, the homology model has significantly better Ramachandran and Verify 3D scores than the substrate free CYP2C5 crystal structure, in which the F/G loop was not observed. The CYP2D6 homology model shows higher structural similarity to the substrate-bound CYP2C5 crystal structures than to the substrate-free CYP2C5 and CYP2D6 crystal structures, as expressed by lower C alpha atom RMSD values and higher percentages of structurally aligned C-alpha atoms within 2 Å distance from each other.

Geometry optimization

Dextromethorphan was docked in the selected CYP2D6 model and the geometry of this protein-substrate complex was then optimized by subsequent energy minimization and a short MD simulation, using the GROMACS molecular simulation package[53] and the GROMOS force field, parameter set 43A1.[54] Heme parameters were taken from this parameter set, with an additional covalent bond defined between the heme iron atom and the sulfur atom of C443 with an ideal bond length of 0.240 nm. The enzyme, including substrate and heme, was energy minimized in vacuum using the steepest descent method, first, with harmonic position restraints using a force constant of 1000 kJ mol^{-1} nm^{-2} on all non-hydrogen atoms and, subsequently, without position restraints. Finally, the complex was relaxed during a 1 ps MD-simulation *in vacuo*. The MD simulation was carried out with a time step of 2 fs, a twinrange cutoff of 0.8/1.4 nm, and a relative dielectric constant of 1.0. The LINCS algorithm[72] was used to constrain the lengths of all covalent bonds. The temperature was maintained at 300 K by weak coupling to an external bath, with a relaxation time of 0.1 ps.[73] **Figure 3** shows a ribbon model of the final homology model of CYP2D6 in which the secondary structure elements relevant for ligand binding and the amino acid residues investigated by site-directed mutagenesis are indicated.
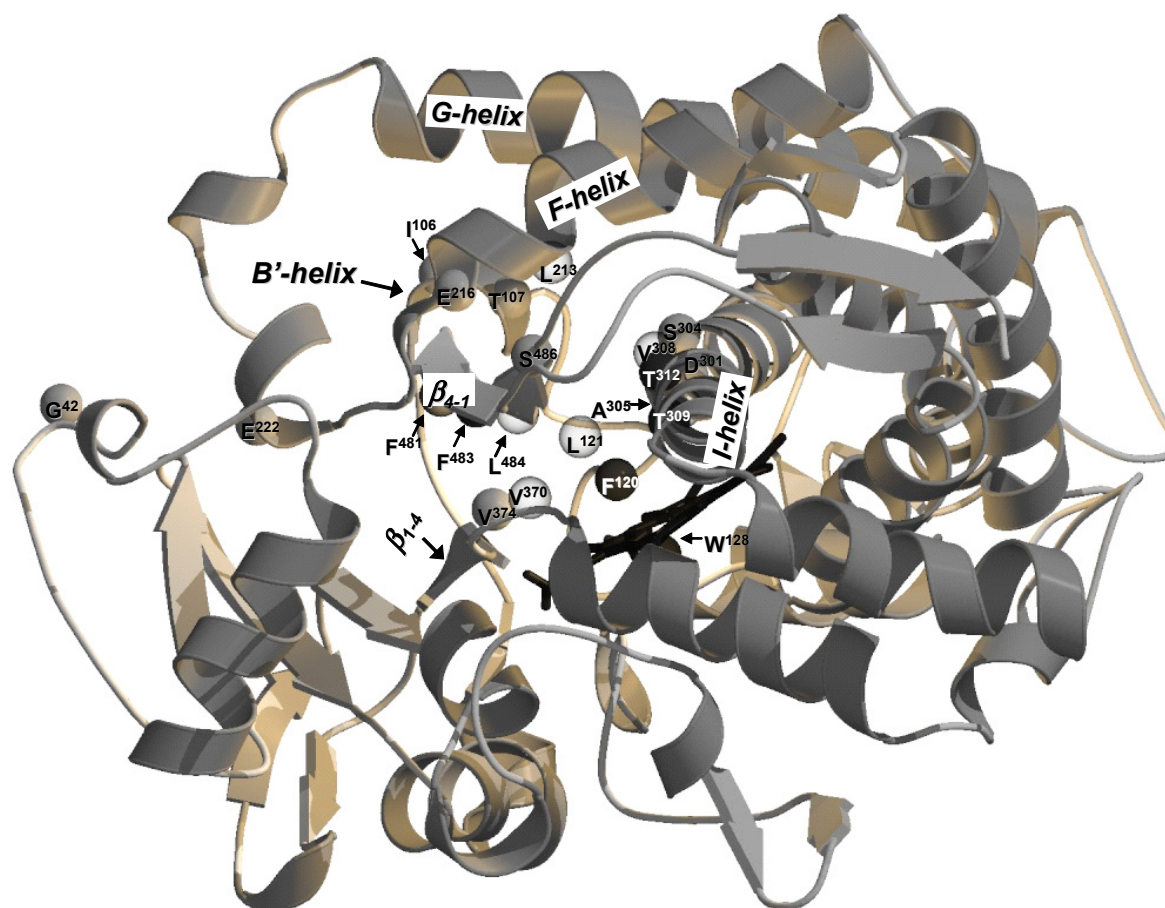
Figure 3: *Three-dimensional structure of CYP2D6 homology model. C-alpha atoms of amino acid residues subjected to site-directed mutagenesis studies are depicted as black (mutants studied in our group) and grey (mutants studied by others) spheres (see also **Figure 2** and **Table 3**); positions of amino acid residues interacting with docked ligands not subjected to mutation studies (see also **Figure 2** and **Table 2**) are indicated by transparent white spheres. The heme catalytic center is depicted in black sticks. Secondary structure elements comprising Substrate Recognition Sites[31] are indicated: SRS 1 (B´-helix), SRS 2 (F-helix), SRS 3 (G-helix), SRS 4 (I-helix), SRS 5 (β_{1-4}), and SRS 6 (β_{4-1}).*

This model of wild-type CYP2D6 was used as a template to model the I106E, F120A, T309V, T312V, and F483A mutants. In order to construct structures for the mutants, the relevant residues were replaced using the homology module of InsightII (Biosym), after which an energy minimization and a 1 ps position restrained MD simulation was carried out with dextromethorphan in the active site as described above. Simulations of wild-type and mutant CYP2D6 with substrates bound in the active site, starting from different automatically docked poses and pre-equilibrated using the scheme described in ref 74, remained stable during 10 ns of unrestrained MD at 300 K, in explicit water. In these simulations the pressure was maintained at 1 bar using the weak coupling scheme with relaxation time 1.0 ps.[73] Final backbone atom-positional root-mean-square deviation (RMSD) values of 1.3-3.8 Å and minor fluctuations in secundary structure organization were obtained.[37,74]

Substrate Pharmacophore Fitting

Pharmacophore models derive information on the active site from similarities in shape, electronic properties, and conformations of substrates, inhibitors, or metabolic products. The so-called "classical" pharmacophore models are constructed by superimposing these molecules onto each other by matching chemically similar groups. Current approaches, however, utilize the three-dimensional quantitative structure-activity relationship (3D-QSAR) approaches in which abstracted descriptors of essential functional groups and topologies of the molecules are superimposed, possibly supplemented by details of the surrounding active site. (This distinction is not fundamental; a molecule consisting of atoms and bonds in a "classical" pharmacophore model can be viewed as a type of "abstracted descriptor", although rather detailed, in a 3D-QSAR model.) Recent reviews on CYP pharmacophore and 3D-QSAR methods are provided by Ekins and de Groot.[75,76]

The main features of the currently known CYP2D6 pharmacophore models are highlighted in (**Figure 4** of **Chapter 2**) where molecular interaction fields (MIFs[52]) calculated for *S*-bufuralol are overlaid on the substrate docked into the CYP2D6 active site. A clear correlation with the specific residues listed in **Table 2**, responsible for substrate interaction, can be obtained. CYP2D6 substrates possess a positive electrostatic site (i.e. a positively charged nitrogen atom) 5, 7 or 10 Å away from the site of catatalysis, two hydrophobic regions (i.e. aromatic rings) at ca. 6-8 Å from the positive electrostatic site, and possibly also a H-bond acceptor ca. 6 Å from the positive electrostatic site.[55-63] Carboxylate group(s) in the protein (i.e. aspartate (D), glutamate (E)) are suggested to interact with the positively charged site in the substrate, while aromatic amino acid residues (phenylalanine (F), tyrosine (Y), tryptophane (W)) or other hydrophobic residues (alanine (A), valine (V), isoleucine (I), leucine (L)) are thought to interact with the hydrophobic regions in the substrate. The MIF due to a carboxylic probe clearly corresponds to the position of E216, while hydrophobic probe MIFs correspond to the position of F120 and a hydrophobic region composed by L213, V308 and F483. D301 was postulated to be involved in polar/ionic substrate interactions by earlier models, based on bacterial crystal structure templates[56,77,78], but this interaction was not observed in our models (**Table 2**).

Interaction Fingerprint Analysis using Automated Docking

Automated molecular docking is a computational method to predict energetically favorable conformations and orientations of ligands in the interior of protein structures. By analysing the binding modes of protein-ligand complexes derived from automated docking simulations, the role of specific amino acid residues in ligand-binding can be predicted. Automated docking strategies successfully applied to binding mode predictions of ligands in CYP crystal structures[79], were used for the catalytic site prediction (CSP) of 65

substrates in the CYP2D6 homology model.[80] For each ligand, the highest ranked docking solutions according to SCORE[51] corresponding to experimentally determined primary metabolites, selected from 200 pooled poses generated by AutoDock[47], FlexX[48], GOLD-Goldscore[49], and GOLD-Chemscore[50] docking programs was used for an analysis of CYP2D6-ligand interactions. The results of this interaction fingerprint analysis are shown in **Table 2** and **Figure 4**.

***Table 2**: Occurrence of ligand-protein interactions based on the analysis of 65 automatically docked protein-substrate complexes. Interactions are divided in hydrophobic interactions (aliphatic or aromatic carbon-carbon interactions) and H-bond interactions (H-bond donor (amine/amide nitrogen, hydroxyl oxygen) or H-bond acceptor (amine/carbonyl/carboxylate/hydroxyl oxygen)). Amino acid residues mediating hydrophobic interactions with less than 10% of the ligands are not listed. A hydrophobic interaction is observed if aliphatic or aromatic carbons are within a distance of 4 Å. A H-bond interaction is defined through contacts between H-bond donors and H-bond acceptors within 3.5 Å.*

residue	hydrophobic (%)	H-bond donor(ligand)-acceptor (protein) (%)	H-bond acceptor(ligand)-donor (protein) (%)
V104	45	0	0
T107	0	8	6
F120	91 (61[a])	0	0
L121	37	0	0
L213	50	0	0
E216	0	62 (50[b])	0
S304	37	37	18
A305	83	3	0
V308	62	0	0
V370	32	0	0
V374	90	0	0
F483	70 (42[a])	0	0
L484	52	0	0

a) Aromatic interactions (%). It should be noted that 97% of the substrates has an aromatic moiety
b) Ionic interactions (%). It should be noted that 97% of the substrates has a positively charged (nitrogen) atom.
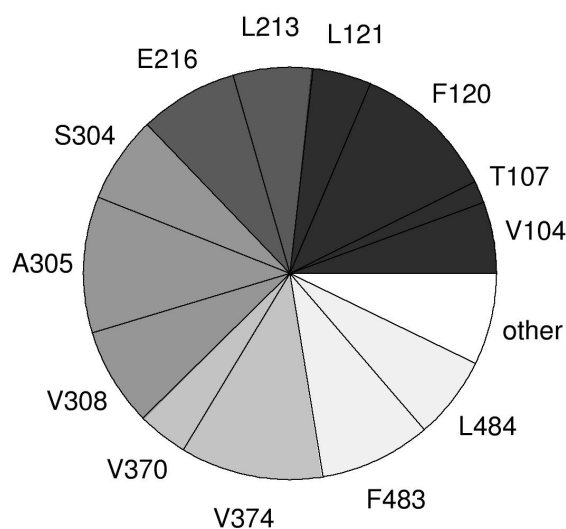
Figure 4: *Relative contributions of amino acid residues to ligand-protein interactions (hydrophobic interactions + H-bond interactions). Residues are colored by Substrate Recognition Sites[31]: SRS1 (V104, T107, F120, L121), SRS2 (L213, E216), SRS4 (S304, A305, V308), SRS5 (V370, V374), and SRS6 (F483, L484).*

The analysis was used to compare the active sites of our CYP2D6 homology model and the recently reported substrate-free CYP2D6 crystal structure, to explain reported site directed mutagenesis data, and to design new site-directed mutagenesis experiments performed in our laboratory, as will be described in the next paragraphs. From the data in **Table 2** it becomes clear that in this model residues F120, L213, A305, V308, V374, F483 and L484 are responsible for most of the hydrophobic interactions, while the H-bond/ionic interactions are most prominently seen through E216. Interacting residues from all substrate recognition sites except for SRS3 could be identified. As can be seen from **Figure 3**, this SRS is positioned at a relatively large distance from the active site, making substrate interactions in the bound state unlikely. However, this SRS may be involved in recognition during substrate entrance (see **Chapter 8**). Interestingly, no mutants were reported involving residues in SRS3 (**Figure 2**), so that no direct evidence for this SRS is available.

Comparison between substrate-bound CYP2D6 homology model and substrate-free crystal structure

Very recently a X-ray crystal structure of substrate-free CYP2D6 was resolved at 3.0 Å resolution (pdb code 2F9Q).[19] Like other crystallized mammalian CYPs, CYP2D6 was solubilized by truncation (residues 1-22) and mutation (residues 23-33) of the N-terminus and small mutations in the G-helix (L230D/L231R). Residues 23-33 of the N-terminus and 42-51 of the A' helix were not resolved in the crystal structure and residues 229-239 of the F-G loop were built only as alanines, as the electron density maps were insufficiently clear to unambiguously assign the correct residue side chains.

The CYP2D6 homology model corresponds closely to the substrate-free CYP2D6 crystal structure in many regions, with 88% of the C alpha carbons within a root-mean-square deviation (RMSD) of 2 Å (**Figure 5**). The heme is anchored in the binding site in the same way (hydrogen bonding interactions with the side chains of R101, W128, R132, H376, S437 and R441, and the heme is pentacoordinated with C443), there is a good overlap with the proposed CYP reductase binding domain around residue R440, and the positions and orientations of most active site residues are similar (**Figure 6**). Both the homology model and the crystal structure show a so-called solvent channel, which passes between the E-helix, SRS2 in the F-helix, SRS4 in C-terminal end of the I-helix and the SRS6 turn region of β sheet 4.

However, significant changes are seen in the N terminus (residues preceding residue 54); between strands 1-1 and 1-2 (residues 73-76); in the helix B to helix C region (SRS 1, residues 104-116); in the C-D loop (residues 141-147); in the region encompassing helix F through the beginning of helix G (SRS 3, residues 217-240); in the G-H loop (residues 264-266); in the H-I loop (residues 285-287); in strand 4-1 (SRS 6, residues 470-482). These differences are:

i)   far away from the active site (N terminus, strands 1-1 and 1-2, C-D, G-H and H-I loops).

ii) not resolved (residues 23-33 and 42-51 in the N-terminus) or not well resolved (residues 229-239 of the F-G loop) in the crystal structure.

iii) Mutated in the crystal structure construct to improve solubility (L230D/L231R).

iv) the same structural differences observed upon substrate binding to other CYPs: CYP102 (F-G helical region), CYP119 (F-G helical region), CYP2C5 (B' helix, F-G helical region, H/I loop, and strand 4-1) and CYP2C9 (strand 4-1).

v)   the same structural differences observed when comparing open and closed forms of CYP2B4 and when comparing CYP2C9 crystal structures determined by different laboratories (N-terminus, B' helix, F-G helical region).

This suggests that the most significant structural differences between the substrate-free CYP2D6 crystal structure and our CYP2D6 homology are caused by substrate-induced

conformational changes. In the homology model, which is based on substrate-bound CYP2C5 structures and is optimized as a substrate-bound complex, the B' helix (SRS1) has moved towards the heme in the homology model to maximize hydrophobic contacts with the substrate. This movement increases the contact with the G-helix (SRS3) and facilitates hydrogen bonding of Q244 to the backbone of L110. This hydrogen bond is not observed in the crystal structure. The same movement of the B' helix and facilitation of hydrogen bonding between the B' helix and G helix are for example observed when comparing substrate-bound and substrate-free structures of CYP2C5 (formation of a hydrogen bond between the homologues K241 and V106 residues). In the CYP2D6 homology model, the F helix has two turns less than in the crystal structure and has moved across the surface of the I helix perpendicular to the axis of the I helix. This movement closes the solvent channel between helices F and I close to V308 and also results in reorientation of the side chain of residue F483, located in the loop between strands 4-1 and 4-2 (SRS6). In the homology model, the phenyl-ring of F483 has turned towards the active site, maximizing hydrophobic contacts with the substrate, while in the crystal structure, F483 forms an intermediate substrate binding pocket with E216 at the entrance of the solvent channel. Substrate binding to CYP102 and CYP2C5 also shows a rearrangement of the region from helix F through the N-terminal end of helix I. This is reflected by a lateral movement of the F and G helices across the surface of the I helix, closing a cleft above the substrate binding site. Substrate binding to CYP2C5 and CYP2C9 results in the same reorientation of the homologues substrate binding phenylalanine residue in SRS6 (F473 (CYP2C5), F476 (CYP2C9)).

Figure 5 shows the results of a comparison between the substrate free crystal structure (pdb-code 2F9Q) and substrate bound homology model of CYP2D6 and of a comparison between substrate free (1DT6) and substrate bound (1NR6) CYP2C5 crystal structures. The same structural differences are observed between substrate-bound and substrate-free CYP2C5 as the differences between the CYP2D6 homology model and substrate-free crystal structure.

Figure 6 shows an overlay of the active sites of the CYP2D6 crystal structure and the homology model active site with dextromethorphan docked into the respective active sites according to the method described in refs 79,80. Almost all residues predicted to be involved in substrate binding in our homology model (see **Table 2**) are also located in the active site of the CYP2D6 crystal structure, with the exception of V104 and T107. L220 and Q244 are two residues located in the active site of the CYP2D6 crystal structure but are not part of the substrate binding pocket of the homology model. The location and orientation of most of the residues subjected to the mutagenesis studies carried out in our laboratory is about the same (e.g. F120, T309, T312). The location and orientation of I106 is significantly different between the crystal structure and the homology model, but in both cases, I106 is within 5 Å distance from E216. The backbone of F483 is located at the same position, but the phenyl ring is differently directed towards the active site. In the homology model, F483 has turned towards the active site while in the crystal structure, F483 forms an intermediate substrate binding pocket with E216 at the entrance of the solvent channel. The E216 residue is slightly shifted but orientated in the same way towards the active site. Both in the crystal structure and in the homology model D301 forms hydrogen bonds with the backbone NH groups of V119 and F120, and its carboxylate side-chain is not directed into the active site, suggesting that this residue primarily plays an important role in positioning the B-C loop, rather than being directly involved in substrate binding. S304 has approximately the same conformation in the homology model and the crystal structure, with its hydroxyl group pointing towards the carbonyl backbone of A305 as well as into the active site. In the crystal structure the residue at position 374 is a methionine, which is a valine (the actual residue in wild-type CYP2D6) in the homology model [81,82]. The bigger methionine residue slightly displaces

F120 into a position more above the heme in the crystal structure. In the crystal structure, aromatic moieties of substrates can stack between F120 and L213, L220, A305 and V308, while in the homology model a hydrophobic pocket is formed by F120, L213, A305, V308 and F483.
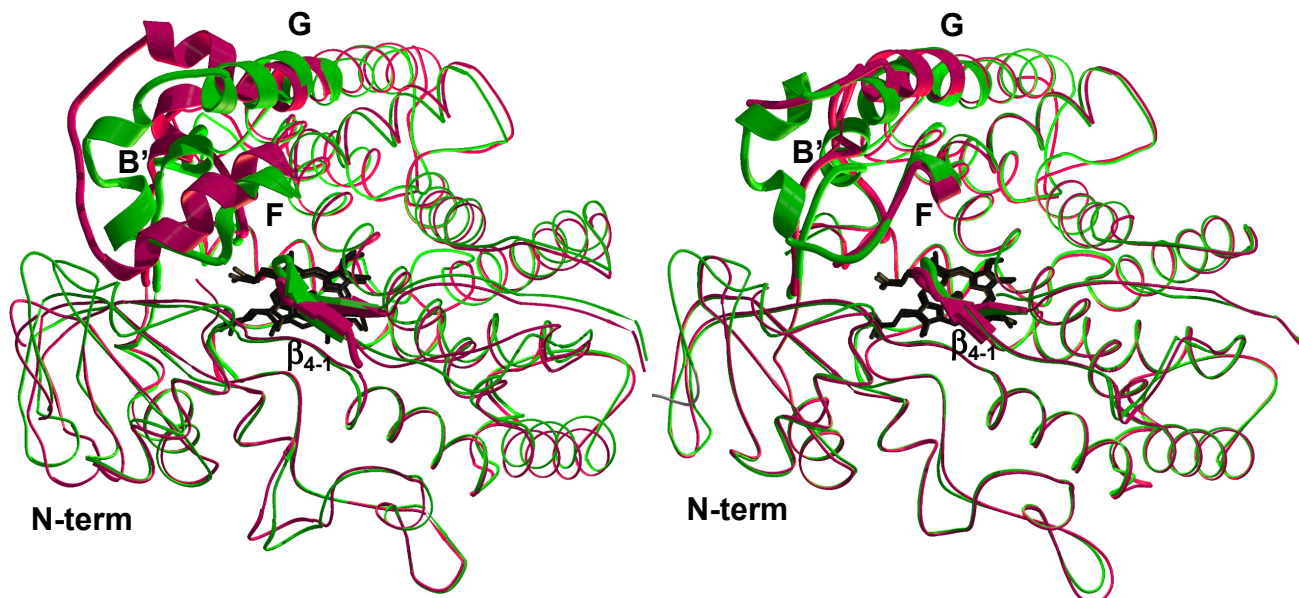
Figure 5: *Comparison of the substrate free crystal structure (pdb-code 2F9Q, magenta) and substrate bound homology model of CYP2D6 (green) in panel **A**, and a comparison between substrate free (1DT6, magenta) and substrate bound (1NR6, green) CYP2C5 crystal structures in panel **B**. The heme is depicted as sticks in black. Regions showing large differences and involved in substrate binding (B' helix, and F-G helical region) are shown as cartoons. Part of the N-terminus in the substrate-free CYP2D6 crystal structure and of the F-G loop in the substrate-free CYP2C5 are not solved.*
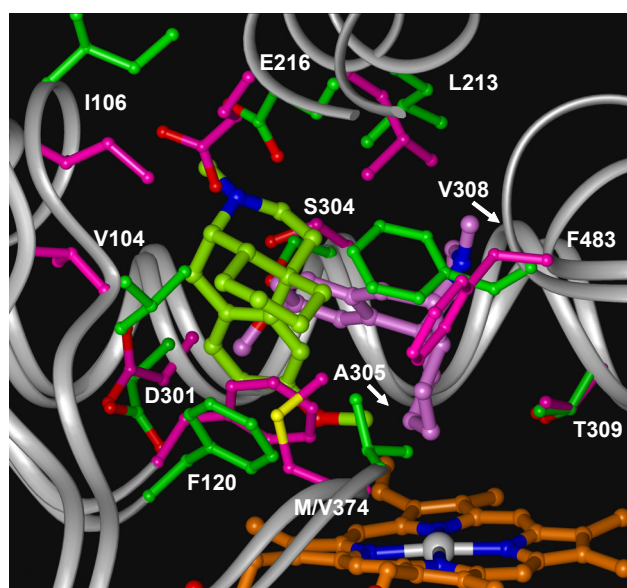
Figure 6: *Overlay of the active sites of the active sites of the CYP2D6 crystal structure (magenta) and the CYP2D6 homology model (green). Most of the residues involved in substrate binding in the crystal-structure and/or homology model are shown as sticks. Dextromethorphan is docked into the active sites considering active-site water molecules and using GOLD-Chemscore in combination with the SCORE scoring function, as described in ref 80. In the homology model, the orientation of dextromethorphan (in light green) with its O-methyl pointing towards the heme is favoured, in the crystal an orientation supporting 7-hydroxylation (in magenta) is the energetically most favourable docking orientation.*

3.3 Site-directed mutagenesis studies in CYP2D6

For CYP2D6, homology models based on crystal structures of CYPs 101, 102, 107A, 108 and 2C5 have been constructed separately as well as in conjunction with each other.[38,56,57,71,77,78,83-86] Ligand-based CYP2D6 pharmacophore models suggest that ligands bind two or three aromatic/hydrophobic sites and a carboxylic site for which F120 and F483, respectively E216 or D301 were suggested in combination with homology models. Extensive reviews on CYP(2D6) modeling have appeared very recently [20,21]. Active site residue predictions based upon substrate and protein models of CYP2D6 were often followed by mutagenesis studies to validate their roles. An overview of site-directed mutagenesis studies is presented in **Table 3**. The residues of interest are listed together with the mutations and the effects of the mutations. From a historical point of view, D301 is the most studied residue.[87-92] It was first indicated as the active site carboxylate involved in fixation of the basic nitrogen commonly observed in typical CYP2D6 substrates and it was the first residue subjected to site directed mutagenesis. Although mutation led to a decreased CYP2D6 activity, the carboxylate group of D301 is not directed into the active site in our CYP2D6 homology model or in the recently reported substrate-free CYP2D6 crystal structure.[19] Another active site carboxylate, E216 was rather suggested to be the main actor responsible for interaction with the basic nitrogen atom in the substrate .[91,92] In the CYP2D6 crystal structure and in our CYP2D6 homology model the side-chain of this residue is indeed directed into the active site and is predicted to be an important residue involved in substrate binding in the homology model (**Table 2**). Our homology model and the CYP2D6 crystal structure indicate that the negative effect on the catalytic activity of mutating D301 is indirect. D301 is suggested to position residue F120, which interacts with the aromatic moiety of the substrate. A mutation in position 301 (with the exception of D301E) disturbs the position of F120 and affects substrate binding through the aromatic interaction. Mutation of active site residue S304 was found to have no effect on substrate turnover. In the homology model as well as in the crystal structure the hydroxyl group of this residue is in a conformation that enables formation of a H-bond with substrates bound in the active site (**Table 2**). However, the conformation of S304 also favors H-bond formation with the backbone carbonyls of A300 in the crystal structure and with the carbonyl backbone of D301 in the homology model, leaving the hydrogen bond donor site of S304 unavailable for substrate hydrogen acceptor groups. Mutation of M374 into V374 (the latter being the actual residue in wild-type CYP2D6 [81,82]) was shown to have an effect on the stereoselectivity of metropolol metabolism. V374 is suggested to be an important substrate-binding residue in the homology model (**Table 2**), and M374 is part of the active site of the crystal structure.

Table 4 and **Figures 2** and **3** show that site-directed mutagenesis studies have primarily focused on residues located in or near the active site (e.g., SRSs 1, 2, 4, 5 and 6) and near the N-terminus. So far, no studies have been described on residues located in SRS3. As shown in **Table 2** and **Figure 3**, our homology model suggest that this substrate recognition site does not directly interact with ligands when positioned in the vicinity of the heme catalytic center. This part of CYP2D6 may however, be involved in substrate access and/or product exit. In the crystal structure Q244 is directed towards the active site, while this residue forms a H-bond with the backbone of L110 in the homology model. **Table 2** and **Figure 4** suggest that several other active site residues are probably interacting with ligands as well (e.g., V104 (SRS1), L213 (SRS2), and V308 (SRS4)) opening the way for further mutagenesis studies.

Table 3: *Structural features and experimentally determined effects of site directed mutations in CYP2D6.*

Residue	Mutated to[a]	secondary structure element[b]	SRS[c]	Residues in close contact homology model (X-ray structure)[d]	Substrates tested	k_{cat}[e]	R[f]	S[g]	ref
P34	S (SNP)[h]	N-terminus		ND[i]	bunitrolol	~			101
					debrisoquine	~			
G42	R K E *VFS*	N-terminus		ND[i]	bunitrolol	→		S	101
					debrisoquine	→			
I106	**E**	B′-helix	SRS1	E216[j]	bufuralol	→			
					dextromethor.	→			
					MAMC	→			
					MDMA	→			
T107	I (SNP)	B′-helix	SRS1	E216	bufuralol	~			6
					codeine	~			
F120	**A**	B′-C loop	SRS1	R68, heme	bufuralol	~	R	S	32,101,10
					dextromethor.	~	R		2
					MAMC	→			
					MDMA	→	R		
W128	**F**	C-helix		Y124, R132, R441, heme[HB]	MAMC	~			37
E216	*D Q* A H F K	F-helix	SRS2	I106[j], T107, F483	bufuralol	→			91,92,103
					dextrometh.	→			
					diclofenac	←			
					nifedipine	←			
					spirosulfonam	~			
					testosterone	~			
					quinidine	←	R		
					tolbutamide	←			
E222	A	F-G loop		R221[HB], T394[HB]	bufuralol	→		S	101
R296	C (SNP)[h]			D292	debrisoquine	~			6
					bunitrolol	~			

Residue	Mutated to[a]	secondary structure element[b]	SRS[c]	Residues in close contact homology model (X-ray structure)[d]	Substrates tested	k_{cat}[e]	R[f]	S[g]	ref
D301	*E*N G *Q*A δ[k]	I-helix	SRS4	L111, V119$_{HB}$, F120$_{HB}$	bufuralol	→			87-92
					dextromethor.	→			
					diclofenac	←			
					nifedipine	←			
					spirosulfonam	→			
					testosterone	~			
S304	A	I-helix	SRS4	D301$_{HB}$	debrisoquine	~			104
					metoprolol	~			
					propranolol	~			
T309	**V**	I-helix	SRS4	V308, A305$_{HB}$, G306$_{HB}$, T310$_{HB}$, heme	bufuralol	→	R		35
					dextromethor.	~			
					MAMC	→			
					MDMA	→			
T312	**V**	I-helix	SRS4[i]	V308	bufuralol	~			35
					dextromethor.	~			
					MAMC	~			
					MDMA	~			
M374	V	β$_{1-14}$ sheet	SSR5	R101$_{HB}$, F120, F483	metoprolol	~		S	105
R440	H			E99$_{HB}$		~		S	105
F481	*Y*N G	β$_{4-1}$ sheet	SSR6[i]	H477, S486	debrisoquine	→			87
					dextromethor.	→			
					metoprolol	→		S	
F483	**A** I W	β$_{4-2}$-β$_{4-1}$ loop	SSR6	E216, L220, V374	bufuralol	→			33,106
					dextromethor.	→	R		
					MAMC	→			
					MDMA	→			
					testerone	→			

Residue	Mutated to [a]	secondary structure element [b]	SRS [c]	Residues in close contact homology model (X-ray structure) [d]	Substrates tested	k_{cat} [e]	R [f]	S [g]	ref
S486	T (SNP) [h]	β_{4-2} sheet		F484	bunitrolol	~			6
					debrisoquine	~			

a) Mutations depicted in *Italics* did not have a significant effect on metabolism, mutations constructed in our laboratory are underlined and **bold**; **b**) Secondary structure element in which the residue is located in our CYP2D6 homology model; **c**) Substrate recognition sites (SRSs) as defined by Gotoh [31]; **d**) Contact between amino acid residue side chains within 4 Å. Hydrogen bonding interactions (within 3.5 Å) are indicated with a "HB" subscript; **e**) Increase ($\uparrow$), decrease ($\downarrow$), or no significant change ($\sim$) in k_{cat} (catalytic efficiency (V_{max}/K_m)); **f**) R= Change in regiospecificity of catalysis compared to wild-type; **g**) S = Change in stereospecificity of catalysis compared to wild-type; S = No change in stereospecificity of catalysis; **h**) SNP = Single nucleotide polymorphism; **i**) N-terminus is not modeled as this part of the protein is not present in the CYP2C5 crystal structure templates; **j**) Possibly ionic interaction in I106E mutant (within 4 Å); **k**) δ = deletion of residue; **l**) Not inside active site

Table 4: *Experimentally determined effects of site-directed mutations on the catalytic efficiency of metabolism of four typical CYP2D6 substrates. Significant increase and decrease (with at least a factor of 5) in catalytic efficiency ($V_{max(mutant)}/K_{m(mutant)})/(V_{max(WT)}/K_{m(WT)}$) of metabolite formation compared to wild-type CYP2D6 are coloured black and grey, respectively. Data are derived from refs 32-35*

Substrate	Metabolite	WT	I106E	F120A	T309	T312	F483
MAMC	HAMC	1.0	-	-	0.1	0.8	-
dextromethorphan	Dextrorphan	1.0	-	2.3	0.3	1.1	0.1
	3-OMe-morphinan	+	-	nd	+	+	+(Q)
	3-OH-Morphinan	-	-	+(Q)	-	-	+
	OH-Dextrorphan	-	-	+	-	-	+
Bufuralol	1-OH	1.0	0.06	2.3	0.1	0.4	0.03
	4-OH	1.0	+	2.1	0.8	0.8	+
	6-OH	1.0	-	+	-	0.9	nd
	$\Delta^{1,2}$	1.0	-	+	-	0.7	-
MDMA	3,4-OH-MA	1.0	0.05	0.05	0.1	3.6	0.04
	MDA	+	-	+(Q)	+(Q)	+	+
	N-OH MDMA	-	-	+(Q)	-	-	-

nd = not determined
+ = present but not quantifiable
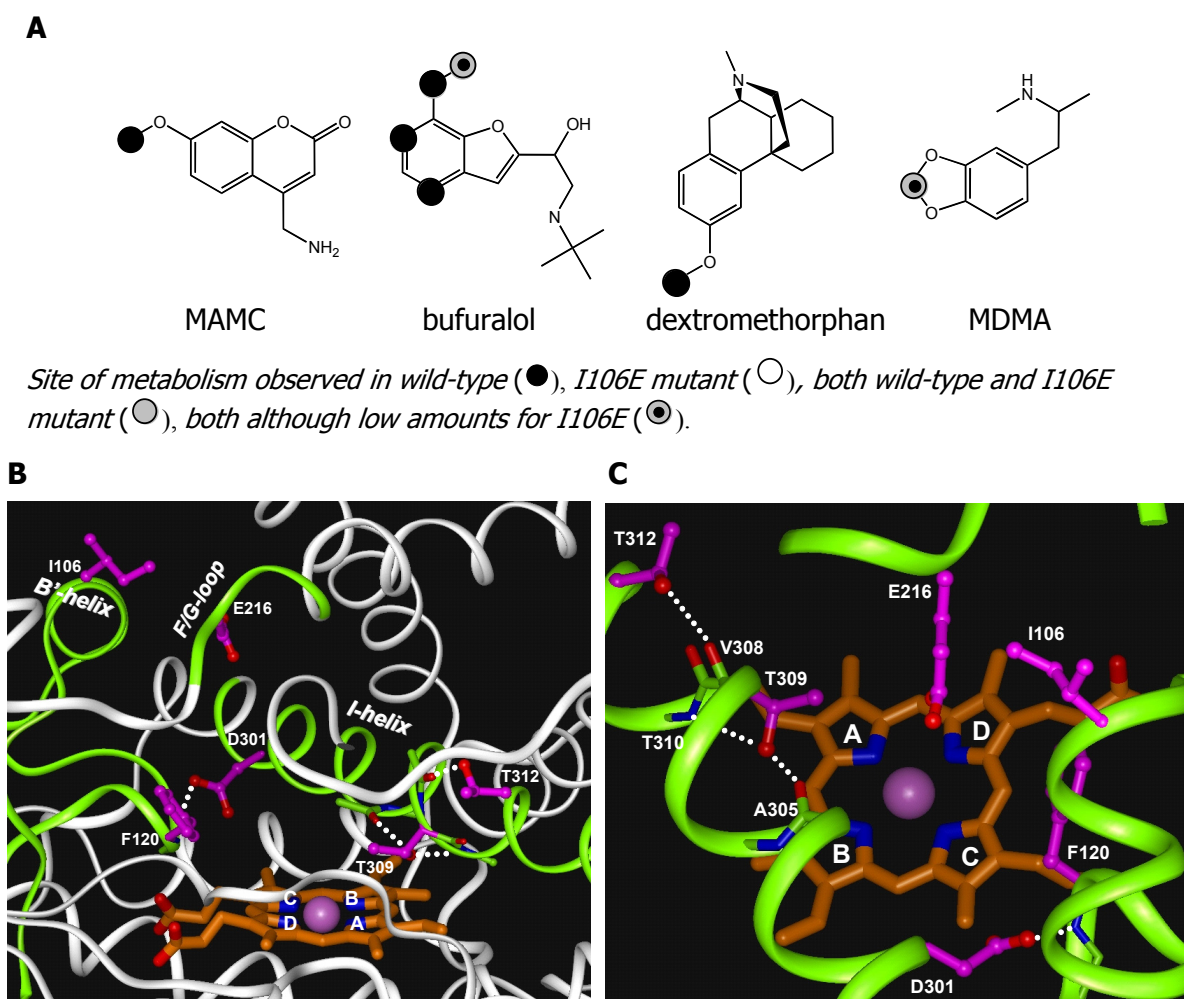+(Q) = present and quantified
– = not detectable

Recently, a series of site-directed mutagenesis studies were performed in our group, which were guided by the CYP2D6 homology model described above. We selected two phenylalanines, F120 and F483, suggested to be involved in substrate binding, as indicated by the interaction fingerprint analysis between the CYP2D6 protein structure and typical CYP2D6 substrates (**Table 2**). F120 is predicted to interact with the aromatic moiety present in most CYP2D6 substrates, while F483 is predicted to more sterically influence substrate binding. Introduction of a negatively charged residue on position 106 (isoleucine in wild-type CYP2D6) was predicted to influence the positioning of E216, alterating the shape and character of the active site cavity, and therefore affecting substrate binding to CYP2D6. Two active-site threonines, T309 and T312, were subjected to site-directed mutagenesis experiments although they were not predicted to be directly involved in substrate binding. Multiple amino acid sequence alignments of CYPs (drnelson.utmem.edu/CytochromesP450.html) indicate T309 to be a highly conserved threonine among fairly all CYPs and in a series of CYPs this residue was experimentally shown to be involved in oxygen activation.[93-99]

Table 4 summarizes the influence of these active site mutations on the catalytic efficiency of CYP2D6 to metabolize four typical CYP2D6 substrates, 7-methoxy-4-(aminomethyl)-coumarin (MAMC), dextromethorphan, bufuralol and 3,4-methylenedioxymethamphetamine (MDMA). In the following paragraphs, these mutants will be described in more detail. In addition to enzyme kinetics studies, we have used aryldiazenes to gain information on CYP2D6 active site topology. These probes react with the heme iron to form σ-bonded aryl–iron complexes. Upon oxidation of the complex with ferricyanide, the aryl group migrates from the heme iron to one of the neighboring four porphyrin nitrogens.[100] Studies with CYPs for which crystal structures are available have shown that the regiochemistry of the migration is largely dependent on the steric constraints within the enzyme's active site and that quantification of the N-arylprotoporphyrin IX thus formed provides an estimate of the available space above each of the four heme pyrroles.[100] The effects of the mutants on the catalytic efficiency and active-site shape of CYP2D6 were again rationalized by comparing automated docking poses obtained in our wild-type homology model to poses obtained in protein models of the mutants.

I106E

The metabolism of MAMC and MDMA by the CYP2D6 I106E mutant was greatly reduced as compared to wild-type and although products were detected, they were in such low quantities that kinetic parameters for these reactions could not be determined [34]. The metabolism of dextromethorphan and bufuralol was also affected by the I106E mutation albeit to a lesser extent. Trace amounts of other metabolites of bufuralol by CYP2D6 (i.e. 6-OH bufuralol, 4-OH bufuralol and $\Delta^{1',2'}$ bufuralol[89]) were also observed with the I106E mutant but could not be accurately quantified.

Recent reports have highlighted the importance of E216 in substrate binding and orientation of amine substrates through an electrostatic interaction.[91,92] According to our homology model, E216 is located above pyrrole D of the heme moiety at a vertical distance of roughly 12 Å. In the wild-type, the sidechain of I106 lies within 5 Å of the carboxylate group of E216. Also in the recently solved substrate-free CYP2D6 crystal structure, E216 is located at 12 Å above pyrrole D and is I106 located within 5 Å of E216, although the position of I106 and the B' helix is significantly different (**Figure 7**).

A

MAMC bufuralol dextromethorphan MDMA

Site of metabolism observed in wild-type (●), I106E mutant (○), both wild-type and I106E mutant (◉), both although low amounts for I106E (◉).

B

C

Figure 7: *Panel A schematically depicts the catalytic profiles for four typical CYP2D6 substrates and changes in metabolism upon the I106E mutation. Panels B and C show three-dimensional representations of the CYP2D6 homology model structure. In the close up view of the active site (panel C) only the green coloured protein backbone ribbon segments presented in the side view (panel B) are shown. In panel B, the positions of the B'-helix, F-G loop, and I-helix are indicated. Side chains of residues mutated in this study (with the exception of E216) are shown in purple ball-and-stick form. The backbone atoms of residues forming H-bonds with the side-chains of D301 (F120), T309 (A305 and T310) and T312 (V308) (indicated as dotted white lines) are shown in green stick form. Letters A-D represent the corresponding pyrrole rings in the heme.*

F120A

Strikingly, the substitution of phenylalanine at position 120 into an alanine led to a completely abolished MAMC O-demethylating activity of CYP2D6.[32] On the other hand, bufuralol metabolism was hardly affected, nor was the dextromethorphan O-demethylation. The F120A mutant did form 3-hydroxymorphinan, the double demethylated form of dextromethorphan, which was not detected using wild-type CYP2D6. MDMA was demethylenated by both mutant and wild-type CYP2D6 to 3,4-dihydroxymethamphetamine. In addition the mutant formed two additional metabolites; 3,4-methylenedioxyamphetamine (MDA) and N-hydroxy-3,4-methylenedioxymeth-amphetamine (N-OH-MDMA). Inhibition experiments of dextromethorphan O-demethylation showed a decreased affinity of the F120A mutant for quinidine, while IC_{50}

values for quinine were equal. These data indicate the importance of F120 in the selectivity and regiospecificity of substrate binding and catalysis by CYP2D6.
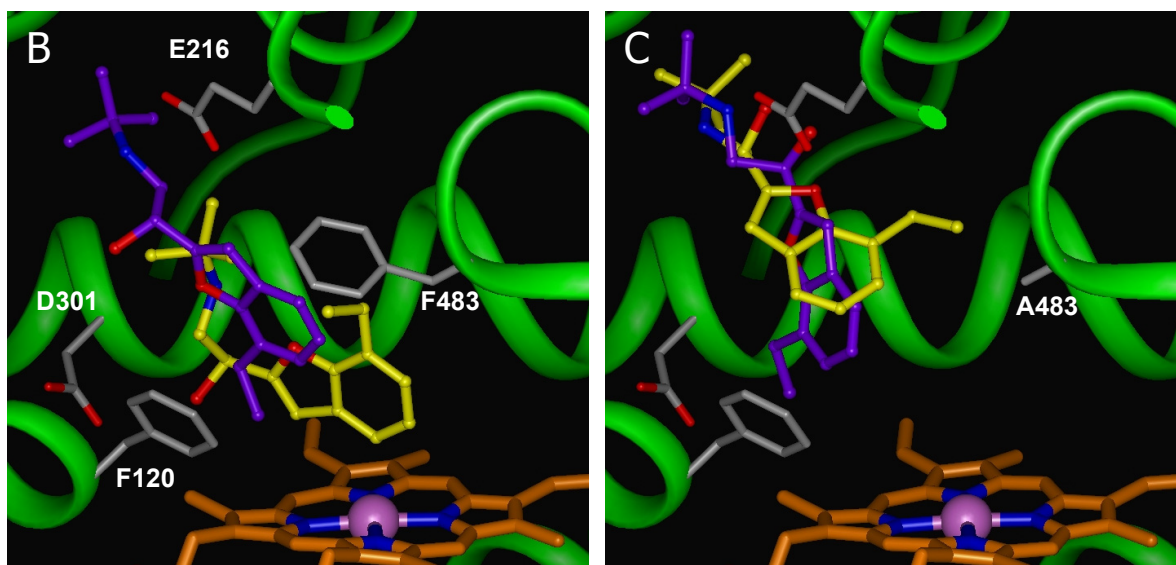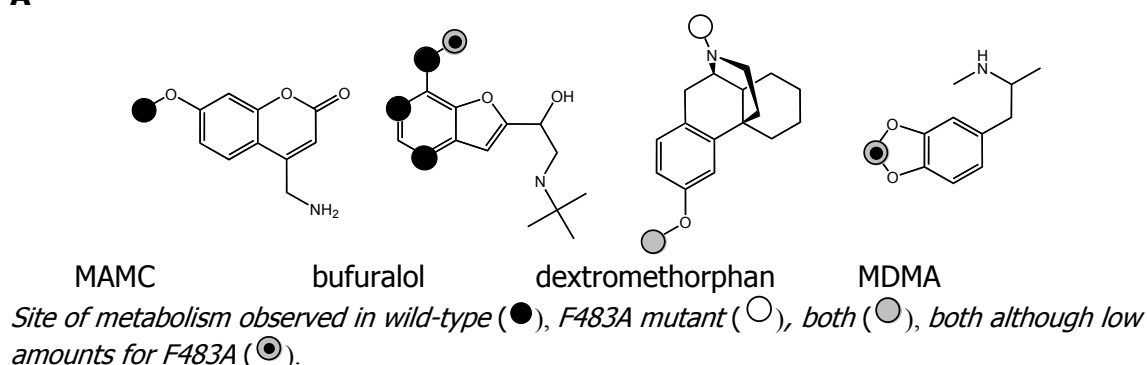
Figure 8 summarizes the changes in the metabolic profiles for these compounds due to the F120A mutation and offers a rationalization in terms of docking poses for dextromethorphan. In the mutant additional docking poses were observed, in which the orientations of dextromethorphan allows for N-demethylation. In the CYP2D6 crystal structure F120 is positioned and oriented in the same way as in our homology model (**Figure 6**). In agreement with the positioning of F120 indicated by the crystal structure and the homology model, the *N*-aryl-PPIX ratio obtained with the CYP2D6 F120A mutant featured an increased migration of the probes to the nitrogen atoms of pyrrole C and D which in the wild-type enzyme remained inaccessible.[34] The greater extent of migration to N_D, suggests that in the wild-type protein the phenyl group of F120 extends significantly more over pyrrole D than it does over pyrrole C.

Site of metabolism observed in wild-type (●), F120A mutant (○), in both (◐).
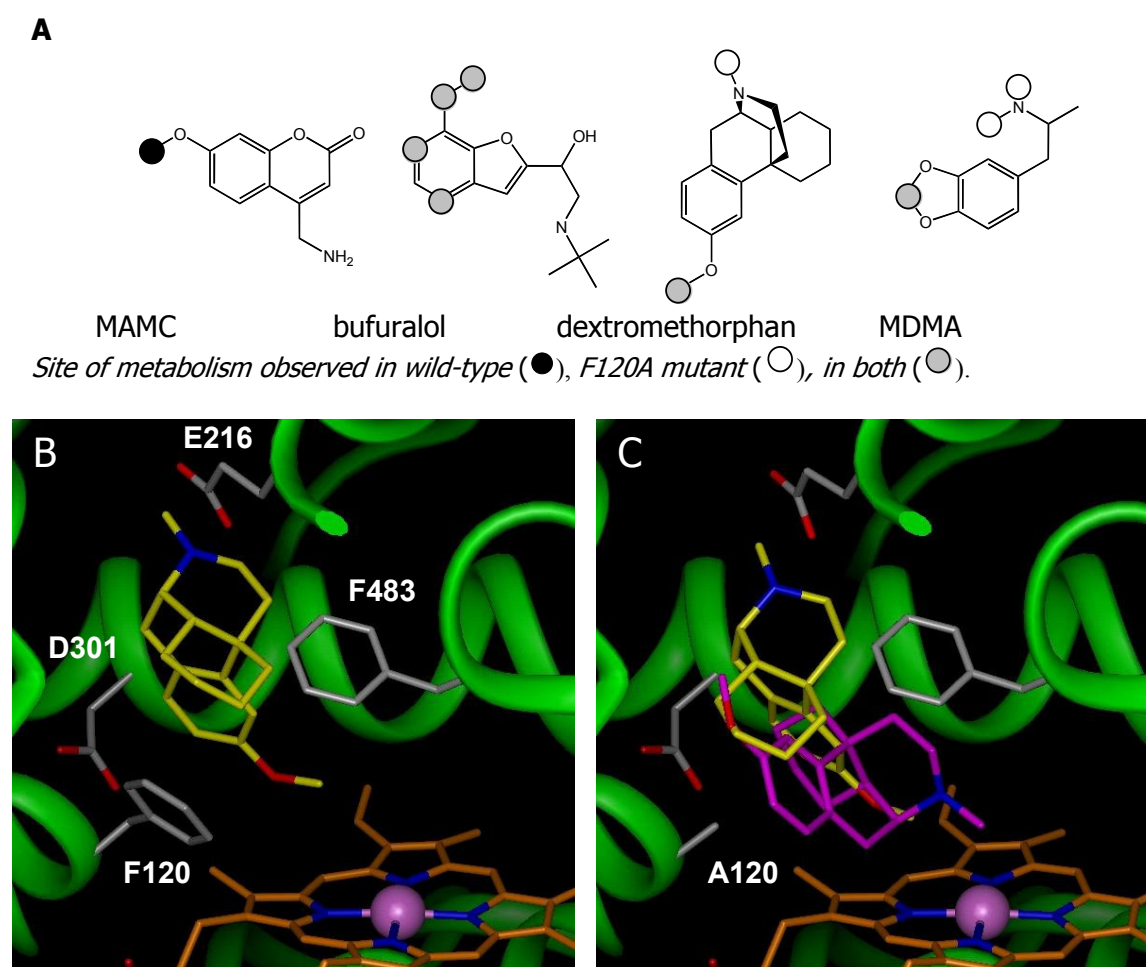
Figure 8: *In panel A, differences in regiospecificity of metabolism between wild-type and F120A mutant CYP2D6 is presented. In panels B and C, binding orientation of dextromethorphan automatically docked in the active site of the homology models of CYP2D6 using Autodock with standard conditions [47]. In wild-type CYP2D6, the orientation of dextromethorphan (in yellow) with its O-methyl pointing towards the heme is favoured (panel B), in the F120A mutant two distinct orientations, supporting both N-demethylation (in magenta) as O-demethylation (in yellow) were suggested (panel C).*

F483A

Replacement of F483 by A resulted in a 30-fold lower V_{max} for bufuralol 1'-hydroxylation, while the K_m was hardly affected.[33] Mutation of F483 to isoleucine or tryptophan did not affect the 1'-hydroxylation of bufuralol at all in an earlier study.[106] The V_{max} for 3,4-methylenedioxy-methylamphetamine O-demethylenation on the other hand decreased only two-fold, whereas the effect on the K_m was much larger. For dextromethorphan, in addition to dextrorphan (O-demethylation) and 3-methoxymorphinan (N-demethylation), two other metabolites were formed that could not be detected for the wild-type. The substrate MAMC was not metabolised at all by CYP2D6 F483A mutant. These data show that next to F120, residue F483 plays a very important role in the metabolism of typical CYP2D6 substrates.

Site of metabolism observed in wild-type ($\bullet$), F483A mutant ($\bigcirc$), both ($\circ$), both although low amounts for F483A ($\circledcirc$).

Figure 9: *In panel A, differences in regiospecificity of metabolism between wild-type and F483A mutant CYP2D6 are presented. In panels B and C, energetically most favourable binding orientations of R- (in yellow) and S-bufuralol (in magenta) in the active site of the homology models of CYP2D6 as determined by automated docking studies.[80] In wild-type CYP2D6 the orientation of R-bufuralol corresponding to 4-hydroxylation is favoured, while the orientation of S-bufuralol corresponds to C1'-hydroxylation (panel B). In the F483A mutant the orientation of S-burfuralol corresponds to C1'-hydroxylation and R-bufuralol is non-reactive (panel C).*

The influence of F483 on metabolism was found to be strongly substrate-dependent, as is summarized in **Figure 9**. For *S*-bufuralol the changes in metabolic profile due to the mutation is rationalized based on automated docking experiments presented in panels B and C of this figure. In the F483A mutant, only the ethyl chain can get in a position with

respect to the heme iron that is catalytically active, while more of such conformations are observed in the wild-type. In our homology model, the backbone of F483 is located at the same position as in the recently solved substrate-free crystal structure of CYP2D6, but the phenylring is directed differently towards the active site. In the homology model, the phenyl-ring of F483 has turned towards the active site, maximizing hydrophobic contacts with the substrate, while in the crystal structure, F483 forms an intermediate substrate binding pocket with E216 at the entrance of the solvent channel.

T309V

Mutation of T309 into a valine led to up to twenty-fold decreases in turnover rates of bufuralol 1-hydroxylation, dextromethorphan and MAMC O-demethylation and MDMA O-demethylenation, while binding affinities and formation of uncoupling products did not change.[35] The T309V mutant CYP2D6 showed increased dextromethorphan and MDMA N-dealkylation rates and thus product ratios of dextromethorphan and MDMA metabolism but also of bufuralol oxidation were altered. The use of cumene hydroperoxide instead of NADPH to support oxidation however, returned O-demethylating and O-demethylenating activity of the T309V mutant to the same level as in wild-type. These results indicate that T309 is a key determinant in maintaining the relative ratios of the potential reactive oxygen species that determine the ultimate products formed by CYP2D6.

A

MAMC   bufuralol   dextromethorphan   MDMA
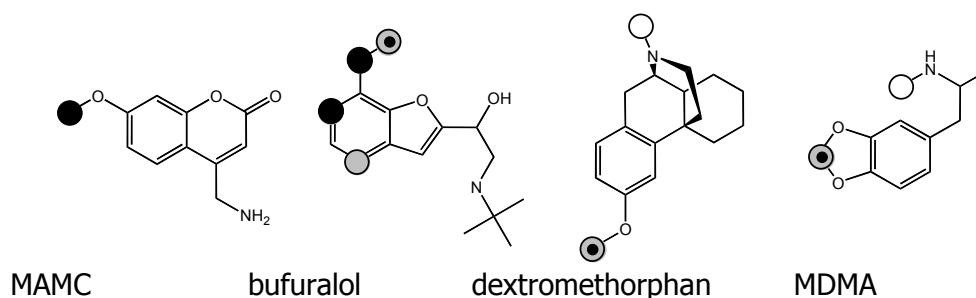
Site of metabolism observed in wild-type (●), T309V mutant (○), both (◓), both although low amounts for T309V (◉).

B

CYP	Species	I-helix sequence	Last resid
101	*P. putida*	S D E A K R M C G L L L V G G L D T V V N F L S F S M E F L A K S P	268
102A2	*B. Megaterium*	D E N I R Y Q I I T F L I A G H E T T S G L L S F A L Y F L V K N P	284
107A	*S. erythraea*	A D E L T S I A L V L L L A G F E A S V S L I G I G T Y L L L T H P	261
55A1	*F. oxysporum*	K S D A V Q I A F L L L V A G N A T M V N M I A L G V A T L A Q H P	258
1A2	*Rat*	Q E K I V N I V N D I F G A G F E T V T T A I F W S I L L L V T E P	335
2B4	*Rabbit*	H Q N L I L T V L S L F F A G T E T T S T T L R Y G F L L M L K Y P	318
2E1	*Rabbit*	L E N I A V T V A D M F F A G T E T T S T T L R Y G L L I L L K H P	319
2C5	*Rabbit*	L E S L V I A V S D L F G A G T E T T S T T L R Y S L L L L L K H P	314
2D6	*Human*	D E N L R I V V A D L F S A G M V T T S T T L A W G L L M I L H P	325

Figure 10: *In panel A, differences in regiospecificity of metabolism between wild-type and T309V mutant CYP2D6 are presented. Panel B presents the alignment of the I-helix protein sequences (drnelson.utmem.edu/CytochromesP450.html) showing the conserved threonine residue (indicated with a triangle) of CYPs in which this residue was experimentally shown to be involved in oxygen activation.[93-99]*

The T309V mutant is instable over time and large amounts of P420 are formed during purification.[35] This observation also suggests a structural role for T309 supported by our

homology model that shows that it is involved in stabilizing a distortion in the I helix by providing hydrogen bonds to the carbonyl groups of A305 and G306 (**Figure 7**). In the CYP2D6 crystal structure, T309 is positioned and oriented in the same way as in our homology model (**Figure 6**). It has been shown that mutation of the homologous T268 residue in CYP102[95] or T252 in CYP101[5,94] to Ala or Ile cause various degrees of conformational changes in the I-helix. Our model likewise shows that T312 is involved in stabilizing the distortion in the I helix as well by providing a hydrogen bond to the carbonyl group of V308. Mutation of that residue could have a conformational effect on the I helix similar to that observed with the T309V mutant and hence give the same N-alkylprotoporphyrin-IX ratios.[34] **Figure 10** summarizes the changes in metabolic profiles for the test compounds and depicts the sequence alignment of the I-helices of several iso-enzymes for which the structural role of the homologue of T309 was experimentally determined. The threonine mutants of the bacterial CYPs 101, 102A1, and fungal CYP55A1 were crystallized[94,95,97] and so are mammalian CYPs 2B4[9] and 2C5[12], the latter structure was used to build the CYP2D6 model. In CYP107A, the A245T mutant was shown to oxidize alternative substrates[99], and in CYPs 1A2, 2B4, and 2E1 the conserved threonine residues were also subjected to mutagenesis studies.[93,96,98]

3.4 Summary and perspectives

In this chapter we described the construction, refinement and validation of the CYP2D6 homology model that is used in our laboratory for the prediction and rationalisation of CYP2D6 substrate metabolism and CYP2D6-ligand interactions as described in this thesis and other work.[32-37] The model could explain reported site-directed mutagenesis data. For example, the model indicates that E216, and not D301, is the main actor responsible for interaction with the basic nitrogen atom in the substrate. D301 however, is involved in stabilizing the B/C-loop (containing F120) and mutation of this residue can thus affect substrate binding indirectly. Furthermore, based on the model, new CYP2D6 mutants were constructed and studied in our lab, and also for these mutants a rationalization of experimentally observed characteristics could be achieved. The influence of two active phenylalanines, F120 (located in substrate recognition site (SRS 1) and F483 (SRS6), on metabolism by CYP2D6 were found to be strongly substrate-dependent. Our homology model suggests that F120 is involved in direct π-π interactions with aromatic rings of substrates, while F483 is less involved in direct substrate interaction but is rather sterically influencing substrate binding. The F120A and F483A are studied in more detail in refs 32,33,36. I106 (SRS1) is not directly interacting with substrates, but the I106E mutation introduces an electrostatic repulsion with the negative charge of E216 (SRS2)[34], an active site residue already shown to be crucial for metabolism of typical CYP2D6 substrates. In this way, the mutation potentially causes shifts in the F- and G-helices leading to an alteration of the shape and character of the active site cavity. The highly conserved T309 (SRS4) also does not directly interact with ligands, but is a key determinant in maintaining the relative ratios of the potential reactive oxygen species that determine the ultimate products formed by CYP2D6.[35] Future combined Resonance Raman and QM/MM studies will focus on the elucidation of the role of T309 in CYP2D6-mediated metabolism [Bonifacio & Groenhof *et al.,* unpublished results]. **Table 4** summarizes the influence of the I106E, F120A, T309V, and F483A mutations on the catalytic efficiency of CYP2D6 to metabolize four typical CYP2D6 substrates, 7-methoxy-4-(aminomethyl)-coumarin (MAMC), dextromethorphan, bufuralol and 3,4-methylenedioxymethamphetamine (MDMA). CYP2D6-substrate interaction fingerprint analysis of docked substrates in our homology model (**Table 2**) suggest that several other active site residues are probably interacting with ligands as well (e.g., V104 (SRS1), L213 (SRS2), and V308 (SRS4)), opening the way for further mutagenesis studies.

Table 5: *Computational predictions, refinement and validation approaches and extra features of the CYP2D6 homology modeling described in this thesis and other work (references between brackets).*

Templates	predicted properties[a]	refinement and validation[b]	extra features
CYP2C5 (DMZ-bound) [11] + CYP2C5 (diclofenac-bound) [13]	CSP (chapters 3,5,6) K_s (chapter 7) IC_{50} (chapter 6)	AD (chapters 3,5,6,7,8 and ref 36) MD (chapter 3,6,7,8 and ref 37) SM (chapters 3,6,7 and refs 32-37) P (this chapter)	Active site chemical probe studies (chapter 3, ref [34]) Refinement and validation using NMR-derived ligand heme distances (chapters 3, 6) Investigation substrate and product egress channels (chapter 8) Fluorescence spectroscopy (ref 37) RR spectroscopy (ref [107])

a) Predicted property: 'Affinity' (spectral dissociation constant K_s or concentration required for 50% inhibition of the metabolism of a specific substrate (IC_{50})); Catalytic Site Prediction (CSP)
b) Method of prediction and validation: Automated docking (AD) of substrates in homology model to predict site of catalysis; Molecular dynamics (MD) simulations to predict site of catalysis or affinity (Thermodynamic Integration method); Combined with pharmacophore model (P); Site-directed mutagenesis (SM) studies used for refinement and validation of sequence alignment and/or homology model.

Further validation and applications of the CYP2D6 homology model are described this thesis and in refs 32-37 and involve the prediction of metabolic sites in substrates, the affinity of ligands to the protein and K_s and IC_{50} values. As is described in **Table 5**, a variety of methods was applied, such as automated docking experiments, molecular dynamics simulations, free energy calculations using Thermodynamic Integration (TI), random acceleration molecular dynamics simulations (RAMD) and statistical methods aimed at predicting affinities.

Very recently the crystal structure of substrate-free CYP2D6 was solved.[19] Our homology model was found to agree with most of the details of this crystal structure. Structural differences between the homology model and crystal structure were the same differences observed between substrate-free and substrate-bound structures of other CYPs, suggesting that these conformational changes are required upon substrate binding. The recently solved CYP2D6 crystal structure further validates our homology modeling approach and demonstrates on one hand the success and accuracy of the state of the art of CYP homology modeling. On the other hand it shows that computational chemistry is a useful and valuable tool to provide models for substrate-bound complexes of CYPs which give insight into CYP-ligand interactions. This information is essential for successful pre-experimental virtual screening, as well as accurate hypothesis generation for *in vitro* studies in drug discovery and development.

References

(1)   Goeptar, A. R.; Scheerens, H.; Vermeulen, N. P. Oxygen and xenobiotic reductase activities of cytochrome P450. *Crit Rev Toxicol* **1995**, *25*, 25-65.
(2)   Guengerich, F. P. Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem Res Toxicol* **2001**, *14*, 611-650.
(3)   Zanger, U. M.; Raimundo, S.; Eichelbaum, M. Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. *Naunyn Schmiedebergs Arch Pharmacol* **2004**, *369*, 23 - 37.
(4)   Bertilsson, L.; Dahl, M. L.; Dalen, P.; Al-Shurbaji, A. Molecular genetics of CYP2D6: clinical relevance with focus on phychotropic drugs. *Br J of Clin Pharmacol* **2002**, *53*, 111 - 122.
(5)   Hishiki, T.; Shimada, H.; Nagano, S.; Egawa, T.; Kanamori, Y. et al. X-ray crystal structure and catalytic properties of Thr252Ile mutant of cytochrome P450cam: roles of Thr252 and water in the active center. *J Biochem (Tokyo)* **2000**, *128*, 965-974.
(6)   Oscarson, M. Pharmacogenetics of drug metabolising enzymes: importance for personalised medicine. *Clin Chem Lab Med* **2003**, *41*, 573 - 580.

(7)   Ingelman-Sundberg, M. Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity. *Pharmacogenomics J* **2005**, *5*, 6-13.

(8)   Yano, J. K.; Hsu, M. H.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen. *Nat Struct Mol Biol* **2005**, *12*, 822-823.

(9)   Scott, E. E.; He, Y. A.; Wester, M. R.; White, M. A.; Chin, C. C. et al. An open conformation of mammalian cytochrome P450 2B4 at 1.6-A resolution. *Proc Natl Acad Sci U S A* **2003**, *100*, 13196-13201.

(10)   Scott, E. E.; White, M. A.; He, Y. A.; Johnson, E. F.; Stout, C. D. et al. Structure of mammalian cytochrome P450 2B4 complexed with 4-(4-chlorophenyl)imidazole at 1.9-A resolution: insight into the range of P450 conformations and the coordination of redox partner binding. *J Biol Chem* **2004**, *279*, 27294-27301.

(11)   Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dansette, P. M.; Mansuy, D. et al. Structure of a substrate complex of mammalian cytochrome P450 2C5 at 2.3 A resolution: evidence for multiple substrate binding modes. *Biochemistry* **2003**, *42*, 6370-6379.

(12)   Williams, P. A.; Cosme, J.; Sridhar, V.; Johnson, E. F.; McRee, D. E. Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol Cell* **2000**, *5*, 121-131.

(13)   Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dijols, S.; Dansette, P. M. et al. Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 A resolution: evidence for an induced fit model of substrate binding. *Biochemistry* **2003**, *42*, 9335-9345.

(14)   Schoch, G. A.; Yano, J. K.; Wester, M. R.; Griffin, K. J.; Stout, C. D. et al. Structure of human microsomal cytochrome P450 2C8. Evidence for a peripheral fatty acid binding site. *J Biol Chem* **2004**, *279*, 9497-9503.

(15)   Williams, P. A.; Cosme, J.; Ward, A.; Angove, H. C.; Matak Vinkovic, D. et al. Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* **2003**, *424*, 464-468.

(16)   Wester, M. R.; Yano, J. K.; Schoch, G. A.; Yang, C.; Griffin, K. J. et al. The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-A resolution. *J Biol Chem* **2004**, *279*, 35630-35637.

(17)   Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D. et al. The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-A resolution. *J Biol Chem* **2004**, *279*, 38091-38094.

(18)   Williams, P. A.; Cosme, J.; Vinkovic, D. M.; Ward, A.; Angove, H. C. et al. Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* **2004**, *305*, 683-686.

(19)   Rowland, P.; Blaney, F. E.; Smyth, M. G.; Jones, J. J.; Leydon, V. R. et al. Crystal structure of human cytochrome P450 2D6. *J Biol Chem* **2006**, *281*, 7614-7622.

(20)   de Graaf, C.; Vermeulen, N. P.; Feenstra, K. A. Cytochrome p450 in silico: an integrative modeling approach. *J Med Chem* **2005**, *48*, 2725-2755.

(21)   de Groot, M. J.; Kirton, S. B.; Sutcliffe, M. J. In silico methods for predicting ligand binding determinants of cytochromes P450. *Curr Top Med Chem* **2004**, *4*, 1803-1824.

(22)   Li, H. Y.; Poulos, T. L. The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. *Nature Structural Biology* **1997**, *4*, 140-146.

(23)   Haines, D. C.; Tomchick, D. R.; Machius, M.; Peterson, J. A. Pivotal role of water in the mechanism of P450BM-3. *Biochemistry* **2001**, *40*, 13456-13465.

(24)   Li, H.; Poulos, T. L. Conformational dynamics in cytochrome P450-substrate interactions. *Biochimie* **1996**, *78*, 695-699.

(25)   Cupp-Vickery, J. R.; Garcia, C.; Hofacre, A.; McGee-Estrada, K. Ketoconazole-induced conformational changes in the active site of cytochrome P450eryF. *J Mol Biol* **2001**, *311*, 101-110.

(26)   Podust, L. M.; Poulos, T. L.; Waterman, M. R. Crystal structure of cytochrome P450 14alpha -sterol demethylase (CYP51) from Mycobacterium tuberculosis in complex with azole inhibitors. *Proc Natl Acad Sci U S A* **2001**, *98*, 3068-3073.

(27)   Park, S. Y.; Yamane, K.; Adachi, S.; Shiro, Y.; Weiss, K. E. et al. Thermophilic cytochrome P450 (CYP119) from Sulfolobus solfataricus: high resolution structure and functional properties. *J Inorg Biochem* **2002**, *91*, 491-501.

(28)   Hasemann, C. A.; Kurumbail, R. G.; Boddupalli, S. S.; Peterson, J. A.; Deisenhofer, J. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure* **1995**, *3*, 41-62.

(29)   Mestres, J. Structure conservation in cytochromes P450. *Proteins* **2005**, *58*, 596-609.

(30)   Li, H. Cytochrome P450. *Handbook of metalloproteins*; Wiley: Chichester, 2001; pp 267-282.

(31)   Gotoh, O. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem* **1992**, *267*, 83-90.

(32)   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.

(33)   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.

(34)   van Waterschoot, R. A.; Keizers, P. H.; de Graaf, C.; Vermeulen, N. P.; Tschirret-Guth, R. A. Topological role of cytochrome P450 2D6 active site residues. *Arch Biochem Biophys* **2006**, *447*, 53-58.

(35)   Keizers, P. H.; Schraven, L. H.; de Graaf, C.; Hidestrand, M.; Ingelman-Sundberg, M. et al. Role of the conserved threonine 309 in mechanism of oxidation by cytochrome P450 2D6. *Biochem Biophys Res Commun* **2005**, *338*, 1065-1074.

(36)   Keizers, P. H.; van Dijk, B. R.; de Graaf, C.; van Vught-Lussenburg, B. M.; Vermeulen, N. P. et al. Metabolism of N-substituted 7-methoxy-4-(aminomethyl)-coumarins by cytochrome P450 2D6 mutants indicates additional substrate interaction points. *Xenobiotica* **2006**, *in press*.

(37)   Stortelder, A.; Keizers, P. H.; Oostenbrink, C.; De Graaf, C.; De Kruijf, P. et al. Binding of 7-methoxy-4-(aminomethyl)-coumarin to wild-type and W128F mutant cytochrome P450 2D6 studied by time-resolved fluorescence spectroscopy. *Biochem J* **2006**, *393*, 635-643.

(38)   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.

(39)   Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **1993**, *234*, 779-815.

(40)   Notredame, C.; Higgins, D. G.; Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **2000**, *302*, 205-217.

(41)   Rost, B.; Liu, J. The PredictProtein server. *Nucleic Acids Res* **2003**, *31*, 3300-3304.

(42)   Cuff, J. A.; Clamp, M. E.; Siddiqui, A. S.; Finlay, M.; Barton, G. J. JPred: a consensus secondary structure prediction server. *Bioinformatics* **1998**, *14*, 892-893.

(43)   Kneller, D. G.; Cohen, F. E.; Langridge, R. Improvements in protein secondary structure prediction by an enhanced neural network. *J Mol Biol* **1990**, *214*, 171-182.

(44)   Laskowski, R. A.; Macarthur, M. W.; Moss, D. S.; Thornton, J. M. Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography* **1993**, *26*, 283-291.

(45)   Colovos, C.; Yeates, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* **1993**, *2*, 1511-1519.

(46)   Luthy, R.; Bowie, J. U.; Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* **1992**, *356*, 83-85.

(47)   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. *Journal of Computational Chemistry* **1998**, *19*, 1639-1662.

(48)   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.

(49)   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.

(50)   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.

(51)   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. *Journal of Molecular Modeling* **1998**, *4*, 379-394.

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

(53)   Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* **2001**, *7*, 306-317.

(54)   van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hunenberger, P. H.; Kruger, P. et al. *Biomolecular simulation: The GROMOS96 manual and user guide* **1996**.

(55)   Strobl, G. R.; von Kruedener, S.; Stockigt, J.; Guengerich, F. P.; Wolff, T. Development of a pharmacophore for inhibition of human liver cytochrome P-450 2D6: molecular modeling and inhibition studies. *J Med Chem* **1993**, *36*, 1136-1145.

(56)   de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A.; Jones, B. C. Novel approach to predicting P450-mediated drug metabolism: Development of a combined protein and pharmacophore model for CYP2D6. *Journal of Medicinal Chemistry* **1999**, *42*, 1515-1524.

(57)   de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A.; Jones, B. C. A novel approach to predicting P450 mediated drug metabolism. CYP2D6 catalyzed N-dealkylation reactions and qualitative metabolite predictions using a combined protein and pharmacophore model for CYP2D6. *Journal of Medicinal Chemistry* **1999**, *42*, 4062-4070.

(58)   Meyer, U. A.; Gut, J.; Kronbach, T.; Skoda, C.; Meier, U. T. et al. The molecular mechanisms of two common polymorphisms of drug oxidation--evidence for functional changes in cytochrome P-450 isozymes catalysing bufuralol and mephenytoin oxidation. *Xenobiotica* **1986**, *16*, 449-464.

(59)   Islam, S. A.; Wolf, C. R.; Lennard, M. S.; Sternberg, M. J. A three-dimensional molecular template for substrates of human cytochrome P450 involved in debrisoquine 4-hydroxylation. *Carcinogenesis* **1991**, *12*, 2211-2219.

(60)   Koymans, L.; Vermeulen, N. P.; van Acker, S. A.; te Koppele, J. M.; Heykants, J. J. et al. A predictive model for substrates of cytochrome P450-debrisoquine (2D6). *Chem Res Toxicol* **1992**, *5*, 211-219.

(61) Wolff, T.; Distelrath, L. M.; Worthington, M. T.; Groopman, J. D.; Hammons, G. J. et al. Substrate specificity of human liver cytochrome P-450 debrisoquine 4-hydroxylase probed using immunochemical inhibition and chemical modeling. *Cancer Res* **1985**, *45*, 2116-2122.

(62) de Groot, M. J.; Bijloo, G. J.; Martens, B. J.; van Acker, F. A.; Vermeulen, N. P. A refined substrate model for human cytochrome P450 2D6. *Chem Res Toxicol* **1997**, *10*, 41-48.

(63) de Groot, M. J.; Bijloo, G. J.; van Acker, F. A.; Fonseca Guerra, C.; Snijders, J. G. et al. Extension of a predictive substrate model for human cytochrome P4502D6. *Xenobiotica* **1997**, *27*, 357-368.

(64) Afzelius, L.; Masimirembwa, C. M.; Karlen, A.; Andersson, T. B.; Zamora, I. Discriminant and quantitative PLS analysis of competitive CYP2C9 inhibitors versus non-inhibitors using alignment independent GRIND descriptors. *J Comput Aided Mol Des* **2002**, *16*, 443-458.

(65) Lüdemann, S. K.; Lounnas, V.; Wade, R. C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. *Journal of Molecular Biology* **2000**, *303*, 797-811.

(66) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E. et al. Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Computer Physics Communications* **1995**, *91*, 1-41.

(67) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577-2637.

(68) Livingstone, C. D.; Barton, G. J. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput Appl Biosci* **1993**, *9*, 745-756.

(69) Engh, R. A.; Huber, R. Accurate Bond and Angle Parameters for X-Ray Protein-Structure Refinement. *Acta Crystallographica Section A* **1991**, *47*, 392-400.

(70) Morris, A. L.; MacArthur, M. W.; Hutchinson, E. G.; Thornton, J. M. Stereochemical quality of protein structure coordinates. *Proteins* **1992**, *12*, 345-364.

(71) 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.

(72) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **1997**, *18*, 1463-1472.

(73) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular-dynamics with coupling to an external bath. *Journal of Chemical Physics* **1984**, *81*, 3684-3690.

(74) Keizers, P. H.; de Graaf, C.; de Kanter, F. J. J.; Oostenbrink, C.; Feenstra, K. A. et al. Metabolic regio- and stereoselectivity of Cytpchrome P450 2D6 towards 3,4-methylenedioxy-N-alkyl-amphetamines: In silico Predictions and experimental validation. *J. Med. Chem.* **2005**, *48*, 6117-6127.

(75) 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 Metabolism and Disposition* **2001**, *29*, 936-944.

(76) de Groot, M. J.; Ekins, S. Pharmacophore modeling of cytochromes P450. *Advanced Drug Delivery Reviews* **2002**, *54*, 367 - 383.

(77) Koymans, L. M.; Vermeulen, N. P.; Baarslag, A.; Donne-Op den Kelder, G. M. A preliminary 3D model for cytochrome P450 2D6 constructed by homology model building. *J Comput Aided Mol Des* **1993**, *7*, 281-289.

(78) de Groot, M. J.; Vermeulen, N. P.; Kramer, J. D.; van Acker, F. A.; Donne-Op den Kelder, G. M. A three-dimensional protein model for human cytochrome P450 2D6 based on the crystal structures of P450 101, P450 102, and P450 108. *Chem Res Toxicol* **1996**, *9*, 1079-1091.

(79) 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.

(80) de Graaf, C.; Oostenbrink, C.; Keizers, P. H.; van der Wijst, T.; Jongejan, A. et al. Catalytic Site Prediction of and Virtual Screening Accuracy of Cytochrome P450 2D6 Substrates by Consideration of Water and Rescoring in Automated Docking. *J Med Chem* **2006**, *49*, 2417-2430.

(81) Gonzales, F. J.; Vilbois, F.; Hardwick, J. P.; McBride, O. W.; Nebert, D. W. et al. Human debrisoquine 4-hydroxylase (P450IID1): cDNA and deduced amino acid sequence and assignment of the CYP2D locus to chromosome 22. *Genomics* **1988**, *2*, 174 - 179.

(82) Kimura, S.; Umeno, M.; Skoda, R. C.; Meyer, U. A.; Gonzalez, F. J. The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and identification of the polymorphic CYP2D6 gene, a related gene, and a pseudogene. *Am J Hum Genet* **1989**, *45*, 889-904.

(83) Modi, S.; Paine, M. J.; Sutcliffe, M. J.; Lian, L. Y.; Primrose, W. U. et al. A model for human cytochrome P450 2D6 based on homology modeling and NMR studies of substrate binding. *Biochemistry* **1996**, *35*, 4540-4550.

(84) Modi, S.; Gilham, D. E.; Sutcliffe, M. J.; Lian, L. Y.; Primrose, W. U. et al. 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine as a substrate of cytochrome P450 2D6: allosteric effects of NADPH-cytochrome P450 reductase. *Biochemistry* **1997**, *36*, 4461-4470.

(85) De Rienzo, F.; Fanelli, F.; Menziani, M. C.; De Benedetti, P. G. Theoretical investigation of substrate specificity for cytochromes P450 IA2, P450 IID6 and P450 IIIA4. *J Comput Aided Mol Des* **2000**,*14*, 93-116.

(86) Snyder, R.; Sangar, R.; Wang, J. B.; Ekins, S. Three-dimensional quantitative structure activity relationship for CYP2D6 substrates. *Quantitative Structure-Activity Relationships* **2002**, *21*, 357-368.

(87) Ellis, S. W.; Hayhurst, G. P.; Smith, G.; Lightfoot, T.; Wong, M. M. et al. Evidence that aspartic acid 301 is a critical substrate-contact residue in the active site of cytochrome P450 2D6. *J Biol Chem* **1995**, *270*, 29055-29058.

(88) Mackman, R.; Tschirret-Guth, R. A.; Smith, G.; Hayhurst, G. P.; Ellis, S. W. et al. Active-site topologies of human CYP2D6 and its aspartate-301 --> glutamate, asparagine, and glycine mutants. *Arch Biochem Biophys* **1996**, *331*, 134-140.

(89) Hanna, I. H.; Kim, M. S.; Guengerich, F. P. Heterologous expression of cytochrome P450 2D6 mutants, electron transfer, and catalysis of bufuralol hydroxylation: the role of aspartate 301 in structural integrity. *Arch Biochem Biophys* **2001**, *393*, 255-261.

(90) Guengerich, F. P.; Miller, G. P.; Hanna, I. H.; Martin, M. V.; Leger, S. et al. Diversity in the oxidation of substrates by cytochrome P450 2D6: lack of an obligatory role of aspartate 301-substrate electrostatic bonding. *Biochemistry* **2002**, *41*, 11025-11034.

(91) 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.

(92) 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.

(93) Furuya, H.; Shimizu, T.; Hirano, K.; Hatano, M.; Fujii-Kuriyama, Y. et al. Site-directed mutageneses of rat liver cytochrome P-450d: catalytic activities toward benzphetamine and 7-ethoxycoumarin. *Biochemistry* **1989**, *28*, 6848-6857.

(94) Raag, R.; Poulos, T. L. Crystal-Structures of Cytochrome-P-450cam Complexed with Camphane, Thiocamphor, and Adamantane - Factors Controlling P-450 Substrate Hydroxylation. *Biochemistry* **1991**, *30*, 2674-2684.

(95) Yeom, H.; Sligar, S. G.; Li, H.; Poulos, T. L.; Fulco, A. J. The role of Thr268 in oxygen activation of cytochrome P450BM-3. *Biochemistry* **1995**, *34*, 14733-14740.

(96) Vaz, A. D.; Pernecky, S. J.; Raner, G. M.; Coon, M. J. Peroxo-iron and oxenoid-iron species as alternative oxygenating agents in cytochrome P450-catalyzed reactions: switching by threonine-302 to alanine mutagenesis of cytochrome P450 2B4. *Proc Natl Acad Sci U S A* **1996**, *93*, 4644-4648.

(97) Park, S. Y.; Shimizu, H.; Adachi, S.; Nakagawa, A.; Tanaka, I. et al. Crystal structure of nitric oxide reductase from denitrifying fungus Fusarium oxysporum. *Nat Struct Biol* **1997**, *4*, 827-832.

(98) Vaz, A. D.; McGinnity, D. F.; Coon, M. J. Epoxidation of olefins by cytochrome P450: evidence from site-specific mutagenesis for hydroperoxo-iron as an electrophilic oxidant. *Proc Natl Acad Sci U S A* **1998**, *95*, 3555-3560.

(99) Xiang, H.; Tschirret-Guth, R. A.; Ortiz De Montellano, P. R. An A245T mutation conveys on cytochrome P450eryF the ability to oxidize alternative substrates. *J Biol Chem* **2000**, *275*, 35999-36006.

(100) Ortiz de Montellano, P. R. Arylhydrazines as probes of hemoprotein structure and function. *Biochimie* **1995**, *77*, 581-593.

(101) Masuda, K.; Tamagake, K.; Okuda, Y.; Torigoe, F.; Tsuzuki, D. et al. Change in enantioselectivity in bufuralol 1''-hydroxylation by the substitution of phenylalanine-120 by alanine in cytochrome P450 2D6. *Chirality* **2005**, *17*, 37-43.

(102) 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.

(103) McLaughlin, L. A.; Paine, M. J.; Kemp, C. A.; Marechal, J. D.; Flanagan, J. U. et al. Why is quinidine an inhibitor of cytochrome P450 2D6? The role of key active-site residues in quinidine binding. *J Biol Chem* **2005**, *280*, 38617-38624.

(104) Lightfoot, T.; Ellis, S. W.; Mahling, J.; Ackland, M. J.; Blaney, F. E. et al. Regioselective hydroxylation of debrisoquine by cytochrome P4502D6: implications for active site modelling. *Xenobiotica* **2000**, *30*, 219-233.

(105) Ellis, S. W.; Rowland, K.; Ackland, M. J.; Rekka, E.; Simula, A. P. et al. Influence of amino acid residue 374 of cytochrome P-450 2D6 (CYP2D6) on the regio- and enantio-selective metabolism of metoprolol. *Biochem J* **1996**, *316*.

(106) Smith, G.; Modi, S.; Pillai, I.; Lian, L. Y.; Sutcliffe, M. J. et al. Determinants of the substrate specificity of human cytochrome P-450 CYP2D6: design and construction of a mutant with testosterone hydroxylase activity. *Biochem J* **1998**, *331 (Pt 3)*, 783-792.

(107) Bonifacio, A.; Keizers, P. H.; Commandeur, J. N.; Vermeulen, N. P.; Robert, B. et al. Binding of bufuralol, dextromethorphan, and 3,4-methylenedioxymethylamphetamine to wild-type and F120A mutant cytochrome P450 2D6 studied by resonance Raman spectroscopy. *Biochem Biophys Res Commun* **2006**, *343*, 772-779.

CYP(2D6) structure-based virtual screening

Eautomated docking

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 (LACDR)/Division of Molecular Toxicology, Department of Chemistry and Pharmacochemistry, Vrije Universiteit, The Netherlands
[§]Institute of Pharmaceutical Sciences, Department of Chemistry and Applied Biosciences, ETH Zurich, Switzerland.
[*]Both authors contributed equally

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. Considering 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.

4.1 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 significantly 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* (CYPs) are hemoproteins which catalyse 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 non-toxic parent compounds are bioactivated into toxic metabolites, and procarcinogens into their ultimate carcinogens.[29] Automated docking has primarily been applied to refine and validate CYP pharmacophore models and protein homology models, as was for example recently done for CYP2D6.[30,31] Regarding CYPs, 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 CYP1A2 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 analogs 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 (CYPs) and thymidine kinases (TKs). Concretely, 19 crystallized P450- and 19 TK-ligand complexes were selected for docking. In order 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 un-arbitrarily 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.

4.2 Results

Prediction of 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 analysed to examine the correctness of the new water prediction method applied in this study, both for CYPs and TKs (**Figure 1A** and **B**).

  <u>CYP</u> – In many cases, the locations of X-ray active site water molecules and their interactions with CYP ligands (filled balls, see **Figure 1A**) were well predicted ("c" $\geq$ 1), concretely with 2-phenylimidazole (pdb code: 1phe, CYP101), 6-deoxyerythronolide B (1jio, CYP107), dimethylsulphophenazole derivative (1n6b, CYP2C5), diclofenac (1nr6, CYP2C5) and epothilone D (1pkf, Cyt P450$_{EpoK}$) (see **Figure 1A**). In the case of almost all CYP101-terpene structures (2cpp, 2cp4, 4cpp, 6cpp, 7cpp, 8cpp), up to 2 water molecules were predicted to be close to the respective ligand, while not present in the X-ray structure ("p - c" = nr. of false positives). Water molecules ligated to the heme Fe-atom according to crystallographic studies (4cpp, 6cpp, 7cpp, 1phe) are not predicted ("x - c" = nr. 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*).

  <u>TK</u> – 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" $\geq$ 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 CYP- 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) x 3 (docking algorithms) x 2 (scoring functions)) docking strategies were tested by assessing their ability to reproduce the experimental binding orientations of protein-ligand complexes of five CYP 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.

Figure 1A,B: *Molecular structures of ligands, their water-mediated interactions to the protein and correctness of water location prediction. Symbols are explained in the legend below.*

→ site of phosphorylation (TK) and site of oxidation correponding to major metabolic route (Cyt P450)

⇢ site of oxidation correponding to major metabolic route (Cyt P450)

∿ covalent bond with the heme iron atom (Cyt P450)

◯ ligand atom within 4 Å distance from a X-ray water molecule

⬚ ligand atom within 4 Å distance from a predicted water molecule

◐ ligand atom within 4 Å distance from both a X-rayand a predicted water molecule

x number of X-ray water molecules

p number of predicted water molecules

c number of correctly predicted water molecules

x Number of water molecules within4 Å distance from the ligand as: observed in the X-ray structure (x), as predicted (p), and correctly predicted (c)

1n6b
x p c
5 2 2

1pkf
x p c
5 3 2

1izo
x p c
2 2 0

7cpp
x p c
1 1 0

8cpp
x p c
1 1 0

1akd
x p c
0 1 0

2cp4
x p c
0 1 0

2cpp
x p c
0 1 0

1nr6
x p c
3 2 2

1phg
x p c
0 1 0

1jio
x p c
5 3 2

5cpp
x p c
0 1 0

4cpp
x p c
1 2 0

1mpw
x p c
0 0 0

1phf
x p c
0 1 0

1phe
x p c
2 1 1

1phd
x p c
0 0 0

1py2
x p c
0 1 0

6cpp
x p c
3 1 0

1A) *P450 pdb codes correspond to the following compounds:* **CYP101 (cam):** *2cpp/2cp4(T252A mutant), 1R-camphor; 1akd, 1S-camphor; 8cpp, thiocamphor; 7cpp, camphene; 6cpp, 1R-norcamphor; 4cpp, adamantane; 5cpp adamantanone; 1p2y, 1S-nicotine; 1phd, 1-phenylimidazole; 1phe, 2-phenylimidazole; 1phf, 4-phenylimidazole; 1phg, metyrapone;* **CYP107 (EryF):** *1jio, 6-deoxyerythronolide B;* **Cyt P450_EpoK:** *1pkf, epothilone D;* **Cyt P450_BSβ :1izo, palmitoleic acid;* **CYP2C5:** *1n6b, 4-methyl-N-methyl-N-(2-phenyl-2H-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.*

1B) TK pdb codes correspond to the following compounds: _TK EHV4: 1p6x and1p72, thymidine; 1e2i, 9-hydroxypropyladenine; 1e2k, (North)-methanocarba-thymidine; 1e2m, 6-hydroxypropylthymidine; 1e2n, 6-[(4-hydroxymethyl)-5-methyl-2,6-dioxohexahydropyrimmethyl]-5-methyl-2,3(1H, 3H)-pyrimidinedione; 1e2p, 6-[3-hydroxy-2-(hydroxymethyl)propyl]-5-methyl-2,4(1H,3H)-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)-N2-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. 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 nr. 1 ranked solutions according to the program implemented scoring functions (**I**)*
- *only nr. 1 ranked solutions according to the stand-alone SCORE scoring function (S)*
- *nr. 1 ranked solutions according to both the program implemented and stand-alone SCORE scoring functions (**B**)*
- *any of the docking solutions, but not considered as nr. 1 ranked solutions by either the program implemented or stand-alone SCORE scoring functions (•).*

CYP

	AutoDock			FlexX			GOLD		
	N	X	P	N	X	P	N	X	P
cam									
1akd									
2cp4			•	*I*	*I*	*I*	**B**	**B**	**B**
2cpp				*S*	*S*	*S*	**B**	**B**	**B**
4cpp		*S*					*S*	**B**	**B**
5cpp				*S*	*S*	*S*	**B**	**B**	**B**
6cpp	*S*	*S*	*S*	•	•	•		•	
7cpp		•		**B**	**B**	**B**		**B**	
8cpp		•	*S*	•		•			
1mpw	*S*		*I*	**B**	**B**	**B**			
1phd			•	•	*I*	*I*	*S*	*I*	
1phe			**B**	*S*	**B**	**B**			
1phf			•	•	•	*I*			*S*
1phg			**B**				**B**	**B**	**B**
1p2y	•	•	•	**B**	**B**	**B**	**B**	**B**	**B**
EryF									
1jio	*S*	**B**	**B**		*I*	*S*	**B**	**B**	**B**
βSB									
1pkf	**B**	**B**	**B**	**B**		**B**	**B**	**B**	**B**
EpoK									
1izo	**B**	**B**	**B**	*S*	•	**B**	**B**	**B**	**B**
2C5									
1n6b	*S*	**B**	**B**		*S*		*I*	**B**	*S*
1nr6	*I*	**B**	**B**			•	•	•	•

TK

	AutoDock			FlexX			GOLD		
	N	X	P	N	X	P	N	X	P
HSV1									
1e2iA		**B**	**B**		*S*		*S*	•	*S*
1e2iB		•							
1e2j	•	*S*	*S*	•			•	*I*	
1e2k	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**
1e2m	**B**	**B**	**B**	**N**	**X**	*S*	**B**	**B**	**B**
1e2n	**B**	**B**	**B**						**B**
1e2p	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**
1ki2	**B**	**B**	**B**	*I*	**B**	*S*	**B**	**B**	**B**
1ki3	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**
1ki4	**B**	**B**	**B**				**B**	**B**	**B**
1ki6	**B**	**B**	**B**	*I*	**B**	**B**	**B**	**B**	**B**
1ki7	**B**	**B**	**B**				**B**	**B**	**B**
1ki8	*S*	**B**	**B**		**B**	**B**	**B**	*S*	**B**
2vtk	**B**	**B**	**B**	**B**	**B**	**B**	•	**B**	**B**
1qhi	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**
2ki5	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**
VZV									
1osn	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**
EHV4									
1p72	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**
1p6x			**B**					**B**	**B**

Incorporation of active site water molecules in automated docking studies

  <u>*CYP*</u> - 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*), 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 CYP 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 CYP101 (*cam*) terpene (4cpp, 8cpp) and imidazole (1phe, 1phg) analogs, and the ligands of CYP107 (*EryF*) (1jio, shown in **Figure 3**) and CYP2C5 (1n6b, 1nr6), improved by including X-ray and/or predicted active site water molecules. For several ligand molecule types and different CYP 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.

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), and 5-bromovinyldeoxyuridine (1ki8, TK HSV1, shown in **Figure 3**), and the *R*-stereoisomer of 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

CYP – 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 CYP-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 CYP 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*).

Figure 2 (next page): *Docking accuracy of different automated docking approaches for CYP and TK in terms of RMSD (2A and B) and CSP accuracy (2C 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 nr. 1 ranked solutions according to the program implemented scoring function; S, RMSD distributions and CSP accuracy for nr. 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. Upper bars and lower bars in panels A and B correspond to percentages RMSD ≤ 1.0 Å and RMSD >1.0 Å, ≤ 2.0 Å, respectively. Bars in panels C and D correspond to percentages correctly predicted catalytic site.*

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

CYP – 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 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 re-ranking 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 (Ā) 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.

TK – Re-ranking 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 *Ā* 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.

Figure 3 (next page): *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 CYP107 (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 nr. 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 [38] and Raster3D.[39]*

P   X   N
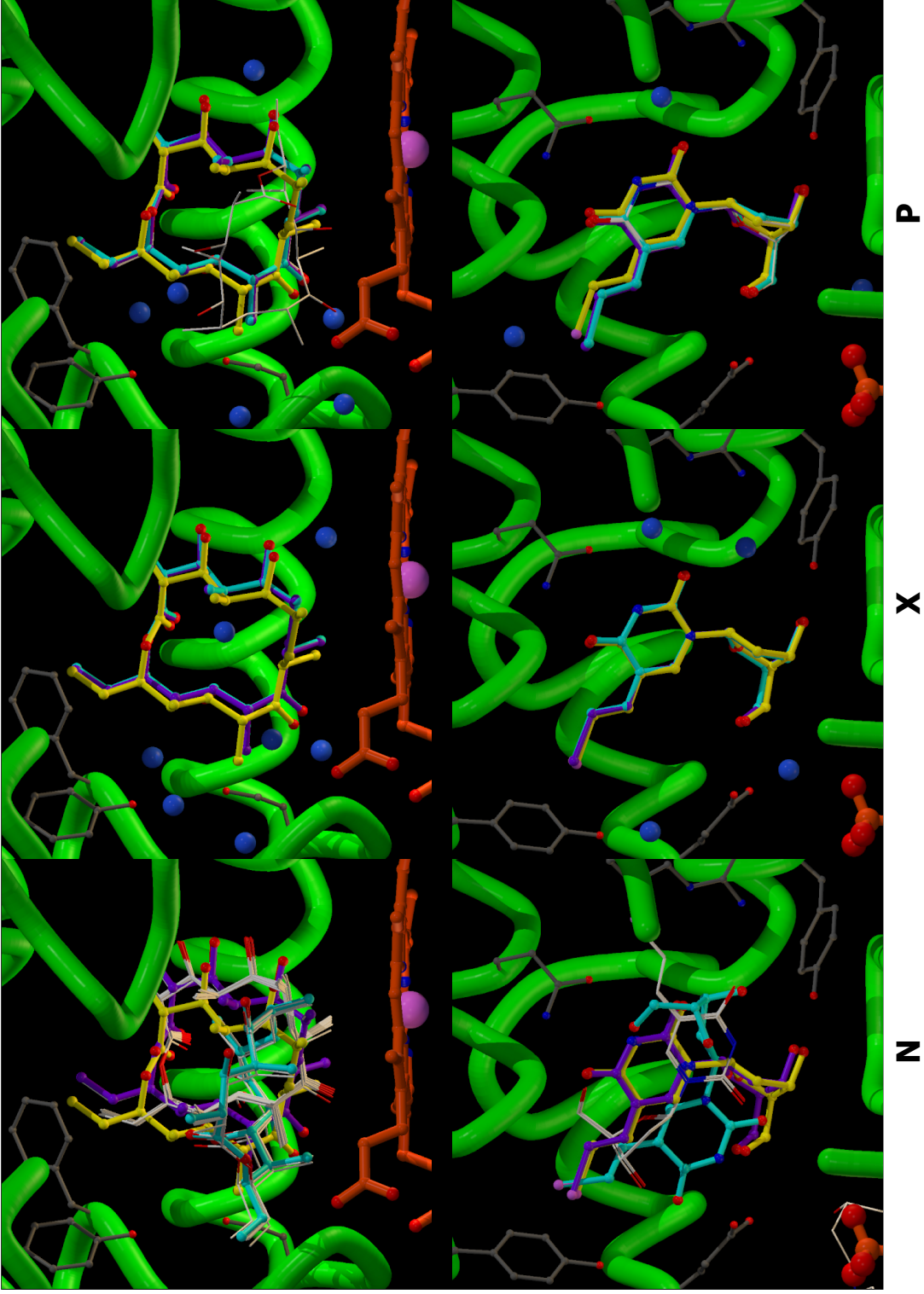
4.3 Discussion

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 CYP 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 CYP, 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[40]. 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 CYP complexes, by mediating protein-ligand interactions and 'fixating' ligand molecules close to the centre of the protein active site (**Table 2**). The present study shows that for five CYP 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 or not 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 CYP-ligand complexes. This might be explained by the fact that the binding pockets of the respective CYP isoenzymes are larger than those of the TK isoenzymes. Concerning CYP 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 CYP-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[41,42] 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 CYP ligands included in this work are small and rigid apolar terpene substrates and phenylimidazole inhibitors covalently bound to the heme Fe-atom of CYP101, as well as larger and more polar macrocycles of CYP107 and Cyt $P450_{EpoK}$, flexible long-chained aliphatic palmitoleic acid substrate of Cyt $P450_{BS\beta}$, and moderately flexible aromatic substrates of CYP2C5. 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 CYPs, 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 CYP101 (**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 sulphaphenazole 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 CYP 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 CYPs[10,20,21,32,42,43], as well as for TK[2,41], be it 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)[44] 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]. In order 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 non-linear 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 modelling methods for this purpose[45].

Rescoring

The CYP and TK docking performance of each of the three 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 CYP. Docking accuracy (RMSD and CSP) was found to be even more improved (in the case of CYP), 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[2,44,46]. 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 a, here 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[47], 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[46]. The present study proved SCORE to be very suitable for this purpose.

4.4 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 (CYPs) 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.

4.5 Computational Methods

Preparation of ligand and target molecules

Ligand input files have been generated with Sybyl 6.8 (TRIPOS Inc., St. Louis). First, ligand structures were extracted from the Protein Databank (pdb) file (containing only non-hydrogen atoms). After proper assignment of Tripos atom types[48], 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 centre of the active site as determined by PASS[49]. 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 CYP and TK enzyme systems. 19 Structures of CYP-ligand complexes were selected from 45 complexes deposited at the Protein databank. The excluded complexes are either variations on CYP-terpene, or CYP-imidazoles, complexes occupying two ligands simultaneously, large covalently bound inhibitors in the CYP protein structure, CYP-alkylcyanide complexes. 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[50]. 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. In order 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 (CYP), cofactor ADP (TK) or sulphate 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

In order 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 x 21.75 x 21.75 $Å^3$ with grid points separated by 0.333 Å, centred on the midpoint of the ligand binding pocket, was automatically hydrated with 25 water molecules (using an energy cut-off 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 centre, were excluded from the docking studies. The positions of the hydrogen

atoms of both crystallized (scenario X) and predicted water (scenario P) molecules were optimised using DOWSER[51]. 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 re-scoring: the process of re-prioritisation 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 CYP and TK, showed that among different "stand-alone" scoring functions (SCORE[11], scoring functions included in the Sybyl CScore module (TRIPOS Inc., St. Louis) and X-SCORE suite[12]), SCORE was either the best one (for CYPs) or very well performing (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 CYPs[44] 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 supplemental material*). For all docking parameters, standard values were used as described for AutoDock[10], FlexX[52] 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[44]. 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 centre determined by PASS is taken as AutoDock affinity grid centre, 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

Two 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[53]. 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[3,21,32,54,55].

- 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 (CYP substrates) or covalent binding (CYP phenylimidazole inhibitors) when they were within 5.5 Å from the CYP 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 nr. 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 nr. 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., CYP: 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:

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

where $\bar{A}$, *I*, and *S* indicate the 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)).

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-Structure Function and Genetics* **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. *Current Science* **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. *Journal of Computational Chemistry* **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. *Journal of Molecular Modeling* **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. *Journal of Computer-Aided Molecular Design* **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. *Journal of Molecular Biology* **2001**, *308*, 377-395.

(15) Schnecke, V.; Kuhn, L. A. Virtual screening with solvation and ligand-induced complementarity. *Perspectives in Drug Discovery and Design* **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-Structure Function and Genetics* **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. *Journal of Medicinal Chemistry* **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. *Journal of Computational Chemistry* **2001***, 22*, 1750-1771.

(25) Clark, D. E.; Grootenhuis, P. D. J. Progress in computational methods for the prediction of ADMET properties. *Current Opinion in Drug Discovery & Development* **2002***, 5*, 382-390.

(26) Hou, T.; Xu, X. 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. *Journal of Computer-Aided Molecular Design* **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) Tang, H. X.; Ye, Y. Z.; Ding, D. F. Flexible Docking of Proteins and "Drug-like" Ligands. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)* **1998***, 30*, 623-630.

(39) Merritt, E. A.; Murphy, M. E. P. Raster3d Version-2.0 - a Program for Photorealistic Molecular Graphics. *Acta Crystallographica Section D-Biological Crystallography* **1994***, 50*, 869-873.

(40) 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.

(41) Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of docking performance: comparative data on docking algorithms. *J Med Chem* **2004***, 47*, 558-565.

144

(42)   Bursulaya, B. D.; Totrov, M.; Abagyan, R.; Brooks, C. L. Comparative study of several algorithms for flexible ligand docking. *Journal of Computer-Aided Molecular Design* **2003**, *17*, 755-763.

(43)   Vigers, G. P.; Rizzi, J. P. Multiple active site corrections for docking and virtual screening. *J Med Chem* **2004**, *47*, 80-89.

(44)   Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J Med Chem* **2003**, *46*, 2287-2303.

(45)   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.

(46)   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.

(47)   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.

(48)   Clark, M.; Cramer, R. D.; Vanopdenbosch, N. Validation of the General-Purpose Tripos 5.2 Force-Field. *Journal of Computational Chemistry* **1989**, *10*, 982-1012.

(49)   Brady, G. P.; Stouten, P. F. W. Fast prediction and visualization of protein binding pockets with PASS. *Journal of Computer-Aided Molecular Design* **2000**, *14*, 383-401.

(50)   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.

(51)   Zhang, L.; Hermans, J. Hydrophilicity of cavities in proteins. *Proteins-Structure Function and Genetics* **1996**, *24*, 433-438.

(52)   Rarey, M.; Kramer, B.; Lengauer, T. Docking of hydrophobic ligands with interaction-based matching algorithms. *Bioinformatics* **1999**, *15*, 243-250.

(53)   Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* **2001**, *7*, 306-317.

(54)   Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* **2000**, *295*, 337-356.

(55)   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. *Journal of Medicinal Chemistry* **2004**, *47*, 45-55.

Chapter 5

CYPs, drugs, and dock'm all

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 (LACDR)/Division of Molecular Toxicology, [§]Division of Pharmaceutical Sciences, Department of Chemistry and Pharmacochemistry, Vrije Universiteit, The Netherlands

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 refined and validated CYP2D6 homology model. Consideration of water molecules at predicted positions in the active site and rescoring of pooled ligand docking results from four different docking programs (AutoDock, FlexX, GOLD-Goldscore, GOLD-Chemscore) with the SCORE scoring function enabled successful predictions of the experimentally reported sites of catalysis of more than 80% of the substrates. Three different database docking algorithms (FlexX, GOLD-Goldscore, GOLD-Chemscore) were subsequently used in combination with six scoring functions (Chemscore, DOCK, FlexX, GOLD, PMF, SCORE) to assess the ability of docking-based virtual screening methods to prioritize 20 known CYP2D6 substrates seeded into a drug-like chemical database of 980 compounds (in the absence and presence of active site water molecules). Finally, the optimal docking strategy in terms of virtual screening accuracy, GOLD-Chemscore with 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 (19.619 entries) and to discriminate between high affinity and medium affinity ligands.

5.1 Introduction

Automated molecular docking has become an important computational method for predicting protein-ligand interactions, guiding lead finding and optimisation in drug discovery.[2,3] 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.[4] 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.[1,5-17] Accuracies of docking (prediction of binding orientation), scoring (prediction of absolute binding free energy) and ranking (discrimination of active from random compounds) of different combinations of docking algorithms and scoring functions still depend on the protein target and the physicochemistry of the protein-ligand interactions.[1,18] This suggests that a docking-scoring strategy should be specifically optimised 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 Cytochrome P450-ligand crystal structures in an earlier study,[20] will be used for binding mode prediction and structure-based virtual screening of substrates of one of the most relevant drug metabolizing Cytochromes P450 (CYP) isoenzymes, being the human CYP2D6. Cytochromes P450 are hemoproteins which catalyse the oxidation and reduction of a wide variety of endogenous and xenobiotic compounds.[21,22] They generally detoxify potentially hazardous compounds, but in a number of cases non-toxic parent compounds are bioactivated into toxic metabolites, or procarcinogens 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 involved in the metabolism of about 30% of the currently marketed drugs.[24,25] Large interindividual differences exist in CYP2D6 activity, due to gene multiplicity and polymorphisms, thus further increasing its clinical importance.[26,27] The early identification of potential CYP2D6 substrates and 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 for 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 binding mode prediction of crystallized CYP-ligand complexes were described, which considered 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 binding mode prediction of CYP-ligand complexes, structure-based virtual screening studies of chemical databases against Cytochromes P450 are scarce. Experimentally determined binding affinities of 11 different CYP101(*cam*)-ligand complexes showed no clear correlation with values from different C-Score scoring functions on CYP101(*cam*)-ligand crystal structures nor on complexes produced by the automated docking program FlexX.[35] Nevertheless, the docking 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 3.508 compound databases, respectively. Very

recently, no significant correlations were found between two training sets of respectively 21 and 30 experimental K_i values and 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 docking scores of a small database of 33 compounds, but nevertheless docking scores could be successfully used to discriminate between weak (IC_{50} > 10 μM) and strong inhibitors (IC_{50} < 10 μM).[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 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 binding mode prediction of Cyt P450-ligand crystal structures,[20] were used for the catalytic site prediction (CSP) of 65 substrates of CYP2D6 (see **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.[1] Docking was performed with four different docking algorithms: AutoDock[39] ('Lamarckian' genetic algorithm, AD), FlexX[40] (incremental construction algorithm, F), and GOLD-GOLD (GG) and GOLD-Chemscore (GC) (in which the GOLD[41] ('Darwinian' genetic algorithm) docking simulation is guided by the Goldscore[41] and the Chemscore[42] scoring function, respectively[14]). The active site of our carefully refined and validated CYP2D6 homology model[30,43,44] was either considered to be 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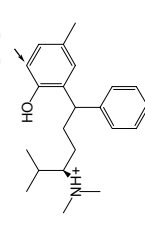 of stand-alone scoring functions (those implemented in the CScore module (Chemscore[42], D-Score[46], F-Score[40], G-Score[41], PMF[47]) and SCORE[48]). Finally, the optimal docking strategy was verified experimentally: it was used to identify high-affinity CYP2D6 ligands among a larger proprietary database (19.619 entries) and to discriminate between high affinity and medium affinity ligands.

5.2 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 (see *Materials and Methods*), was used as the criterium for determining 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.[30,43,44] 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 x 2 water scenarios x 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 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 during docking studies was shown to strongly improve CSP accuracy of all docking-scoring combinations (see **Table 1** and **Figure 1**). CSP accuracies

using FlexX (RAI$_{water}$ = 43%) and AutoDock (30%) were more improved by the consideration of water than the docking performance 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 molecular weight or the number of rotational bonds. Docking results were in many cases sensitive towards small differences in chemical structure of the substrate. This is most clearly examplified by the fact that for all docking programs stereoisomer-dependent (amiflamine, metoprolol, fluoxetine, tolterodine, venlafaxine, carteolol, propranolol, bufuralol, MDMA, MDEA, MDPA, promethazine, citalopram, azelastine), regioisomer-dependent (m- and p-tyramine) and MDMA- and MAMC-derivative-dependent CSP accuracies were observed (see **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 **Figure1** and **Table 1**) as compared to ranking with the scoring function implemented in the respective docking program.
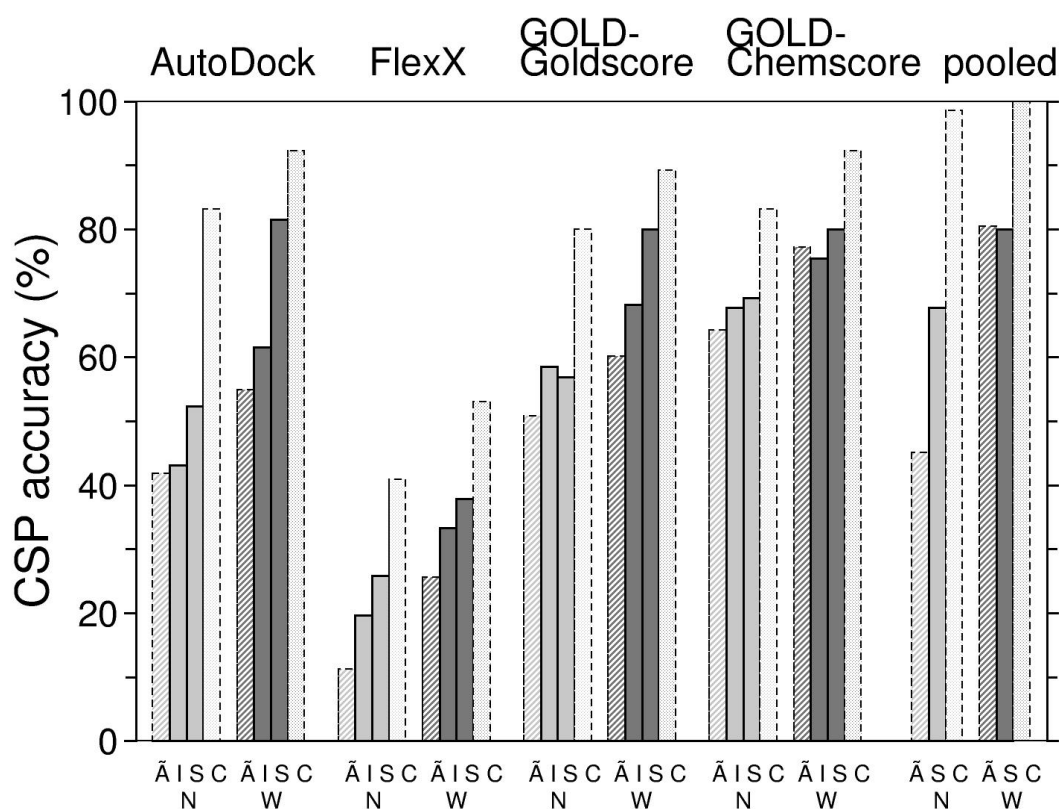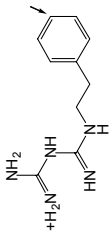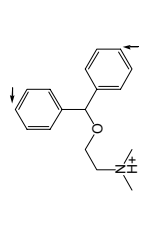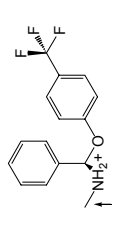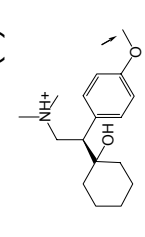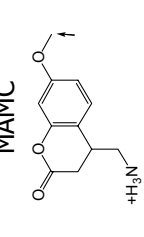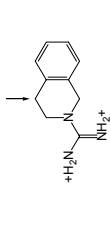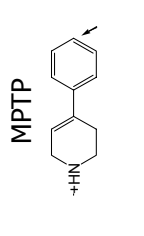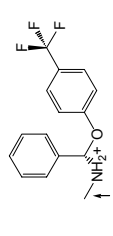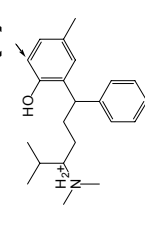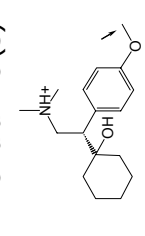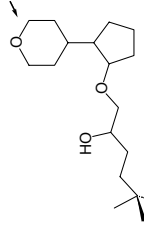
Figure 1: ***Cataltytic site prediction accuracy of different automated docking approaches*** *for 65 CYP2D6 substrates, considering different scenarios with respect to the presence of water (N: no active-site water (light grey); W: water at predicted positions in the active site (dark grey)). Abbreviations on the x-axis correspond to: Ā, Average CSP accuracy for all solutions of each docking study; I, CSP accuracy for nr. 1 ranked solutions according to the scoring function implemented in the docking program; S, CSP accuracy for nr. 1 ranked solutions according to SCORE; C, CSP accuracy for poses accurately predicting the experimentally determined catalytic site, whatever its ranking.*
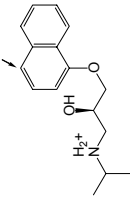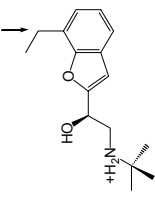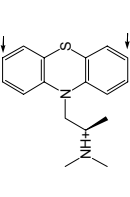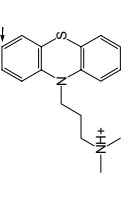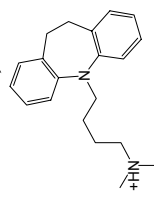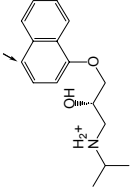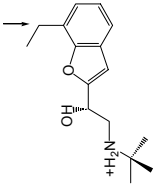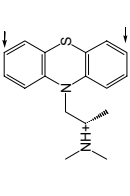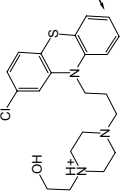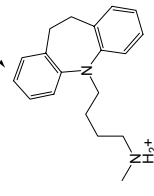
Table 1: Catalytic Site Prediction (CSP) of 65 CYP2D6 substrates. *Letters AD (AutoDock), F (FlexX), GG (GOLD guided by Goldscore), and GC (GOLD guided by Chemscore) indicate the docking programs used, "pooled" indicate that all pooled poses generated by the four docking programs are rescored with SCORE. 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 nr. 1 ranked solutions according to the program implemented scoring functions (superscript 'I'); are only correctly predicted by the nr. 1 ranked solutions according to the stand-alone SCORE scoring function only (superscript 'S'); are correctly predicted by both the nr. 1 ranked solutions according to the program implemented and the nr. 1 ranked solutions according to stand-alone SCORE scoring functions (**bold**); are correctly predicted by any of the docking solutions, but not considered as nr. 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 (*).*

Compound	AD		F		GG		GC		pooled	
m-tyramine	**N**	**W**			N[I]	**W**	**N**	**W**	●[S]	W[S]
2-methoxyphenamine (R)	●	**W**		**W**	**W**	**N**	**N**	**W**	N[S]	W[S]
amiflamine (R)	**N**	**W**	**N**	N[S]	**W**	**W**	**N**	**W**	N[S]	W[S]
mexelitine	**N**	**W**	**N**	**W**	**W**	**N**	**N**	**W**	N[S]	W[S]
metoprolol (R)	**N**	**W**	**N**	**W**	**W**	**N**	**N**	**W**	●[S]	W[S]
p-tyramine	**N**	**W**	●		**N**	**W**	**N**	**W**	N[S]	W[S]
2-methoxyphenamine (S)	N[S]	W[S]	●		**N**	**W**	**N**	**W**	N[S]	W[S]
amiflamine (S)	N[I]	**W**	N[S]	**W**	**N**	**W**	W[S]	**N**	N[S]	W[S]
bunitrolol	●		W[S]	●	●	●	**N**	**W**	N[S]	W[S]
metoprolol (S)	●	**W**	**W**	**W**	**N**	**W**	**N**	**W**	N[S]	W[S]
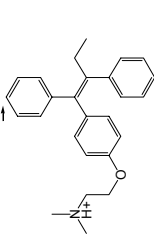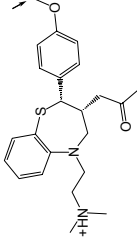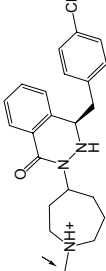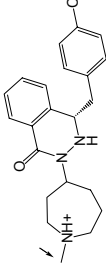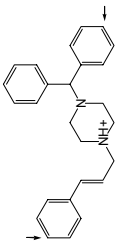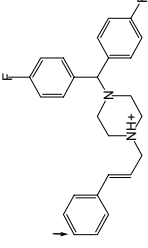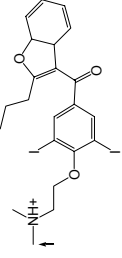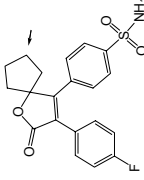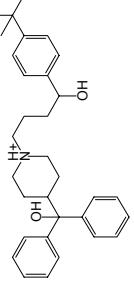
Compound	AD	F	GG	GC	pooled
phenphormine	N W	W^S	W N^S	N^S W	N^S W^S
diphenhydramide*	• W^S	N W^I	• W^S	N •	N^S W^S
fluoxetine (R)	•			W •^{S}	• •^{S} S
tolterodine (R)*			W^S	W^S N	N^S W^S
venlafaxine (R)	• W		• W^S	• W^S	• S W^S
MAMC	N^I W	W N^S W^S		W N W	N N^S W^S

Compound	AD	F	GG	GC	pooled
debrisoquine*	N^I W^S	W	W N^S	N W	•^{S} W^S
MPTP	N W^S	•	W N^I	N W	N^S W^S
fluoxetine (S)	W^S N^I			W	• •^{S} S
tolterodine (S)*	W		W	W W^S	N^S W^S
venlafaxine (S)		N	N^I W^I	N^I W	•^{S} W^S
timolol	• W		N^S W^S	W W^S	•^{S} W^S

Compound	AD	F	GG	GC	pooled
EMAMC*	N^S **W**	•	**N**	**W**	N^S W^S
PMAMC	**N W**	**N** N	**W**	N^S **W**	N^S W^S
MDMA (*R*)	**N W**	**W**	**N** **W**	**N** **W**	N^S W^S
MDEA (*R*)	**N W**	**W**	**N** **W**	**N** **W**	N^S W^S
MDPA (*R*)*	N^S W^S N^S	**W**	N^S **W**	**N** W^I	N^S •^{S}
carteolol (*R*)	W^S	**W**	**W**	**N W**	•^{S} W^S

Compound	AD	F	GG	GC	pooled
diMAMC	**N W**	•	• W^S	N^I **W**	•^{S} •^{S} W^S
BMAMC	**N W**		N^I **W**	N^S **W**	N^S W^S
MDMA (*S*)	**N W**	•	**N W**	**N W**	•^{S} •^{S} W^S
MDEA (*S*)	**N** W^I	• W^S	N^S N^I **W**	N^I **W**^{I}	•^{S} •^{S} •^{S}
MDPA (*S*)*	N^S W^S	•	**N W**	N^S **W**^{I}	•^{S} •^{S} •^{S}
carteolol (*S*)	N^S	**W**	**N W**	**N W**	N^S W^S

Compound	AD	F	GG	GC	pooled
propranolol (R)*	• W	W	W •	N W	N^S W^S
bufuralol (R)*	N^I •	W^S	W •	N W	• ^{S} W^S
promethazine (R)	N^S W	N	N W	N W	N^S W^S
chloropromazine*	N^S W	N W	•	N W	N^S **W^S**
imipramine*	N W	N	W •	W •	N^S W^S

Compound	AD	F	GG	GC	pooled
propranolol (S)*	N^S W	•	•	W	N^S W^S
bufuralol (S)*	N^S W^S	•	•	W	N^S W^S
promethazine (S)	- W		N	W	N^S W^S
pherphenazine*	W	W	N^I W^I	W^S	^{S} ^{S} •
desipramine	N^S W	W	N^S	W	N^S W^S

Compound	AD	F	GG	GC	pooled
nortryptiline*	N **W**	•	N W^S	**W** N^S	**W**^S
mianserine (R)*	N **W**		N **W**	N **W** N^S	N^S W^S
clozapine	• **W**		N **W**	N^I **W** N^S	• ^S N^S W^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 W^S	N^S W^S
citalopram (R)	• W^S			•	• ^S • ^S

Compound	AD	F	GG	GC	pooled
E-doxepin*	• W^S	W^S N^S **W**		N **W** **W**	N^S W^S
mianserine (S)*	•			N **W**	N^S W^S
biperiden	• •	W^S	•		• ^S • ^S
progesterone	N **W**	•	N **W**	N **W** **W**	N^S W^S
codeine	N W^I	N •	N **W**	N **W** **W**	N^S W^S
citalopram (S)	• •	W^S	N^S N	N^I W^S **W** **W**	• ^S • ^S

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

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

Compound	AD	F	GG	GC	pooled
diltiazem	• •	**N**	**N W**	**N N W**	W^S N^S W^S
azelastine (*S*)	•				•^{S}
flunarizine	**W** •	•	**N** •	N^S •	N^S •^{S}
spirosulfonamide	W^I •	•	• •	**N W** •^{S}	•^{S} W^S W^S

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) vs. 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) vs. 33% (*F-F*) to 82% (*AD-S and GC-S*) with predicted water molecules.
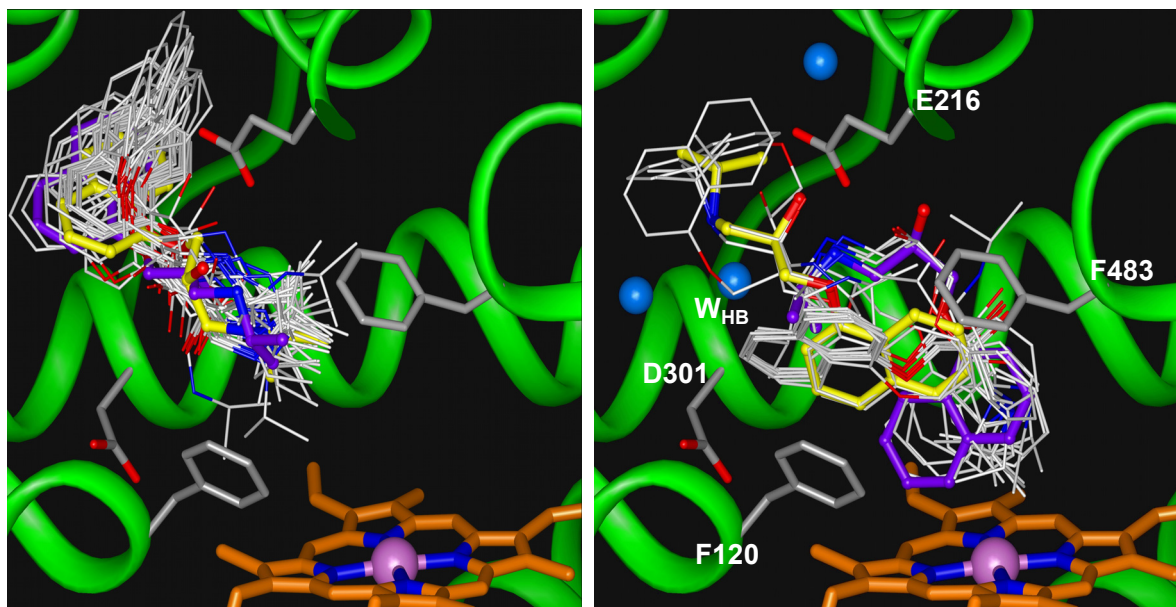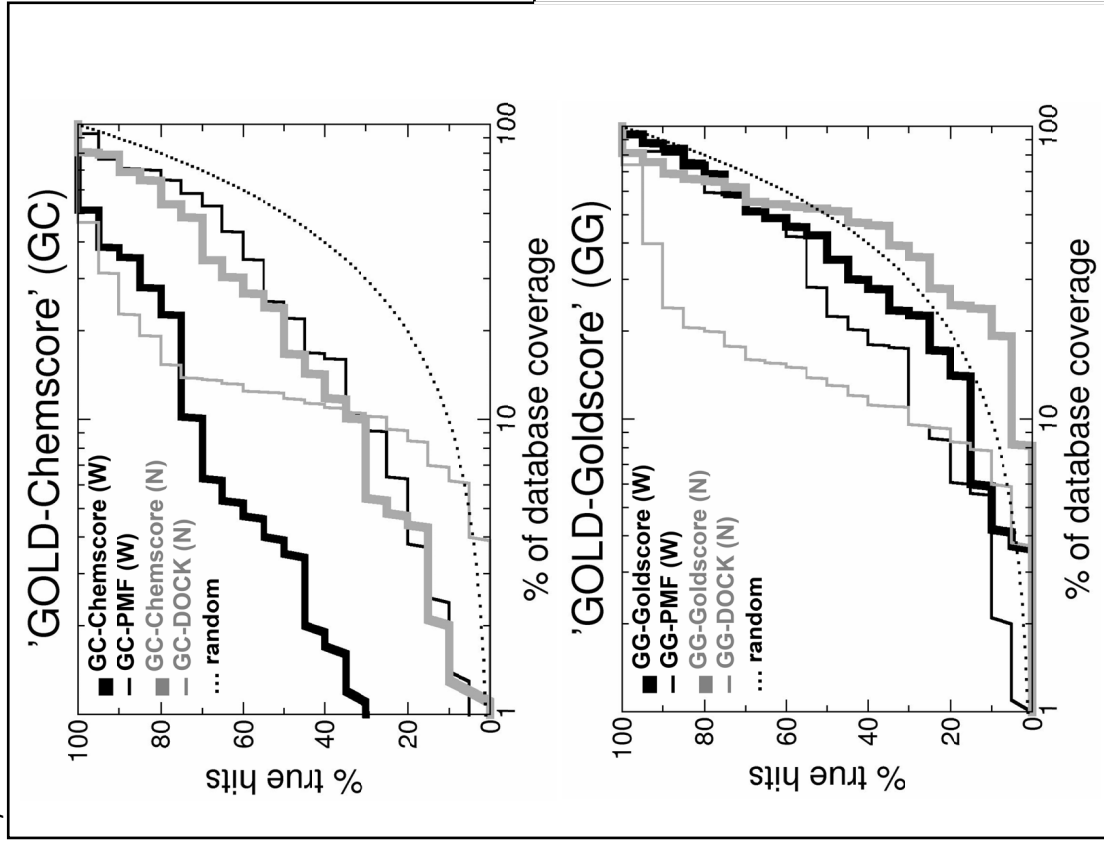

Figure 2: *R-propranolol docked in the binding pocket of CYP2D6, using FlexX in the absence (left panel) and presence (right panel) of active site water. Orientations of 50 docking solutions (white) ranked as nr. 1 with the scoring function of FlexX (yellow) and SCORE (purple) are compared. Amino acid residues involved in substrate binding (F120, E216, F483, D301) are shown. Water oxygen atoms are depicted in blue and the active-site water molecule observed to mediate protein-ligand hydrogen bond is indicated (W_{HB}).*

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 (see *Materials and Methods*) of 20 known CYP2D6 substrates (see **Table 1**) among 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, SCORE), with and without the consideration of active site water molecules, yielding in total 3 x 6 x 2 = 36 different docking strategies. AutoDock was not applied in our virtual screening studies as 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**. Results for all docking-scoring combinations are available as supplementary material. 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 consideration of active site water, the virtual screening accuracy of FlexX docking resulted in hit rates and yields of only 12% and 30%, respectively (considering top 5% ranked compounds).

Docking program	Scoring function	Water scenario	Top 2.5% r_H	y	Top 5% r_H	y	Top 10% r_H	y
FlexX	FlexX	N	0	0	0	0	0	0
	DOCK[a]	N	0	0	4	10	14	70
	FlexX	W	0	0	2	5	5	25
	DOCK[a]	W	0	0	12	30	15	75
GOLD-Goldscore	Goldscore	N	0	0	0	0	0	0
	DOCK[a]	N	0	0	2	5	6	30
	Goldscore	W	0	0	4	10	3	15
	PMF[a]	W	8	10	4	10	6	30
GOLD-Chemscore	Chemscore	N	12	15	10	25	7	35
	Goldscore[a]	N	0	0	4	10	3	15
	Chemscore	W	36	45	24	60	15	75
	PMF[a]	W	12	15	8	20	7	35

[a]As implemented in Cscore

FlexX (F)

F-FlexX (W)
F-DOCK (W)
F-FlexX (N)
F-Goldscore (N)
random

% true hits — % of database coverage

'GOLD-Chemscore' (GC)

GC-Chemscore (W)
GC-PMF (W)
GC-Chemscore (N)
GC-DOCK (N)
random

% true hits — % of database coverage

'GOLD-Goldscore' (GG)

GG-Goldscore (W)
GG-PMF (W)
GG-Goldscore (N)
GG-DOCK (N)
random

% true hits — % of database coverage
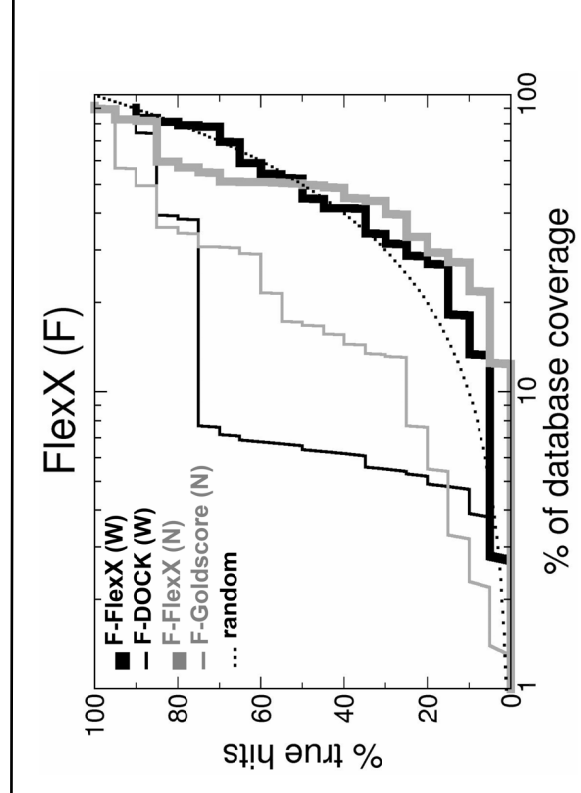
The virtual screening accuracy of GOLD-Goldscore docking was even lower, resulting in a maximum yield and hit rate of 4% and 10%, respectively, including active-site water and using either the Goldscore or PMF scoring function (at top 5%). GOLD-Chemscore docking in combination with the Chemscore scoring 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 5.760 structures was enriched by generating stereo-isomers 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[30,43,44,49] (see **Figure 4**).
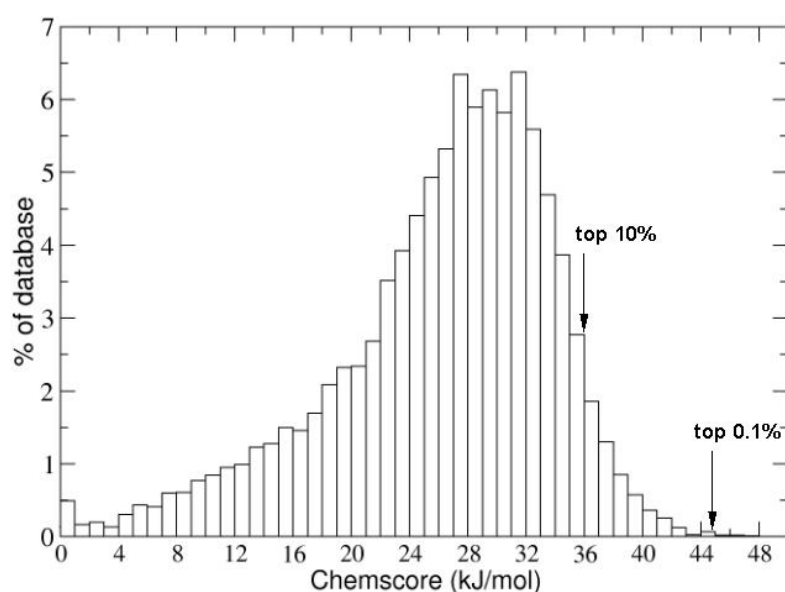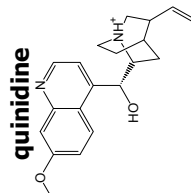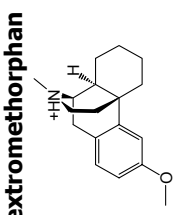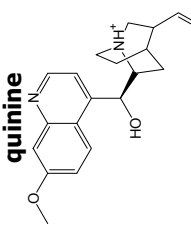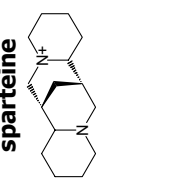
Figure 4: *Distribution of Chemscore scores of the 19.619 compound proprietary database. The experimentally tested samples (of eight compounds) at the top 0.1% (~45-48 kJ/mol) and top 10% (~36 kJ/mol) of the database, shown in Tables 4-5, are indicated with arrows.*

The affinities (as reflected by $IC_{50}'s$) of four typical CYP2D6 inhibitors and substrates (**Table 3**), and eight top 0.1% (**Table 4**) and eight top 10% (**Table 5**) ranked compounds were determined experimentally. Stereo-isomers (of 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% 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), two comparable to that of sparteine (10-100 µM), while 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 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% ranked scorers were 63% and 13% respectively.
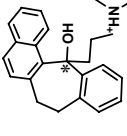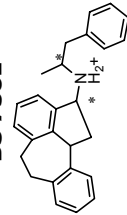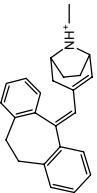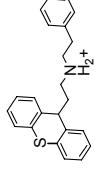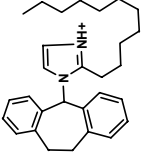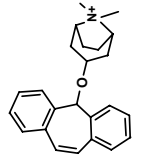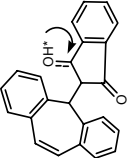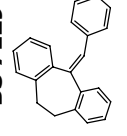
Table 3: *Affinities of known CYP2D6 inhibitors, measured as their IC50 in MAMC O-demethylation.*

compound	IC$_{50}$ (µM)	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
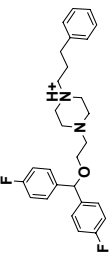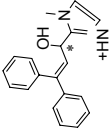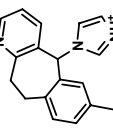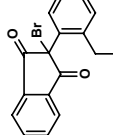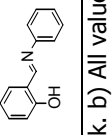
All values are the means of at least three independent experiments ± standard deviations as described in the *Materials and Methods* section.

Table 4: *CYP2D6 affinities of compounds ranked at the top 0.1% of virtual screening studies (see figure 4), measured as their IC50 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.

Table 5: *CYP2D6 affinities of compounds ranked at the top 10% of virtual screening studies (see figure 4), measured as their IC50 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

5.3 Discussion

The primary aim of the present study was to find optimal docking strategies for catalytic site 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 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,[50-52] 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,[39,48,53,54] but only few docking programs allow for explicit water-mediated interactions between proteins and ligands during the docking simulation.[7,55-57] Furthermore, only a few docking studies are reported which evaluate the effects of water molecules at specific locations in the 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. The current study shows that water molecules placed on energetically favorable locations in the CYP2D6 binding pocket, improve the CSP and virtual screening accuracy 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 (see **Figure 2**). In an earlier paper 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 accomodate ligands of variable size and topology.[20] Linked to this are increased computational efforts and the necessary use of scoring functions for comparison 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 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.[1,18] The set of 65 CYP2D6 substrates (among which 32 stereo-isomers) included in this work has a broad chemical and structural diversity (see **Table 1**). The substrates contain up to 4 hydrogen-bond acceptors, up to 5 hydrogen bond donors, and have 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 molecular weight or the number of rotational bonds. Docking results were in many cases sensitive towards small differences in 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 M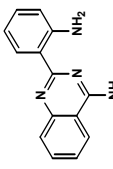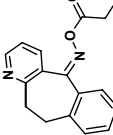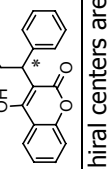DMA- and MAMC-derivative-dependent CSP accuracy. This indicates that very subtle differences may play a crucial role in accurately predicting the site of metabolism. It should be noted that most experimental CYP2D6 metabolism studies have only reported results on racemic mixtures. Ligand descriptors 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 (see **Table 1** and **Figure 1**). Rescoring all 'pooled' AutoDock, FlexX, GOLD-Goldscore and GOLD-Chemscore 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.[1,67,68] Comparing the overall CSP accuracy of the optimal docking strategies for each docking algorithm (including active-site water and rescoring with SCORE), AutoDock is shown to be 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,[1,9,18] while 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 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 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 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 physico-chemical properties similar to the actives.[15] For the training of different docking-based virtual screening methods, we used the Rognan database,[1] because this database is used in previous comparative docking studies.[1,9] 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, as this is not likely the case in a physical screen 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.[1,5,6,8,9,12,15,69] The best docking-based virtual screening strategy, i.e. GOLD-Chemscore with the Chemscore scoring function and including active-site water, was found to be superior to all other strategies.

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 is 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 actually the only false-positively screened hit (IC_{50} > 200 μM) among the top 0.1% ranked scorers. It should be noticed that compound BS7129 does not contain a positively charged nitrogen atom. Therefore a combined CYP2D6 2D/3D-pharmacophore-search[70] and structure-based virtual screening might exclude this

false positive from the top ranked hits. On the other hand, compound GBR 30111 does not have a positively charged nitrogen atom either, but it does contain a hydrogen bond donor which forms a hydrogen bond with E216 (like *R*-propranolol in **Figure 2**) and does show affinity for CYP2D6 (IC$_{50}$ value of 60 μM). This illustrates that docking-based virtual screening methods might enable the discovery of leads with chemical features and structures dissimilar from known actives, thus extending the possibilities of drug design. False-positive hits nevertheless may highlight particular weaknesses of a docking scoring function.[69] All top 0.1% scorers were predicted to be binding to the CYP2D6 binding pocket in about the same binding mode, with their tricyclic dibenzo moiety stacked between the F120 and F483 and (with the exception of BS7129) forming an electrostatic interaction with the negatively charged carboxylic group of E216. These amino acid residues were previously found to be key determinants in CYP2D6-mediated metabolism in site-directed mutagenesis studies.[44,49,71-73] BS7129 was found to have a tight steric fit in the CYP2D6 binding pocket, expressed by a high lipophilic score (E$_{lip}$) and relatively low protein-ligand clash (E$_{clash}$) and ligand-internal energy (E$_{int}$) scores in the GOLD-Chemscore docking score.[14]

Two of the eight hits sampled at the top 10% scorers contained a positively charged nitrogen atom, GB12819 (IC$_{50}$ = 1.21 μM) and BS 9086 (IC$_{50}$ = 40 μM), via which chemical group these compounds were found to interact with E216 in the docked CYP2D6-ligand complex structure. The first compound is actually the only high affinity (<10 μM[38]) CYP2D6 inhibitor found in this sample. The top 10% 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 performance.[1,69] 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, while 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, 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 accuracy (82%, 80% and 80%, respectively). AutoDock was not applied in our virtual screening studies, as 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, this CSP and virtual screening strategy was 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 accuracy of docking-scoring combinations are often shown to vary considerably with the selected target protein, physicochemical details

of target-ligand interactions[1,5,6,13,15,16] 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 based on a training set of ligand-bound protein crystal structures. In the absence of such a training set (which is the case for CYP2D6), other sorts of experimental data, like 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[43,44,49]) can be used to validate docking strategies.

5.4 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 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 CSP accuracy and virtual screening accuracy 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 docking program and 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.

5.5 Methods

Preparation of CYP2D6 substrate and protein input structures
A set of 65 known CYP2D6 substrates (including 32 stereoisomers) was selected for this study, based on the availability of experimental information on CYP2D6-catalysed product formation and affinity for CYP2D6 (K_m < 200 μM). The protein homology model of CYP2D6 was constructed, refined and validated as described in.[30,43,44] Also the preparation of protein and substrate input structures, definition of the binding pocket, and the GRID-based[45] prediction of energetically favorable positions of active-site water molecules was performed as described earlier.[20] The substrate dextromethorpan was used to determine the ligand-based cut-off for the exclusion of water molecules too close to the active-site centre.

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] As scoring is a very important second aspect of automated docking methodologies, it was decided to investigate the effect of re-scoring: the process of re-prioritisation of 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 re-ranking the 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 percentage of docking solutions with binding modes corresponding to experimentally determined major biotransformation products. Ligand atoms were considered to be potential sites of oxidation when they were within 6.0 Å from the CYP2D6 heme Fe-atom.

CSP accuracies are presented as:
- Average percentages of successful catalytic site predictions for all solutions (50) of each docking study (illustrating the chance of finding a reliable solution) ($\bar{A}$).
- CSP accuracy for nr. 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).
- CSP accuracy for nr. 1 ranked solutions according to the scoring function SCORE (reflecting the ability of SCORE to properly rank poses after the docking procedure) (S).
- 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).

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

$$RI_{score} = (X_S - X_I)/X_I$$

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

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 three-dimensional chemical databases

For the evaluation of the virtual screening performance of different docking strategies a database was prepared consisting of 980 randomly selected compounds from a 3D database (in mol2 format) of 1.000 drug like molecules previously described[1] and 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 5.760 compounds was converted into 3D structures (in mol2 format) using the program MOE (v. 2004. Chemical Computing Group:. Montreal, Canada). Subsequently, structures were washed (filtering counter ions and solvents), energy minimized and MOE was used for the generation of all possible stereoisomers for the database (yielding 9.706 entries). The database was tautomerised using AGENT2.0.[79] AGENT generated approximately two tautomers per compound in the database (yielding 19.619 entries in total). For some compounds, AGENT did not generate any tautomer, for others 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 virtual screening of the two databases (respectively the combined database of 980 drug-like molecules and 20 known CYP2D6 substrates and the proprietary database of in total 19.619 molecules). 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], PMF[47]) and the scoring function SCORE (as part of a spl-script used by Bissantz et al.[1]). It should be

noted that FlexX scores calculated either from FlexX or Cscore are very similar,[1] while Chemscore and GOLD scores calculated by the GOLD program differ from those calculated by Sybyl, and thus cannot be compared.[80] 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 accuracy 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 in 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 Prof. Dr. Ingelman-Sundberg. *E. coli* JM109 were 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.[44] 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 before as described.[81] Dextromethorphan hydrobromide, debrisoquine sulfate and quinidine sulfate dihydrate were obtained from Sigma (St Louis, MD, USA). All other chemicals were of analytical grade and obtained from standard suppliers.

MAMC *O*-demethylation reactions by CYP2D6 were carried out in 96 wells plates, in a total volume of 200 µl.[44] 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 inhibitor. The inhibitors were dissolved in DMSO and stored as 20 to 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 fluorescence of the samples was subsequently measured on a Victor[2] 1420 multilabel counter (Wallac, Oy, Finland) using λ_{ex}= 405nm, λ_{em}= 460nm.

References

(1) 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.

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

(3) Congreve, M.; Murray, C. W.; Blundell, T. L. Keynote review: Structural biology and drug discovery. *Drug Discovery Today* **2005**, *10*, 895-907.

(4) 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.

(5) Bursulaya, B. D.; Totrov, M.; Abagyan, R.; Brooks, C. L., 3rd 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. *Current Science* **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 Schmiedebergs 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) Vermeulen, N. P. E. Prediction of drug metabolism: the case of cytochrome P450 2D6. *Curr Top Med Chem* **2003**, *3*, 1227-1239.

(33) Snyder, R.; Sangar, R.; Wang, J. B.; Ekins, S. Three-dimensional quantitative structure activity relationship for CYP2D6 substrates. *Quantitative Structure-Activity Relationships* **2002**, *21*, 357-368.

(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 Comp 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.; de Graaf, C.; de Kanter, F. J. J.; Oostenbrink, C.; Feenstra, K. A. et al. Metabolic regio- and stereoselectivity of Cytpchrome P450 2D6 towards 3,4-methylenedioxy-N-alkyl-amphetamines: In silico Predictions and experimental validation. *J. Med. Chem.* **2005**, *in press*.

(44) 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.

(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 (Online)* **1998**, *4*, 379-394.

(49) 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.

(50) 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.

(51) 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.

(52) 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.

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

(54) 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.

(55) Rarey, M.; Kramer, B.; Lengauer, T. The particle concept: placing discrete water molecules during protein-ligand docking predictions. *Proteins* **1999**, *34*, 17-28.

(56) Schnecke, V.; Kuhn, L. A. Virtual screening with solvation and ligand-induced complementarity. *Perspectives in Drug Discovery and Design* **2000**, *20*, 171-190.

(57) 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.

(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-Structure Function and Genetics* **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 Transduct Res* **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., 3rd 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 Graph Model* **2002**, *20*, 281-295.

(68)   Wang, R.; Lu, Y.; Wang, S. 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. *Chemical Physics Letters* **2004**, *390*, 500-505.

(79)   Pospisil, P.; Ballmer, P.; Scapozza, L.; Folkers, G. Tautomerism in computer-aided drug design. *J Recept Signal Transduct Res* **2003**, *23*, 361-371.

(80)   Cole, J. C.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. Comparing protein-ligand docking programs is difficult. *Proteins* **2005**.

(81)   Onderwater, R. C.; Venhorst, J.; Commandeur, J. N.; Vermeulen, N. P. 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.

Supplementary Data

Table I: *Validation of virtual screening strategies for the selection of high-affinity CYP2D6 substrates. Description of CYP2D6 hit lists generated by three docking programs in combination with the program implemented scoring function and stand-alone scoring functions (at top 5%) and 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 Methods section.*

Docking program	Scoring function	Water scenario	Top 2.5%		Top 5%		Top 10%	
			r_H (%)	y (%)	r_H (%)	y (%)	r_H (%)	y (%)
FlexX	Chemscore[a]	N	0	0	0	0	2	10
	DOCK[a]	N	0	0	4	10	14	70
	FlexX[a]	N	0	0	0	0	0	0
	Goldscore	N	8	10	6	15	5	25
	PMF[a]	N	4	5	4	10	2	10
	SCORE	N	0	0	0	0	1	5
	Chemscore[a]	W	0	0	2	5	5	25
	DOCK[a]	W	0	0	12	30	15	75
	FlexX	W	0	0	2	5	5	25
	Goldscore	W	4	5	6	15	7	35
	PMF[a]	W	0	0	2	5	2	10
	SCORE	W	0	0	0	0	0	0
GOLD-Goldscore	Chemscore[a]	N	4	5	2	5	1	5
	DOCK[a]	N	0	0	2	5	6	30
	FlexX[a]	N	0	0	2	5	1	5
	Goldscore	N	4	5	2	5	1	5
	PMF[a]	N	0	0	0	0	0	0
	SCORE	N	0	0	0	0	0	0
	Chemscore[a]	W	0	0	0	0	2	10
	DOCK[a]	W	0	0	2	5	6	30
	FlexX[a]	W	0	0	0	0	2	10
	Goldscore	W	0	0	4	10	3	15
	PMF[a]	W	8	10	4	10	6	30
	SCORE	W	0	0	0	0	0	0
GOLD-Chemscore	Chemscore	N	12	15	10	25	7	35
	DOCK[a]	N	0	0	2	5	6	30
	FlexX[a]	N	0	0	0	0	0	0
	Goldscore[a]	N	0	0	4	10	3	15
	PMF[a]	N	0	0	0	0	0	0
	SCORE	N	0	0	0	0	0	0
	Chemscore	W	36	45	24	60	15	75
	DOCK[a]	W	0	0	0	0	5	25
	FlexX[a]	W	0	0	0	0	3	15
	Goldscore[a]	W	0	0	2	5	1	5
	PMF[a]	W	12	15	8	20	7	35
	SCORE	*W*	*0*	*0*	*2*	*5*	*1*	*5*

[a]As implemented in Cscore

Dynamics of CYP2D6-ligand interactions

Chapter 6

Dynamic dancing with XTC

Metabolic regio- and stereoselectivity of cytochrome P450 2D6 towards 3,4-methylenedioxy-N-alkyl-amphetamines: in silico predictions and experimental validation

Chris de Graaf,[‡*] Peter H.J. Keizers,[‡*] Frans J.J. de Kanter,[§] Chris Oostenbrink,[‡] K. Anton Feenstra,[‡] Jan N.M. Commandeur,[‡] and Nico P.E. Vermeulen[‡*]

[‡]Leiden Amsterdam Center for Drug Research (LACDR)/Division of Molecular Toxicology, [§]Division of Organic and Inorganic Chemistry, Department of Chemistry and Pharmacochemistry, Vrije Universiteit, The Netherlands
[*]Both authors contributed equally

A series of 3,4-methylenedioxy-N-alkylamphetamines (MDAAs) were automatically docked and subjected to molecular dynamics (MD) simulations in a Cytochrome P450 2D6 (CYP2D6) protein model. The predicted substrate binding orientations, sites of oxidation and relative reactivities were compared to experimental data of wild-type and F120A mutant CYP2D6. Automated docking results were not sufficient to accurately rationalize experimental binding orientations of 3,4-methylenedioxy-N-methylamphetamine (MDMA) in the two enzymes as measured with spin lattice relaxation NMR. Nevertheless, the docking results could be used as starting structures for MD simulations. Predicted binding orientations of MDMA and sites of oxidation of the MDAAs derived from MD simulations matched well with the experimental data. It appeared that the experimental results were best described in MD simulations considering the nitrogen-atoms of the MDAAs in neutral form. Differences in regioselectivity and stereoselectivity in the oxidative metabolism of the MDAAs by the F120A mutant CYP2D6 were correctly predicted and the effects of the F120A mutation could be rationalized as well.

6.1 Introduction

Cytochromes P450 (CYPs) are heme containing enzymes which can be found in virtually all living organisms. This large family of enzymes is capable of oxidizing and reducing a broad range of endogenous and exogenous substrates, such as steroids, carcinogens and drugs.[1,2] In humans one of the most relevant drug metabolizing CYPs is CYP2D6. Although the expression levels of CYP2D6 are only 2% of all hepatic CYPs, following CYP3A4, it is the second most important drug metabolizing enzyme, involved in the metabolism of about 30% of the currently marketed drugs, including β-blockers, neuroleptics, antidepressants and antiarythmics.[3-5] Large interindividual differences exist in CYP2D6 activity, due to gene multiplicity and polymorphisms, thus further increasing its clinical importance.[6,7] The rationalization and prediction of potential CYP2D6 substrates is therefore advantageous in the discovery and development of new drugs. Nowadays the first crystal structures of mammalian CYPs are becoming available.[8-10] However, so far a crystal structure of CYP2D6 remains to be resolved and any structural information on this enzyme still depends on homology modeling, mutagenesis and spectroscopic studies.

CYP2D6 is one of the CYP isoforms studied most extensively using molecular modeling.[11] Several homology model structures of CYP2D6 have been built, refined and validated experimentally. The homology models of CYP2D6 suggest substrates to interact with two or three aromatic/hydrophobic residues, F120 and Phe[483] and a carboxylic acid residue, E216 or D301.[12,13] The relevance of these residues was supported by site-directed mutagenesis studies,[14-18] and these interactions are consistent as well with those derived from pharmacophore models of inhibitors and substrates.[19,20] Molecular modeling can provide information on active site characteristics and the importance of specific amino acid residues in enzyme-substrate interactions, but it can also be used to rationalize and predict regio- and stereoselectivity in metabolism by CYPs. It has recently also been shown that automated docking can successfully be applied to predict sites of oxidation in substrates using CYP crystal structures.[21] MD simulations in addition can account for distributions of multiple binding conformations and thus give a more comprehensive description of multiple sites of oxidation in substrates catalyzed by CYPs.[22-25]

In this study a molecular modeling approach has been used to study the binding orientation and the sites of oxidation of a series of 3,4-methylenedioxy-N-alkylamphetamines (MDAAs), or XTC analogues, by CYP2D6. The primary aim was to evaluate an integrated molecular modeling approach to rationalize and predict substrate binding and metabolism by CYP2D6. When regioselectivity and stereoselectivity in drug metabolism and the effect of active site mutations can be rationalized for a series of closely related compounds, such an approach can be considered as accurate and reliable. The molecular modeling predictions were based on a homology model of CYP2D6. Reactive substrate poses were generated using automated docking and were obtained from MD simulations of wild-type and mutant CYP2D6 protein containing all three MDAAs. Two different protonation states of the basic nitrogen-atom were studied, as it is known that this can drastically influence the oxidation by CYP2D6.[26] The computational predictions were validated experimentally. The effects of mutating F120 into alanine on the binding modes and metabolic products were studied for five MDAAs, i.e. 3,4-methylenedioxy-N-methylamphetamine (MDMA) and its stereoisomers, 3,4-methylenedioxy-N-ethylamphetamine (MDEA) and 3,4-methylenedioxy-N-propylamphetamine (MDPA). For the recently described F120A mutant large effects on the regioselective oxidation of MDMA were demonstrated.[16] Dissociation constants for the three MDAAs were determined by spectral titration. The metabolic products were analyzed and furthermore, the binding orientations of MDMA in the wild-type and mutant active sites were explored experimentally by NMR spin lattice relaxation rate measurements, a technique previously used to determine hydrogen atom to heme distances in CYPs.[27-30]

176

6.2 Results

Expression and Purification of Wild-type and F120A Mutant CYP2D6

The expression levels of both wild-type and F120A mutant CYP2D6 were similar as described previously.[16] The enzymes which contained a C-terminal 6 x histidine tag were purified using Ni-NTA-agarose. Eluting the enzymes from the Ni-NTA-agarose using imidazole led to a large amount of enzyme in the high spin state, even after overnight dialysis high spin remained the predominant state. Therefore 0,2 mM *L*-histidine was preferred as eluting agent over imidazole, because it did not cause a change in spin state after dialysis yielding pure enzymes which were stable for at least four hours at 24° C in the presence of 5% glycerol and 100 μM of substrate as assessed by CO-difference spectroscopy. Hardly any P420 (i.e. inactive CYP detected as an absorbance maximum at 420 nm in the CO-difference spectrum) was present in the batches of purified enzyme.

Spectral Titration of Substrates to the Enzymes

The addition of all MDAAs to purified wild-type or F120A mutant CYP2D6 led to a type I change in the visible absorbance spectrum. By titration of the compounds to the solutions of enzyme the spectral dissociation constants K_s could be estimated (using Equation 1, data shown in **Table 1**). While there were minor differences in affinity between the wild-type and the mutant, for both enzymes a slightly lower K_s was found for MDEA and MDPA than for MDMA. The enantiomers of MDMA had about equal affinity for the wild-type, but a slight preference towards *S*-MDMA was observed for the mutant enzyme.

Table 1: *Estimated dissociation constants, K_s, (μM) of the substrates for wild-type and F120A mutant CYP2D6.*

Substrate	Wild-type K_s	F120A K_s
R/S-MDMA	28 ± 3	28 ± 2
R-MDMA	43 ± 6	53 ± 2
S-MDMA	40 ± 5	38 ± 4
MDEA	19 ± 2	15 ± 4
MDPA	18 ± 1	12 ± 2

All values are the means ± S.D. of at least two independent experiments as described in the Materials and Methods. *R/S*-MDMA refers to the racemic mixture of 3,4-methylenedioxy-N-methylamphetamine (MDMA), R-MDMA and S-MDMA to the pure *R*- and *S*-isomer of MDMA respectively. For the corresponding N-ethyl and N-propyl amphetamines (MDEA and MDPA) racemic mixtures were used.

Metabolism of MDAAs

As reported previously, [16] MDMA is oxidized by the F120A mutant to MDA and N-OH-MDMA, in addition to 3,4-OH-MA, the only product formed by the wild-type enzyme (**Figure 1**). Elongation of the N-methyl chain to an ethyl or propyl yielded the corresponding products after incubation with both enzymes. MDEA was O-demethylenated by wild-type and mutant CYP2D6 and additionally the mutant N-dealkylated and N-hydroxylated MDEA. A fourth oxidation product of MDEA by the mutant was detected at t_R 18.5 min, with a *m/z* of 222, fragmenting to *m/z* 163. This product could be the aldehyde- or keto-form of ω- or ω-1-hydroxylation respectively.
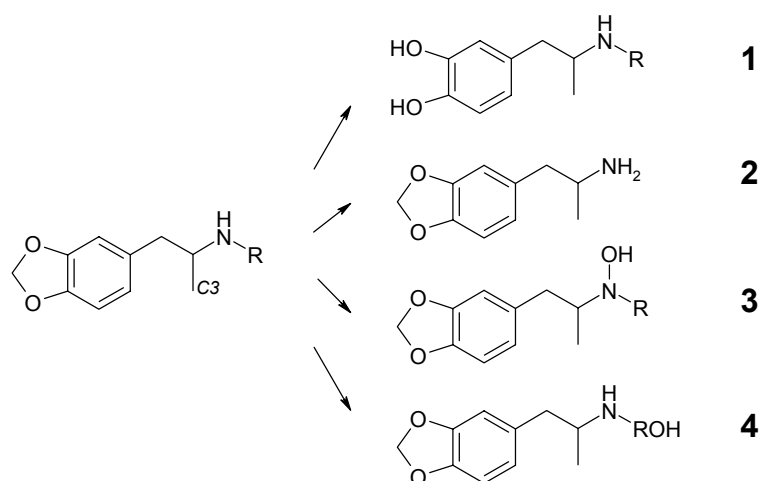
Figure 1: *Scheme of 3,4-methylenedioxy-N-alkylamphetamines (MDAAs) oxidation by F120A mutant CYP2D6. The MDAAs are O-demethylenated to 3,4,-dihydroxy-N-alkylamphetamine (1), N-dealkylated to 3,4-methylenedioxyamphetamine (2), N-hydroxylated to 3,4-methylenedioxy-N-hydroxy-N-alkylamphetamine (3) or ω/ω-1-hydroxylated to 3,4-methylenedioxy-ω/ω-1-hydroxy-N-alkylamphetamine (4). The C3 methyl is indicated, R is methyl, ethyl and propyl for MDMA, MDEA and MDPA respectively.*

MDPA was O-demethylenated by wild-type and the mutant CYP2D6, the latter enzyme also N-dealkylated and N-hydroxylated this substrate. A peak with *m/z* of 254 was found at 19.2 min with the mutant, indicating a double hydroxylation. No fragmentation took place so the positions of the hydroxyl groups remain unknown. Product formation from all substrates was linear for at least 10 min for both enzymes under the conditions used, so the kinetic parameters were determined for the major products (**Table 2**).

Table 2: *Product formation of MDAAs oxidation by wild-type and F120A mutant CYP2D6.*

Substrate	Product	Structure	Wild-type		F120A	
			K_m	V_{max}	K_m	V_{max}
R/S-MDMA	3,4-OH-MA	1	1.9 ± 0.5	1.2 ± 0.4[a]	11.8 ± 3.8	2.2 ± 0.0[a]
	MDA	2	-[c]	-[c]	14.5 ± 2.4	3.5 ± 0.1[b]
	N-OH-MDMA	3	-[c]	-[c]	11.2 ± 1.4	7.1 ± 0.5[a]
R-MDMA	3,4-OH-MA	1	2.7 ± 0.4	1.2 ± 0.2[a]	13.2 ± 0.6	2.4 ± 0.1[a]
	MDA	2	-[c]	-[c]	26.1 ± 3.3	5.9 ± 0.7[b]
	N-OH-MDMA	3	-[c]	-[c]	20.6 ± 2.5	10.7 ± 1.3[a]
S-MDMA	3,4-OH-MA	1	3.3 ± 0.5	1.2 ± 0.0[a]	4.9 ± 0.1	2.4 ± 0.0[a]
	MDA	2	-[c]	-[c]	10.3 ± 0.6	3.8 ± 0.1[b]
	N-OH-MDMA	3	-[c]	-[c]	6.9 ± 0.6	8.4 ± 0.4[a]
MDEA	3,4-OH-EA	1	1.1 ± 0.1	1.2 ± 0.2[a]	3.3 ± 0.0	2.9 ± 0.2[a]
	MDA	2	-[c]	-[c]	4.5 ± 0.8	4.2 ± 0.1[b]
	N-OH-MDEA	3	-[c]	-[c]	5.7 ± 0.3	1.4 ± 0.1[a]
MDPA	3,4-OH-PA	1	1.0 ± 0.1	1.3 ± 0.1[a]	5.2 ± 0.6	7.5 ± 0.6[a]
	MDA	2	-[c]	-[c]	9.1 ± 0.5	7.3 ± 1.8[b]
	N-OH-MDPA	3	-[c]	-[c]	28.0 ± 6.5	3.0 ± 0.2[a]

All values are the means of at least two independent experiments $\pm$ S.D. as described under Materials and Methods. Structure numbers refer to products drawn in Figure 1, K_m expressed in μM, [a] V_{max} expressed in 1 x 10^5 fluorescence units/min/nmol CYP, [b] V_{max} expressed in min^{-1}, [c] not observed.

The overall rate of oxidation by the F120A mutant was higher than that of wild-type CYP2D6 for all MDAAs. The rates for O-demethylenation were two to five-fold higher for the mutant and the fact that more products are formed indicate an even higher increase in oxidation rates. To quantify the turnover rates, synthetic reference compound was only available for MDA, for the other products turnover rates were only compared in relative terms. Because the fluorescence peak areas of equal concentrations of MDA, MDMA, MDEA and MDPA are equal, it is likely that the N-alkyl chain has no influence on the fluorescence yield of the compounds, and therefore reasonable this holds true for the O-demethylenated and N-hydroxylated products of the MDAAs as well.

The turnover rates of MDMA, MDEA and MDPA were equal for wild-type CYP2D6 (**Table 2**). The catalytic efficiency (V_{max}/K_m) indicated that MDPA is oxidized most efficient and MDMA the least efficient (**Figure 2A**). The mutant showed a higher catalytic efficiency in O-demethylenating MDPA over MDMA and MDEA, while the efficiency in N-hydroxylation was the highest for MDMA (**Figure 2B-D**). MDEA and MDPA were N-dealkylated more efficiently than MDMA. So elongation of the N-alkyl chain leads to an altered catalytic regioselectivity by the mutant enzyme.
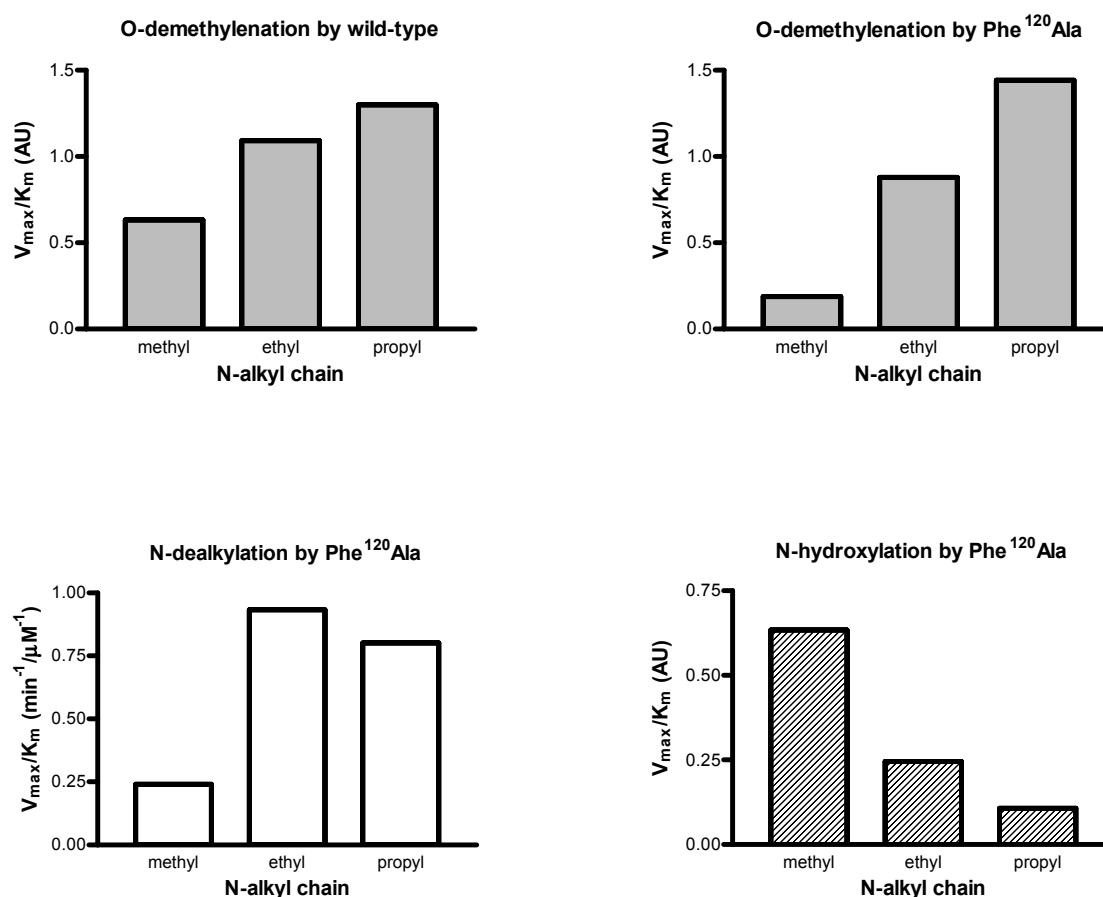
Figure 2: V_{max} / K_m product ratio's ($min^{-1} \cdot \mu M^{1}$ for N-dealkylation, AU for the other products) for the O-demethylation by wild-type (grey left), and O-demethylation (grey right), N-dealkylation (blank) and N-hydroxylation (diagonal stripe) by F120A mutant CYP2D6 of the different MDAAs.
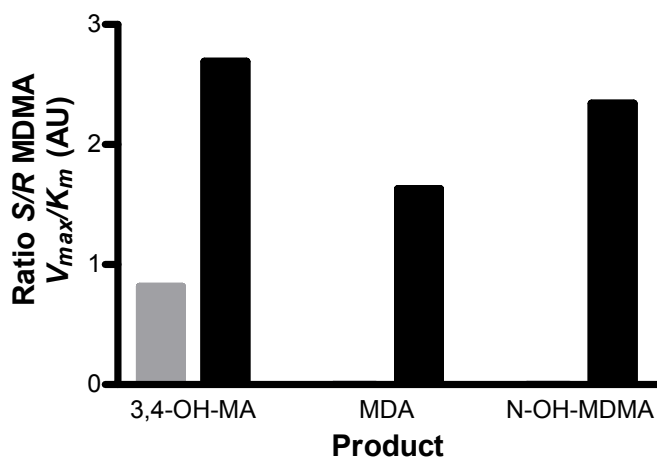
Figure 3: *Ratio's of S-MDMA catalytic efficiency (V_{max} / K_m) over R-MDMA catalytic efficiency (V_{max}/K_m) by wild-type (in gray), and by F120A mutant CYP2D6 (in black) for the products 3,4,-dihydroxy-N-alkylamphetamine (3,4-OH MA), 3,4-methylenedioxyamphetamine (MDA) and 3,4-methylenedioxy-N-hydroxy-N-alkylamphetamine (N-OH-MDMA).*

Stereoselectivity in MDMA oxidation by wild-type CYP2D6 was hardly found (**Figure 3**). However, the mutant enzyme showed at least two-fold lower K_m values for all three products of *S*-MDMA when compared to *R*-MDMA, while the V_{max} values for the products of both enantiomers did not differ (**Table 2**). Eventually these changes lead to a different product ratio for the enantiomers: O-demethylenation is more prominent for *S*-MDMA then for *R*-MDMA, followed by N-hydroxylation and N-dealkylation (**Figure 3**).

NMR Spin Lattice Relaxation Rate Measurements

Enzyme bound MDMA hydrogen atom to heme iron distances were determined by spin lattice relaxation NMR to validate the computationally predicted distances. The ^{1}H-NMR spectrum of MDMA in presence of glycerol and each of the two enzyme was well resolved, all signals of the different hydrogen atoms were clearly visible (**Figure 4**). Because the doublet of the *C3* methyl (δ 1.18), and the singlets of the N-methyl (δ 2.61), and methylene (δ 5.89) hydrogen atoms could be best quantified and are the most distal groups in MDMA, these were used to determine the substrate orientation in the active site of the two enzymes (inversion recovery spectra shown in Supplementary Data). An increase in T_l of the MDMA hydrogen atoms was found after reducing the enzymes with dithionite, indicating that the paramagnetic effect of the enzymes was diminished. From the T_l the distances from the methylene, *C3* and N-methyl hydrogen atoms to the heme iron atom were calculated (using Equation 2) and tabulated (in **Table 3**). The fast exchange condition (see Methods section) was validated by measuring the temperature dependence of R_{IP} from 281° K to 307° K (see Supplementary data). For the three groups of hydrogen atoms studied, there was a linear increase of R_{IP} with the reciprocal temperature (1/T), showing that the exchange between substrate molecules in the enzyme active site and in solution is fast on the NMR timescale. Only minor differences were observed for the hydrogen atoms to iron distances, between the enantiomers of MDMA and each of the two enzymes. The measured hydrogen atoms to iron distances indicate that the methylene moiety is closest to the heme iron, at about 6.4 Å, the N-methyl is the most distal to the heme at about 7.3 Å and the *C3* hydrogen atoms are at a distance of about 6.7 Å.
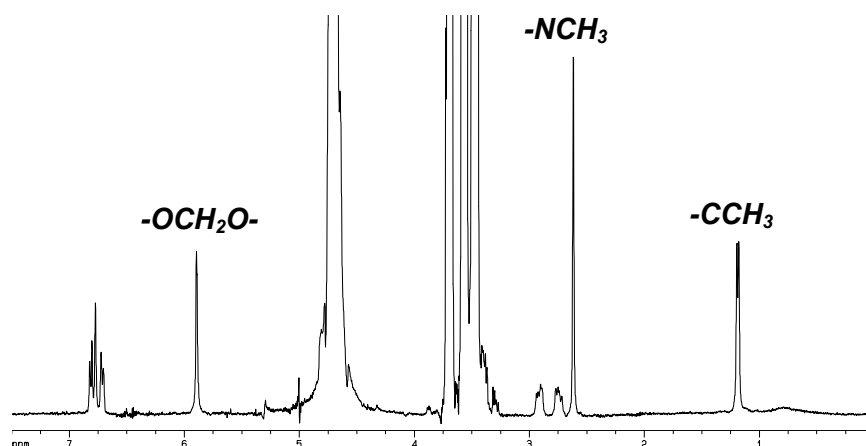
Figure 4: ^{1}H NMR spectrum of 20 mM R-MDMA in deuterated Kpi-glycerol in presence of 5 µM F120A mutant CYP2D6 showing the resonance assignments.

Table 3: Hydrogen atom to heme iron distances (Å) of the MDMA enantiomers to wild-type and F120A mutant CYP2D6 as determined by NMR spin lattice relaxation rate measurements and by molecular dynamics (MD) simulations of neutral and charged MDMA.

Position		wild-type		F120A	
		R-MDMA	*S*-MDMA	*R*-MDMA	*S*-MDMA
Methylene (δ 5.89)	NMR	6.4	6.3	6.4	6.5
	MD (charged)	3.5	3.6	4.6	3.1
	MD (neutral)	4.3	4.5	5.0	5.5
N-methyl (δ 2.61)	NMR	7.3	7.2	7.5	7.2
	MD (charged)	11.0	10.1	10.1	9.4
	MD (neutral)	5.9	6.5	8.4	6.0
C3 (δ 1.18)	NMR	6.9	6.7	6.6	6.7
	MD (charged)	11.8	11.0	8.7	11.5
	MD (neutral)	5.3	6.6	6.9	6.5

NMR derived values are the means of at least two independent measurements with S.D. less then 5%, as described in the Materials and Methods. MD simulations derived values are averaged over time and for three individual runs, starting from different docking poses, according to equation 4.

Automated Docking Studies

All automatically docked poses of the MDAAs, in two protonation states and in both enzymes were found to be predominantly within a reactive distance of 6 Å from the heme iron atom (69% and higher, **Table 4A**). Of these reactive poses, the large majority corresponded to O-demethylenation (62-100%). The relative occurrence of docking poses corresponding to N-dealkylation (0-38%) and N-hydroxylation (0-13%) of the MDAAs was found to be enzyme, stereoisomer, and protonation state dependent. In both enzymes, the neutral forms of the compounds showed the widest range of reactive substrate orientations, whereas the wild-type allowed for more diversity. This observation is in contrast with the experimental finding that wild-type CYP2D6 only catalyzes O-demethylenation, while the F120A mutant allows for alternative product formation. Three distinct, energetically most favorable automatically docked poses were selected of all MDAAs as starting conformations for MD calculations, an example of charged *R*-MDMA being shown (**Figure 5**). These three distinct docked poses could be generally described as one binding orientation corresponding to O-demethylenation, with the substrate nitrogen atom in close contact with E216 (pose 1, generated by AutoDock), another corresponding to O-demethylenation, with the substrate nitrogen atom placed between E216 and D301 (pose 2, generated by GOLD) and one corresponding to N-dealkylation or N-hydroxylation (pose 3, generated by AutoDock). In cases where pose 3 was not

observed for the substrate in the charged form (see **Table 4A**), the corresponding pose of the neutral state was protonated and taken as starting conformation for MD simulations.
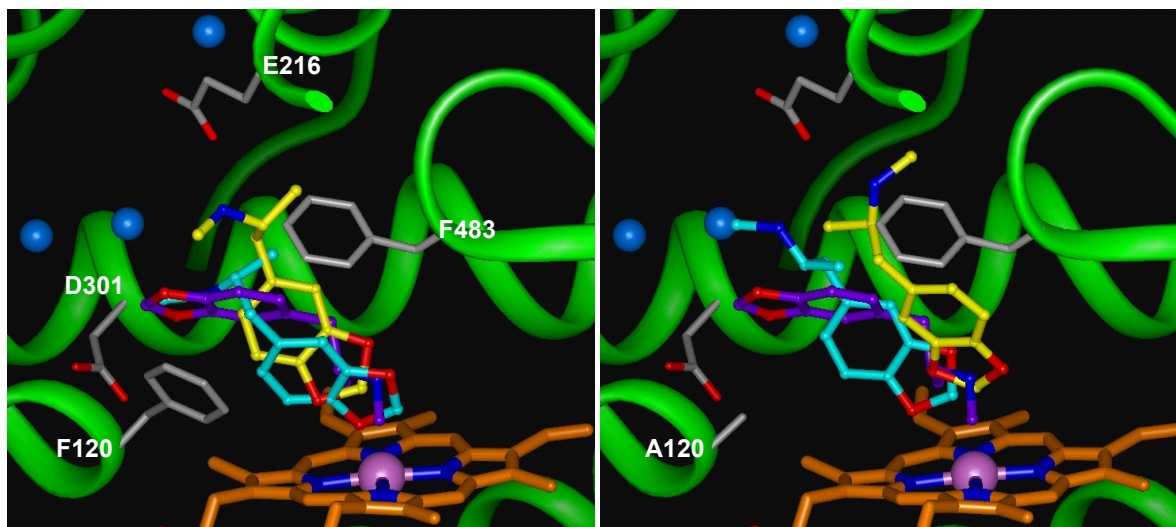
Figure 5: *Molecular dynamics (MD) simulations starting orientations of charged R-MDMA, generated by automated docking studies in wild-type (A) and F120A mutant (B) CYP2D6 homology models. Two orientations are corresponding to demethylenation, with the substrate nitrogen-atom either in close contact with E216 (pose 1, in yellow) or placed between E216 and D301 (pose 2, in cyan). A third orientation is corresponding to N-demethylation or N-hydroxylation (pose 3, in purple). Water oxygen atoms, as predicted by GRID (see Materials and Methods section), are depicted in blue.*

MD Simulations

Simulations of the enzymes with substrates bound in the active site, starting from the three different automatically docked poses, remained stable during 10 ns unrestraint MD at 300° K, with final atom-positional root-mean-square deviation values of 1.3 to 3.8 Å. Throughout the simulations structure determining hydrogen-bonds were observed and the majority of the orientations of the substrates in the enzymes were within a reactive distance to the iron atom (74% and higher, **Table 4B**). The substrates appeared to have considerable freedom within the binding site. Complete reorientation from the initial coordinates was frequently observed, indicating that the unrestrained MD simulations sampled the substrate conformational space sufficiently. The distances between substrate hydrogen atoms and the heme iron atom of wild-type and F120A mutant CYP2D6 were derived for two different protonation states of the MDMA enantiomers, by averaging the distances over the three simulations (using Equation 4) and these were subsequently compared to experimentally determined distances (**Table 3**). Relatively large differences in averaged hydrogen atom to iron distances were found within individual MD simulations starting from different substrate docking poses. However, averaged distances of individual MD simulations of the same protonation state and starting from the same docking pose, were approximately equal. An exception was the difference in averaged distance of the methylenedioxy hydrogen atoms of both neutral MDMA enantiomers in simulations starting from pose 3 (corresponding to N-demethylation/hydroxylation) in wild-type CYP2D6 (~ 7.5 Å) compared to the F120A mutant (~ 11.0 Å).

Table 4: *Distributions of substrate orientations corresponding to specific sites of oxidation (in %), according to automated docking studies (A) and MD simulations (B).*

A

Substrate			reactive	wild-type ODM	NDM	NOH	ωOH	reactive	F120A ODM	NDM	NOH	ωOH
MDMA	R	charged	100	100	0	0	-	100	100	0	0	-
		neutral	88	86	10	4	-	93	90	3	7	-
	S	charged	92	100	0	0	-	100	100	0	0	
		neutral	99	73	14	13	-	80	99	0	1	-
MDEA	R	charged	91	99	1	0	0	100	99	1	0	0
		neutral	100	91	9	0	0	100	98	2	0	0
	S	charged	69	99	1	0	0	100	99	1	0	0
		neutral	99	78	22	0	0	100	93	7	0	0
MDPA	R	charged	70	79	21	0	0	99	99	1	0	0
		neutral	92	62	38	0	0	100	97	3	0	0
	S	charged	74	100	0	0	0	95	62	38	0	0
		neutral	82	99	1	0	0	100	73	27	0	0

B

Substrate			reactive	wild-type ODM	NDM	NOH	ωOH	reactive	F120A ODM	NDM	NOH	ωOH
MDMA	R	charged	97	100	0	0	-	89	100	0	0	-
		neutral	77	83	3	12	-	96	69	3	28	-
	S	charged	99	100	0	0	-	100	100	0	0	-
		neutral	80	88	1	11	-	92	64	23	13	-
MDEA	R	charged	74	99	0	0	1	64	99	0	0	1
		neutral	95	85	2	2	11	93	71	18	9	2
	S	charged	100	100	0	0	0	76	100	0	0	0
		neutral	82	92	2	6	0	79	61	32	5	2
MDPA	R	charged	97	100	0	0	0	99	100	0	0	0
		neutral	97	100	0	0	0	100	58	22	1	20
	S	charged	100	100	0	0	0	98	100	0	0	0
		neutral	100	100	0	0	0	98	99	1	0	0

The columns 'reactive' indicate the percentage of all binding orientations that displayed a site of oxidation within 6 Å of the heme iron. The subsequent columns show the distribution of the reactive conformations for the different sites of oxidation, corresponding to O-demethylenation (ODM), N-demethylation (NDM), N-hydroxylation (NOH), and 'omega-hydroxylation' (ωOH). The automated docking studies are based on 50 independent AutoDock and GOLD docking runs, the MD simulations are Boltzmann weighted averages over three independent simulations, starting from three different docking poses. Substrates were simulated in neutral and charged form.

Also the averaged distances derived from all three simulations of individual protonation states were approximately the same for the different MDMA enantiomers in wild-type and F120A mutant CYP2D6. The averaged hydrogen atom to heme iron distances of the neutral MDMA enantiomers agreed well with the distances determined by NMR spin lattice relaxation rate measurements. Apparently, considering the substrates to be neutral describes the experimental situation best. With the exception of the N-methyl and *C3* hydrogen atoms of *R*-MDMA in the F120A mutant, all hydrogen atom to heme iron distances calculated from MD simulations of neutral MDMA enantiomers were shorter than the experimentally derived distances. The methylenedioxy hydrogen atom to heme iron distances were slightly lower than those of the N-methyl and *C3* hydrogen atoms. In contrast, the averaged N-methyl and *C3* hydrogen atom to heme iron distances of the charged MDMA enantiomers violated the experimental distances consistently.

For all MDAAs, differences between the averaged enzyme-substrate interaction energies calculated using SCORE, were within 3 kJ/mol (~kT) for all three MD simulations of the same protonation states and within 5 kJ/mol for all six MD simulations. The relative weights of the MD simulations of the three starting orientations used in the Boltzmann averaging (Equation 5, see Methods section) over the simulations ranged between 0.11 and 0.54.

Table 5: *Enzyme-substrate interactions (%) during MD simulations.*

Substrate		Wild-type		F120A	
		VdW/arom.F120	H-bond E216	VdW/arom. A120	H-bond E216
MDMA	*R*	81	65	12	54
	S	80	69	23	47
MDEA	*R*	55	51	2	46
	S	66	70	1	57
MDPA	*R*	71	74	5	52
	S	81	52	7	61

Enzyme-substrate interactions during the MD simulations were monitored in terms of the occurrence of atom-atom distances of 3.5 Å or less between the F/A120 side-chain and the MDAA ring system, representing aromatic/van der Waals interactions and hydrogen bonds between the E216 carboxylate oxygen atoms and the MDAA nitrogen-atom. Values of neutral and charged substrates were averaged.

Already during the equilibration time, or soon after the start of the MD simulations with charged MDAAs, reorientation of pose 3 into pose 1 (corresponding to O-demethylenation) occurred. In the case of MDEA, this led to non-reactive binding orientations outside of the binding pocket, forming a hydrogen bond between its protonated nitrogen and the carboxylate group of E216 (**Table 4B** and **Figure 5**). Relatively high probabilities of O-demethylenation poses were found for neutral MDAAs (>58%), although orientations corresponding to N-demethylation/hydroxylation were also observed, except for MDPA in wild-type CYP2D6. N-demethylation or N-hydroxylation of neutral MDAAs had significant higher probabilities in the F120A mutant than in wild-type CYP2D6. Furthermore, substrate dependent catalytic regioselectivity was observed for the mutant: elongation of the N-alkyl chain led to a decreased N-hydroxylation efficiency with an optimal N-dealkylation efficiency for MDEA (**Table 4B**). These trends were equal to those found experimentally and became even more pronounced when the MDAAs were averaged over enantiomers, over protonation states and Boltzmann weighted (**Figure 6**).

Significant stereoselective differences in the probabilities of O-demethylenation vs. N-demethylation/hydroxylation were not observed for wild-type CYP2D6. The F120A mutant however, did in all cases discriminate the enantiomers of the neutral MDAAs. Differences between the probabilities of N-demethylation and N-hydroxylation of MDMA enantiomers were found and *R*-MDPA was found in orientations leading to N-demethylation and ϖ-hydroxylation more often than *S*-MDPA (**Table 4B**).

A

B
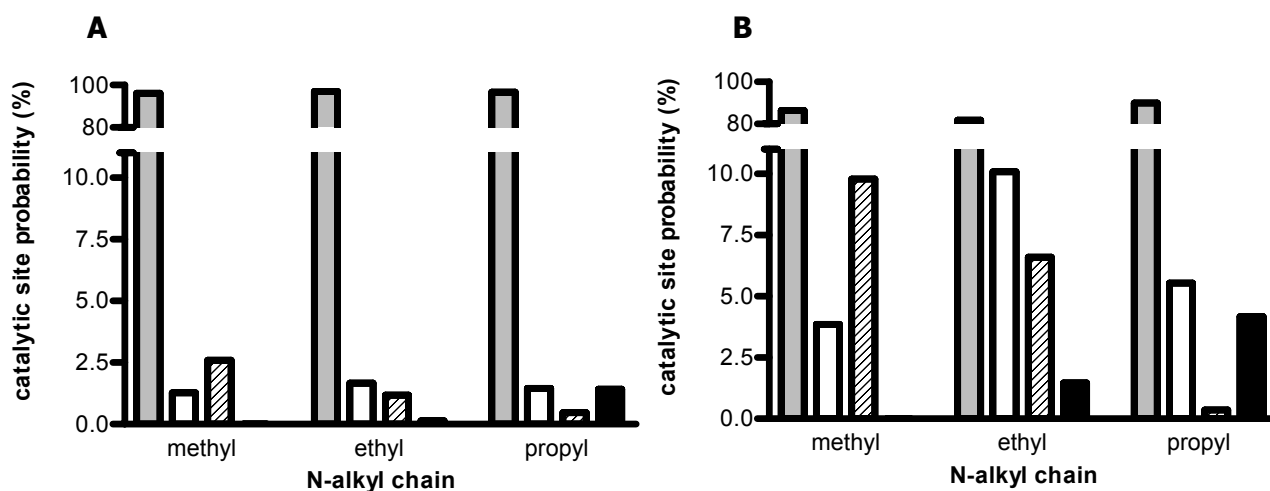
***Figure 6**: Probabilities (%) of reactive binding orientations corresponding to O-demethylenation (grey), N-dealkylation (blank), N-hydroxylation (diagonal stripe), and ω-hydroxylation (black), of the MDAAs in wild-type (A) and in F120A mutant (B) CYP2D6, as derived from 12 MD simulations. Shown are the Boltzmann weighted averages of two different protonation states and three different docking poses (examples in Figure 5), and assuming a 1:1 ratio between R- and S-enantiomers.*

Enzyme-substrate interactions observed in the MD simulations of neutral MDAAs showed a significant decrease in van der Waals/aromatic interactions with residue 120 in the F120A mutant compared to wild-type CYP2D6 (**Table 5**). Furthermore, a small decrease in hydrogen-bond interactions with the carboxylic acid of E216 was observed for the F120A mutant as a direct result of the relatively high probability of binding orientations corresponding to N-dealkylation or N-hydroxylation. In both wild-type and F120A mutant CYP2D6 however, hydrogen-bond interactions were found to be significantly decreased for neutral MDAAs (up to 23% on average for each enantiomer) when compared to charged MDAAs (up to 92% on average for each enantiomer). During the MD-simulations, water molecules were primarily observed to fill the part of the active site which is connected with a so called "water channel", [31] which was already partially occupied by predicted water molecules in the automated docking starting structures (**Figure 5**). These water molecules were found to form H-bonds with the substrate. At most, one water molecule was observed in other regions of the active site.

6.3 Discussion

The primary aim of this study was to develop an integrated molecular modeling approach to analyze substrate oxidation by CYP2D6. When regioselectivity and stereoselectivity in metabolic oxidation and the effect of enzyme active site mutations can be visualized and quantified using a series of related compounds, such an approach could be a useful tool in the rationalization and prediction of metabolism. Using this approach the experimentally determined binding modes and the product formation of a series of MDAAs by wild-type and F120A mutant CYP2D6 could be rationalized.

Substrate Binding Orientations

In the present study, molecular modeling was used to explain the apparent discrepancy between experimentally observed differences in regioselective oxidation and the stereoselective preference by the F120A mutant towards MDMA (**Figure 1)** and the average substrate orientation in the mutant and wild-type active sites as determined by NMR spin lattice relaxation rate measurements (**Table 3**). The average hydrogen atom to iron distances calculated from the T_1 of the MDMA enantiomers in both enzymes were equal, with the methylene moiety in both cases closest to the heme iron atom. Although

the distances between the hydrogen atoms at potential sites of oxidation in MDMA and the heme iron atom were somewhat larger than comparable distances found in CYP X-ray structures (4-6 Å) they were within the ranges of distances determined in other NMR studies.[27-30] This difference can be explained by the fact that experimentally determined distances are averages of all possible substrate orientations during access to, residence in and exit from the active site. As a result they contain more information then just that of a reactive substrate orientation. NMR spin lattice relaxation rate measurements have been used before to determine the orientation of codeine in CYP2D6, where the measured hydrogen atoms to iron atom matched the orientation expected from codeine's metabolic profile.[28] Codeine is a more rigid and bulky substrate than MDMA which may explain the good match. Caffeine, a small and low affinity substrate of CYP1A2 was also subject of NMR spin lattice relaxation rate measurements, but the orientation found of caffeine did not match the metabolic profile.[30] From other experimental studies it is known that substrates can be quite mobile in CYP active sites,[32] and the existence of multiple substrate binding modes and substrate mobility in CYP2D6 is clearly supported by our molecular modeling studies. Averaged distances derived from the present studies, considering three distinct automatically docked substrate binding modes of the MDAAs in combination with long MD simulations, were in good agreement with the experimentally determined metabolic profiles when considering the MDAAs to be in their neutral form (**Table 3**). Still, this does not explain the experimentally observed change in regioselective oxidation and stereoselective preference of the F120A mutant towards MDMA compared to wild-type. So it is likely that a small substrate like MDMA is very mobile in the CYP2D6 active site and that the reactivity of possible sites of oxidation determines also what products are being formed.

Regioselectivity and Stereoselectivity in Oxidation

Experimentally observed trends in regioselectivity of the oxidation of the MDAAs by wild-type and F120A mutant CYP2D6 (**Table 2**) were in good agreement with the relative probabilities of different binding modes of the MDAAs observed in the MD simulations (**Table 4B**). In the wild-type enzyme, only O-demethylenation was observed experimentally, while in the F120A mutant also N-demethylation and N-hydroxylation occurred. The molecular modeling studies showed that N-demethylation/hydroxylation probabilities for neutral MDAAs were significantly higher in F120A mutant than in the wild-type enzyme (**Table 4B**). Both the experimental and modeling studies showed a decreased N-hydroxylation efficiency with elongation of the N-alkyl chain and the highest MDA formation efficiency for MDEA. Stereoselectivity in oxidation of MDMA by wild-type CYP2D6 was hardly found experimentally neither predicted by molecular modeling. However, the experimentally observed preference for *S*-MDMA by the F120A mutant could not be rationalized by molecular modeling. This may be due to the fact that the stereoselectivity is predominantly caused by a difference in K_m and not in V_{max}, while the affinities of substrates to enzymes were not predicted in a rigorous manner in this study. However, the small violations of the calculated to the experimentally determined hydrogen atom to iron distances for neutral *R*-MDMA bound to the F120A mutant do indicate that this enantiomer experiences more difficulties to find its optimum orientation in the binding pocket.

Influence of F120A Mutation on Oxidation Reactions

By removing the phenyl ring of F120 more space is created in the active site of CYP2D6 and a specific aromatic interaction point is eliminated.[16] From the MD simulations it became clear that this space was not occupied by the MDAAs, nor by water molecules, but that critical substrate interactions were lost after mutating F120 into an alanine (**Table 5**). This observation and the fact that the stereo- and regioselectivity of the mutant enzyme

differed from wild-type shows that F120 is important for specific interactions with the MDAAs in CYP2D6. There is not sufficient space in the active sites of wild-type or the F120A mutant CYP2D6 to accommodate two molecules of the MDAAs at the same time. However, it could be possible that two substrate molecules bind simultaneously to different parts of the protein e.g. in the active site and in substrate access channels,[31] or in a peripheral binding site.[9]

Combination of Automated Docking and MD simulations

In earlier studies it was shown that automated docking can successfully be applied to predict sites of oxidation by CYPs.[21] In contrast, the current study shows that automated docking alone is not suitable for accurate determination of relative probabilities of different substrate binding modes and for discrimination between substrates and enzymes of high similarity, such as regio/stereoisomers and mutants of enzymes. In order to reproduce the experimental data, which are an average over many different substrate and enzyme orientations in time, a dynamic treatment of both substrate and enzyme is required. Only when averaging over different positions and orientations was taken into account, experimentally determined substrate hydrogen atoms to heme iron atom distances could be reproduced accurately. The experimentally observed differences in substrate oxidation by wild-type and F120A mutant CYP2D6 could only be explained from the different binding modes observed in the MD simulations. Multiple and extensive MD simulations are needed to catch subtle differences between substrates, enzyme structures and their dynamics.

Substrate Protonation States.

The molecular modeling was performed with neutral and charged substrates because this markedly influenced the binding and mobility of the MDAAs in the active site of the two enzymes. Even though the MDAAs in physiological solution will occur predominantly in their charged form, the preferred protonation state within the binding pocket can differ significantly, thus reductions of apparent pK_a values for substrates upon binding to CYP2D6 of up to 2 pH units have been reported.[26] From a direct comparison between the experimentally derived hydrogen atom to heme iron distances and the computationally determined values it became clear that the MD simulations reproduced the experimental data best when the MDAAs were represented in their neutral form.

6.4 Conclusions

The approach of combining automated docking and MD simulations in a protein model was a successful way to rationalize the oxidative metabolism of a series of 3,4-methylenedioxy-N-alkylamphetamines (MDAAs) by CYP2D6. Differences in oxidative metabolism of these closely related substrates could be predicted in details like regioselectivity and stereoselectivity. Furthermore, the effects of the active site F120A mutation on the substrate selectivity could be rationalized. The presented integrated modeling method is a promising tool in the prediction of the metabolic properties of new (drug-like) compounds by CYP2D6 and might be applicable for other drug metabolizing enzymes as well.

6.5 Methods
Materials
The pSP19T7LT plasmid containing human 2D6 with a C- terminal His$_6$-tag bicistronically co-expressed with human NADPH-cytochrome P450 reductase was kindly provided by Prof. Dr. Ingelman-Sundberg. 3,4-methylenedioxy- N-alkylamphetamines (MDAAs) were synthesized as described before. [33,34] S- and R-MDMA.HCl were obtained from the division of Neurioscience and Behavioral Research of the National Institute on Drug Abuse (Bethesda, MD USA). Emulgen 911 was purchased at KAO Chemicals (Tokyo, Japan). Ni-NTA-agarose was from Qiagen (Westburg, Leusden, The Netherlands). All other chemicals were of analytical grade and obtained from standard suppliers.

Expression and Purification of CYP2D6

Both the F120A mutant and the wild-type pSP19T7LT plasmids were transformed into *Escherichia coli*, strain JM109. Expression and membrane isolation was carried out as described. [16] Membranes were resuspended in 0.4% of the original culture volume of KPi-glycerol buffer (50 mM potassium phosphate buffer, pH 7.4, 10% glycerol) and CYP contents were determined by CO difference spectra. [35]

   Enzymes were solubilized from membranes by stirring in KPi-glycerol supplemented with 0.5% Emulgen 911 for 2 h at 4°C. Insoluble parts were removed by centrifugation (60 min, 120,000 g at 4 °C). The supernatant was incubated, gently rocking, with Ni-NTA agarose for 30 min at 4 °C. The Ni-NTA agarose was retained in a polypropylene tube with porous disc (Pierce, Perbio Science, Ettenleur, The Netherlands), washed with KPi-glycerol buffer containing 2 mM histidine. CYP2D6 was eluted with 0.2 M histidine. After overnight dialysis in KPi-glycerol buffer the sample was concentrated on a Vivaspin 20 filtration tube (10.000 MWCO PES, Sartorius, Nieuwegein, The Netherlands). For NMR spin lattice relaxation rate measurements the buffer was exchanged for deuterated KPi-glycerol buffer by repeated rounds of adding a larger volume of deuterated buffer and concentrating. Deuterated KPi-glycerol buffer was made by repeated rounds of evaporating water or D_2O from a KPi buffer and dissolving the residue again in D_2O. Then 5% of glycerol was added and traces of iron were removed with Chelex 100 (Biorad, Richmond CA).

Optical Titrations

   Dissociation constants of substrates to the enzymes were determined by spectral titration, according to the method of Jefcoate.[36,37] Spectra were taken at room temperature on a Pharmacia Ultrospec 2000 spectrometer. In short, 1 ml of 0.5 μM purified enzyme in KPi-glycerol buffer was divided over two cuvettes, to the first 5 μl of a 1 mM solution of compound in the same buffer was added, to the second the same volume of just buffer. Difference spectra from 350 to 450 nm were taken before, and after each addition of compound. The difference in absorbance at 390 and 425 was plotted against the substrate concentration to estimate the dissociation constant (K_s) using Equation (1) in Graph Pad Prism 4.0 software, where B is the absorbance difference (390-425), B_{max} is the absorbance difference when [S] = ∞.

$$B = \frac{B_{max}[S]}{K_s + [S]} \qquad\qquad (1)$$

Metabolism of MDAAs

Metabolic reactions were carried out in 200 μl 50 mM KPi pH 7.4, 5 mM $MgCl_2$ supplemented with 10 concentrations ranging from 0 to 200 μM of one of the MDAAs and *E. coli* membranes corresponding to 25 nM wild-type or F120A mutant CYP2D6. After 5 min of pre-incubation at 37 °C, the reactions were initiated with an NADPH regenerating system, resulting in final concentrations of 0.2 mM NADPH, 0.3 mM glucose-6-phosphate, and 0.4 units/ml glucose-6-phosphate dehydrogenase. The reaction was allowed to proceed for 10 min before it was stopped by the addition of 20 μl 23% $HClO_4$. After centrifugation (10 min, 6800 g), 30 μl aliquots of the supernatant were analyzed by HPLC. To measure O-demethylenation, analytes were separated isocratically using a C18 column (Phenomenex Luna 5u 150 x 4.6) with a mobile phase consisting of 22% acetonitrile and 0.1% triethylamine, set to pH 3 with $HClO_4$ at a flow rate of 0.6 ml/min. Other oxidation products were found using the same column with a gradient (A: 5% ACN with 20 mM ammonium acetate, B: 90% ACN with 10 mM ammonium acetate). All oxidation products were detected by fluorescence (λ_{ex}= 280 nm, λ_{em}= 320 nm) and identified by LC-MS. Peak areas of all products were quantified by Shimadzu Class VP 4.3 software and analyzed using non linear regression with one site binding in Graph Pad Prism 3.0.

LC-MS

To identify the metabolic products of the MDAAs, incubations were carried out for 15 min as described above with 100 μM of MDMA, MDEA or MDPA and *E. coli* membranes corresponding to 50 nM wild-type or F120A CYP2D6. Volumes of 100 μl supernatant were injected and separated using a C18 column (Phenomenex Luna 5u 150 x 4.6) with a flow rate of 0.6 ml/min and analyzed by MS.

The analytes were eluted using a gradient starting with a 5% ACN eluens, supplemented with 20 mM ammonium acetate for 7 min, then increasing linearly to 90% ACN with 10 mM ammonium acetate in 14 min and kept there for 5 min. APCI positive ionization was used on a LCQ Deca mass spectrometer (Thermo Finnigan, Breda, The Netherlands), vaporizer temperature 450 °C, N_2 as sheath (40 psi) and auxiliary gas (10 psi), needle voltage 6000 V, heated capillary 150 °C.

Products of MDMA metabolism: 3,4-OH-MA (t_R: 9.5 min., *m/z* 182, MS/MS: *m/z* 151), MDA (t_R: 17.0 min., *m/z* 180, MS/MS: *m/z* 163) and N-OH-MDMA (t_R: 20.4 min., *m/z* 210, MS/MS: *m/z* 163). Products of MDEA metabolism: 3,4-OH-EA (t_R: 13.0 min., *m/z* 196, MS/MS: *m/z* 151), MDA (t_R: 17.0 min., *m/z* 180, MS/MS: *m/z* 163) N-OH-MDEA (t_R: 21.6 min., *m/z* 224, MS/MS: *m/z* 163) a fourth product was detected for the mutant (t_R 18.5 min., *m/z* 222, MS/MS: *m/z* 163). Products of MDPA metabolism: 3,4-OH-PA (t_R: 16.0 min., *m/z* 210, MS/MS: *m/z* 151), MDA (t_R: 17.0 min., *m/z* 180, MS/MS: *m/z* 163), N-OH-MDPA (t_R: 23.4 min., *m/z* 238, MS/MS: *m/z* 163) a fourth product was detected for the mutant (t_R: 19.2 min *m/z* 254, no fragmentation observed).

NMR Spin Lattice Relaxation Rate Measurements.

Distances from substrate hydrogen atoms to the enzyme heme iron atom were calculated from their longitudinal relaxation rates determined using ^{1}H NMR.[27-30] ^{1}H NMR measurements were performed on a Bruker Avance 400 MHz spectrometer, signals were referenced to HDO at 4.75 ppm. Longitudinal relaxation times (T_l) of substrate hydrogen atoms were measured with a standard inversion recovery sequence of 180°-τ-90°. The spectra were recorded with 20 values of τ, ranging from 0.1 to 12.5 s and a relaxation delay of 5-10 x T_l. The temperature was kept at 300 K, except for the variable temperature experiments. T_l was determined by plotting peak areas against delay times τ and exponential fitting ($I_t = A + Be^{-\tau/T_l}$) with Bruker XWINNMR software. The measured T_l was transformed into the longitudinal paramagnetic relaxation rate (R_{lP}) according to equation 2.

$$R_{lP} = (T_{l,Fe^{3+}})^{-1} - (T_{l,Fe^{2+}+CO})^{-1} \qquad (2)$$

Where $T_{l,Fe^{3+}}$ is the relaxation time of substrate hydrogen atoms in the presence of ferric enzyme, and $T_{l,Fe^{2+}+CO}$ is the relaxation time in the presence of ferrous CO-bound enzyme. The hydrogen atoms to iron atom distance, r^{NMR} can be calculated from R_{lP} using the simplified Solomon and Bloembergen equation [27,38] (Equation 3)

$$r^{NMR} = C(\alpha R_{lP} F(\tau_c))^{1/6} \qquad (3)$$

Where C is a constant taking into account the spin state as well as other nuclear and electronic factors, with a value of 813 for high spin CYP, [30] α is the fraction of enzyme bound substrate, determined as the ratio of the concentration of enzyme and the sum of the substrate concentration and its spectral dissociation constant (K_s). $F(\tau_c)$ is a function of the correlation time τ_c, for which the value of 3×10^{-10} s is taken, as determined for CYP2D6 by others.[28] In the calculations of hydrogen atoms to heme distances according to eqs. 2 and 3, the assumption has been made that on the NMR time-scale there is a fast exchange between substrate molecules in solution and in the enzyme active site. This was verified by measuring the temperature dependence of $1/T_l$, which should show a linear increase with the reciprocal value of the temperature.[29,30]

In a typical experiment 5 μM of enzyme and 20 mM of substrate was used in a volume of 500 μl. Recording the spectra for relaxation times determination with oxidized enzyme took about an hour, then sodium dithionite was added and the sample was treated for 30 sec with CO to do the measurements with the reduced CO-bound enzyme.

Homology Modeling

A protein homology model of CYP2D6 was constructed based on the crystal structures of dimethylsulphophenazole derivative and diclofenac bound rabbit CYP2C5 (PDB codes 1N6B and 1NR6, respectively). [39,40] Homology modeling, model refinement and model validation was performed according to the approach described previously.[12] Shortly, a hundred CYP2D6 models were generated with the restraint-based comparative modeling program Modeller,[41] using the same amino acid alignment as before.[12] Subsequently, three models were selected with the best loop

conformations as determined by visual inspection, stereochemical parameters using PROCHECK, [42] and side-chain environment using Errat[43] and Verify3D.[44] Homology models had approximately the same protein quality check scores as the crystal structure templates. Finally, one single model was selected which could accommodate codeine best in the experimental binding orientation.[28] This final model was validated on its ability to reproduce substrate binding orientations corresponding to metabolic profiles with AutoDock,[45] FlexX,[46] and GOLD[47] automated docking studies.[48] The final model of wild-type CYP2D6 was used as a template for modeling of the F120A mutant. The Phe residue at position 120 was mutated to Ala using the homology module of Insight II (Biosym. San Diego, USA), after which an energy minimization and a 1 ps MD simulation with position restraints on the protein backbone was carried out as described under *"MD Simulations"*.

Automated Docking Studies.

The *R-*, and *S*-enantiomers of MDMA, MDEA and MDPA in their neutral and charged forms were docked into the active site of the CYP2D6 homology model using AutoDock and GOLD. Active site water molecules, predicted using a protocol based on the program GRID[49] and a ligand-based cut off were included. The preparation of enzyme and substrate structures in the automated docking studies were performed using default settings and predicted water molecules as described before. [21] For each substrate, 50 independent AutoDock and GOLD docking runs were performed using default parameter values. Docking poses generated by both docking algorithms were clustered (maximum root-mean square deviations between members of the cluster of 1 Å) and re-scored with the SCORE scoring function.[50] The three energetically most favorable, distinct docking poses of each MDAA species were selected as starting structures for MD calculations, yielding in total 12 (3 docking poses times 2 enatiomers, times 2 protonation states) simulations for each substrate in both wild-type and F120A mutant CYP2D6.

MD Simulations

Substrate conformations selected from the automated docking studies were used as starting structures for MD simulations, using the GROMACS molecular simulation package[51] and the GROMOS-96 force field, parameter set 43A1.[52,53] Heme parameters were taken from this parameter set, and additionally, a covalent bond was defined between the heme iron atom and the sulfur atom of C443 with an ideal bond distance of 0.240 nm. Enzyme including substrate and heme, was energy minimized in vacuum using the steepest descent method, first with harmonic position restraints using a force constant of 1000 kJ mol^{-1} nm^{-2} on all non-hydrogen atoms, then without position restraints. Subsequently, pre-equilibrated water was added in a dodecahedral box with a minimum distance of 1 nm between the enzyme and box edges, followed by minimization. For 1 ps, the water was allowed to relax, while the whole enzyme was position restrained. Then, first only the side chains of the residues not involved in substrate binding were released for 1 ps, next also the backbone of these residues for 10 ps. Finally, only the backbone of the binding residues were restrained for 100 ps. The equilibration scheme was followed by an unrestrained 10 ns MD production run. MD simulations were carried out with a time step of 1 fs, a cut-off of 1 nm and a relative dielectric constant of 1.0. The LINCS algorithm was used to constrain the length of all covalent bonds.[54] Pressure was maintained at 1 bar and temperature at 300° K by weak coupling to an external bath,[55] with relaxation times of 0.1 and 1.0 ps respectively. Coordinates were saved at a 1.0 ps interval for subsequent analysis. Enzyme-substrate interactions were estimated using the SCORE scoring function and averaged over the trajectories.

MD Simulations Analysis

Hydrogen atoms to iron atom distances r^{MD} for different protonation states of the MDMA enantiomers were derived from MD simulations by averaging distances (r) observed in the three independent simulations with different starting positions as:

$$r^{MD} = (\langle r^{-6} \rangle)^{-1/6} \qquad (4)$$

These distances could directly be compared to the experimentally obtained distances r^{NMR},(Equation 3).

Sites of oxidation were predicted based on enzyme-substrate conformations generated by automated docking and MD simulations. Distances of 6.0 Å or less between the heme iron and an

atom in the substrate (e.g. corresponding to O-demethylenation, N-dealkylation, N-hydroxylation, or ω-hydroxylation) were considered as reactive in the sense that activation of the C-H bond by the heme-Fe-O moiety is possible.[56] Site of oxidation predictions were performed to determine both substrate reactivity (expressed in the number of reactive MD configurations out of the total) and regioselectivity. The catalytic regioselectivity was calculated by determining which site of oxidation was closest to the heme iron atom for each reactive docking pose or MD configuration. Enzyme-substrate interactions during the MD simulations were monitored in terms of the occurrence of atom-atom distances of 3.5 Å or less between the F/A120 side-chain and the MDAA ring system (representing aromatic/van der Waals interactions) and hydrogen bonds between the E216 carboxylate oxygen atoms and the MDAA N-atom. Substrate reactivity, regioselectivity and enzyme substrate interactions were calculated for any given protonation state, stereo- and regio-isomer, as a Boltzmann weighted average over the three independent MD simulations with different substrate starting orientations. The average of an individual simulation i was weighted by a relative probability p_i, which was calculated from the average enzyme substrate interaction estimates as obtained from the program SCORE (E^i_{SCORE}).

$$p_i = (e^{-\Delta E\,SCORE_i} / k_B T)/(\sum_i e^{-\Delta E\,SCORE_i} / k_B T) \qquad (5)$$

In this equation, k_B is the Boltzmann constant (8,3441 kJ mol^{-1} K^{-1}) and T is the temperature (300 K).

Acknowledgement

References

(1)   Goeptar, A. R.; Scheerens, H.; Vermeulen, N. P. E. Oxygen and xenobiotic reductase activities of cytochrome P450. *Crit Rev Toxicol* **1995**, *25*, 25-65.

(2)   Guengerich, F. P. Cytochrome P450: what have we learned and what are the future issues? *Drug Metab Rev* **2004**, *36*, 159-197.

(3)   Zanger, U. M.; Raimundo, S.; Eichelbaum, M. Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. *Naunyn Schmiedebergs Arch Pharmacol* **2004**, *369*, 23-37.

(4)   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.

(5)   Shimada, T.; Yamazaki, H.; Mimura, M.; Inui, Y.; Guengerich, F. P. Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians. *J Pharmacol Exp Ther* **1994**, *270*, 414-423.

(6)   Oscarson, M. Pharmacogenetics of drug metabolising enzymes: importance for personalised medicine. *Clin Chem Lab Med* **2003**, *41*, 573-580.

(7)   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.

(8)   Williams, P. A.; Cosme, J.; Ward, A.; Angove, H. C.; Matak Vinkovic, D. et al. Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* **2003**, *424*, 464-468.

(9)   Williams, P. A.; Cosme, J.; Vinkovic, D. M.; Ward, A.; Angove, H. C. et al. Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* **2004**, *305*, 683-686.

(10)   Williams, P. A.; Cosme, J.; Sridhar, V.; Johnson, E. F.; McRee, D. E. Mammalian microsomal cytochrome P450 monooxygenase:structural adaptations for membrane binding and functional diversity.*Mol Cell* **2000**,*5*,121-131.

(11)   de Graaf, C.; Vermeulen, N. P.; Feenstra, K. A. Cytochrome p450 in silico: an integrative modeling approach. *J Med Chem* **2005**, *48*, 2725-2755.

(12)   Venhorst, J.; ter Laak, A. M.; Commandeur, J. N. M.; 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.

(13)   Kemp, C. A.; Flanagan, J. U.; van Eldik, A. J.; Marechal, J. D.; 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.

(14)   Ellis, S. W.; Hayhurst, G. P.; Smith, G.; Lightfoot, T.; Wong, M. M. et al. Evidence that aspartic acid 301 is a critical substrate-contact residue in the active site of cytochrome P450 2D6. *J Biol Chem* **1995**, *270*, 29055-29058.

(15) 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.

(16) Keizers, P. H. J.; Lussenburg, B. M. A.; de Graaf, C.; Mentink, L. M.; Vermeulen, N. P. E. 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.

(17) 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.

(18) 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.

(19) Vermeulen, N. P. E. Prediction of drug metabolism: the case of cytochrome P450 2D6. *Curr Top Med Chem* **2003**, *3*, 1227-1239.

(20) 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.

(21) de Graaf, C.; Pospisil, P.; Pos, W.; Folkers, G.; Vermeulen, N. P. E. Binding Mode Prediction of Cytochrome P450 and Thymidine Kinase Protein-Ligand Complexes by Consideration of Water and Rescoring in Automated Docking. *Journal of Medicinal Chemistry* **2005**.

(22) Harris, D.; Loew, G. Prediction of Regiospecific Hydroxylation of Camphor Analogs by Cytochrome-P450(Cam). *Journal of the American Chemical Society* **1995**, *117*, 2738-2746.

(23) Keseru, G. M.; Kolossvary, I.; Bertok, B. Cytochrome P-450 catalyzed insecticide metabolism. Prediction of regio- and stereoselectivity in the primer metabolism of carbofuran: A theoretical study. *Journal of the American Chemical Society* **1997**, *119*, 5126-5131.

(24) Harris, D. L.; Loew, G. H. Investigation of the proton-assisted pathway to formation of the catalytically active, ferryl species of P450s by molecular dynamics studies of P450eryF. *J Am Chem Soc* **1996**, *118*, 6377-6387.

(25) Park, J. Y.; Harris, D. Construction and assessment of models of CYP2E1:Predictions of metabolism from docking, molecular dynamics, and density functional theoretical calculations. *J Med Chem* **2003**, *46*,1645-1660.

(26) Miller, G. P.; Hanna, I. H.;Nishimura, Y.; Guengerich, F. P. Oxidation of phenethylamine derivatives by cytochrome P450 2D6: the issue of substrate protonation in binding and catalysis. *Biochemistry* **2001**, *40*, 14215-14223.

(27) van de Straat, R.; de Vries, J.; de Boer, H. J.; Vromans, R. M.; Vermeulen, N. P. E. Relationship between paracetamol binding to and its oxidation by two cytochromes P-450 isozymes--a proton nuclear magnetic resonance and spectrophotometric study. *Xenobiotica* **1987**, *17*, 1-9.

(28) Modi, S.; Paine, M. J.; Sutcliffe, M. J.; Lian, L. Y.; Primrose, W. U. et al. A model for human cytochrome P450 2D6 based on homology modeling and NMR studies of substrate binding. *Biochemistry* **1996**, *35*, 4540-4550.

(29) Poli-Scaife, S.; Attias, R.; Dansette, P. M.; Mansuy, D. The substrate binding site of human liver cytochrome P450 2C9: an NMR study. *Biochemistry* **1997**, *36*, 12672-12682.

(30) Regal, K. A.; Nelson, S. D. Orientation of caffeine within the active site of human cytochrome P450 1A2 based on NMR longitudinal (T1) relaxation measurements. *Arch Biochem Biophys* **2000**, *384*, 47-58.

(31) Wade, R. C.; Winn, P. J.; Schlichting, I.; Sudarko A survey of active site access channels in cytochromes P450. *J Inorg Biochem* **2004**, *98*, 1175-1182.

(32) Prasad, S.; Mitra, S. Role of protein and substrate dynamics in catalysis by Pseudomonas putida cytochrome P450cam. *Biochemistry* **2002**, *41*, 14499-14508.

(33) 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.

(34) Braun, U.; Shulgin, A. T.; Braun, G. Centrally active N-substituted analogs of 3,4-methylenedioxyphenylisopropylamine (3,4-methylenedioxyamphetamine). *J Pharm Sci* **1980**, *69*, 192-195.

(35) Omura, T.; Sato, R. The Carbon Monoxide-Binding Pigment of Liver Microsomes. Ii. Solubilization, Purification, and Properties. *J Biol Chem* **1964**, *239*, 2379-2385.

(36) Jefcoate, C. R. Measurement of substrate and inhibitor binding to microsomal cytochrome P-450 by optical-difference spectroscopy. *Methods Enzymol* **1978**, *52*, 258-279.

(37) Hanna, I. H.; Krauser, J. A.; Cai, H.; Kim, M. S.; Guengerich, F. P. Diversity in mechanisms of substrate oxidation by cytochrome P450 2D6. Lack of an allosteric role of NADPH-cytochrome P450 reductase in catalytic regioselectivity. *J Biol Chem* **2001**, *276*, 39553-39561.

(38) Solomon, I.;Bloembergen, N. Nuclear magnetic interactions in the HF molecule.*J Chem Phys* **1956**, *25*,261-266.

(39) Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dansette, P. M.; Mansuy, D. et al. Structure of a substrate complex of mammalian cytochrome P450 2C5 at 2.3 A resolution: evidence for multiple substrate binding modes. *Biochemistry* **2003**, *42*, 6370-6379.

192

(40) Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dijols, S.; Dansette, P. M. et al. Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 A resolution: evidence for an induced fit model of substrate binding. *Biochemistry* **2003***, 42*, 9335-9345.

(41) Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **1993***, 234*, 779-815.

(42) Laskowski, R. A.; Macarthur, M. W.; Moss, D. S.; Thornton, J. M. Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography* **1993**, *26*, 283-291.

(43) Colovos, C.; Yeates, T. O. Verification of Protein Structures - Patterns of Nonbonded Atomic Interactions. *Protein Science* **1993***, 2*, 1511-1519.

(44) Luthy, R.; Bowie, J. U.; Eisenberg, D. Assessment of Protein Models with 3-Dimensional Profiles. *Nature* **1992***, 356*, 83-85.

(45) 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. *Journal of Computational Chemistry* **1998***, 19*, 1639-1662.

(46) 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.

(47) 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.

(48) de Graaf, C.; Oostenbrink, C.; Keizers, P. H.; van der Wijst, T.; Jongejan, A. et al. Catalytic Site Prediction of and Virtual Screening Accuracy of Cytochrome P450 2D6 Substrates by Consideration of Water and Rescoring in Automated Docking. *J Med Chem* **2006***, 49*, 2417-2430.

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

(50) 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. *Journal of Molecular Modeling* **1998***, 4*, 379-394.

(51) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* **2001***, 7*, 306-317.

(52) Daura, X.; Mark, A. E.; van Gunsteren, W. F. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *J Comput Chem* **1998***, 19*, 535-547.

(53) Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R. et al. The GROMOS biomolecular simulation program package. *Journal of Physical Chemistry A* **1999***, 103*, 3596-3607.

(54) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **1997***, 18*, 1463-1472.

(55) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *Journal of Chemical Physics* **1984***, 81*, 3684-3690.

(56) Meunier, B.; de Visser, S. P.; Shaik, S. Mechanism of oxidation reactions catalyzed by cytochrome p450 enzymes. *Chem Rev* **2004***, 104*, 3947-3980.

Supplementary Data
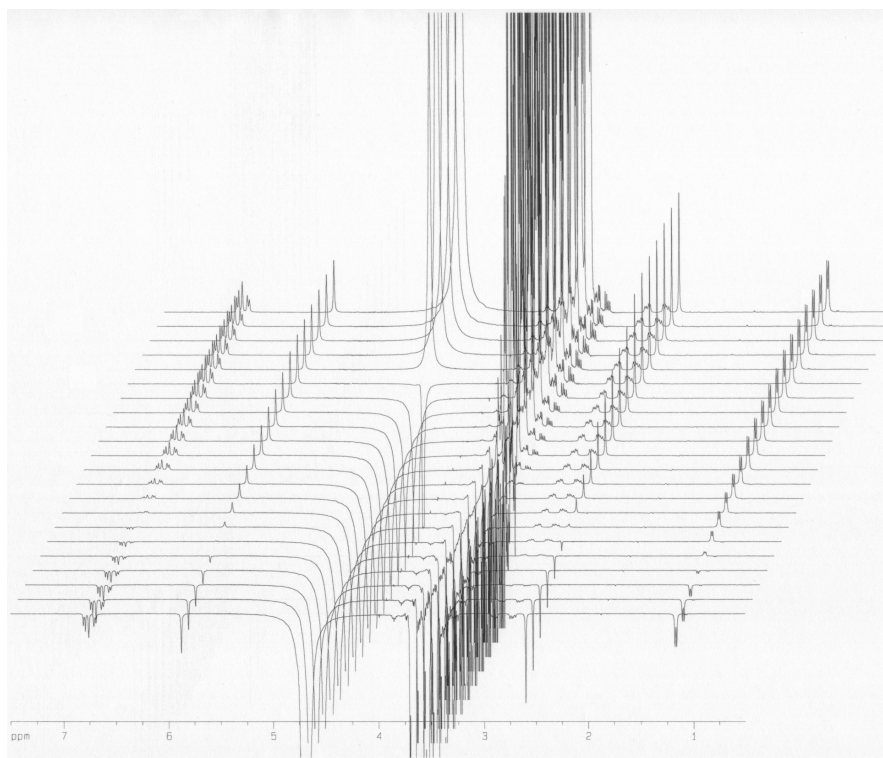
Figure I: ^1H NMR spectra after the inversion recovery sequence of 180°-τ-90°, of 20 mM racemic MDMA in deuterated Kpi-glycerol in presence of 5 μM wild-type CYP2D6 with τ varying from 0.1 to 12.5 s from front to back.
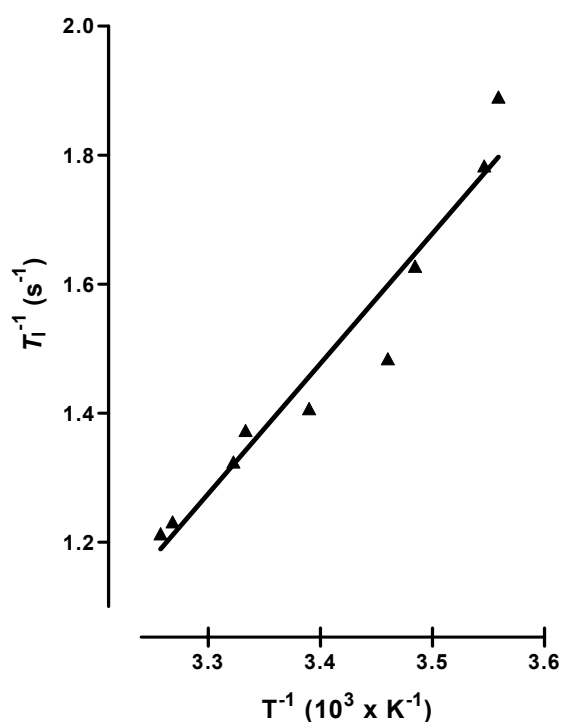
Figure II: Temperature dependence of T_1 relaxation rates for the hydrogen atoms of N-methyl (δ 2.61) of +/- MDMA in wild-type CYP2D6. The positive slope indicates that the fast exchange condition is satisfied.

Chapter 7

Stereochemical alchemy to predict affinity

Free energies of binding of *R*- and *S*-propranolol to wild-type and F483A mutant Cytochrome P450 2D6 from molecular dynamics simulations

Chris de Graaf, Chris Oostenbrink, Peter H.J. Keizers, Barbara M.A. van Vugt-Lussenburg, Jan N.M. Commandeur, and Nico P.E. Vermeulen

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

Detailed molecular dynamics (MD) simulations have been performed to reproduce and rationalize the experimental finding that the F483A mutant of CYP2D6 has lower affinity for *R*-propranolol than for *S*-propranolol. Wild-type (WT) CYP2D6 does not show this stereospecificity. The thermodynamic integration approach was used to calculate the free energy differences between all combinations of *R*- and *S*-propranolol complexed to wild-type and F483A mutant CYP2D6 (end-points). Based on differences between multiple calculations connecting the same end-point and the closure of the thermodynamic cycles, it was clear that not all simulations have converged sufficiently. The approach that calculates the free energies of exchanging *R*-propranolol with *S*-propranolol in the F483A mutant relative to the exchange free energy in WT was able to reproduce the experimental binding data. Careful inspection of the end-points of the simulations involved in this approach, allowed for a molecular interpretation of the observed differences. R-propranolol forms fewer hydrogen bonds than *S*-propranolol, both in WT and in the F483A mutant. The mutation causes a loss of favourable hydrophobic interactions, which can be compensated by increased hydrogen bond formation by *S*-propranolol, but not by R-propranolol.

7.1 Introduction

The superfamily of Cytochrome P450 (CYP) enzymes constitutes the most important phase I metabolic enzymes. CYPs generally detoxify pharmacologically active and potentially hazardous compounds, but in a number of cases non-toxic parent compounds are bioactivated into pharmacologically or toxicologically active metabolites, or procarcinogens into their ultimate carcinogens.[1,2] Although the expression levels of CYP2D6 represent only 2% of all hepatic CYPs, it is the second most important drug metabolizing enzyme after CYP3A4, and involved in the metabolism of about 15-30% of the currently marketed drugs.[3,4] Large interindividual differences exist in CYP2D6 activity, due to gene multiplicity and polymorphisms further emphasizing its clinical importance.[5-7] The early identification of potential CYP2D6 substrates and prediction of their metabolism is therefore advantageous in the discovery and development of new drugs.

Experimentally, the active site of CYP2D6 can be probed by site directed mutagenesis studies as described in **Chapter 3** and e.g., refs 8-12. Such experiments can be designed from structural models of enzyme-substrate complexes and the observed effect of mutations on metabolism and inhibition can subsequently be rationalized at an atomic resolution by modeling approaches. One mutant of CYP2D6 that was proposed based on computational models, is the F483A mutant.[8] The effect of this mutation on substrate binding affinities and orientations was shown to be strongly dependent on the substrate. For example, the metabolism of MAMC was no longer observed in the F483A mutant and metabolism of bufuralol was reduced. On the other hand no effects on the MDMA metabolism were found, while for dextrometorphan several alternative metabolites were observed. The binding affinity of these compounds did not seem to be influenced significantly by the mutation. Only for the substrate R-propranolol was the affinity reduced by a factor 20 upon mutation. Interestingly, a large stereospecific affinity in the mutant CYP2D6 was observed, but not in wild-type (WT). The metabolites formed from both R- and *S*-propranolol remained the same upon mutation of the enzyme.

In this study, we explore the stereospecific binding of propranolol to wild-type CYP2D6 and the F483A mutant. The spectral dissociation constants obtained were computationally rationalized. Computational tools may offer an atomic resolution often lacking in experiments. In the present case, one expects that subtle differences in the protein-ligand interactions in WT CYP2D6 or in the F483A mutant have a determining influence on the observed affinities. In order to catch such subtle differences a static approach based on protein-ligand structures only will not likely suffice, but rather the structure *and* dynamics of protein-ligand interactions have to be taken into account, for instance by molecular dynamics (MD) simulation. In CYPs it was shown that docking and scoring strategies can be applied to accurately predict binding modes and to select high affinity compounds from chemical databases (e.g. CYP101, CYP2D6 and CYP3A4) (**Chapter 5**, refs [13-17]). However, the scoring functions available in docking programs show little or no correlation with (experimental) binding affinity for these CYPs.[16,17] In the case of MDMA binding to CYP2D6 (WT and F120A mutant) it was shown that a dynamic view on the interaction was required to rationalize experimental findings (**Chapter 6**). Detailed MD simulations also offer the possibility to calculate the free energy difference between complexes of different proteins and ligands.

Previously, successful free energy calculations have been reported using different methods. The linear interaction energy (LIE) method[18] was used for CYP101,[19] CYP1A1,[20] free energy perturbation (FEP) calculations[21,22] for CYP101[23,24] and a combination of MD simulations, entropy estimates and Poisson-Boltzmann calculations (MMPB-SA)[25] for CYP2B4.[26]

A remaining hurdle in the accurate calculation of CYP ligand binding affinity, however, is the diversity of binding modes of substrates, stemming from a flexible binding cavity that is large in relation to the size of the substrates. Subsequently, for many substrates the

binding mode in CYPs is not as strictly defined as is the case for many other more substrate specific enzymes. This introduces sampling problems, prohibiting the application of MD based methods that can provide the most accurate free-energy calculations.

approach 1:

$$\Delta\Delta G^{WT}_{bind}(R,S) = \Delta G^{WT}_{bind}(S) - \Delta G^{WT}_{bind}(R) = \Delta G^{WT}_{inv} - \Delta G^{free}_{inv}$$

$$\Delta\Delta G^{F483}_{bind}(R,S) = \Delta G^{F483}_{bind}(S) - \Delta G^{F483}_{bind}(R) = \Delta G^{F483}_{inv} - \Delta G^{free}_{inv}$$
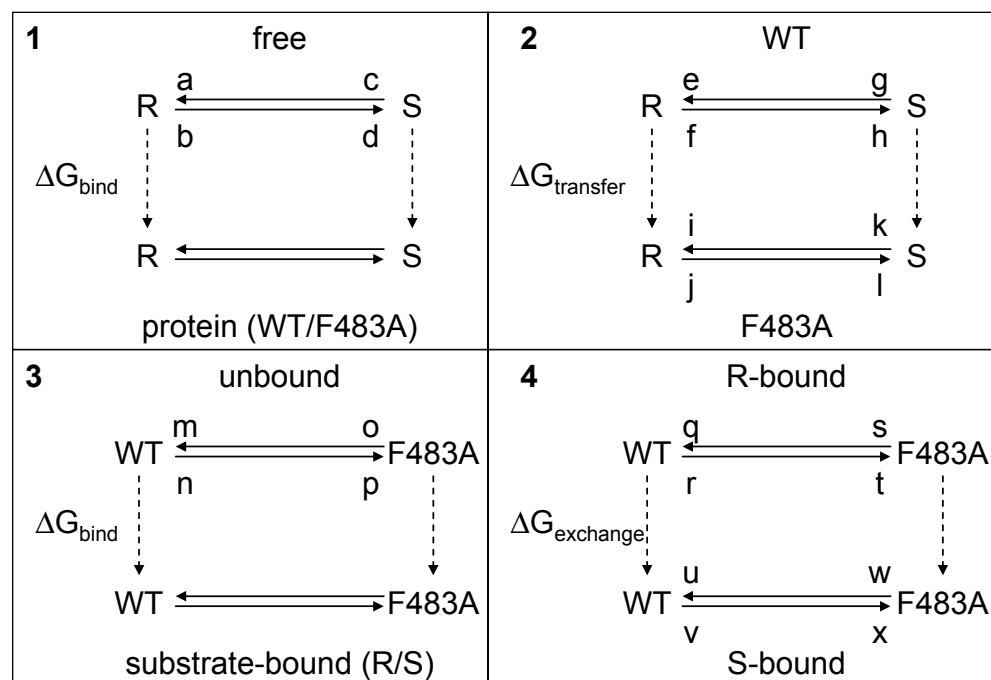
approach 2:

$$\Delta\Delta G_{transfer}(R,S) = \Delta G_{transfer}(S) - \Delta G_{transfer}(R) = \Delta G^{F483}_{inv} - \Delta G^{WT}_{inv}$$

approach 3:

$$\Delta\Delta G^{R}_{bind}(WT,F483A) = \Delta G^{R}_{bind}(F483) - \Delta G^{R}_{bind}(WT) = \Delta G^{R}_{mut} - \Delta G^{unbound}_{mut}$$

$$\Delta\Delta G^{S}_{bind}(WT,F483A) = \Delta G^{S}_{bind}(F483) - \Delta G^{S}_{bind}(WT) = \Delta G^{S}_{mut} - \Delta G^{unbound}_{mut}$$

approach 4:

$$\Delta\Delta G_{exchange}(WT, F483A) = \Delta G_{exchange}(F483) - \Delta G_{exchange}(WT) = \Delta G^{S}_{mut} - \Delta G^{R}_{mut}$$

$$\Delta G_{cycle} = \Delta G^{WT}_{inv} + \Delta G^{S}_{mut} - \Delta G^{F483}_{inv} - \Delta G^{R}_{mut}$$

Figure 1: *Four different approaches to calculate the differences in binding affinities of R-propranol (R) and S-propranolol (S) to wild-type (WT) and F483A mutant (F483A) CYP2D6, making use of different thermodynamic cycles (described in the text). Simulations at the initial and final stages along the solid arrows are labeled with a small letter (also referred to in the text and in* **Table 2**).

In the case of *R*- and *S*-propranolol binding to WT CYP2D6 and the F483A mutant, several approaches can be made, making use of different thermodynamic cycles.[27] We distinguish four of such approaches, which are exemplified in **Figure 1**.

1. Calculate the free energy of inversion of *R*-propranolol into *S*-propranolol in solution and when bound to protein (WT or F483): ΔG^{free}_{inv}, ΔG^{WT}_{inv} and ΔG^{F483A}_{inv}. From this approach the relative binding affinities of *R*- and *S*-propranolol in protein (WT or F483A) can be obtained ($\Delta\Delta G^{WT}_{bind}(R,S)$ and $\Delta\Delta G^{F483A}_{bind}(R,S)$).
2. Calculate the free energy of inversion of *R*-propranolol into *S*-propranolol when bound to WT and when bound to the F483A mutant: ΔG^{WT}_{inv} and ΔG^{F483A}_{inv}. From this

the relative free energy of transfer from WT to F483A for R- and S-propranolol can be obtained ($\Delta\Delta G_{transfer}(R,S)$).

3. Calculate the free energy of mutating the phenylalanine sidechain into an alanine at position 483 in a substrate free protein model and a substrate (R- or S-propranolol) bound protein model: $\Delta G^{unbound}_{mut}$ ΔG^{R}_{mut} and ΔG^{S}_{mut}. From this the relative binding affinities of the ligand (R- or S-propranolol) to WT and F483A can be obtained ($\Delta\Delta G^{R}_{bind}(WT,F483A)$ and $\Delta\Delta G^{S}_{bind}(WT,F483A)$).

4. Calculate the free energy of mutating the phenylalanine sidechain into an alanine at position 483 in a R-propranolol bound protein model and in a S-propranolol bound protein model: ΔG^{R}_{mut} and ΔG^{S}_{mut}. From this the relative free energy of exchanging R-propranolol by S-propranolol in WT and the F483A mutant can be obtained ($\Delta\Delta G_{exchange}(WT,F483A)$).

Each of these approaches yields information that can be compared directly to the experimentally determined differences in binding affinities of R- and S-propranolol to WT and F483A CYP2D6. ΔG^{free}_{inv}, representing the inversion of the stereocenter of propranolol free in solution, should equal zero. Because free energy is a state function, combinations of individual free energy terms defined above can be found that should sum up to exactly zero (ΔG_{cycle}). Comparisons of calculated to experimental values and to theoretically defined values will be used to determine which of the approaches is most accurate and can be used to rationalize the stereospecificity of the F483A mutant.

7.2 Methods

Optical Titrations

Dissociation constants of the substrates to the enzymes were determined by spectral titration, according to the method of Jefcoate.[28] Spectra were taken at room temperature on a Pharmacia Ultrospec 2000 spectrometer. In short, 1 ml of 0.5 μM purified enzyme in a 50 mM potassium phosphate buffer, pH 7.4 containing 10% glycerol, was divided over two cuvettes, to the first 5 μl of a 1 mM solution of R- or S-propranolol in the same buffer was added, to the second the same volume of just buffer. Difference spectra from 350 to 450 nm were taken before, and after each addition of compound. The substrates themselves did not show absorbance in this spectral range at the concentrations used. The difference between the absorbance at 390 nm and 425 nm was plotted against the substrate concentration to estimate the dissociation constant (K_s) using equation (1):

$$B = \frac{B_{max}[S]}{K_s + [S]}$$

(1)

where B is the absorbance difference (390-425) and B_{max} is the absorbance difference when $[S] \to \infty$.

CYP2D6-propranolol model

Coordinates for a ligand bound model of cytochrome P450 were obtained from homology modeling as described in **chapter 3**. Very recently a model based on the crystal structure of this isoenzyme has been published, but this involves a ligand-free conformation of the protein. For other CYPs (e.g. 2C5) it was shown that significant conformational changes within the active site take place upon substrate binding.[29-31] For this reason we chose to rather use the ligand-bound homology model, which was validated in earlier simulations studies.[32] For a detailed description of the homology model and a comparison with the crystal structure we refer to **Chapter 3**. Initial coordinates for R-propranolol in complex with the protein model were obtained by an automated docking approach as described in **Chapter 4**.

Molecular dynamics simulations

Molecular dynamics simulations and free energy calculations were performed along all solid angles in the thermodynamic cycles in **Figure 1**. All molecular dynamics simulations were performed using

a preliminary version of the GROMOS05 programme.[33] Interaction parameters were taken from the GROMOS 43A1 force field.[34,35] For propranolol, parameters were determined according to analogy to protein functional groups (see supplementary material). The GROMOS force field is a united atom force field, treating aliphatic hydrogen atoms together with the carbon atoms to which they are bound as a single interaction site. The wild-type protein was solvated in a rectangular periodic box containing 5314 (ligand free simulation) or 5306 (with *R*-propranolol bound) SPC water molecules.[36] For the simulations of propranolol free in solution, the simulation box contained 2367 water molecules. After an initial energy minimization in which the protein was positionally restrained, simulations were started at a temperature of 300 K, maintaining harmonic restraints with force constant 2500 kJ/mol/nm^2 on all solute atoms.

The system was carefully heated up and equilibrated as described in **Chapter 6**. Bond lengths were constrained at their minimum energy values using the SHAKE algorithm[37] allowing for a timestep of 2 fs. Nonbonded interactions were calculated using a triple range cutoff scheme. Interactions within a short range cutoff of 0.8 nm were calculated at every time step from a pairlist that was generated every 5 steps. Longer range interactions within a cutoff of 1.4 nm were calculated at pairlist generation and kept constant between updates. To account for electrostatic interactions beyond the long range cutoff, a reaction field contribution to the energies and the forces was calculated.[38] The temperature was kept constant by coupling solute and solvent degrees of freedom separately to a temperature bath of 300 K and using a relaxation time of 0.1 ps.[39] Constant pressure was maintained by isotropically coupling to a weak pressure bath of 1 atm using a relaxation time of 0.5 ps and an estimated isothermal compressibility of $4.575 \cdot 10^{-4}$ (kJ mol^{-1} nm^{-3})$^{-1}$.

Free energy calculations

Free energies were calculated along the solid arrows in **Figure 1**, using the thermodynamic integration method.[22,40] The arrows in this figure connect an initial (A) and a final (B) state, which can be defined as: 1) *R*-propranolol (A) vs. *S*-propranolol (B) free in solution; 2) *R*-propranolol (A) vs. *S*-propranolol (B) bound to wild-type and F483A mutant CYP2D6, respectively; 3) unbound wild-type (A) vs. unbound F483A mutant CYP2D6; 4) wild-type (A) vs. F483A mutant (B) binding *R*- and *S*-propranolol, respectively. The Hamiltonians describing the initial and final states of each process were related using a coupling parameter λ, such that at $\lambda = 0$ the Hamiltonian described the initial state and at $\lambda = 1$, it described the final state of the process. Simulations were performed at a number of discreet λ-values and the free energy was obtained by numerical integration of the derivative of the Hamiltonian with respect to λ:

$$\Delta G_{BA} = \int_0^1 \left\langle \frac{H(r,\rho,\lambda)}{\lambda} \right\rangle_\lambda d\lambda \qquad (2)$$

Where $<>_\lambda$ represents an ensemble average obtained at the specified value of λ, r and p are the positions and momenta of all particles in the system. Free energies of mutating residue 483 and of *R*/*S* inversion of propranolol were calculated as shown **Figure 2**. Processes involving the inversion of the chiral center of propranolol were performed in a two-step process.[41,42] First one stereoisomer was transformed into a planar intermediate (I), using 6 simulations for the integration of equation (2). The intermediate state was subsequently transformed into the alternative stereoisomer, using again 6 equidistant values of λ. Processes involving the transformation of residue 483 were obtained from 11 simulations at equidistant λ-values. Aromatic atoms in F483 were changed into non-interacting dummy atoms, while C_β was changed from a CH2 group into a CH3 group. The nonbonded interactions at intermediate λ-values were described using the soft-core approach to avoid singularities in the derivative.[43] No changes were made in the covalent interactions. 20 ps of equilibration were followed by 200 ps of data collection at every λ-value. Statistical error estimates on the ensemble averages were obtained using block averages.[44] Simulations with relative large statistical errors were prolonged for another 200 ps, and additional points were added wherever needed to obtain a smooth profile of $<\partial H/\partial \lambda>$ as a function of λ. A more empirical error estimate was obtained by comparing calculated free energy values associated with a forward process and the corresponding reverse process, also called hysteresis. An average free energy estimate from the forward and backward processes can be calculated using boltzmann weighted averages using:[45,46]

$$\Delta G_{average} = -k_B T \ln\left(\tfrac{1}{2}\left[e^{-\Delta G_{forward}/k_B T} + e^{-\Delta G_{backward}/k_B T}\right]\right) \qquad (3)$$

A total of 76 simulations (6 x 2 x 6 + 4 added λ-points) were performed involving R/S inversions and 72 simulations (6 x 11 + 6 added λ-points) were performed involving F/A mutations. This leads to an overall simulation time of 33 ns.
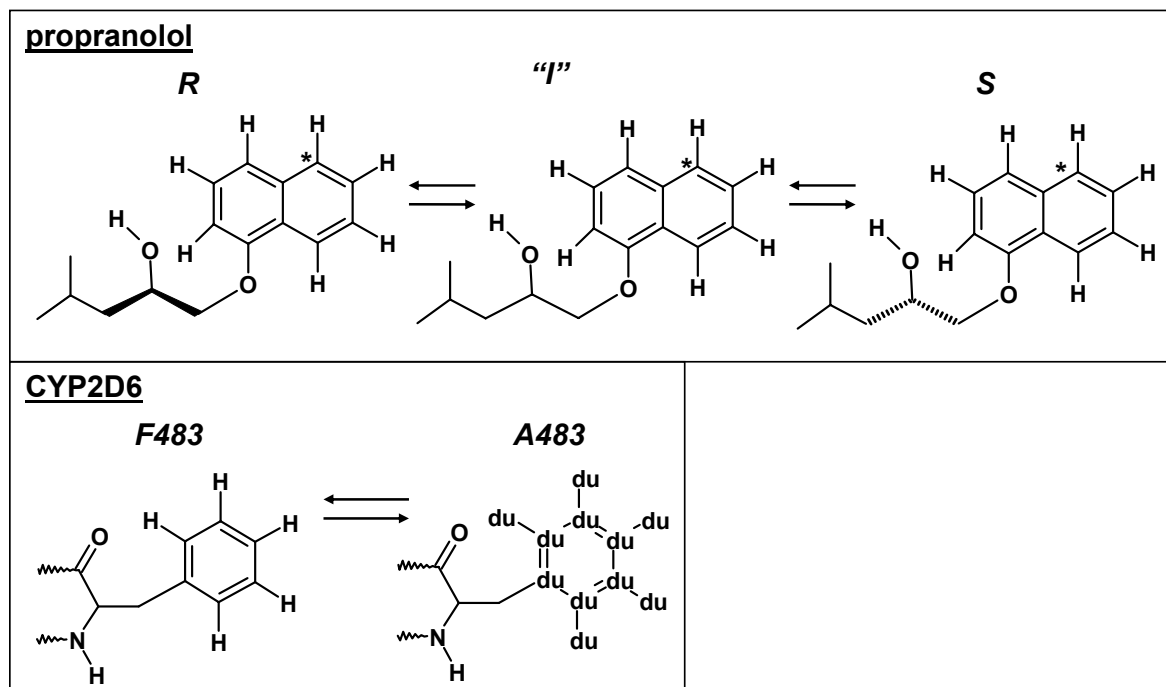
Figure 2: *Schemes describing the (reversible) inversion of R- into S-propranolol and the (reversible) mutation of the residue on position 483 in CYP2D6 (F (WT) into A (mutant)). The planar propranolol intermediate is indicated by "I", non-interacting dummie atoms in the A483 residue are indicated by "du". An asterix indicates atom C5 of propranolol.*

7.3 Results
Experimental results
Both the propranolol enantiomers displayed a type I binding spectrum when added to wild-type CYP2D6, as derived from a decrease of the Soret maximum at 417 nm, and the appearance of a shoulder at 380 nm, which is typical for type I binding of a substrate.[28] The affinity of the wild-type enzyme for the R- and S-enantiomers was similar (**Table 1**).

Table 1: K_s values (μM) of S- and R-propranolol for wild-type and F483A mutant CYP2D6.

Ligand	K_s (μM)	
	wild-type	F483A
S-propranolol	26 $\pm$ 6	23 $\pm$ 2
R-propranolol	19 $\pm$ 3	376 $\pm$ 26

All values are the means of at least two independent experiments $\pm$ S.D as described in the experimental procedures.

The addition of S-propranolol to the F483A mutant gave the same results as obtained for wild-type CYP2D6. The addition of R-propranolol to the F483A mutant of CYP2D6 did also lead to a type I spectrum, however, the affinity was 20-fold lower compared to the wild-type enzyme.

Computational approaches
Simulations at the initial and final stages along the solid arrows in **Figure 1** have been analyzed in detail. For every combination of R- or S-propranolol and WT or F483A mutant,

four of such simulations of 200 ps were performed, labeled with small letters in **Figure 1**. The four simulations included the end points of at least two different processes (*R*/*S* inversion or F/A mutation) allowing for a comparison of convergence at every state.

The atom-positional root-mean-square deviation of Cα atoms with respect to the initial homology model, remained stable throughout the simulations, with maximum values of 0.25 nm. For a 464 residue protein (the N-terminus of the homology model is truncated by 33 residues, see **chapter 3**) such values are reasonable and indicate simulations of a stable protein structure.

Binding mode orientation

The active site of CYP2D6 is formed by substrate recognition sites (SRS) 1 (B-C loop), 2 (F-helix), 4 (I-helix), 5 (β_{1-4}) and 6 (β_{4-1}). **Figure 3** displays a top view of the active site with interacting residues displayed in stick. The protonated nitrogen atom and the hydroxyl group of propranolol can form a hydrogen bond network with the negatively charged carboxylate of E216 (not visible in **Figure 3**, but shown in **Figure 5**). Two binding modes for the naphthalene moiety can be distinguished (see **Figure 3**). In WT CYP2D6, propranolol occupies a hydrophobic cavity formed by F120, V308, V370, and F483 (binding mode 1). In the F483A mutant, a larger cavity becomes available and SRS5 slightly shifts. The naphthalene moiety of propranolol now occupies a cavity formed by F120, V374, V370 and A483 (binding mode 2).
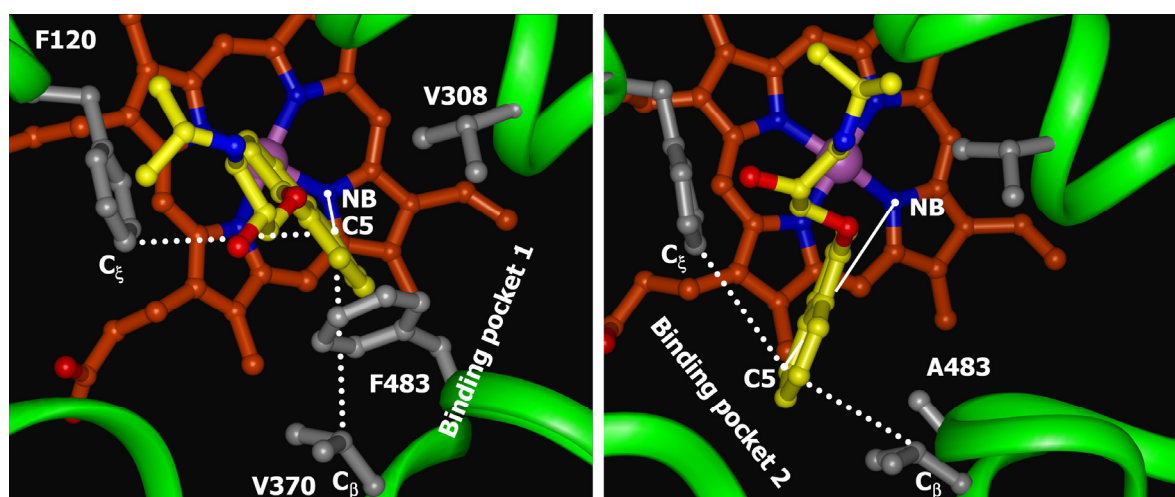
Figure 3: *Representative snapshots of two distinct binding modes observed in simulations of WT CYP2D6 (binding mode 1, left panel) and the F483A mutant (binding mode 2, right panel). Propranolol is depicted in yellow sticks and hydrophobic substrate binding amino acid residues are depicted in grey ball-and-sticks. Two geometric features used to discriminate between the two binding modes are displayed: the distance between the nitrogen atom of heme pyrrole ring B (NB) and the C5 atom of R-propranolol and the angle defined by the C_ζ atom in F120, the C5 atom in R-propranolol and the C_β in V370.*

In order to distinguish these binding modes, an atomic distance and an angle are defined in **Figure 3**. The distance between the heme NB atom and the C5 atom of propranolol will be larger in binding mode 2. The angle defined by the C_ζ atom in F120, the C5 atom of propranolol and the C_β in V370 will approach 180 in binding mode 2. **Table 2** presents the average values of the NB-C5 distance and the C_ζ-C5-C_β angle over the end-point simulations, together with the average number of hydrogen bonds between propranolol and E216. The presence of a hydrogen bond was determined using geometric criteria: a donor-acceptor distance of at most 0.25 nm and a H-bond angle of at least 135°.

Alternative hydrogen bonds partners of propranolol were not observed for significant periods of time.

Table 2: *The average values of the geometric features defined in Figure 3 together with the average number of hydrogen bonds between propranolol and E216 for the different MD-simulations at the initial and final stages of the inversion and mutation processes defined in Figure 1. The results belonging to approach 2 (defined in Figure 1) are displayed in bold.*

sim.[a]	propranolol	CYP2D6	average H-bonds	NB-C5 distance	F120.C_ζ-C5-V370C_β angle
e	R-propranolol	WT	**2.2**	**4.7**	**90**
f			**2.2**	**5.4**	**111**
g	S-propranolol	WT	**2.4**	**4.7**	**92**
h			**2.6**	**4.5**	**87**
i	R-propranolol	F483A	**2.1**	**8.3**	**159**
j			**2.4**	**8.0**	**138**
k	S-propranolol	F483A	**2.7**	**10.5**	**117**
l			**2.1**	**8.6**	**134**
q	R-propranolol	WT	2.4	6.6	157
r			2.2	4.8	103
s	S-propranolol	WT	1.5	4.6	88
t			2.2	4.8	83
u	R-propranolol	F483A	2.5	4.9	104
v			2.5	5.2	101
w	S-propranolol	F483A	1.7	5.0	90
x			1.7	4.9	89
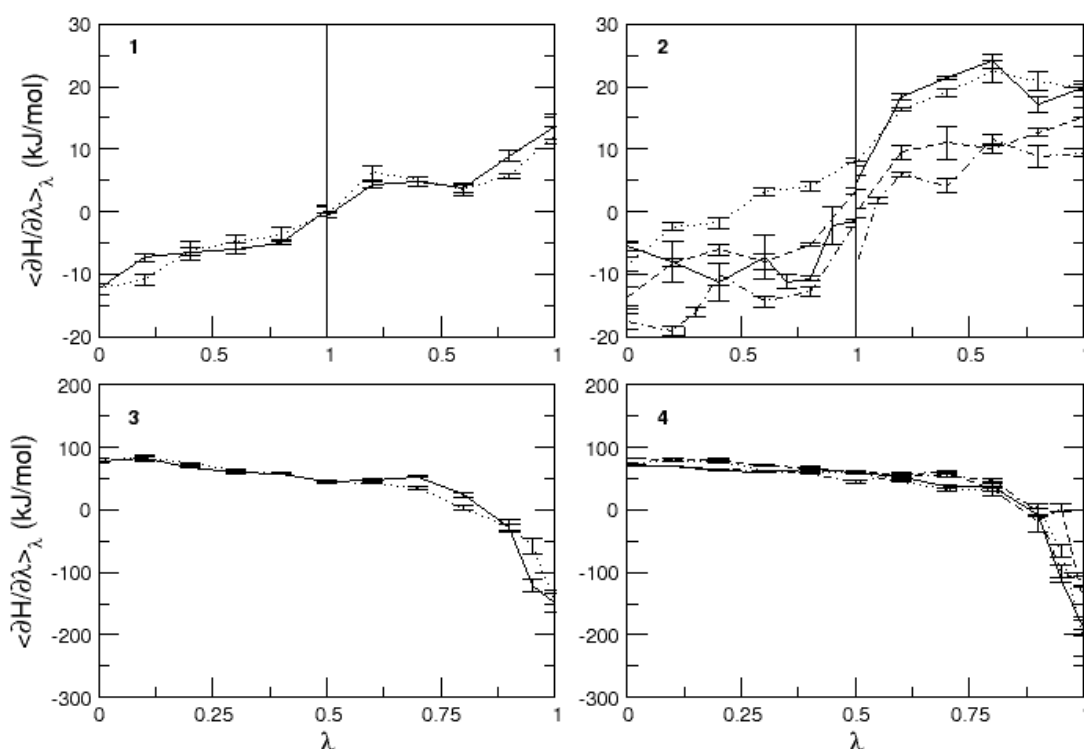
Figure 4: *The values of $<\partial H/\partial\lambda>$ as a function of λ for the calculation of the following free energy values. Panel 1: ΔG^{free}_{inv} forward (solid) and backward (dotted). Panel 2: ΔG^{WT}_{inv} forward (solid) and backward (dotted), ΔG^{F483A}_{inv} forward (dashed) and backward (dot-dash). Panel 3: $\Delta G^{unbound}_{mut}$ forward (solid) and backward (dotted). Panel 4: ΔG^{R}_{mut} forward (solid) and backward (dotted); ΔG^{S}_{mut} forward (dashed) and backward (dot-dash).*

Free energy calculations

The values of $\langle\partial H/\partial\lambda\rangle$ as a function of λ are displayed in **Figure 4** for all thermodynamic integration simulations. Values for the backward processes are multiplied by -1, for ease of comparison to the forward processes. Free energy estimates for all processes were obtained by a simple trapezoid integration of these curves and the resulting free energies are presented in **Table 3**. Using the definitions in **Figure 1**, the different relative free energies are calculated and compared to experimental values.

7.4 Discussion

The primary aim of this study was to reproduce and rationalize the experimental finding that the F483A mutant of CYP2D6 has lower affinity for *R*-propranolol than for *S*-propranolol, using detailed molecular dynamics (MD) simulations. From these simulations, free energy estimates for all processes in **Figure 1** have been derived, which are presented in **Table 3**. The error estimates in **Table 3** should be considered as statistical error estimates, indicating the statistical precision of the calculation. The absolute accuracy, however, is more difficult to determine. Three criteria that may give an indication of the accuracy can be distinguished. First, the hysteresis, defined as the difference between performing a process in the forward and backward directions, yields information on the convergence of the directions. At infinite sampling times, the hysteresis will be equal to zero.

Table 3: *Free energy estimates, the resulting average free energies, and the relative free energies for all processes defined in **Figure 1** and compared to experimentally determined values.*

	forward	backward	average[a]	hysteresis[b]
ΔG^{free}_{inv}	-0.4 ± 1.5	-1.0 ± 1.8	-0.7	0.6
ΔG^{WT}_{inv}	10.6 ± 1.5	19.1 ± 1.7	12.2	8.5
ΔG^{F483A}_{inv}	3.6 ± 2.1	6.8 ± 1.9	4.7	3.2
$\Delta G^{unbound}_{mut}$	35.8 ± 3.0	36.0 ± 3.5	35.9	0.2
ΔG^{R}_{mut}	36.8 ± 2.0	40.2 ± 2.7	38.0	3.4
ΔG^{S}_{mut}	50.5 ± 4.3	41.3 ± 2.5	42.9	9.2
ΔG_{cycle}	20.7 ± 11.4	13.4 ± 8.8	12.5	

	average[a]	experimental
approach 1:		
$\Delta\Delta G^{WT}_{bind}(R,S)$	12.9	0.8
$\Delta\Delta G^{F483A}_{bind}(R,S)$	5.4	-6.9
approach 2:		
$\mathbf{\Delta\Delta G_{transfer}(R,S).}$	**-7.5**	**-7.7**
approach 3:		
$\Delta\Delta G^{R}_{bind}(WT,F483A)$	2.1	7.4
$\Delta\Delta G^{S}_{bind}(WT,F483A)$	7.0	0.3
approach 4:		
$\Delta\Delta G_{exchange}(WT,F483A)$	5.0	-7.1

a) Average determined according to equation 3
b) Hysteresis, defined as the difference between performing a process in the forward and backward directions, yields information on the convergence of the two independent calculations.

In **Table 3**, two cases can be observed with relatively large hystereses, the *R*/*S* inversion of propranolol in WT CYP2D6 and the F/A mutation in *S*-propranolol bound protein. In the other cases the hysteresis falls within the statistical uncertainty. Second, using the studied processes, one can obtain cyclic pathways along which the sum of free energies should be equal to zero. In **Table 3**, such a pathway is presented as ΔG_{cycle}. It is clear that the

present calculations do not meet this criterion, with values of ΔG_{cycle} larger than the statistical uncertainty. This indicates that not all free energy calculations have converged properly. The third criterion is the ability to reproduce experimental data. Knowing that not all simulations have reached convergence, this last criterion allows us to determine if any of the four approaches mentioned in the introduction can still be used to rationalize the experimental data at an atomic level.

Free energies of binding: approaches 1 and 3

In the first approach, the relative free energies of binding R- and S-propranolol are determined for both proteins. This approach uses a simulation of the R/S-inversion free in solution, of which the free energy value should fundamentally equal zero. As can be seen in **Table 3**, indeed ΔG^{free}_{inv} is calculated to be 0 within the error estimate. However, the relative free energies $\Delta\Delta G^{WT}_{bind}(R,S)$ and $\Delta\Delta G^{F483A}_{bind}(R,S)$ do not correspond to the experimentally determined value. The inversion process of propranolol when bound to the protein seems to be too unfavourable. Still, this may be due to a systematic contribution in the protein simulations, which would cancel in approach 2, where only free energies resulting from simulations of the propranolol inversion when bound to the protein are compared.

In the third approach, the relative free energies of binding R- and S-propranolol to the WT and mutant CYP2D6 protein are calculated and compared. The mutation in the ligand-free protein shows the smallest hysteresis in **Table 3**. The obtained value of $\Delta\Delta G^{R}_{bind}$ (WT,F483A) is within 4 kJ/mol of the experimental value, which is within the statistical uncertainty, but not satisfactory. The value of $\Delta\Delta G^{S}_{bind}$(WT,F483A) is even further from the experimental value. Note that the value of ΔG^{S}_{mut} shows the largest hysteresis, indicating that particularly these simulations may not have converged properly. An indication of the convergence can be obtained by comparing simulations k and l to simulations w and x in **Figure 1**. in terms of the structural properties. From the average distance and angle in **Table 2**, we can see that in simulations w and x, the naphthalene moiety of S-propranolol binds in binding mode 2, while in simulations k and l, S-propranolol still orients according to binding mode 1 (see **Figure 2**). This difference indicates that the processes (R/S inversion in the mutant protein, or F/A mutation with S-propranolol bound) do not converge to the same end point. This is also reflected by significant differences in H-bond counts (see **Table 3**): S-propranolol forms more H-bonds with E216 in the F483A mutant in simulations k and l than it does in simulations w and x. A similar difference in the average number of hydrogen bonds between S-propranolol and E216 in WT is observed when comparing simulations g and h with simulations u and v. This indicates that S-propranolol occupies a different area in phase space during the F/A mutation than during the other simulations leading to the same end-points. With these observations, we also find an explanation for the non-zero value of ΔG_{cycle}. Since not all end-points of the simulations are connected, the calculated cycle is not complete.

Free energies of transfer and exchange: approaches 2 and 4

Approach 4 uses the same mutational processes with propranolol bound to the protein as were used in approach 3. From these values the free energies of exchanging R-propranolol by S-propranolol in the mutant relative to WT ($\Delta\Delta G_{exchange}$(WT,F483A)) can be obtained. As was discussed above, the F/A mutation with S-propranolol bound does not converge to the same end-points. For this reason it should not be expected that the value of $\Delta\Delta G_{exchange}$(WT,F483A) will correspond to the experimental data (**Table 3**).

The only approach for which the calculated free energies match the experimental data appears to be approach 2. In this approach the free energy of transferring R-propranolol from WT CYP2D6 to the F483 mutant is calculated relative to the free energy of transferring S-propranolol ($\Delta\Delta G_{transfer}(R,S)$). The obtained value lies only 0.2 kJ/mol from

the experimental value, which is well within the statistical error estimate of the calculations. It is also within k_BT (~2.5 kJ/mol), corresponding to thermal fluctuations, which is the accuracy that can be expected from free energy calculations.[47,48]

In summary, the approaches using protein mutations (i.e., 3 and 4) do not converge sufficiently to reproduce the experimental data. A systematic error in the *R/S* inversion for protein-bound propranolol disqualifies approach 1, but still allows for a comparison with approach 2 (the inversion in the different proteins), which does yield free energy values in agreement with the experimental data. From **Figure 4** it can be seen that <∂H/∂λ> takes much larger values in the processes involving a mutation than in the processes involving an inversion. It has been shown earlier that free energy calculations involving smaller structural changes and smaller energy barriers yield more accurate free energy estimates than integrations involving large changes in <∂H/∂λ>[41,49]

Possibly, better results with lower hysteresis and cycle closure, could be obtained by further extending the length of the simulations and by adding more λ-points.[49,50] Another possible source of error could be the fact that a substrate bound homology model was used, which has slightly different properties than the recently solved substrate-free crystal structure. However, we conclude that the changes required in the active site of CYP2D6 upon ligand binding are so large[51,52] that the use of the homology model is justified. In addition, CYP2D6 will remain a challenging target for computational approaches, because of the broad range of inhibitors and substrates and the multiple binding modes observed for these.[53]

Now that we have determined that approach 2 meets the third accuracy criterion best, we can investigate the molecular basis for the lower affinity of the F483A mutant for *R*-propranolol at an atomic resolution.

Structural rationalization

In order to rationalize the difference in affinity for *R*- and *S*-propranolol upon mutation of residue 483, we can compare the observed binding modes in simulations *e-l* in **Figure 1**. Representative orientations of propranolol in the binding cavities of WT and mutant protein in these simulations are presented in **Figure 5**. In all simulations propranolol is involved in a salt bridge between the protonated nitrogen atom and the negatively charged E216 in the protein. Together with the propranolol hydroxyl group, these three moieties form a dynamic hydrogen bond network. The average number of hydrogen bonds observed throughout the simulations, however varies for different combinations of *R/S*-propranolol and WT/F483A proteins. Already in WT (simulations *e-h*), we observed fewer hydrogen bonds for *R*-propranolol than for *S*-propranolol (see **Table 2** and **Figure 5**).

For both stereoisomers, the naphthalene moiety orients itself according to binding mode 1, it fits snugly into a hydrophobic pocket formed by V308, V370, and F483 (pocket 1, see **Figures 3** and **5**). Upon the F483A mutation, the size of the pocket increases, which opposes tight hydrophobic interactions. The naphthalene moiety now occupies the pocket formed by F120, V370, V374 and A483 (pocket 2), according to binding mode 2, as indicated by a small value of the heme-NB to propranolol-C5 distance (see **Table 2**). A large value of the C_ζ-C5-C_β angle indicates that a position between F120 and V370 is occupied. In the F483A mutant, *R*-propranolol binds in pocket 2 without significant decrease in the average number of hydrogen bonds (simulations *i* and *j*), compared to the simulations performed in WT (*e* and *f*). When *S*-propranolol binds in pocket 2 in F483A, hydrogen bond interactions are lost, in particular the propranolol-OH to E216-O_ε hydrogen bond is observed only for 50 % of the time in simulation *l*. This hydrogen bond can be regained by moving away from F120 and V370 (angle C_ζ-C5-C_β decreases) in simulation *k*. In fact the average number of hydrogen bonds in this simulation is even larger than for *S*-propranolol in WT, compensating the lack of tight hydrophobic interactions at the naphthalene moiety. *R*-propranolol cannot compensate for this loss.
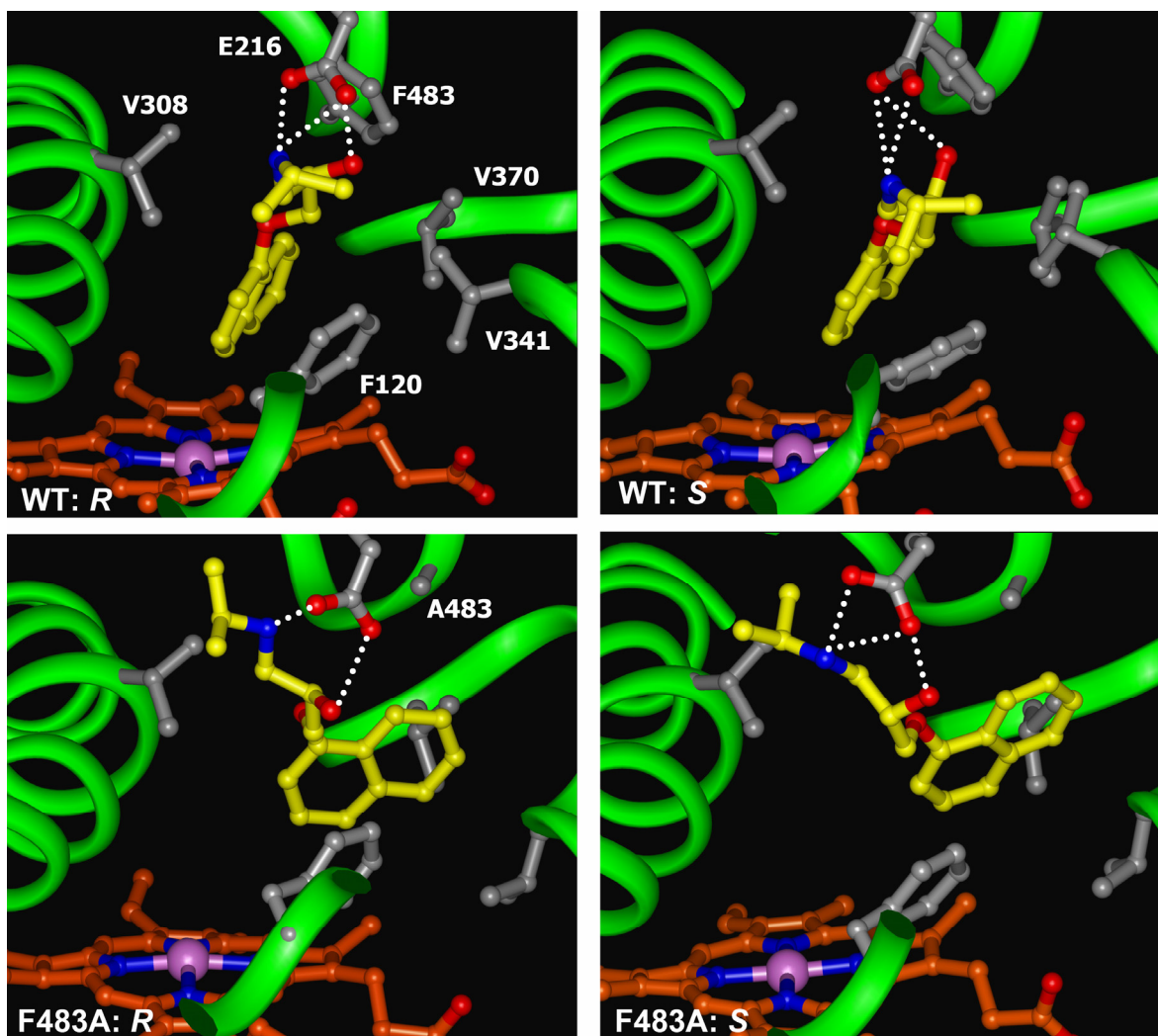
Figure 5: *Representative orientations of R- and S-propranolol in the binding cavities of WT and mutant CYP2D6 observed during the MD-simulations at points defined in **Figure 1** and included in approach 2: e (WT:R), g (WT:S), i (F483A:R), and k (F483A:S). R-propranolol forms fewer hydrogen bonds than S-propranolol, both in WT and in the F483A mutant. The mutation causes a loss of favourable hydrophobic interactions, which can be compensated by increased hydrogen bond formation by S-propranolol, but not by R-propranolol.*

In summary, the following picture arises. *R*-propranolol forms fewer hydrogen bonds than *S*-propranolol, both in WT and in the F483A mutant. The mutation causes a loss of favourable hydrophobic interactions, which can be compensated by increased hydrogen bond formation by *S*-propranolol, but not by *R*-propranolol.

7.5 Conclusions

Detailed molecular dynamics (MD) have been performed to reproduce and rationalize the experimental finding that the F483A mutant of CYP2D6 has lower affinity for *R*-propranolol than for *S*-propranolol. Wild-type (WT) CYP2D6 does not show this stereospecificity. Four different approaches to calculate the free energy differences have been investigated and were compared to the experimental binding data. From the differences between calculations based on forward and backward processes and the closure of the thermodynamic cycles, it was clear that not all simulations have converged sufficiently. The approach that calculates the free energies of exchanging *R*-propranolol with *S*-

propranolol in the F483A mutant relative to the exchange free energy in WT accurately reproduced the experimental binding data. Careful inspection of the end-points of the simulations involved in this approach, allowed for a molecular interpretation of the observed differences.

References

(1)   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.

(2)   Sheweita, S. A. Drug-metabolizing enzymes: Mechanisms and functions. *Current Drug Metabolism* **2000**, *1*, 107-132.

(3)   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.

(4)   Zanger, U. M.; Raimundo, S.; Eichelbaum, M. Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. *Naunyn Schmiedebergs Arch Pharmacol* **2004**, *369*, 23-37.

(5)   Ingelman-Sundberg, M. Human drug metabolising cytochrome P450 enzymes: properties and polymorphisms. *Naunyn Schmiedebergs Arch Pharmacol* **2004**, *369*, 89-104.

(6)   Oscarson, M.; Ingelman-Sundberg, M. CYPalleles: a web page for nomenclature of human cytochrome P450 alleles. *Drug Metab Pharmacokinet* **2002**, *17*, 491-495.

(7)   Wormhoudt, L. W.; Commandeur, J. N.; Vermeulen, N. P. Genetic polymorphisms of human N-acetyltransferase, cytochrome P450, glutathione-S-transferase, and epoxide hydrolase enzymes: relevance to xenobiotic metabolism and toxicity. *Crit Rev Toxicol* **1999**, *29*, 59-124.

(8)   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.

(9)   Keizers, P. H.; Schraven, L. H.; de Graaf, C.; Hidestrand, M.; Ingelman-Sundberg, M. et al. Role of the conserved threonine 309 in mechanism of oxidation by cytochrome P450 2D6. *Biochem Biophys Res Commun* **2005**, *338*, 1065-1074.

(10)   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.

(11)   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.

(12)   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.

(13)   DeVoss, J. J.; Demontellano, P. R. O. Computer-Assisted, Structure-Based Prediction of Substrates for Cytochrome P450(Cam). *J Am Chem Soc* **1995**, *117*, 4185-4186.

(14)   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. *Journal of the American Chemical Society* **1997**, *119*, 5489-5498.

(15)   Zhang, Z.; Sibbesen, O.; Johnson, R. A.; Ortiz de Montellano, P. R. The substrate specificity of cytochrome P450cam. *Bioorg Med Chem* **1998**, *6*, 1501-1508.

(16)   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.

(17)   Kemp, C. A.; Flanagan, J. U.; van Eldik, A. J.; Marechal, J. D. 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.

(18)   Aqvist, J.; Medina, C.; Samuelsson, J. E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng* **1994**, *7*, 385-391.

(19)   Paulsen, M. D.; Ornstein, R. L. Binding free energy calculations for P450cam-substrate complexes. *Protein Eng* **1996**, *9*, 567-571.

(20)   Szklarz, G. D.; Paulsen, M. D. Molecular modeling of cytochrome P450 1A1: enzyme-substrate interactions and substrate binding affinities. *J Biomol Struct Dyn* **2002**, *20*, 155-162.

(21)   Zwanzig, R. W. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* **1954**, *22*, 1420-1426.

(22)   Beveridge, D. L.; DiCapua, F. M. Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu Rev Biophys Biophys Chem* **1989**, *18*, 431-492.

(23)   Jones, J. P.; Trager, W. F.; Carlson, T. J. The Binding and Regioselectivity of Reaction of (R)-Nicotine and (S)-Nicotine with Cytochrome-P-450cam - Parallel Experimental and Theoretical-Studies. *J Am Chem Soc* **1993**, *115*, 381-387.

(24)   Helms, V.; Wade, R. C. Thermodynamics of water mediating protein-ligand interactions in cytochrome P450cam: a molecular dynamics study. *Biophys J* **1995**, *69*, 810-824.

(25)   Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* **2000**, *33*, 889-897.

(26)   Harris, D. L.; Park, J. Y.; Gruenke, L.; Waskell, L. Theoretical study of the ligand-CYP2B4 complexes: effect of structure on binding free energies and heme spin state. *Proteins* **2004**, *55*, 895-914.

(27)   Tembe, B. L.; McCammon, J. A. Ligand-receptor interaction. *Computers & Chemistry* **1984**, *8*, 281-283.

(28)   Jefcoate, C. R. Measurement of substrate and inhibitor binding to microsomal cytochrome P-450 by optical-difference spectroscopy. *Methods Enzymol* **1978**, *52*, 258-279.

(29)   Williams, P. A.; Cosme, J.; Sridhar, V.; Johnson, E. F.; McRee, D. E. Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol Cell* **2000**, *5*, 121-131.

(30)   Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dansette, P. M.; Mansuy, D. et al. Structure of a substrate complex of mammalian cytochrome P450 2C5 at 2.3 A resolution: evidence for multiple substrate binding modes. *Biochemistry* **2003**, *42*, 6370-6379.

(31)   Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dijols, S.; Dansette, P. M. et al. Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 A resolution: evidence for an induced fit model of substrate binding. *Biochemistry* **2003**, *42*, 9335-9345.

(32)   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.

(33)   Christen, M.; Hunenberger, P. H.; Bakowies, D.; Baron, R.; Burgi, R. et al. The GROMOS software for biomolecular simulation: GROMOS05. *J Comput Chem* **2005**, *26*, 1719-1751.

(34)   Van Gunsteren, W. F.; Billeter, S. R.; Hünenberger, P. H.; Krüger, P. M., A.E. et al. *Biomolecular simulations: The GROMOS96 manual and user guide*; Vdf blochschulverlag AG an der ETH Zürich: Zürich, 1996.

(35)   Daura, X.; Mark, A. E.; van Gunsteren, W. F. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *Journal of Computational Chemistry* **1998**, *19*, 535-547.

(36)   Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Hermans, J. Ìnteraction models for water in relation to protein hydration. *Intermolecular forces*; Reidel: Dordrecht, The Netherlands, 1981; pp 331-342.

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

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

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

(40)   Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300-313.

(41)   Zhou, Y.; Oostenbrink, C.; Van Gunsteren, W. F.; Hagen, W. R.; De Leeuw, S. W. et al. Relative stability of homochiral and heterochiral dialanine peptides. Effects of perturbation pathways and force-field parameters on free energy calculations. *Molecular Physics* **2005**, *103*, 1961-1969.

(42)   Zhou, Y.; Oostenbrink, C.; Jongejan, A.; Van Gunsteren, W. F.; et al. Computational study of ground-state chiral induction in small peptides: Comparison of the relative stability of selected amino acid dimers and oligomers in homochiral and heterochiral combinations. *J Comput Chem* **2006**, *27*, 857-867.

(43)   Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; Van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529-539.

(44)   Allen, M. P.; Tildesley, D. J. *Computer simulations of liquids*; Clarenden press: Oxford, 1987.

(45)   Jarzinsky, C. Non equilibrium equality for free energy differences. *Phys. Rev. Lett.* **1997**, *87*, 2690-2693.

(46)   Lu, N.; Adhikari, J.; Kofke, D. A. Variational formula for the free energy based on incomplete sampling in a molecular simulation. *Phys Rev E Stat Nonlin Soft Matter Phys* **2003**, *68*, 026122.

(47)   van Gunsteren, W. F.; Daura, X.; Mark, A. E. Computation of free energy. *Helv. Chim. Acta.* **2002**, *85*, 3113-3129.

(48)   Brandsdal, B. O.; Osterberg, F.; Almlof, M.; Feierberg, I.; Luzhkov, V. B. et al. Free energy calculations and ligand binding. *Protein Simulations* **2003**, *66*, 123-+.

(49)   Mark, A. E.; Van Gunsteren, W. F.; Berendsen, H. J. C. Calculation of relative free energy via indirect pathways. *J. Chem. Phys.* **1991**, *94*, 3808-3816.

(50)   Oostenbrink, C.; Van Gunsteren, W. F. Calculating zeros: Non-equilibrium free energy calculations. *Chem Phys* **2006**, *323*, 102-108.

(51)   Rowland, P.; Blaney, F. E.; Smyth, M. G.; Jones, J. J.; Leydon, V. R. et al. Crystal structure of human cytochrome P450 2D6. *J Biol Chem* **2006**, *281*, 7614-7622.

(52)   de Graaf, C.; Oostenbrink, C.; Keizers, P.; van Vugt-Lussenburg, B.; van Waterschoot, R. et al. Molecular Modeling-Guided Site-Directed Mutagenesis of Cytochrome P450 2D6. *Curr Drug Metab* **2006**, *in press*.

(53)   de Graaf, C.; Vermeulen, N. P.; Feenstra, K. A. Cytochrome p450 in silico: an integrative modeling approach. *J Med Chem* **2005**, *48*, 2725-2755.

Supplementary Material
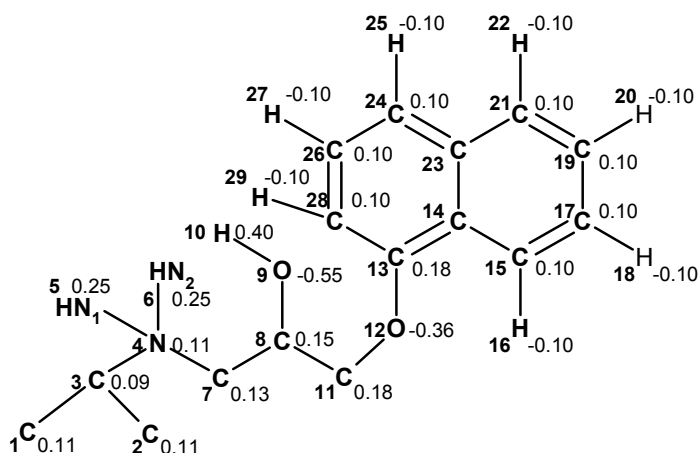
Figure I: Chemical structure of propranolol. Atom numbers (left label) and partial charges in e (right label). If no charge is quoted, it is zero. Force field parameters for propranolol are listed in **Tables I-V**.

Table I: Integer (van der Waals) atom type codes (IACs)

atom$_i$	IAC	atom$_i$	IAC
1	14	16	17
2	12	17	11
3	14	18	17
4	5	19	11
5	18	20	17
6	18	21	11
7	13	22	17
8	12	23	11
9	3	24	11
10	18	25	17
11	13	26	11
12	3	27	17
13	11	28	11
14	11	29	17
15	11		

Table II: Bond types

atom$_i$	atom$_j$	bond type	atom$_i$	atom$_j$	bond type	atom$_i$	atom$_j$	bond type
1	2	26	11	12	17	19	21	15
2	3	26	12	13	12	19	20	3
2	4	20	13	14	15	21	23	15
4	7	20	13	28	15	21	22	3
4	5	2	14	15	15	23	24	15
4	6	2	14	23	15	24	26	15
7	8	26	15	17	15	24	25	3
8	9	17	15	16	3	26	28	15
8	11	26	17	19	15	26	27	3
9	10	1	17	18	3	28	29	3

Table III: *Bond angle types*

atom_i	atom_j	atom_k	angle type		atom_i	atom_j	atom_k	angle type
1	2	3	14		17	19	20	24
1	2	4	12		20	19	21	24
2	4	7	20		19	21	22	24
2	4	5	17		22	21	23	24
2	4	6	17		23	24	25	24
3	2	4	12		25	24	26	24
4	7	8	12		24	26	27	24
5	4	6	9		27	26	28	24
5	4	7	10		13	28	29	24
6	4	7	10		26	28	29	24
8	11	12	12		26	28	29	24
8	9	10	11		26	28	29	24
11	12	13	11		26	28	29	24
12	13	14	26		26	28	29	24
12	13	28	26		26	28	29	24
14	13	28	26		26	28	29	24
14	15	16	24		15	17	19	26
13	14	15	26		17	19	21	26
13	14	23	26		19	21	23	26
15	14	23	26		14	23	21	26
14	15	17	26		14	23	24	26
15	17	19	26		21	23	24	26
16	15	17	24		23	24	26	26
15	17	18	24		24	26	28	26
18	17	19	24		13	28	26	26

	atom_i	atom_j	atom_k	angle type
R/S-propranolol	7	8	9	14
R/S-propranolol	7	8	11	14
R/S-propranolol	9	8	11	14
I-propranolol	7	8	9	25
I-propranolol	7	8	11	25
I-propranolol	9	8	11	25

Table IV: *Improper dihedral types*

atom_i	atom_j	atom_k	atom_l	angle type		atom_i	atom_j	atom_k	atom_l	angle type
15	14	16	17	1		14	15	17	19	1
17	15	18	19	1		14	23	24	26	1
19	17	20	21	1		15	14	23	21	1
21	19	22	23	1		15	17	19	21	1
24	23	25	26	1		17	19	21	23	1
26	24	27	28	1		19	21	23	14	1
28	13	26	29	1		23	14	15	17	1
2	1	3	4	2		23	14	21	24	1
13	12	14	28	1		23	24	26	28	1
13	14	23	24	1		24	26	28	13	1
14	13	15	23	1		28	13	14	23	1
14	13	28	26	1						

	atom_i	atom_j	atom_k	atom_l	angle type
R-propranolol	8	7	11	9	2
S-propranolol	8	7	9	11	2
I-propranolol	8	7	11	9	1

Table V: Dihedral types

atom_i	atom_j	atom_k	atom_l	angle type		atom_i	atom_j	atom_k	atom_l	angle type
7	8	9	10	12		7	8	11	12	17
1	2	4	7	14		8	11	12	13	17
2	4	7	8	14		11	12	13	28	2
4	7	8	11	17						

Chapter 8

Finding your way into and out of cytochrome P450 2D6

Random acceleration molecular dynamics investigation of substrate and product egress channels and mechanisms from cytochrome P450 2D6

Chris de Graaf,[‡] Chris Oostenbrink,[‡] Tim Johann,[§] Nico P.E. Vermeulen,[‡] and Rebecca C. Wade[§]

[‡]Leiden Amsterdam Center for Drug Research (LACDR)/Division of Molecular Toxicology, Department of Chemistry and Pharmacochemistry, Vrije Universiteit, Amsterdam, The Netherlands
[§]Molecular and Cellular Modeling, EML Research, Heidelberg, Germany

Substrate specificity of cytochrome P450 enzymes may be determined not only by the characteristics of the active site itself but also by the (dynamic) selectivity of access routes to the active site. In addition, enzyme kinetics may be influenced by the substrate binding and product unbinding process and associated protein dynamics. We have used random acceleration molecular dynamics (RAMD) simulations to investigate the routes and mechanisms via which substrates may enter and products may exit the buried active site of CYP2D6. Three dominant pathways were identified, which were rather substrate and substrate protonation state dependent. E216 was shown to play an important role in substrate egress, suggesting that E216 is not only crucial for substrate binding in the active site, but that it is also involved in directing the substrate towards the active site via electrostatic steering. Substrate/product egress via the three pathways is furthermore accompanied by large displacement of segments in the BC loop, FG loop and F and G helices and involve rotation of aromatic and aliphatic residues, and loss of internal protein H-bonds. We propose the following mechanism for substrate access into and product exit from CYP2D6: Hydrophobic substrates coming from the membrane enter the protein via a pathway termed pw2a, and their hydrophilic products leave via pw2c and/or the solvent channel, while soluble substrates enter (and their products leave) via pw2c and/or the solvent channel. From our RAMD studies hypotheses can be derived which can be tested experimentally, such as site-directed mutagenesis studies of non-active site residues lining the substrate egress pathways.

8.1 Introduction

In **chapters 3** to **7** we have shown how specific residues in the CYP2D6 active (e.g. E216, F120, and F483) influence substrate affinity and specificity, and metabolic regiospecificity. Substrate specificity of Cytochrome P450 (CYP) enzymes however, may be determined not only by the characteristics of the active site itself but also by the (dynamic) selectivity of access routes to the active site. In addition, enzyme kinetics may be influenced by the substrate binding and product unbinding process and associated protein dynamics.[1] The buried active site of many CYPs implies that the protein must undergo dynamic motion to allow substrates to enter the active site and products to leave the active site.[2,3]

In the past years, more and more experimental and molecular modeling studies have tried to explore the pathways and mechanisms via which substrates enter and products exit CYPs. Experimental methods applied to investigate substrate entrance (product exit) into (from) CYPs, are site-directed mutagenesis[1,4-7], antibody labeling[8], and fluorescence spectroscopy.[9-11] Simulating ligand egress from the CYP active site computationally is a challenging exercise because the timescale of ligand egress is orders of magnitude greater than the times (typically ~1-10 ns) for which the protein dynamics can be simulated using standard molecular dynamics simulation techniques. Nevertheless, several modeling studies on ligand egress from CYPs have been described in the past years. Table 1 gives an overview of the substrate entrance / producs exit channels observed in CYP structures. A thermal motion pathway analysis of B-factors in the crystal structure of CYP101 (P450 *cam*) with camphor bound revealed the existence of three distinct pathway families. Exit channels were explored by molecular dynamics (MD) simulations by imposing an artificial expulsion force on the substrate in addition to the forces due to the standard force field.[12] This random acceleration molecular dynamics (RAMD) method (formerly known as random expulsion MD) was used in simulations of the palmitoleic acid bound structure of CYP102 (P450 *BM3*), a system for which a substrate-access channel is closed in the substrate-bound and open in the substrate-free crystal structure.[13] RAMD simulations reproduced the closing motion.[14,15] It was further applied to complexes of CYP101 with camphor, endo-borneol allyl ether and 5-hydroxy camphor (metabolic product of camphor). The three ligand expulsion pathways of CYP101 observed by RAMD matched well with three putative ligand channels found by thermal motion pathway analysis.[16] In a complementary study, estimates of the relative probabilities of ligand passage through the three channels were computed using steered molecular dynamics (SMD) and adiabatic mapping.[17] The REMD and SMD approaches combined with essential dynamics (ED) analysis were used for 6-deoxyerythronolide B bound CYP107A (P450 *EryF*), and the results obtained for this complex were compared to those for CYP101 and 102.[18] Several possible exit/entrance routes in each enzyme were identified and one route, located between the F/G loop, the β1 sheet and either the B' helix (CYP101 and 102) or B'/B loop (CYP107), was common to all three proteins. The mechanisms of substrate entrance and/or product exit along this pathway, however, were different for each of the three isoforms and apparently adapted to the properties of the substrate. Based on this observation, it is suggested that the characteristics of the active site itself may not be the only determinant of substrate specificity. The access channel properties may thus play a key role by kinetic modulation of substrate recognition.[18] Recently, RAMD was applied to rabbit CYP2C5. In contrast, in this CYP, egress is predominantly by pathway 2c. This route requires the rupture of an intra-protein hydrogen bond between a lysine amino group in the G-helix and a valine backbone carbonyl oxygen in the B'-helix. Taken together, the results for these four P450s suggest that while substrate entrance and product egress may both be by pw2a in the three soluble bacterial P450s, in the mammalian CYP2C5, substrate entrance and product egress may take place by pw2c. Very recently, open channels leading to the active site of currently available CYP crystal structures were analyzed, providing experimental evidence

for the existence of four substrate exit/entrance routes that were identified by MD simulation.[3] Recent MD-studies have also identified ligand egress channels in insect CYP6B1 (pw1, solvent channel)[1], rat CYP2B1 (pw2c, pw2e, solvent channel)[19], and human CYP2C9 (pw2a, solvent channel).[20]

The aim of the current study is to investigate the routes and mechanisms via which substrates may enter and products may exit the active site of CYP2D6. For these investigations we have used the RAMD simulation protocol developed by Wade and co-workers.[16] According to this method, acceleration with a regularly updated random direction is applied to the ligand to find expulsion routes during a relatively short but otherwise standard MD simulation. In the current study RAMD is for the first time applied to a human CYP2D6 homology model (**Chapter 3**), based on CYP2C5 crystal structures.[21,22] **Figure 1** displays potential exit channels in the 2D6 homology model observed in the current RAMD study or by mapping the pathways observed in earlier studies to the corresponding residues in the 2D6 model. We have described the egress pathways of two compounds: dextromethorphan (DXM) and *R*-MDMA (MDR), the latter compound in its neutral form (MDR^0) as well as its positively charged form (MDR^+). Three dominant pathways were identified, and based on detailed analysis of protein structural changes and protein-ligand interactions, molecular mechanisms have been proposed for substrate egress via these pathways. From these studies hypotheses can be derived which can be tested by experiments, such as site-directed mutagenesis studies of non-active site residues lining the substrate egress pathways. Based on initial simulation data, one such mutant is proposed and included in the calculations. In the L110C/Q244C mutant a cysteine brigde can be formed, possibly blocking product release through pw2c.

Table 1: *Overview of open access channels observed in CYP crystal structures[3] and earlier (RA)MD simulations with CYPs 101,[16] 102,[14-16] 107[18], 6B1,[1] 2B1,[19] 2C5,[23] and 2C9[20], and the secondary structure elements lining these pathways.*

pw	Egress via secondary structure elements	method	CYP
1	C/C' and H or L helices	X-ray MD	51 101, 6B1
2a	B' helix/BB' loop/BC loop, FG loop, β_1 (and a/A helix)*	X-ray MD	101, 102, 108, 119, 152, 154, 55, 2B4, 2C5, 2C9, 3A4 101, 102, 107, 2C9
2b	B/B' loop, β_1 and β_3 sheets	MD	101
2c	BC loop, G and I helices	X-ray MD	2C5, 2D6 101, 107, 2B1, 2C5
2d	N-terminus, a and A helices	MD	102
2e	through BC loop	X-ray MD	epok, 51, 2B4, 2C5, 3A4 107, 2B1, 2C5
3	F and G helices or EF loop	MD	2C5
solv	E, F, and I helices, β_4 sheet	X-ray MD	CYP102, 119, 175, 2C5, 2C8, 2C9, 2D6, 3A4 2B1, 2C5, 2C9, 6B1
other[a]		MD	2C5

[a]Other routes not previously identified in crystal structures or simulations as potential active site channels for substrate or product passage.

Figure 1: *Potential substrate and product egress pathways in CYP2D6, and the secondary structure elements and Cα atoms (colored balls) lining them. The highlighted residues are either those observed in the current RAMD study (pws 2a.I, 2a.II, 2c, 2e, 3.I, 3.II and the solvent channel), or those observed in RAMD-studies with CYP101, CYP102 and CYP107 and projected on the CYP2D6 structure (pws 1, 2b, and 2d).*
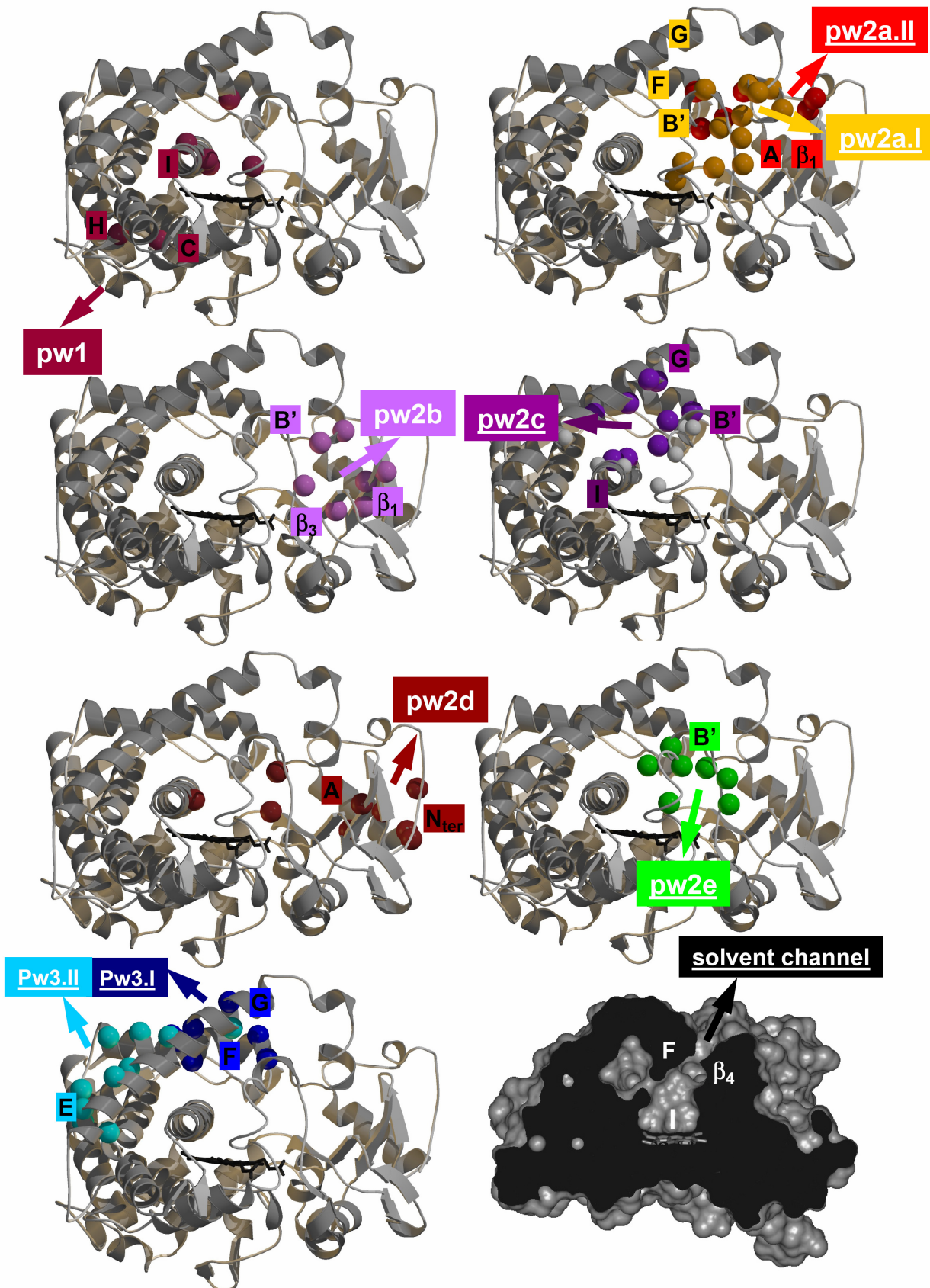
8.2 Results and Discussion

Substrate egress pathways in CYP2D6 are substrate and substrate protonation state dependent

For every ligand (DXM, MDR$^+$, MDR^0) RAMD simulations were performed in both the wildtype and in the L110C/Q244C mutant, using a set of six different simulation settings, defined in **Table 2**. These parameters were chosen on the basis of preliminary simulations of CYP2C5. RAMD-simulations performed with higher force constants showed very large distortion of the secondary structure of CYP2D6 (data not shown here). This leads to a total of 36 simulations, as described in the methods section. The details of these simulations, resulting in ligand expulsion from wild-type and L110C/Q244C mutant CYP2D6, are given in **Tables 3** and **4**.

Table 2: *Summary of RAMD-simulations performed with wildtype and L110C/Q244C mutant CYP2D6*

simulation	RAMD settings[a]		seed[b]
	N	r$_{min}$	
I	20	0.001	seed1
II	40	0.001	seed1
IIIa	20	0.005	seed1
IVa	40	0.005	seed1
IIIb	20	0.005	seed2
IVb	40	0.005	seed2

a) A, the magnitude of the acceleration resulting from the random expulsion force acting on the centre of mass of the ligand, is kept on a constant value of 0.05 kcal/g/Å for all simulations. The direction of the force is kept for N time steps of 2 fs. If, during this time period, a specified distance, r$_{min}$, is covered by the substrate molecule, the direction of the force is maintained, otherwise a new direction is chosen randomly. These parameters were chosen on the basis of preliminary simulations of CYP2C5, so that artefactual distortion of the secondary structure was minimal while still permitting ligand egress to be observed within a reasonable simulation time; **b**) Simulation settings III and IV were performed with two different random number seeds for generation of initial velocities.

Table 3: *The residues and secondary structure elements lining egress pathways observed in RAMD-simulations performed in wild-type (WT) and L110C/Q244C mutant (MUT) CYP2D6 together with the observation frequencies of the pathways in the simulations.*

pw	residues	frequency
2a [a]	I: F51, Q52, (A helix), L73, A74 (β_1 sheet), E216, G218, E222 (F-G loop), F483 (β_4 sheet) II: P102, P103, V104, I106, F120, L121 (BB'-loop), L210, Q211, L220, R221, V223, L224, N225 (F-G loop), V374, T375, T394 (β_1 sheet), F483 (β_4 sheet)	5 (WT) 4 (MUT)
2c	T107 ^{c}, V109, L110, G111, F112 ^{c}, Q117 ^{c} (B' helix), E216 (F-helix), K239, V240, L241, Q244, F247, D252 ^{bc} (G-helix), R296 ^{c}, I297 ^{c}, P300, D301, S304 (I-helix)	4 (WT) 4 (MUT)^b
2e	P102, P103, V104, T107, Q108, G111, P114, L121 (B-C loop)	2 (WT) 2 (MUT)
3 [a]	I: D207, L208, A209, E211, E215, E216 (F-helix), K239, F243 (G-helix) II: C191, R193, F195, R201, R204, E216 (F-helix), V240, A246, F247, Q250, L251, E253, L254, L257 (G-helix)	3 (WT) 3 (MUT)
solvent	L206, A209, Q210, L213, K214, E216 (F-helix), S304, M307, V308 (I-helix), F483, L484 (β_4 sheet)	4 (WT) 2 (MUT)
other [d]	S304, V308, T313, W316 (I-helix), V363, Q364, F366, I369 (K-helix), V485 (β_4 sheet)	1 (MUT)

a) Two different variations on the 2a and 3 pathways could be defined, labelled I and II; **b**) Egress via pathway 2c in the L110C/Q244C mutant is accompanied by large distortion of secondary structure elements; **c**) Residues lining pathway 2c in the mutant which do not line pathway 2c in wild-type CYP2D6; **d**) Other route not previously identified in crystal structures or simulations as potential active site channels for substrate or product passage.

Table 4: *Observed routes and parameters used in the RAMD simulations resulting in substrate expulsion for the complexes of dextromethorphan (DXM), charged R-MDMA (MDR⁺), and neutral R-MDMA (MDR⁰) with wild-type and L110C/Q244C mutant CYP2D6.*

	pw [a]	sim. [b]	length (ps)	Max. C_α RMSF [c]	res. [d]	Exit via [e]
wild-type						
DXM	2a.I	I	74	3.9	F51	F51, Q52, G218
	2a.I	IIIa	289	2.7	G218	F51, A74, E222
	2c	IVb	206	3.7	Q244	L110, Q244
	2c	IIIb	96	2.9	Q244	L110, Q244
	2c	IVa	178	5.4	L110	L110, L241
	3.I	II	140	2.3	E216	E216, K239
MDR⁺	2c	I	76	2.7	V240	I109, Q244, F247
	2e	IIIb	400	2.0	G212	D100, P103, Q108
	solvent	II	175	1.2	D207	L206, L213, Q210
	solvent	IVa	172	2.3	L213	L213, V308, L484
	solvent	IVb	27	2.5	L213	L213, V308
	3.I	IIIa	28	1.4	V229	E211, E215, K239
MDR⁰	2a.II	I	262	1.2	L224	P103, N225
	2a.II	IIIa	76	1.4	P103	P102, V104, R221
	2a.II	IVa	17	0.8	N225	P103, N225
	2e	IIIb	85	1.9	Q108	P102, P103, V104, Q108
	3.II	IVb	39	2.7	L254	R196, E257
	solvent	II	72	1.9	L213	L213, V308, T312, L484
L110C/Q244C mutant						
DXM	2a.I	IIIa	111	2.8	E216	F51, Q52, G218
	2c [f]	I	182	2.6	Q244	G111, Q117, C244, F247
	2c [f]	IIIb	209	2.8	E216	T107, I109, V240
	3.I	IVb	102	1.9	A209	D207, L208, E211, F243
	3.II	II	128	2.6	L254	R201, L251, L254
	solvent	IVa	131	3.9	L213	L206, A209, L213, L484
MDR⁺	2a.I	IVa	193	2.1	E216	P103, V104, N225
	2c [f]	I	268	2.7	L248	C110, L248
	2e	II	280	1.5	Q108	P103, Q108, P114
	solvent	IVb	22	2.2	E216	L213, E215, F481
	none [g]	IIIa	-	-	-	-
	none [g]	IIIb	-	-	-	-
MDR⁰	2a.II	IIIa	32	1.3	P103	P103, P105, N225
	2a.II	IVa	65	2.8	E216	Q51, E216, G218
	2c [f]	I	228	3.9	L248	F112, D252
	2e	IIIb	116	1.1	L121	P103, Q108, P114
	3.II	IVb	199	1.4	R204	L208, F243, A246
	other	II	110	1.2	W316	W316, F366

a) Two different variations on the 2a and 3 pathways could be defined. **b** RAMD simulation settings as defined in **Table 2**; **c**) Maximal C_α root-means-square fluctuation observed during the simulation; **d**) Residue with the highest C_α RMSF; **e**) Residues via which the substrate exits the protein; **f**) Egress via pathway 2c is accompanied by large distortions of secondary structure motifs; **g**) No substrate egress observed after 1600 ps RAMD simulation.

Figure 2 (see next page): *CYP2D6-substrate hydrogen bond interactions (left panels) and C_α root-mean-square-fluctuations (RMSFs, right panels) during the RAMD simulations belonging to egress pathways pw2a, pw2c, and the solvent channel. H-bond and RMSF patterns belonging to the same simulation have the same color. All protein-substrate H bonds are interactions between the E216 carboxylate group and the protonated nitrogen atom of the substrate, with the exception of the second part of the DXM.IIIb simulation (pw2c) where the hydrogen bond switches to the OE1 atom of Q244. The position of secondary structure elements (A helix, BC loop, F and G helices, FG loop, β_4 sheet) and residues showing the largest RMSF fluctuations as a result of substrate egress are indicated.*
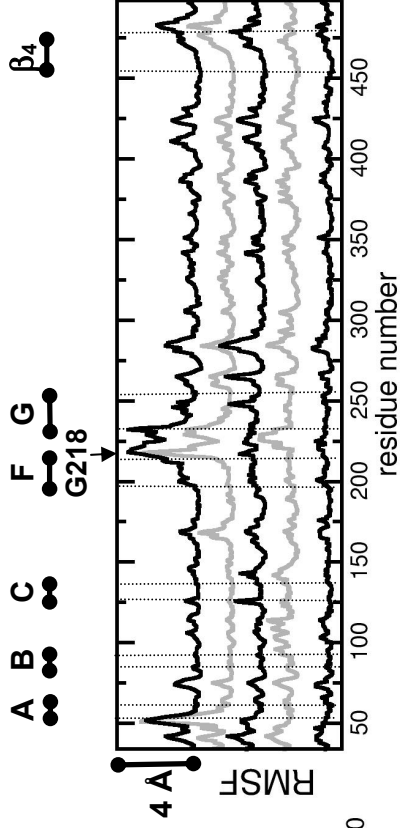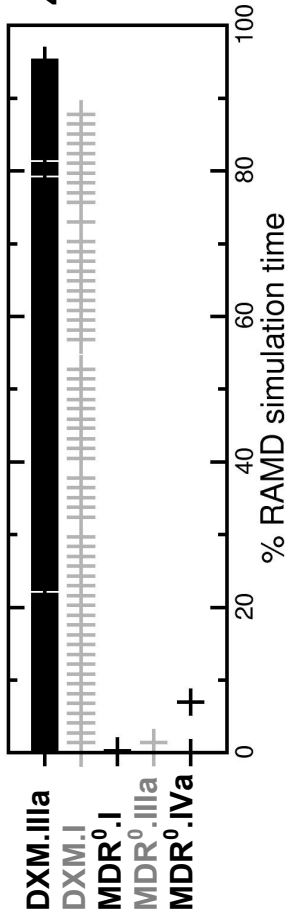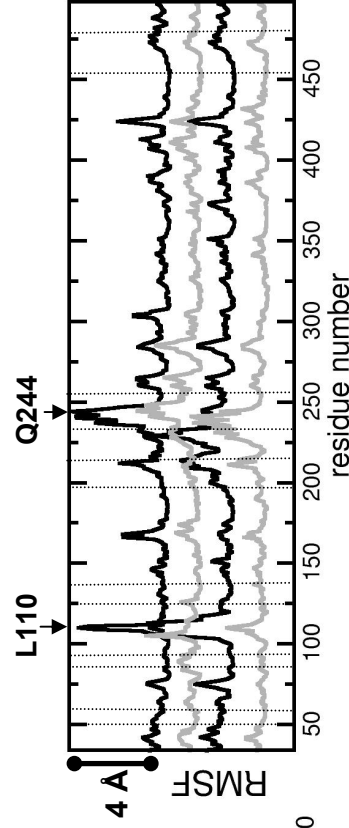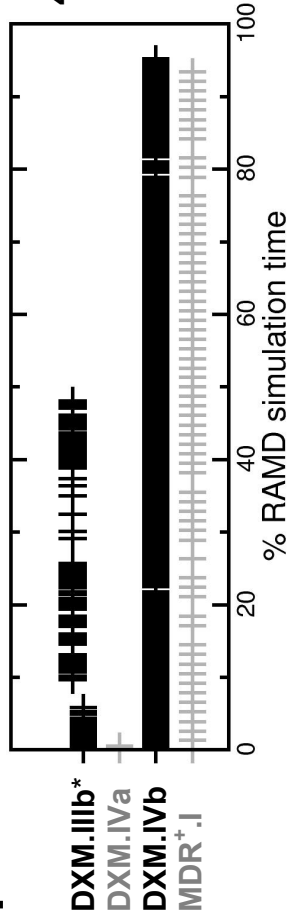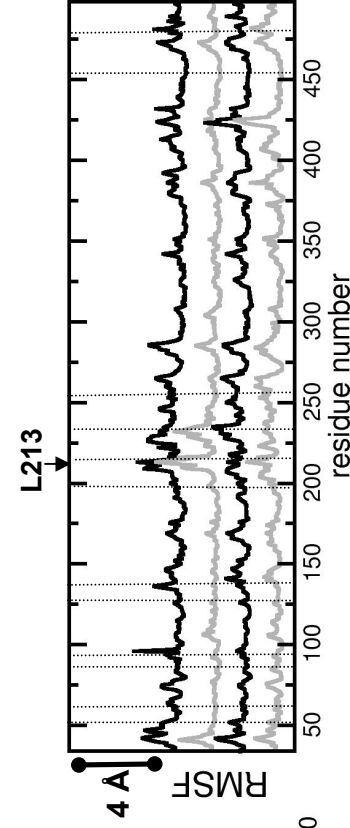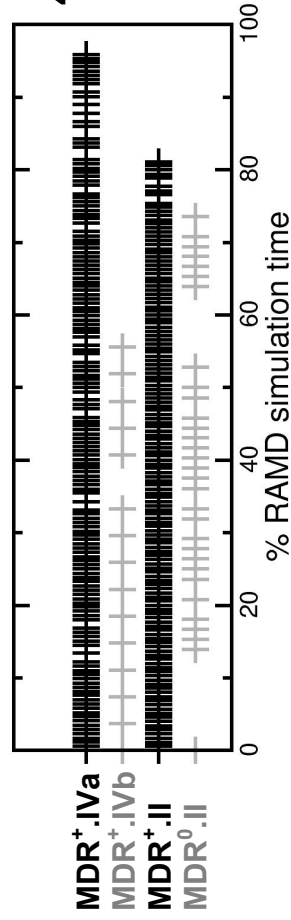
The ligand exit trajectories were clustered into pws based on the secondary structure elements lining them, as defined in earlier RAMD studies (see **Table 1**).[12,16,18,23,24] Two variations on pathways 2a and 3, could be observed and these variations seem to be rather substrate-specific. The exit region for pathway 2a.I (used by DXM in wild-type CYP2D6) is located between the A helix and the F-G loop, and is also observed in RAMD studies performed with palmitoleic acid in CYP102.[16] The exit region for pathway 2a.II (used by MDR[0]) is located between the B′-B loop and the F-G loop, and is also observed in RAMD studies performed with camphor in CYP101[16] and 6-DEB in CYP107.[18] The exit regions of pathways 3.I (used by DXM and MDR[+]) and 3.II (MDR[0]) are located between the F-G loop, and near the E/F loop, respectively.

In wild-type CYP2D6, there are four predominant pathways, pw2a (I and II), 2c, 3, and egress via the solvent channel. Also pw2e was observed. Pathways were found to be rather substrate-dependent: pw2c and 2a.I were predominant pathways for the relatively bulky and rigid dextromethorphan, the solvent channel[25] was the predominant pathway for the smaller and more flexible positively charged *R*-MDMA, while pw2a.II was the predominant pathway for neutral *R*-MDMA. For each substrate, only one simulation via pw3 was observed. For this reason, only egress pathways 2a, 2c and egress via the solvent channel are discussed below. **Figure 2** presents the CYP2D6-substrate hydrogen bond interactions and the C_α root-mean-square-fluctuations (RMSFs) during the RAMD simulations belonging to these three pathways.

The carboxylate group of E216 is believed to form a salt bridge with positively charged nitrogen atoms in typical CYP2D6 substrates.[26,27] The equilibrated starting structures of the RAMD simulations also show this interaction between E216 and *R*-MDMA and dextromethorphan and during most of the substrate egress pathways, the ligand-protein salt bridge is so stable that the initially buried aromatic moiety off the substrate is expelled first and the substrate is only able to leave the protein after breakage of this salt link (**Figure 2**). All simulations show some movement of the HI loop, and most show movement of the L′L loop, which probably can be attributed to the normal global dynamics of the CYP2D6 homology structure. Depending on the pathway, movements are observed in the A helical region, BC loop/B′ helix, F and G helices and FG loop region, and β_4 sheet region, which accommodate the substrate egress process (**Figure 2**). These movements seem to be rather large, and could be an artifact of the MD simulations.[16] Large plasticity of the BC loop and FG loop regions has, however, been observed for other CYPs, e.g. by ligand-induced conformational changes[28,29] or even by an interaction of the FG loop of another CYP protein within the active site.[30]

Egress via pathway 2c involves substrate-protein and intraprotein H-bond breakage

Substrate egress along pw2c was observed to occur via two different mechanisms. According to both mechanisms, final substrate exit occurs via the G helix (L241, Q244, F247) and the B′ helix (I109, L110) (**Table 4** and **Figure 3**), but they differ in the substrate-protein H bond interaction pattern along the egress pathway. The location of the substrate exit pathway and the structural changes occurring during the simulation of DXM using RAMD parameters IIIb (denoted as DXM.IIIb from now on) are shown in **Figure 3**. In simulations DXM.IIIb and DXM.IIIa, the H-bond between the protonated nitrogen atom of the substrate and the E216 carboxylate group is broken in an early stage of the simulation (**Figures 2** and **3**). The positively charged nitrogen moves towards a polar subpocket at the edge of the active site, formed by Q244 (G-helix) and D301 and S304 (I-helix). In the DXM.IIIb simulation, the protein-substrate H-bond is maintained via interaction with Q244, resulting in the loss of an internal H bond between Q244 and the backbone of L110 (**Figures 2** and **3**). In both simulations, the protonated nitrogen is

within long-range electrostatic interaction distance (<6 Å) of the negatively charged D301 carboxylate group.
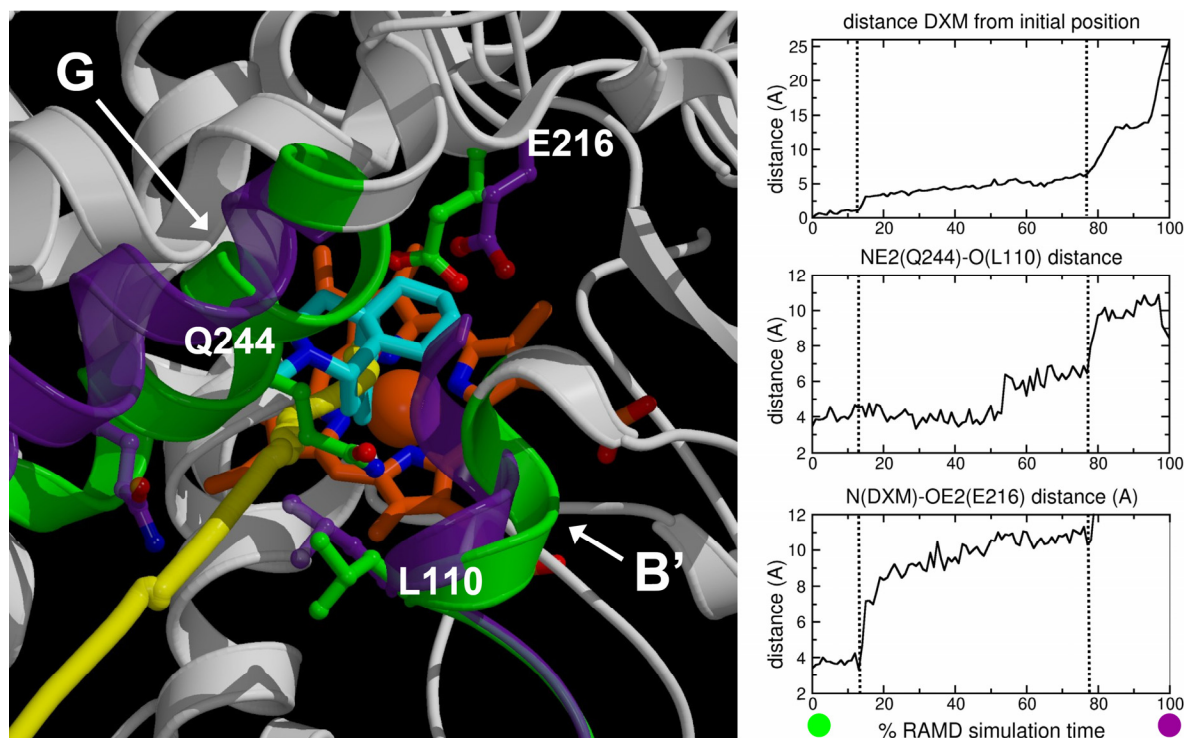
Figure 3: *Egress of dextromethorphan (cyan sticks) via pw2c (following the yellow line), accompanied by breakage of a dextromethorphan(DXM)-E216 H-bond and an internal H-bond between Q244 and the backbone of L110, large movements in the G helix and minor movement of the B' helix. Substrate exit occurs via these two residues (depicted in ball-and-stick) and the secondary structures of the G and B' helices at the beginning (green) and at the end (purple) of the RAMD simulation are highlighted. The view in this figure is rotated ca. 90⁰ about the horizontal axis with respect to the view in* **Figure 1**.

Substrate egress along pw2c involves large movements in the G-helix (**Figure 2**). Final substrate exit occurs via the FG loop region (G218, E222) and the N-terminal region after the A helix (F51, Q52) and/or the β_1 sheet (A74) (**Table 4** and **Figure 3**).

According to an alternative pw2c egress mechanism (simulations DXM.IIIb and MDR⁺.I), the H bond interaction between the substrate and E216 is maintained throughout the simulation, and only after disruption of this H-bond, the substrate is able to leave the protein. Substrate egress is accompanied by large movement of the FG loop region and the B' helix (**Figure 2**).

Based on these RAMD simulations, a double mutant of CYP2D6 was constructed, in which L110 (in the B' helix) and Q244 (in the G-helix) were both mutated into a cysteine, allowing for the formation of a cysteine bridge between the B'- and G-helices. This cysteine bridge in principle blocks pw2c, but also the mutant RAMD-simulations show egress via pw2c, although in these simulations large distortions of secondary structure elements were observed. Cross-linking of C29 and C194 with bis-maleimidohexane in Y29C/A194C double mutant CYP101 showed that these residues are important for modulating substrate access and exit in an earlier study.[4] Experimental site-directed mutagenesis studies to block pw2c and the other identified egress pathways in CYP2D6 (i.e. pw2a and the solvent channel) can be designed to validate these simulations. Evidence for ligand passage via pw2c in other members of the CYP2 subfamily is provided

223

Pw2c is not observed to be open in the substrate-bound CYP2C5 crystal structures (PDB: 1N6B[21] and 1NR6[22]) used as templates for the construction of the CYP2D6 homology model. The pathway is however open to a water probe in the lower resolution substrate-free structure of CYP2C5 (PDB: 1DT6),[32] as well as in the recently solved CYP2D6 crystal structure.[33] (RA)MD simulations identified pw2c as a potential substrate access/product exit pathway in CYP101[16,17] and 107,[18] and suggested pw2c to be the major pathway for CYP2C5[23] and 2B1[19] (**Table 1**).

Egress via pathway 2a involves substrate-protein hydrogen-bond breakage and rotation of aromatic side-chains

The mechanism of substrate egress along pw2a is dependent of the substrate nitrogen formal charge. The location of the substrate exit pathway and the structural changes occurring during simulation DXM.IIIa are shown in **Figure 4**. In simulations with positively charged dextromethorphan (DXM.I and DXM.IIIA), the H-bond/ionic salt link between the protonated nitrogen atom of the substrate and the E216 carboxylate group is maintained throughout the simulation (**Figures 2** and **4**). F483 blocks the pw2a exit and forms aromatic interactions with the substrate. Only after reorientation of the phenyl ring of this residue and disruption of the dextromethorphan-E216 H-bond, the substrate is able to leave the protein. The side chains of F483, and, in the final substrate exit stage, F51, act as doors that are pushed open via transitions of the side-chain torsions χ_1 and χ_2 (**Figure 4**). Flipping of aromatic residues was also observed in RAMD studies with CYP101.[16] In the homology model starting structure, the phenyl-ring of F483 is turned towards the active site, maximizing hydrophobic contacts with the substrate, while in the RAMD snapshot facilatiting substrate exit (**Figure 4**), as well as in the recently solved CYP2D6 crystal structure,[33] F483 forms an intermediate substrate binding pocket with E216 at the entrance of the solvent channel. Substrate egress is accompanied by large movements in the F helix and FG loop, and smaller movements in the β_4 sheet (**Figure 2**). Final substrate exit occurs via the FG loop region (G218, E222) and the N-terminal region after the A helix (F51, Q52) and/or the β_1 sheet (A74) (**Table 4** and **Figure 4**).

In simulations with neutral *R*-MDMA (MDR^0.I, MDR^0.IIIa and MDR^0.IVa) that make use of pw2a, the H-bond/ionic salt link between the protonated nitrogen atom of the substrate and the E216 carboxylate group is already broken at the beginning of the simulation (**Figure 2**). Final substrate exit occurs via the FG loop region (N221, R225) and the BB'-loop region (P103, V104) (**Table 4**) and subtrate egress is accompanied by relatively smaller structural changes, compared to simulations with positively charged substrates (**Figure 2**).

Site directed mutagenesis studies support a substrate access/product exit channel in the region of pw2a in CYP101[4,5], but not in CYP2B1.[7,19] In the crystal structures of CYP101, 102, 108, 119, 152, 154, 55, 2B4, 2C5, 2C9, and 3A4, pw2a is open to a water probe.[3] Molecular wires have been used to trap the open pw2a conformational states of CYP101[34,35], and in the recently determined CYP2B4 structure pw2a can even open wide enough to allow another protein (the other P450 molecule in the crystallographic dimer) to penetrate into the active site.[30] (RA)MD simulations in CYP101[16,17], 102[14-17], 107[18], and 2C9[20] suggest pw2a as the major substrate egress pathway for these CYPs (**Table 1**).
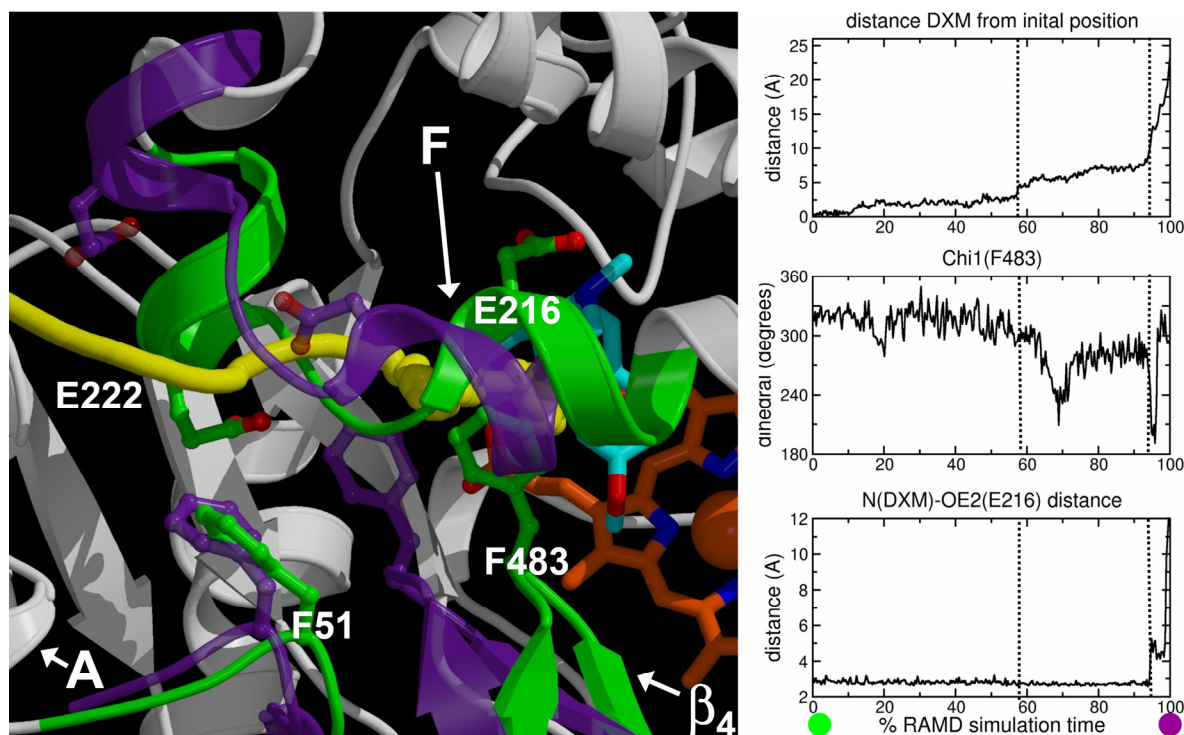
Figure 4: *Egress of dextromethorphan (cyan sticks) via pw2a.I (following the yellow line), accompanied by breakage of dextromethorphan-E216 H-bond, reorientation of F483, and movements in the F helix, FG loop, and the β_4 sheet. Final substrate exit occurs via E222 and F51. The secondary structures of the F helix, FG loop, the β_4 sheet, and the loop between the A helix and the β_1 sheet, as observed at the beginning (green) and at the end (purple) of the RAMD simulation, are highlighted. The view in this figure is rotated ca. 90^0 about the horizontal axis and 180^0 about the vertical axis with respect to the view in* **Figure 1***.*

Egress via the solvent channel involves substrate-ligand hydrogen-bond breakage and rotation of aliphatic side-chains

Like many CYP crystal structures, the CYP2D6 homology model has a small opening between the F and I helices and the β_4 sheet, referred to as the solvent channel.[25] The channel has been suggested to be important for controlling proton access to the active site, and in the current CYP2D6 RAMD it is identified as a potential substrate egress pathway. The location of the substrate exit pathway and the structural changes occurring during simulation MDR$^+$.IVa are shown in **Figure 5**.

The charged E216 residue is positioned in close vicinity to the opening of the solvent channel. In all RAMD trajectories along the solvent channel, the H-bond/ionic salt link between the protonated nitrogen atom of the substrate and the E216 carboxylate group is maintained throughout almost the whole simulation (**Figures 2** and **5**).

Substrate egress involves movements in the F-helix and the FG loop, but these movements are generally smaller than those observed during substrate egress via pw2a and pw2c (**Figure 2**). Final substrate exit occurs via the I helix (V308, M309, T312), the F helix and FG loop region (L206, A209, Q210, L213, E222) and the β_4 sheet (L484). Residue L484 rotates upon the release of the substrate (**Table 4** and **Figure 5**).

In the crystal structures of CYP102, 119, 175, 2C5, 2C8, 2C9, 2D6 and 3A4, the solvent channel is open to a water probe,[3] and (RA)MD simulations in CYP2B1[19], 2C5[23], 2C9[20], and 6B1[1] suggest the solvent channel as a potential substrate egress pathway for these CYPs (**Table 1**).
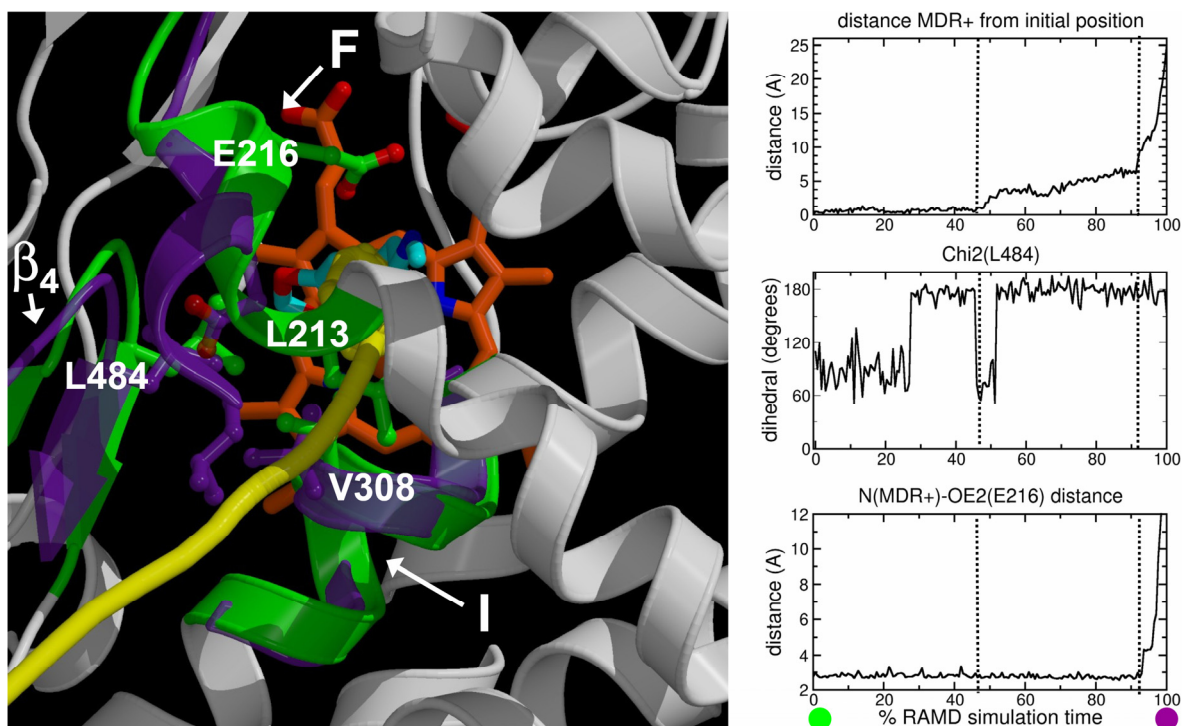
Figure 5: *Egress of positively charged R-MDMA (cyan sticks) via the solvent channel (following the yellow line), accompanied by reorientation of L484, movements in the F helix and FG loop, and loss of an ionic salt link between the substrate and E216 (in ball-and-stick). Substrate exit occurs via L213 in the F helix, V308 in the I helix, and L484 in the β_4 sheet. The secondary structures of the F helix, the FG loop, the β_4 sheet, and the I helix, as observed at the beginning (green) and at the end (purple) of the RAMD simulation, are highlighted. The view in this figure is rotated ca. 90^0 about the horizontal axis and 120^0 counter clockwise about the vertical axis with respect to the view in **Figure 1**.*

Electrostatic steering by E216

For all three dominant egress pathways (pw2a, pw2c and the solvent channel) in CYP2D6, and also for most of the other pws observed in the current RAMD studies, E216 plays an important role in the mechanism of substrate egress. Only in pw2c it seems more likely that the substrate/product interacts with another polar binding pocket in the vicinity of D301 and Q244 at the edge of the active site, after entering or before leaving the protein. Trajectories along pw2c show that pronounced ligand-E216 H-bond interactions involve larger distortion of the FG loop and B' helix (**Figure 2**). Based on the analysis of the RAMD simulations along pw2c and the solvent channel, a common mechanism of substrate entrance or product release can be proposed for these pathways. The positively charged nitrogen atom of the substrate is electrostatically directed to the side-chain of E216 and a stable salt link is formed. The apolar moiety of the substrate diffuses through pw2a or the solvent channel involving shorter ranged hydrophobic interactions. Both pw2a and the solvent channel are flanked by hydrophobic residues, presumably serving as the substrate-docking region. Also based on the currently solved CYP2D6 crystal structure, the substrates have been proposed to dock into such a hydrophobic subpocket (between E216 and F483) before entering the active site.[33] During the egress process along pw2a and the solvent channel, substrate/product egress is accompanied by rotation of aromatic and aliphatic residues, but generally also by large displacement of segments in the BC loop, FG loop and F and G helices. These large structural distortions could be an artifact of the MD simulations,[16] but large flexibility of the BC loop and FG loop regions has been observed for other CYPs, e.g. by ligand-induced conformational changes[28,29] or even by interactions of the FG loop of another CYP protein within the active site.[30] The series of RAMD

simulations performed with neutral *R*-MDMA (MDR^0) show that by removing the salt link between the substrate and E216 (i.e. in the absence of electrostatic steering) other egress pathways are used, involving less distortion of secondary structure elements. This suggests that substrate entrance and product exit in CYP2D6 depends of the substrate/product formal charge.

Earlier experimental studies already supported the idea that the binding mode[36] and affinity[37] of ligands in the active site of CYP2D6 is not only dependening on the charge of substrate binding residues,[27] but that it is also dependent on the pK$_a$ of the ligand. The apparent pK$_a$ for substrates was even reported to decrease upon binding to CYP2D6 (making the substrate less basic).[36] In **chapter 6** we described simulations of MDMA in a charged and in a neutral state. There, it appeared that using the neutral parameters to describe MDMA were better suited to reproduce the experimental NMR data. For this reason, a neutral description of MDMA was taken into account here as well, Interestingly, the neutral form of MDMA again results in the least distortive simulation of the protein. Note, however, that these findings may only give an indication of the quality of the force field parameters and not necessarily imply that MDMA binds in a neutral form to CYP2D6. In summary, our RAMD studies suggest that E216 is not only crucial for substrate binding in the active site,[27] but that it also plays an important role in directing the substrate towards the active site via electrostatic steering.

Substrate passage through the membrane and docking to a hydrophobic subpocket

In vivo, CYP2D6, like other mammalian CYPs, is attached to a lipid membrane. It has been hypothesized that in soluble, bacterial CYPs, pw2a is used as both substrate access and product egress route and that the process of channelling through this passage may contribute to substrate recognition.[18] For membrane bound CYPs, other mechanisms for substrate access/product exit have been proposed: (i) a one-way route by which hydrophobic substrates access the enzyme from the membrane by pw2a and hydroxylated products egress along pw2c or pw2e; and (ii) a two-way route for access and egress, along pw2c, for soluble compounds.[2,3,23]

Trypsin digestion studies in CYP1A2,[38,39] antibody accessibility studies in CYP2B1[40] and CYP2C5[32], and recent tryptophan fluorescence scanning studies in CYP2C2[41] indicate that the following parts of mammalian CYPs are inserted into the membrane: the N-terminus untill the A-helix (corresponding to (approximately) residues 1-50 in CYP2D6), after the A-helix (~65-75), in β sheet 2-2 (~390-395), and the F-G loop (~220-230). This suggests that the residues identified in the current RAMD studies to be located around the entrance of pw2a.I and (most of the residues located) around the entrance of pw2a.II are either inserted into the lipid bilayer, or near or within the phospholipid headgroup region of the membrane. The region around the entrances of pws 2c, 2e, 3.I, and 3.II are expected to be not embedded in the membrane. Lipophilic substrates in the membrane might therefore be extracted into the CYP2D6 active site by pw2a and expelled in their hydroxylated form by pw2c and/or the solvent channel. The current RAMD studies in CYP2D6 also support this hypothesis by indicating pathways 2a and 2c as the most important pathways, with pathway 2a being predominantly used by the neutral (MDR^0) and thus more hydrophobic substrate. In addition, the solvent channel is identified as a potential egress pathway for hydrophobic ligands. These finding do, however, imply that for a proper description of the dynamics of substrate entrance and product exit in CYP2D6, protein-membrane (and substrate-membrane) interactions should be taken into account.

Analysis of the CYP51 crystal structure suggests that pw2e must close in order for pw2a to open, thus providing synchronization between substrate access and product egress via different routes.[42] Based on the CYP152A crystal structure, pw2e is proposed to permit water to escape from the active site after its generation in the peroxygenation reaction.[43]

The 2e pathway has also been observed in the current RAMD studies in CYP2D6, but only in 2 out of 18 simulations in the wild-type protein. Therefore, we propose pw2c and the solvent channel to be potential access and exit channels for hydrophylic substrates and products, respectively, while pw2a may serve as substrate access channel for hydrophobic substrates entering the protein from the membrane.

Perspective: Experimental validation of RAMD studies

From the current CYP2D6 RAMD studies hypotheses can be derived which can be tested experimentally, such as site-directed mutagenesis studies of non-active site residues lining the substrate egress pathways. In the L110C/Q244C mutant a cysteine brigde can be formed, possibly blocking product release through pw2c. A L110C/Q244C mutant potentially forming a cysteine brigde was proposed to block pw2c, but also for the other identified egress pathways in CYP2D6 (i.e. pw2a and the solvent channel) such mutants can be proposed. In addition, experimental approaches such as (time-resolved) fluorescence spectroscopy, measuring the emission of tryptophan residues (and of the substrate), can be used to analyze possible local conformational changes in CYPs upon substrate binding.[9-11] In fact, this method has recently been applied to CYP2D6, showing contraction of the protein caused by substrate binding.[11] Mutagenesis of more tryptophan residues will, however, be needed to localize the flexible parts of CYP2D6.

A complete interpretation of these mutations would also require the calculation of free energy profiles for ligand passage, as was done by using steered MD and adiabatic mapping for CYP101 and CYP102.[17] Nevertheless, the egress pathways identified in the current RAMD studies are in line with earlier RAMD work on CYP crystal structures. Furthermore, we have observed a distribution of possible egress pathways, which is different from the distribution of pathways in CYP2C5 (the template used to construct the CYP2D6 homology model). This suggests that our data are not biased by the homology modeling template, but rather a result of the specific structural and chemical characteristics of CYP2D6 and its substrates.

8.3 Conclusions

We have used RAMD simulations to investigate the routes and mechanisms via which substrates may enter and products may exit the active site of CYP2D6. Three dominant pathways were identified, and based on detailed analysis of structural changes within the protein and protein-ligand interactions, molecular mechanisms have been proposed for substrate egress via these pathways. E216 was shown to play an important role in substrate egress along pw2a and the solvent channel. After entering or before leaving the protein via pw2c, substrates/products are more likely to interact with another polar binding pocket in the vicinity of D301 and Q244 at the edge of the active site. This suggests that E216 is not only crucial for substrate binding in the active site, as was shown in earlier studies, but that it is also involved in directing the substrate towards the active site via electrostatic steering. Substrate/product egress via the three pathways is furthermore accompanied by large displacement of segments in the BC loop, FG loop and F and G helices. Egress via pw2a involves rotation of aromatic residues, egress via the solvent channel involves the rotation of aliphatic residues, while egress via pw2c is accompanied by loss of an internal protein H-bond.

We propose the following general mechanism for substrate access into and product exit from CYP2D6: Hydrophobic substrates coming from the membrane enter the protein via pw2a, and their hydrophilic products leave via pw2c and/or the solvent channel, while soluble substrates enter (and their products leave) via pw2c and/or the solvent channel. From our RAMD studies hypotheses can be derived which can be tested experimentally, such as site-directed mutagenesis studies of non-active site residues lining the substrate egress pathways.

8.4 Methods

The simulations of CYP2D6 were based on a refined and validated homology model (Chapter 3). The energetically most favorable docking poses of DXM, MDR^0, and MDR$^+$, according to a previously reported automated docking protocol (chapters 4 and 5),[44,45] were used as starting configurations for the MD-simulations.

The enzyme complexes, was solvated in periodic boxes containing TIP3P water molecules.[46] In addition four Na$^+$ ions were added to neutralize the system, placed in energetically favourable locations using the GRID program.[47] Interactions were calculated according to the AMBER99 force field,[48] using particle-mesh-ewald summation to include the longrange electrostratic forces. Van der Waals interactions were calculated using a cutoff of 8.5 Å. A modified version of the AMBER program (Version 8, http://amber.scripps.edu) was used, including the RAMD routines.[16,18] Heme parameters were kindly supplied Dr. Danni Harris.[49] Substrate force-field parameters were derived using the antechamber program[50] and partial charges for the substrates were derived using the RESP procedure described in ref [51]. The systems were energy-minimized for 200 steps and then subjected to MD simulation at 300 K and 1 bar for 1 ns, using the Berendsen approach to keep the temperature and the pressure constant.[52] The RAMD simulations were started from the coordinates obtained after the 1 ns equilibration simulation, when the systems were well equilibrated as indicated by the backbone root-mean-square devitation. RAMD trajectories were analysed using the tools provided in the AMBER, GROMACS,[53] and GROMOS[54] packages.

References

(1)   Wen, Z.; Baudry, J.; Berenbaum, M. R.; Schuler, M. A. Ile115Leu mutation in the SRS1 region of an insect cytochrome P450 (CYP6B1) compromises substrate turnover via changes in a predicted product release channel. *Protein Eng Des Sel* **2005**, *18*, 191-199.
(2)   Wade, R. C.; Motiejunas, D.; Schleinkofer, K.; Sudarko; Winn, P. J. et al. Multiple molecular recognition mechanisms. Cytochrome P450--a case study. *Biochim Biophys Acta* **2005**, *1754*, 239-244.
(3)   Wade, R. C.; Winn, P. J.; Schlichting, I.; Sudarko A survey of active site access channels in cytochromes P450. *J Inorg Biochem* **2004**, *98*, 1175-1182.
(4)   Gerber, N. C. Bioorganic activation of cytochrome P450cam; University of Illinois: Urbana, USA, 1994.
(5)   Mueller, E.; Loida, P.; Sligar, S. *in Cytochrome P450: Structure, Mechanism, and Biochemistry*; Plenum Press: New York and London, 2003; pp 83-125.
(6)   Honma, W.; Li, W.; Liu, H.; Scott, E. E.; Halpert, J. R. Functional role of residues in the helix B' region of cytochrome P450 2B1. *Arch Biochem Biophys* **2005**, *435*, 157-165.
(7)   Scott, E. E.; He, Y. Q.; Halpert, J. R. Substrate routes to the buried active site may vary among cytochromes P450: mutagenesis of the F-G region in P450 2B1. *Chem Res Toxicol* **2002**, *15*, 1407-1413.
(8)   Kronbach, T.; Johnson, E. F. An inhibitory monoclonal antibody binds in close proximity to a determinant for substrate binding in cytochrome P450IIC5. *J Biol Chem* **1991**, *266*, 6215-6220.
(9)   Prasad, S.; Mazumdar, S.; Mitra, S. Binding of camphor to Pseudomonas putida cytochrome p450(cam): steady-state and picosecond time-resolved fluorescence studies. *FEBS Lett* **2000**, *477*, 157-160.
(10)   Prasad, S.; Mitra, S. Role of protein and substrate dynamics in catalysis by Pseudomonas putida cytochrome P450cam. *Biochemistry* **2002**, *41*, 14499-14508.
(11)   Stortelder, A.; Keizers, P. H.; Oostenbrink, C.; De Graaf, C.; De Kruijf, P. et al. Binding of 7-methoxy-4-(aminomethyl)-coumarin to wild-type and W128F mutant cytochrome P450 2D6 studied by time-resolved fluorescence spectroscopy. *Biochem J* **2006**, *393*, 635-643.
(12)   Lüdemann, S. K.; Carugo, O.; Wade, R. C. Substrate access to Cytochrome P450cam: A comparison of a thermal motion pathway analysis with molecular dynamics simulation data. *J Mol Mod* **1997**, *3*, 369-374.
(13)   Arnold, G. E.; Ornstein, R. L. Molecular dynamics study of time-correlated protein domain motions and molecular flexibility: cytochrome P450BM-3. *Biophys J* **1997**, *73*, 1147-1159.
(14)   Chang, Y. T.; Loew, G. H. Molecular dynamics simulations of P450 BM3--examination of substrate-induced conformational change. *J Biomol Struct Dyn* **1999**, *16*, 1189-1203.
(15)   Paulsen, M. D.; Ornstein, R. L. Dramatic Differences in the Motions of the Mouth of Open and Closed Cytochrome P450bm-3 by Molecular-Dynamics Simulations. *Proteins* **1995**, *21*, 237-243.
(16)   Ludemann, S. K.; Lounnas, V.; Wade, R. C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. *Journal of Molecular Biology* **2000**, *303*, 797-811.

(17)   Ludemann, S. K.; Lounnas, V.; Wade, R. C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways. *Journal of Molecular Biology* **2000**, *303*, 813-830.

(18)   Winn, P. J.; Ludemann, S. K.; Gauges, R.; Lounnas, V.; Wade, R. C. Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine. *PNAS* **2002**, *99*, 5361-5366.

(19)   Li, W.; Liu, H.; Scott, E. E.; Grater, F.; Halpert, J. R. et al. Possible pathway(s) of testosterone egress from the active site of cytochrome P450 2B1: a steered molecular dynamics simulation. *Drug Metab Dispos* **2005**, *33*, 910-919.

(20)   Seifert, A.; Tatzel, S.; Schmid, R. D.; Pleiss, J. Multiple molecular dynamics simulations of human p450 monooxygenase CYP2C9: The molecular basis of substrate binding and regioselectivity toward warfarin. *Proteins* **2006**, *64*, 147-155.

(21)   Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dansette, P. M.; Mansuy, D. et al. Structure of a substrate complex of mammalian cytochrome P450 2C5 at 2.3 A resolution: evidence for multiple substrate binding modes. *Biochemistry* **2003**, *42*, 6370-6379.

(22)   Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dijols, S.; Dansette, P. M. et al. Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 A resolution: evidence for an induced fit model of substrate binding. *Biochemistry* **2003**, *42*, 9335-9345.

(23)   Schleinkofer, K.; Sudarko; Winn, P. J.; Ludemann, S. K.; Wade, R. C. Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling? *EMBO Rep* **2005**, *6*, 584-589.

(24)   Lüdemann, S. K.; Lounnas, V.; Wade, R. C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. *J Mol Biol* **2000**, *303*, 797-811.

(25)   Haines, D. C.; Tomchick, D. R.; Machius, M.; Peterson, J. A. Pivotal role of water in the mechanism of P450BM-3. *Biochemistry* **2001**, *40*, 13456-13465.

(26)   Vermeulen, N. P. E. Prediction of drug metabolism: the case of cytochrome P450 2D6. *Curr Top Med Chem* **2003**, *3*, 1227-1239.

(27)   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.

(28)   Raag, R.; Li, H.; Jones, B. C.; Poulos, T. L. Inhibitor-induced conformational change in cytochrome P-450CAM. *Biochemistry* **1993**, *32*, 4571-4578.

(29)   Cupp-Vickery, J. R.; Garcia, C.; Hofacre, A.; McGee-Estrada, K. Ketoconazole-induced conformational changes in the active site of cytochrome P450eryF. *J Mol Biol* **2001**, *311*, 101-110.

(30)   Scott, E. E.; He, Y. A.; Wester, M. R.; White, M. A.; Chin, C. C. et al. An open conformation of mammalian cytochrome P450 2B4 at 1.6-A resolution. *Proc Natl Acad Sci U S A* **2003**, *100*, 13196-13201.

(31)   Melet, A.; Marques-Soares, C.; Schoch, G. A.; Macherey, A. C.; Jaouen, M. et al. Analysis of human cytochrome P450 2C8 substrate specificity using a substrate pharmacophore and site-directed mutants. *Biochemistry* **2004**, *43*, 15379-15392.

(32)   Williams, P. A.; Cosme, J.; Sridhar, V.; Johnson, E. F.; McRee, D. E. Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol Cell* **2000**, *5*, 121-131.

(33)   Rowland, P.; Blaney, F. E.; Smyth, M. G.; Jones, J. J.; Leydon, V. R. et al. Crystal structure of human cytochrome P450 2D6. *J Biol Chem* **2006**, *281*, 7614-7622.

(34)   Dunn, A. R.; Dmochowski, I. J.; Bilwes, A. M.; Gray, H. B.; Crane, B. R. Probing the open state of cytochrome P450cam with ruthenium-linker substrates. *Proc Natl Acad Sci U S A* **2001**, *98*, 12420-12425.

(35)   Hays, A. M.; Dunn, A. R.; Chiu, R.; Gray, H. B.; Stout, C. D. et al. Conformational states of cytochrome P450cam revealed by trapping of synthetic molecular wires. *J Mol Biol* **2004**, *344*, 455-469.

(36)   Miller, G. P.; Hanna, I. H.; Nishimura, Y.; Guengerich, F. P. Oxidation of phenethylamine derivatives by cytochrome P450 2D6: the issue of substrate protonation in binding and catalysis. *Biochemistry* **2001**, *40*, 14215-14223.

(37)   Upthagrove, A. L.; Nelson, W. L. Importance of amine pKa and distribution coefficient in the metabolism of fluorinated propranolol analogs: metabolism by CYP1A2. *Drug Metab Dispos* **2001**, *29*, 1389-1395.

(38)   Brown, C. A.; Black, S. D. Membrane topology of mammalian cytochromes P-450 from liver endoplasmic reticulum. Determination by trypsinolysis of phenobarbital-treated microsomes. *J Biol Chem* **1989**, *264*, 4442-4449.

(39)   Dong, M. S.; Bell, L. C.; Guo, Z.; Phillips, D. R.; Blair, I. A. et al. Identification of retained N-formylmethionine in bacterial recombinant mammalian cytochrome P450 proteins with the N-terminal sequence MALLLAVFL...: roles of residues 3-5 in retention and membrane topology. *Biochemistry* **1996**, *35*, 10031-10040.

230

(40)   De Lemos-Chiarandini, C.; Frey, A. B.; Sabatini, D. D.; Kreibich, G. Determination of the membrane topology of the phenobarbital-inducible rat liver cytochrome P-450 isoenzyme PB-4 using site-specific antibodies. *J Cell Biol* **1987***, 104*, 209-219.

(41)   Ozalp, C.; Szczesna-Skorupa, E.; Kemper, B. Identification of membrane-contacting loops of the catalytic domain of cytochrome P450 2C2 by tryptophan fluorescence scanning. *Biochemistry* **2006***, 45*, 4629-4637.

(42)   Podust, L. M.; Poulos, T. L.; Waterman, M. R. Crystal structure of cytochrome P450 14alpha -sterol demethylase (CYP51) from Mycobacterium tuberculosis in complex with azole inhibitors. *Proc Natl Acad Sci U S A* **2001***, 98*, 3068-3073.

(43)   Lee, D. S.; Yamada, A.; Sugimoto, H.; Matsunaga, I.; Ogura, H. et al. Substrate recognition and molecular mechanism of fatty acid hydroxylation by cytochrome P450 from Bacillus subtilis. Crystallographic, spectroscopic, and mutational studies. *J Biol Chem* **2003***, 278*, 9761-9767.

(44)   de Graaf, C.; Oostenbrink, C.; Keizers, P. H.; van der Wijst, T.; Jongejan, A. et al. Catalytic Site Prediction of and Virtual Screening Accuracy of Cytochrome P450 2D6 Substrates by Consideration of Water and Rescoring in Automated Docking. *J Med Chem* **2006***, 49*, 2417-2430.

(45)   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.

(46)   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **1983***, 79*, 926-935.

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

(48)   Cheatham, T. E., 3rd; Cieplak, P.; Kollman, P. A. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J Biomol Struct Dyn* **1999***, 16*, 845-862.

(49)   Harris, D. L.; Park, J. Y.; Gruenke, L.; Waskell, L. Theoretical study of the ligand-CYP2B4 complexes: effect of structure on binding free energies and heme spin state. *Proteins* **2004***, 55*, 895-914.

(50)   Wang, J. M.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comp Chem* **2000***, 21*, 1049-1074.

(51)   Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz Jr., K. M. et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* **1994***, 117*, 5179-5197.

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

(53)   Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* **2001***, 7*, 306-317.

(54)   Christen, M.; Hunenberger, P. H.; Bakowies, D.; Baron, R.; Burgi, R. et al. The GROMOS software for biomolecular simulation: GROMOS05. *J Comput Chem* **2005***, 26*, 1719-1751.

Summary, conclusions, and perspectives

Chapter 9

9.1 Summary

The primary aim of the investigations described in this thesis was to obtain knowledge on the structure and dynamics of CYP2D6-ligand interactions, enabling the rationalization and prediction of drug metabolism by this specific Cytochrome P450 enzyme, and offering new approaches for the study of CYP-ligand interactions in general. Our work particularly focused on three lines of research: **1**) The construction and validation of a high-quality CYP2D6 homology model; **2**) The development of methods for the prediction of CYP-ligand interactions in a high-throughput fashion; **3**) The description of the dynamics of CYP-ligand interactions by using experimental and *in silico* techniques in an integrative way. These three lines are addressed in the different chapters of this thesis.

Chapter 1 gives a general introduction into the background and context of the research presented in the thesis. Early consideration of not only drug potency, but also ADME (absorption, disposition, metabolism and excretion) properties and toxicity is increasingly seen as essential for efficient discovery and development of new drugs and drug candidates. Today, drug lead compounds are discovered and optimized by an integrative process of *in silico* modeling, *in vitro* screening, and *in vivo* testing. Drug metabolism primarily determines the fate of a compound entering the body. Absent, reduced or increased activity of enzymes involved in drug metabolism can lead to therapeutic failure or adverse drug effects. Understanding biotransformation enzymes and their mechanisms of action is therefore invaluable in the process of drug discovery and development. CYPs, and particularly CYP2D6, play an important role in drug metabolism. CYP2D6 is the second most important drug metabolizing enzyme after CYP3A4. Large interindividual differences exist in CYP2D6 activity due to genetic polymorphisms, further emphasizing its clinical importance. The early identification of potential CYP2D6 substrates and prediction of their metabolism is therefore advantageous in the discovery and development of new drugs. Computational chemistry is a most suitable tool to predict metabolic stability and sites of metabolism in potential drugs. Predictive *in silico* database approaches and quantitative structure-activity relationships (QSARs) derived from physicochemical (1D) and topological (2D) properties of CYP ligands provide robust but rough classifications of the metabolic stability and inhibition potency of substrates and inhibitors, respectively, without yielding much insight into the possible (molecular) background of the mode of action involved. The structure-based (3D) molecular modeling methods, on the other hand, aim to be insightful as well as predictive and were applied in this thesis to gather knowledge concerning the structure and dynamics of CYP2D6-ligand interactions. Of all human CYPs the CYP2D6 isoenzyme is a particularly suited model system to investigate the structure and dynamics of protein-ligand interactions.

Chapter 2 presents in an integrative manner structure-based computational approaches used to understand, rationalize and predict the activity and substrate selectivity of CYPs, as well as the possibilities and limitations of these approaches now and in the future.[1] One can divide structure-based models and methods into: i) ligand-based models (QM calculations on substrates, pharmacophore-, and QSAR-models), ii) protein-based models (crystal structures, homology models and various ways in which this structural information is analyzed and applied), and iii) ligand-protein interaction based models (e.g. automated docking and molecular dynamics simulations to describe ligand binding orientation, protein-ligand dynamics and ligand binding affinity). Apart from earlier mechanism-based *ab initio*

calculations on substrates, also pharmacophore modeling of ligands (*i.e.* both substrates and inhibitors) and protein homology modeling in combination with automated docking and molecular dynamics simulations have been used successfully for the rationalization and prediction of metabolite formation by several CYP isoenzymes. In the past six years, the number of available crystal structures, especially for mammalian and human CYP-isoforms, has increased significantly. In addition, the lack of detailed structural knowledge for many other isoforms is being overcome in part by the prediction of CYP structures using computer-aided homology modeling techniques in combination with small-molecule pharmacophore models and mechanism-based *ab initio* and semi-empirical calculations. A great variety of *in silico* modeling approaches have been applied to CYP enzymes, thereby successfully adding to our understanding of CYP structure and function in a way that is complementary to experimental studies. Another clear trend in many recent computational studies is the combination of multiple modeling techniques to arrive at meaningful and successful rationalizations of experimental data, and interpretations and predictions of CYP function.

Chapter 3 focuses on **research line 1**, i.e. the construction and validation of a high-quality CYP2D6 homology model. Experimentally, the active site of CYP2D6 can be probed by site-directed mutagenesis studies. Such studies can be designed from structural models of enzyme-substrate complexes and the observed effect of mutations on metabolism and inhibition of metabolism can subsequently be rationalized at an atomic resolution by modeling approaches. **Chapter 3** presents the construction, refinement and validation of the CYP2D6 homology model used in our laboratory for the prediction and rationalisation of CYP2D6 substrate metabolism and CYP2D6-ligand interactions. The protein model could explain reported site-directed mutagenesis data (for example, mutation of E216 and D301). Furthermore, based on the model, new CYP2D6 mutants were constructed and studied in our laboratory, and also for these mutants a rationalization of experimentally observed characteristics could be achieved (i.e. I106E, F120A, T309V, F483A). An analysis of the CYP2D6-substrate interaction fingerprint of docked substrates in the homology model suggests that several other active site residues are probably interacting with ligands as well, opening the way for further mutagenesis studies. Importantly, our homology model was found to agree with most of the details of the recently solved substrate-free CYP2D6 crystal structure. Structural differences between our homology model and the crystal structure were the same as the differences observed between substrate-free and substrate-bound structures of other CYPs, suggesting that these conformational changes are required upon substrate binding. The CYP2D6 crystal structure further validates our homology modeling approach and clearly demonstrates on one hand the success and accuracy of state of the art CYP homology modeling. On the other hand, we have shown that computational chemistry is a useful and valuable tool to provide models for substrate-bound complexes of CYPs which give insight into CYP-ligand interactions. This conclusion strongly supports the validitiy of homology models as a valuable approach for successful pre-experimental virtual screening, as well as accurate hypothesis generation for *in vitro* studies in drug discovery and development.

Chapters 4 and **5** involve **research line 2**, i.e. the development of improved methods for the prediction of CYP-ligand interactions in a high-throughput fashion. In **Chapter 4** 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.[2] 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. Considering 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.

In **Chapter 5**, the automated docking strategies successfully applied to binding mode predictions of ligands in CYP crystal structures in **chapter 4**, were used for the catalytic site prediction of 65 substrates in our refined and validated CYP2D6 homology model.[3] Consideration of water molecules at predicted positions in the active site and rescoring of pooled ligand docking results from four different docking programs (AutoDock, FlexX, GOLD-Goldscore, GOLD-Chemscore) with the SCORE scoring function enabled successful predictions of the experimentally reported sites of catalysis of more than 80% of the substrates. Three different database docking algorithms (FlexX, GOLD-Goldscore, GOLD-Chemscore) were subsequently used in combination with six scoring functions (Chemscore, DOCK, FlexX, GOLD, PMF, SCORE) to assess the ability of docking-based virtual screening methods to prioritize 20 known CYP2D6 substrates seeded into a drug-like chemical database of 980 compounds (in the absence and presence of active site water molecules). Finally, the optimal docking strategy in terms of virtual screening accuracy, GOLD-Chemscore with 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 (19,619 entries) and to discriminate between high affinity and medium affinity ligands.

Chapters 6, **7**, and **8** focus on **research line 3**, i.e. the description of the dynamics of CYP-ligand interactions by using several different experimental and *in silico* techniques (in an integrative way). In **Chapter 6** a series of 3,4-methylenedioxy-N-alkylamphetamines (MDAAs) were automatically docked and subjected to molecular dynamics (MD) simulations in our CYP2D6 protein model.[4] The predicted substrate binding orientations, sites of oxidation and relative reactivities were compared to experimental data of wild-type and F120A mutant CYP2D6. Automated docking results were not sufficient to accurately rationalize experimental binding orientations of 3,4-methylenedioxy-N-methylamphetamine (MDMA) in the two enzymes as measured with spin lattice relaxation NMR. However, the docking results could be used as starting structures for MD simulations. Predicted binding orientations of MDMA and sites of oxidation of the MDAAs derived from MD-simulations matched well with the experimental data. It appeared that the experimental results were best described in MD-simulations considering the nitrogen-atoms of the MDAAs in neutral form. Differences in regioselectivity and stereoselectivity in the oxidative metabolism of the MDAAs by the F120A mutant CYP2D6 were correctly predicted and the effects of the F120A mutation could be rationalized as well.

Detailed molecular dynamics simulations have also been performed in **Chapter 7** to reproduce and rationalize the experimental finding that the F483A mutant of CYP2D6 has lower affinity for *R*-propranolol than for *S*-propranolol.[5] Wild-type (WT) CYP2D6 does not show this stereospecificity. The thermodynamic integration approach was used to calculate the free energy differences between all combinations of *R*- and *S*-propranolol complexed to wild-type and F483A mutant CYP2D6 (end-points). Based on differences between multiple calculations connecting the same end-point and the closure of the thermodynamic cycles, it was clear that not all simulations have converged sufficiently. The approach that calculates the free energies of exchanging *R*-propranolol with *S*-propranolol in the F483A mutant relative to the exchange free energy in WT was able to reproduce the experimental binding data. Careful inspection of the end-points of the simulations involved in this approach, allowed for a molecular interpretation of the observed differences. *R*-propranolol forms fewer

hydrogen bonds than *S*-propranolol, both in WT and in the F483A mutant. The mutation causes a loss of favourable hydrophobic interactions, which can be compensated by increased hydrogen bond formation by *S*-propranolol, but not by *R*-propranolol.

Substrate specificity of Cytochrome P450 enzymes may be determined not only by the characteristics of the active site itself but also by the (dynamic) selectivity of access routes to the active site. In addition, enzyme kinetics may be influenced by the substrate binding and product unbinding process and associated protein dynamics. In **Chapter 8** we have used random acceleration molecular dynamics (RAMD) simulations to investigate the routes and mechanisms via which substrates may enter and products may exit the buried active site of CYP2D6. Three dominant pathways were identified, which were rather substrate and substrate protonation state dependent. E216 was shown to play an important role in substrate egress, suggesting that E216 is not only crucial for substrate binding in the active site, but that it is also involved in directing the substrate towards the active site via electrostatic steering. Substrate/product egress via the three pathways is furthermore accompanied by large displacement of segments in the BC loop, FG loop and F and G helices and involve rotation of aromatic and aliphatic residues, and loss of internal protein H-bonds. We propose the following mechanism for substrate access into and product exit from CYP2D6: Hydrophobic substrates coming from the membrane enter the protein via a pathway termed pw2a, and their hydrophilic products leave via pw2c and/or the solvent channel, while soluble substrates enter (and their products leave) via pw2c and/or the solvent channel. From our RAMD studies hypotheses can be derived which can be tested experimentally, such as site-directed mutagenesis studies of non-active site residues lining the substrate egress pathways.

9.2 Conclusions and Perspectives

The work presented in this thesis has led to a well validated protein model of CYP2D6, and based on that, to an in-depth exploration and understanding of two different aspects of Cytochrome P450-ligand interactions: binding modes and binding affinities. We have developed methods which can successfully predict in which conformation and orientation substrates and inhibitors bind to the CYP2D6 isoenzyme (binding modes), how tight these ligands bind (binding affinity), and how binding mode and binding affinity are influenced by specific structural and chemical characteristics of the ligand and the protein. The computational approaches in this work have also opened areas that may stimulate and guide further investigations. The most important conclusions with respect to the possibilities and limitations of CYP-ligand binding mode and affinity predictions, and future perspectives for new areas in computational and experimental CYP research are discussed below and summarized in **Table 1**.

Structures, Crystals, and Computers:
Inducing the fit, shaping the mould, carving out key positions

CYPs share the same general three-dimensional fold[6] and for all CYPs six substrate recognition sites (SRSs) have been assigned.[7] However, our modeling and site-directed mutagenesis studies suggest that only specific amino acid residues located in these SRSs are actually directly involved in CYP2D6-ligand binding, while other residues have an indirect effect on ligand binding and catalysis by i) influencing the protein conformation and shape of the active site; ii) stabilizing catalytically active heme-oxygen intermediates; iii) controlling substrate entrance and/or product exit.[8-12] Furthermore, in our studies the roles of specific residues in CYP2D6 mediated metabolism have been shown to be rather substrate-specific (see **Chapters 3**, **6**, and **7**).[4,5,12] **Chapter 2** shows that several CYP isoforms have only been crystallized in a non-active ligand-bound state,[1] and for some it is not known whether the crystal structure corresponds to an active state or not.[13-15] It is important to note that this may render the crystal structure of a certain CYP isoenzyme not directly applicable to

e.g. explaining observed substrate selectivity and product formation by the enzyme. Ligand-induced[16-20] and temperature-dependent[21] changes in the size, shape and hydration of the active sites of several CYPs are also observed, and these are likely to underlie the capacity of many CYPs to metabolize substrates that are diverse in size and structure and to generate different metabolic products from these substrates. Computational chemistry is an especially useful and valuable tool to provide dynamic models for substrate-bound complexes of CYPs which give insight into CYP-ligand interactions, as was demonstrated in this thesis for the CYP2D6 isoenzyme.

Docking Difficulties:
Docking and scoring dependence, water molecules, and protein flexibility

Several docking algorithms and scoring functions have been described in the past few years, but docking (structure prediction) and scoring (prediction of binding free energy) accuracy of docking-scoring combinations still vary with the selected target and the physicochemistry of target-ligand interactions.[22,23] This observation suggests that, prior to automated docking studies, one should find optimal docking-scoring strategies for targeted systems by evaluating their docking and scoring accuracy for a training set of protein-ligand X-ray structures (as was done for CYPs in **Chapter 4**).[2] In the absence of such a training set, other sorts of experimental data, like regio-specificity of metabolism (binding mode prediction/CSP accuracy), binding affinity determinations (scoring/ranking accuracy) and site-directed mutagenesis studies (binding mode prediction/relevance of specific amino acid residues for binding) can be used to validate docking strategies (as was done for CYP2D6 in **Chapter 5**).[3]

In addition to the problem of protein-ligand complex specificity of docking performance, there are also other critical issues in automated docking, such as ligand-induced complementarity (the accommodation of full flexibility of both ligand and protein),[24] and the inclusion, or omission, of explicit water molecules in the ligand binding pocket.[25] In the past several years, automated docking has been established as a standard modeling tool to predict ligand-protein binding modes and affinities in lead discovery and optimization.[26] Benchmarks have been defined,[22,23] and more and more approaches have been developed which tackle the problems of finding the optimal docking and scoring combination for a given target,[27] active site solvation,[28] and protein flexibility.[24] In **Chapters 4** and **5** novel docking strategies are described to predict the locations of active-site water molecules in CYPs and to include them in automated docking simulations.[2,3] However, to be able to describe the docking of ligands to protein targets in an even more accurate way, it is suggested to focus on the development of docking search algorithms that also optimize the positions of several water molecules during the docking process. To our knowledge, no docking tool has been properly validated which, in addition to the ligand and the protein, also considers fully 'flexible' water molecules as a 'third' partner in protein-ligand interactions.

Dynamic Distributions:
Multiple binding modes, multiple entrance/exit channels, multiple substrates

Chapters 4 and **5** concluded that automated docking can successfully be applied to predict sites of oxidation in substrates using CYP crystal structures and CYP homology models. However, **Chapters 6** and **7** show that automated docking alone is not suitable for accurate determination of relative probabilities of different substrate binding modes and protein-ligand interactions for discrimination between substrates and enzymes of high similarity, such as regio/stereoisomers and mutants of enzymes. These subtle differences in protein-ligand binding modes and interactions can have a determining influence on the observed affinities and metabolite distributions. Computational tools offer insight at an atomic resolution often lacking in experiments. In order to reproduce and rationalize experimental data, which are an average over many different substrate and enzyme conformations and orientations over

time, a dynamic treatment of both substrate and enzyme is required. MD simulations in addition can account for distributions of multiple binding conformations and thus give a more comprehensive description of multiple sites of oxidation in substrates catalyzed by CYPs (**Chapter 6**)[4] and of CYP-ligand interactions (**Chapter 7**).[5] Multiple extensive MD simulations are needed to catch subtle differences between substrates, enzyme structures and their dynamics. A remaining hurdle in the accurate calculation of CYP ligand binding modes and affinities is the diversity of binding modes of ligands, stemming from a flexible binding cavity that is large in relation to the size of the substrates. Consequently, for many ligands (i.e. substrates and inhibitors) the binding mode in CYPs is not as strictly defined as is the case for many other more substrate and inhibitor specific enzymes.[29] This potentially introduces sampling problems,[30,31] prohibiting the application of MD based methods that can provide the most accurate free-energy calculations.

Substrate specificity of CYPs may be determined not only by the characteristics of the active site itself (as investigated in **Chapters 3**, **4**, **5**, **6** and **7**)[2-5] but also by the (dynamic) selectivity of access routes to the active site. In addition, enzyme kinetics may be influenced by the substrate binding and product unbinding process and associated protein dynamics. The generally buried active site of many CYPs implies that the protein must undergo dynamic motion to allow substrates to enter and products to leave the active site.[32,33] We have explored the routes and mechanisms via which substrates may exit the active site of CYP2D6 in **Chapter 8**. The hypotheses derived from these studies should still be tested experimentally, e.g. by site-directed mutagenesis studies of non-active site residues lining the substrate egress pathways and spectroscopic techniques to localize regions of conformational flexibility in the protein.

Mammalian CYPs are inserted into the membrane by means of a hydrophobic N-terminal helical anchor and some hydrophobic loop regions located around the entrance of pw2a (**Chapter 8**).[34] To facilitate crystallization of mammalian membrane-bound CYPs residues in these regions have been truncated and mutated into more hydrophilic residues, respectively.[35] Although these mutated constructs have been shown to be catalytically active, the mutations might cause significant conformational changes. Even when one assumes that the actual shape of the active site of CYPs is not much influenced by CYP-membrane interactions, association to the membrane can have a significant effects on protein (FG loop) flexibility, and for a proper description of the dynamics of substrate entrance and product exit in CYP2D6, protein-membrane (and substrate-membrane) interactions should be taken into account.

Some CYPs are capable to bind several substrates simultaneously, occupying the same or different binding sites. This results in unusual enzyme kinetics. This hypothesis is supported by kinetic models[36] and crystal structures,[37,38] but modeling studies describing and predicting allosterism are still very scarce.[39,40] Methodological problems linked to the computational prediction of allosterism are the identification of alternative binding pockets, the sequence of binding events, and the proper description of ligand-ligand interactions. Atypical kinetics have been reported for CYP2D6,[10,41] spectroscopic studies indicate that substrates may associate to CYP2D6 at different sites,[42] and alternative substrate binding pockets can be identified along potential substrate access channels in CYP2D6 (**Chapter 8**). The capability of CYP2D6 to bind multiple ligands at the same time, may potentially have a great impact on the prediction of CYP2D6 mediated metabolism, and computational research is needed to predict and investigate the nature of the observations of allosterism in this and other CYP isoenzymes.

Table 1: Current possibilities, current limitations, and future perspectives of computational CYP-ligand binding mode and affinity predictions, and in silico prediction of CYP mediated metabolism rates. Not mentioned in the table, but essential for any successful computational investigation, is extensive training and validation of the computational model with accurate experimental data.

Prediction	Current possibilities	Current limitations	Future perspectives
Binding mode/ site of metabolism	Identification binding mode determining residues and residues lining substrate access/exit channels	Accurate predictions of the effects of subtle structural differences in residues on global and local protein structure (flexibility)	Analysis of protein-ligand motions in multiple long run MD simulations (possibly with future ligand-bound crystal structures)
	Prediction major sites of metabolism; Qualitative prediction product distribution	Quantitative prediction product distribution	Combined prediction of probabilities of reaction accessibility and intrinsic reactivity of potential metabolism sites
	Description of possible substrate access and product exit mechanisms	Determination dominant substrate access and product exit channels	Experimental validation of proposed ligand acces and exit mechanisms
		Accounting for membrane association and CYP-reductase interactions	Structural model of membrane-embedded CYP-reductase complex
Binding affinity	Selection of high-affinity binders and identification of non-binders from chemical databases	Quantitative prediction binding affinities of chemically diverse compounds	More efficient MD simulation and free energy calculation techniques with improved conformational space sampling and more accurate description of (ionic) protein-ligand interactions
	Quantitative prediction and structural rationalization of differences in affinities between stereo-isomers and mutants	Calculation speed, conformational sampling, parameterization of ionic ligand-protein interactions	
	Identification of protein and ligand characteristics important for protein-ligand recognition and binding	Prediction of relative affinities of drugs for different protein targets	Subsequent and/or parallel virtual screening of chemical databases against multiple protein targets
Rate of metabolism	Description of the energies and structures of reactive heme species	Accounting for the influence of the protein environment	QM/MM calculations on MD snapshots
	Semi-quantitative prediction of the probability of substrate binding and metabolite formation (binding modes + intrinsic reactivity)	Identification of rate-determining steps in reaction pathways; quantitative prediction of (relative) rates of metabolism	Integrative usage of multiple modeling approaches describing the dynamics of binding orientation, affinity, and reactivity along the catalytic cycle
		Prediction of relative kinetics of different reaction pathways mediated by different enzymes	Subsequent and/or parallel prediction of rates of metabolism mediated by different (iso)enzymes

CYP in Silico Velocitas:
Catalytic steps, catalytic components, relative rates

A yet unresolved challenge in computational drug metabolism research is the accurate quantitative prediction of metabolite formation ratios and turnover rates. For accurate prediction of reactivities of substrate catalytic sites, the dependence on kinetic factors and the influence of the active site electrostatic potential should be accounted for. This, in turn, depends on details of the protein structure and the binding conformation and orientation of the substrate. The absence of certain predicted metabolites may be due to slow conversion compared to other metabolic reactions, from competing metabolic pathways or from other enzymes. In many cases it is possible to identify rate-limiting steps (which may vary between CYP isoforms and between substrates of the same isoform). These imply that product formation will be dominated by the characteristics of that particular step, greatly simplifying the complexity of the prediction.[43] In the oxidation of ethanol and acetaldehyde by CYP2E1,[44] a slow step occurs after product formation, although it is not product release perse. In the catalytic cycles of CYP1A2,[45] CYP2D6,[46] and CYP3A4,[47] C-H bond breaking is only partially rate-limiting. However, the rates of all the various individual steps are considerably higher than the overall k_{cat}, and the contributions of several steps must be considered to understand rates of CYP1A2[45] and CYP2D6 reactions.[46] Therefore accurate quantitative computational estimates of (relative) rates of CYP mediated metabolism in principle can only be made when the quantitative description of all the catalytic steps in the CYP catalytic cycle (as depicted in Figure 6 of **Chapter 1** and Figure 1 of **Chapter 2**) are properly combined.

Even when the relative rates of the different catalytic steps can be predicted, it should be kept in mind that chemical reactivity alone, without taking into consideration the orientation of the substrate in the CYP active site (with respect to the heme), is not likely able to predict the correct site of metabolism in a CYP substrate.[48] For CYP2D6 and many other CYPs, the orientation of the substrate in the active site might even be the most important factor determining catalytic regiospecificity, as **Chapters 4**, **5** and **6** showed that automated docking and molecular dynamics simulations are already very successful in predicting the major sites of metabolism and to provide qualitative predictions of product distributions.[2-4] **Chapter 6**, however, also showed that a small substrate like MDMA is very mobile in the CYP2D6 active site and that the reactivity of possible sites of oxidation also determines what products are being formed.[4] For a proper quantitative description of product distributions surely the combined probabilities of reaction accessibility (i.e. proximity to the heme reactive species) and intrinsic reactivity of sites in the substrate should be estimated. Recently, such a combined modeling approach was successfully used to predict sites of metabolism in various CYPs, and indicated that the relative contributions of substrate docking and intrinsic reactivity factors was isoenzyme dependent.[48,49]

Another remaining challenge in predicting product formation in CYPs is the elucidation of the relevant reactive heme-iron-oxygen species that is ultimately responsible for the chemistry of the enzyme reaction, and the coupling between the kinetics of the reaction cycle of the heme-iron-oxygen moiety and the dynamics of substrate and oxygen and release of the product. Elucidation of the reaction cycle of CYPs is complicated by the different influences of the substrate, sequence of events and rate-limiting step between different isoforms.[43] The development of QM and combined QM/MM methodologies to describe catalytic species in the CYP catalytic cycle has evolved rapidly over the past years.[50-54] Recent extensive QM/MM calculations on CYP2C9, CYP2B4, CYP3A4, and CYP101, revealed that the differences in the electronic structure of Compound I for one isoform to another are not significantly larger than those derived from different MD snapshots and from different starting crystal structures of the same isoform.[55] Furthermore, QM/MM calculations of H-abstraction barriers in CYP101 were found to be dependent on subtle differences in protein-water configuration.[56] In addition, recent

combined experimental and modeling studies indicate that the highly conserved I-helix threonine residue in CYPs (T309 in CYP2D6) plays not only an important role in maintaining the balance of multiple oxygenating species (**Chapter 3**, refs [10,57]), but also influences heme spin state distributions by interacting with heme bound water or molecular oxygen [*Groenhof and Bonifacio, unpublished results 2006*, ref [58]].

The studies in this thesis have focused on the description of interactions between CYPs and small ligand molecules. The interaction of CYPs with other proteins, however, is also an important phenomenon to take into account when describing CYP mediated metabolism. Electrons are transferred from NAD(P)H to CYPs via electron transfer proteins (e.g. see Figure 6 in **Chapter 1** and Figure 1 in **Chapter 2**). These mostly consist of either a membrane-bound NADPH reductase (like in mammalian CYPs) or a soluble reductase and a two-iron-sulfur cluster ferredoxin shuttle protein (in most soluble bacterial CYPs) that together deliver the electrons. There is evidence for favorable electrostatic interactions between CYPs and a number of their electron transfer partners.[59] CYPs may also interact with other CYPs and form oligomers.[38] CYP-reductase interactions are even thought to have an allosteric effect on substrate binding and metabolic regiospecificity,[60] although such an effect has not been shown for CYP2D6.[61] Although structures of the individual components, and even a crystal structure of the complex between the heme- and FMN-binding domains of bacterial CYP102 have been solved,[59] no crystal structure has yet been determined of a complex of a mammalian CYP with its non-covalently bound electron transfer protein(s). Thus, modeling studies play an important role in the interpretation of mutational and spectroscopic data on CYP protein–protein interactions.[62,63] With the development of more and more accurate and efficient protein-protein docking methods[64] possibly involving experimentally determined geometrical constraints,[65] more and more accurate structural models of CYP-protein complexes can be generated.[62]

CYP in Silico Applicatus
Which drugs, which CYPs, which metabolites?

The protein model of CYP2D6 and the methods described in this thesis can be used for pre-experimental virtual screening purposes, as well as accurate hypothesis generation for *in vitro* studies in drug discovery and development. From binding mode predictions the site of metabolism and the number and relevance of reactive metabolic positions can be estimated. Moreover, binding affinity predictions give insight into substrate specificity and into the potential of a compound to inhibit the metabolism of other drugs by CYPs (drug-drug interactions).

Since pharmacological activity, toxicity, bioavailability, distribution, and final elimination all depend on metabolic biotransformations, the prediction of metabolic stability and the identification of potential sites of metabolism in a putative drug molecule can be a significant help in designing new compounds with better pharmacokinetic and safety profiles. Metabolic stability of compounds can generally be determined in a high-throughput fashion. Recently, some attempts for high-throughput metabolite identification have been reported,[66,67] but generally, the experimental methods for early identification of metabolites is still low throughput and requires considerable amounts of compound. Computational chemistry is therefore an especially suitable tool to rapidly predict metabolic stability and sites of metabolism in potential drugs.[48] When the site of metabolism is known, metabolically labile compounds can be stabilized by removing, replacing or protecting metabolically susceptible groups.[68] Similarly, the formation of toxic metabolites can be avoided by chemically protecting the metabolically labile moieties. Finally, knowing the site of metabolism is essential for the design of pro-drugs, where the compound needs to be metabolized in order to become active[69], and also for drugs showing excessively long half lives. Modifying functional groups in drug lead compounds

that are essential for binding to metabolic enzymes may affect the binding affinity for these enzymes as well as their metabolic regiospecificity.

Another yet unresolved challenge in computational drug metabolism research is the construction of combined models for multiple biotransformation enzymes (like different CYP isoenzymes), enabling the prediction of the relative kinetics of different reaction pathways and prediction of the relative affinities of drugs for different enzymes. This would give insight into the relative importance of different enzymes in the formation of different metabolites and might finally lead to the prediction of (the kinetics of) complete metabolic pathways of drugs.

In principle such combined biotransformation models could be combined with specific structure-based ligand-receptor interaction models. Drug leads found to have high affinity for receptors are often found to have a high affinity for CYPs as well. For example, serotonin transporters[70] and (aminergic) GPCRs[71] show high affinity for typical CYP2D6 substrates. This can lead to undesired drug-drug interactions and adverse drug effects. Several ideas can be postulated to address the problem of protein specificity in affinity and metabolism of drugs by CYPs. The methods can roughly be divided into two groups. In the first approach one could use known methods to rationalise the binding and/or metabolism by several target proteins, and subsequently try to relate the differences in protein-ligand interactions to molecular differences in the substrates. Such ligand-based pharmacophore models have been defined for several individual CYPs[72] and receptor targets[73,74], but studies to predict and rationalize differences in protein specificity are relatively scarce.[41,75-77] Alternatively, one can use computational methods to train a model on the specificity between targets, i.e. the difference between metabolism or binding affinities. Instead of virtually screening a chemical database against a specific CYP isoenzyme to search for high (or low) affinity substrates or inhibitors, one can for example also try to recover the most likely (CYP iso)enzyme target for a given ligand ((CYP iso)enzyme specificity).[78] Another possibility would be to train 3D-QSAR models on differences in binding affinity of the same compound for different targets, instead of training on differences in affinity of different compounds for the same target. This will automatically point at the molecular elements that are most important to obtain substrates with e.g. a high affinity for one target and a low affinity for another target. Once one would be able to understand and predict the molecular reasons for isoenzyme and receptor specificities in binding and possibly metabolism, even more powerful tools for computer-aided drug design will evolve.

In summary, hit identification and lead development could be helped significantly if during the design process care can be taken to prevent CYP2D6 activity. In this thesis we have described the development of computational methodologies by the integrative use of molecular modeling with site-directed mutagenesis studies, enzymology, and spectroscopical studies. The developed techniques described in this thesis were successfully applied to predict CYP2D6 affinity, to predict regiospecificity of CYP2D6 mediated metabolism, and to define the characteristics of CYP2D6 and its ligands determining CYP2D6-drug interactions, and can be seen as model approaches for other CYPs. Furthermore, the combination of computational and experimental techniques used in this thesis is a promising approach for the investigation of pharmaceutically relevant targets of which structural information is lacking. Combining computational interaction models between proteins and drugs (and their metabolites) will enable the development of even more efficient hit identification and lead optimization protocols.

References

(1)   de Graaf, C.; Vermeulen, N. P.; Feenstra, K. A. Cytochrome p450 in silico: an integrative modeling approach. *J Med Chem* **2005**, *48*, 2725-2755.

(2)   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.

(3)   de Graaf, C.; Oostenbrink, C.; Keizers, P. H.; van der Wijst, T.; Jongejan, A. et al. Catalytic Site Prediction of and Virtual Screening Accuracy of Cytochrome P450 2D6 Substrates by Consideration of Water and Rescoring in Automated Docking. *J Med Chem* **2006**, *49*, 2417-2430.

(4)   Keizers, P. H.; de Graaf, C.; de Kanter, F. J. J.; Oostenbrink, C.; Feenstra, K. A. et al. Metabolic regio- and stereoselectivity of Cytpchrome P450 2D6 towards 3,4-methylenedioxy-N-alkyl-amphetamines: In silico Predictions and experimental validation. *J. Med. Chem.* **2005**, *48*, 6117-6127.

(5)   de Graaf, C.; Oostenbrink, C.; Keizers, P. H.; Lussenburg, B. M.; Vermeulen, N. P. Free energies of binding of R- and S- propranolol to wildtype and mutant Cytochrome P450 2D6 from molecular dynamics simulations. *in preparation* **2006**.

(6)   Mestres, J. Structure conservation in cytochromes P450. *Proteins* **2005**, *58*, 596-609.

(7)   Gotoh, O. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem* **1992**, *267*, 83-90.

(8)   van Waterschoot, R. A.; Keizers, P. H.; de Graaf, C.; Vermeulen, N. P.; Tschirret-Guth, R. A. Topological role of cytochrome P450 2D6 active site residues. *Arch Biochem Biophys* **2006**, *447*, 53-58.

(9)   Keizers, P. H.; van Dijk, B. R.; de Graaf, C.; Vugt-Lussenburg, B. M.; Vermeulen, N. P. et al. Metabolism of N-substituted 7-methoxy-4-(aminomethyl)-coumarins by cytochrome P450 2D6 mutants indicates additional substrate interaction points. *Xenobiotica* **2006**, *in press*.

(10)   Keizers, P. H.; Schraven, L. H.; de Graaf, C.; Hidestrand, M.; Ingelman-Sundberg, M. et al. Role of the conserved threonine 309 in mechanism of oxidation by cytochrome P450 2D6. *Biochem Biophys Res Commun* **2005**, *338*, 1065-1074.

(11)   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.

(12)   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.

(13)   Haines, D. C.; Tomchick, D. R.; Machius, M.; Peterson, J. A. Pivotal role of water in the mechanism of P450BM-3. *Biochemistry* **2001**, *40*, 13456-13465.

(14)   Williams, P. A.; Cosme, J.; Vinkovic, D. M.; Ward, A.; Angove, H. C. et al. Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* **2004**, *305*, 683-686.

(15)   Williams, P. A.; Cosme, J.; Ward, A.; Angove, H. C.; Matak Vinkovic, D. et al. Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* **2003**, *424*, 464-468.

(16)   Cupp-Vickery, J. R.; Garcia, C.; Hofacre, A.; McGee-Estrada, K. Ketoconazole-induced conformational changes in the active site of cytochrome P450eryF. *J Mol Biol* **2001**, *311*, 101-110.

(17)   Li, H.; Poulos, T. L. Conformational dynamics in cytochrome P450-substrate interactions. *Biochimie* **1996**, *78*, 695-699.

(18)   Podust, L. M.; Poulos, T. L.; Waterman, M. R. Crystal structure of cytochrome P450 14alpha -sterol demethylase (CYP51) from Mycobacterium tuberculosis in complex with azole inhibitors. *Proc Natl Acad Sci U S A* **2001**, *98*, 3068-3073.

(19)   Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dijols, S.; Dansette, P. M. et al. Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 A resolution: evidence for an induced fit model of substrate binding. *Biochemistry* **2003**, *42*, 9335-9345.

(20)   Park, S. Y.; Yamane, K.; Adachi, S.; Shiro, Y.; Weiss, K. E. et al. Thermophilic cytochrome P450 (CYP119) from Sulfolobus solfataricus: high resolution structure and functional properties. *J Inorg Biochem* **2002**, *91*, 491-501.

(21)   Jovanovic, T.; Farid, R.; Friesner, R. A.; McDermott, A. E. Thermal equilibrium of high- and low-spin forms of cytochrome P450 BM-3: repositioning of the substrate? *J Am Chem Soc* **2005**, *127*, 13548-13552.

(22)   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.

(23)   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.

(24)   Cavasotto, C. N.; Orry, A. J. W.; Abagyan, R. A. The challenge of considering receptor flexibility in ligand docking and virtual screening. *Curr Comput-Aid Drug Des* **2005**, *1*, 423-440.

(25)   McConkey, B. J.; Sobolev, V.; Edelman, M. The performance of current methods in ligand-protein docking. *Current Science* **2002**, *83*, 845-856.

(26)   Jain, A. N. Virtual screening in lead discovery and optimization. *Curr Opin Drug Discov Devel* **2004**, *7*, 396-403.

(27)   Paul, N.; Rognan, D. ConsDock: A new program for the consensus analysis of protein- ligand interactions. *Proteins-Structure Function and Genetics* **2002**, *47*, 521-533.

(28) 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.

(29) Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dansette, P. M.; Mansuy, D. et al. Structure of a substrate complex of mammalian cytochrome P450 2C5 at 2.3 A resolution: evidence for multiple substrate binding modes. *Biochemistry* **2003**, *42*, 6370-6379.

(30) Oostenbrink, C.; Van Gunsteren, W. F. Calculating zeros: Non-equilibrium free energy calculations. *Chem. Phys. in press* **2006**.

(31) Mark, A. E.; Van Gunsteren, W. F.; Berendsen, H. J. C. Calculation of relative free energy via indirect pathways. *J. Chem. Phys.* **1991**, *94*, 3808-3816.

(32) Winn, P. J.; Ludemann, S. K.; Gauges, R.; Lounnas, V.; Wade, R. C. Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 5361-5366.

(33) Wade, R. C.; Winn, P. J.; Schlichting, I.; Sudarko A survey of active site access channels in cytochromes P450. *J Inorg Biochem* **2004**, *98*, 1175-1182.

(34) Ozalp, C.; Szczesna-Skorupa, E.; Kemper, B. Identification of membrane-contacting loops of the catalytic domain of cytochrome P450 2C2 by tryptophan fluorescence scanning. *Biochemistry* **2006**, *45*, 4629-4637.

(35) Li, H.; Poulos, T. L. Crystallization of cytochromes P450 and substrate-enzyme interactions. *Curr Top Med Chem* **2004**, *4*, 1789-1802.

(36) Atkins, W. M. Non-Michaelis-Menten kinetics in cytochrome P450-catalyzed reactions. *Annu Rev Pharmacol Toxicol* **2005**, *45*, 291-310.

(37) Cupp-Vickery, J.; Anderson, R.; Hatziris, Z. Crystal structures of ligand complexes of P450eryF exhibiting homotropic cooperativity. *Proc Natl Acad Sci U S A* **2000**, *97*, 3050-3055.

(38) Scott, E. E.; Halpert, J. R. Structures of cytochrome P450 3A4. *Trends Biochem Sci* **2005**, *30*, 5-7.

(39) Torimoto, N.; Ishii, I.; Hata, M.; Nakamura, H.; Imada, H. et al. Direct interaction between substrates and endogenous steroids in the active site may change the activity of cytochrome P450 3A4. *Biochemistry* **2003**, *42*, 15068-15077.

(40) Egnell, A. C.; Eriksson, C.; Albertson, N.; Houston, B.; Boyer, S. Generation and evaluation of a CYP2C9 heteroactivation pharmacophore. *Journal of Pharmacology and Experimental Therapeutics* **2003**, *307*, 878-887.

(41) 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.

(42) Stortelder, A.; Keizers, P. H.; Oostenbrink, C.; De Graaf, C.; De Kruijf, P. et al. Binding of 7-methoxy-4-(aminomethyl)-coumarin to wild-type and W128F mutant cytochrome P450 2D6 studied by time-resolved fluorescence spectroscopy. *Biochem J* **2006**, *393*, 635-643.

(43) Guengerich, F. P. Rate-limiting steps in cytochrome P450 catalysis. *Biol Chem* **2002**, *383*, 1553-1564.

(44) Bell-Parikh, L. C.; Guengerich, F. P. Kinetics of cytochrome P450 2E1-catalyzed oxidation of ethanol to acetic acid via acetaldehyde. *J Biol Chem* **1999**, *274*, 23833-23840.

(45) Guengerich, F. P.; Krauser, J. A.; Johnson, W. W. Rate-limiting steps in oxidations catalyzed by rabbit cytochrome P450 1A2. *Biochemistry* **2004**, *43*, 10775-10788.

(46) Guengerich, F. P.; Miller, G. P.; Hanna, I. H.; Sato, H.; Martin, M. V. Oxidation of methoxyphenethylamines by cytochrome P450 2D6. Analysis of rate-limiting steps. *J Biol Chem* **2002**, *277*, 33711-33719.

(47) Krauser, J. A.; Guengerich, F. P. Cytochrome P450 3A4-catalyzed testosterone 6beta-hydroxylation stereochemistry, kinetic deuterium isotope effects, and rate-limiting steps. *J Biol Chem* **2005**, *280*, 19496-19506.

(48) Cruciani, G.; Carosati, E.; De Boeck, B.; Ethirajulu, K.; Mackie, C. et al. MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. *J Med Chem* **2005**, *48*, 6970-6979.

(49) Zhou, D.; Afzelius, L.; Grimm, S. W.; Andersson, T. B.; Zauhar, R. J. et al. Comparison of methods for the prediction of the metabolic sites for cyp3a4-mediated metabolic reactions. *Drug Metab Dispos* **2006**, *34*, 976-983.

(50) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. *Chem Rev* **2005**, *105*, 2279-2328.

(51) Groenhof, A. R.; Swart, M.; Ehlers, A. W.; Lammertsma, K. Electronic ground states of iron porphyrin and of the first species in the catalytic reaction cycle of cytochrome P450s. *Journal of Physical Chemistry A* **2005**, *109*, 3411-3417.

(52) Swart, M.; Groenhof, A. R.; Ehlers, A. W.; Lammertsma, K. QM/MM study on the catalytic cycle of cytochrome P450: The importance of selecting an appropriate density functional theory functional. *Abstracts of Papers of the American Chemical Society* **2004**, *227*, U1449-U1449.

(53) Friesner, R. A.; Guallar, V. Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Annu Rev Phys Chem* **2005**, *56*, 389-427.

(54)   Bathelt, C. M.; Ridder, L.; Mulholland, A. J.; Harvey, J. N. Mechanism and structure-reactivity relationships for aromatic hydroxylation by cytochrome P450. *Org Biomol Chem* **2004**, *2*, 2998-3005.

(55)   Bathelt, C. M.; Zurek, J.; Mulholland, A. J.; Harvey, J. N. Electronic structure of compound I in human isoforms of cytochrome P450 from QM/MM modeling. *J Am Chem Soc* **2005**, *127*, 12900-12908.

(56)   Altun, A.; Guallar, V.; Friesner, R. A.; Shaik, S.; Thiel, W. The effect of heme environment on the hydrogen abstraction reaction of camphor in P450cam catalysis: a QM/MM study. *J Am Chem Soc* **2006**, *128*, 3924-3925.

(57)   Blobaum, A. L.; Harris, D. L.; Hollenberg, P. F. P450 active site architecture and reversibility: inactivation of cytochromes P450 2B4 and 2B4 T302A by tert-butyl acetylenes. *Biochemistry* **2005**, *44*, 3831-3844.

(58)   Keizers, P. H. J. Mutational and spectroscopic analysis of CYP2D6; Vrije Universiteit: Amsterdam, the Netherlands, 2006.

(59)   Sevrioukova, I. F.; Li, H.; Zhang, H.; Peterson, J. A.; Poulos, T. L. Structure of a cytochrome P450-redox partner electron-transfer complex. *Proc Natl Acad Sci U S A* **1999**, *96*, 1863-1868.

(60)   Modi, S.; Sutcliffe, M. J.; Primrose, W. U.; Lian, L. Y.; Roberts, G. C. The catalytic mechanism of cytochrome P450 BM3 involves a 6 A movement of the bound substrate on reduction. *Nat Struct Biol* **1996**, *3*, 414-417.

(61)   Hanna, I. H.; Krauser, J. A.; Cai, H.; Kim, M. S.; Guengerich, F. P. Diversity in mechanisms of substrate oxidation by cytochrome P450 2D6. Lack of an allosteric role of NADPH-cytochrome P450 reductase in catalytic regioselectivity. *J Biol Chem* **2001**, *276*, 39553-39561.

(62)   Wade, R. C.; Motiejunas, D.; Schleinkofer, K.; Sudarko; Winn, P. J. et al. Multiple molecular recognition mechanisms. Cytochrome P450--a case study. *Biochim Biophys Acta* **2005**, *1754*, 239-244.

(63)   Allorge, D.; Breant, D.; Harlow, J.; Chowdry, J.; Lo-Guidice, J. M. et al. Functional analysis of CYP2D6.31 variant: homology modeling suggests possible disruption of redox partner interaction by Arg440His substitution. *Proteins* **2005**, *59*, 339-346.

(64)   Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* **2002**, *47*, 409-443.

(65)   Dominguez, C.; Boelens, R.; Bonvin, A. M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **2003**, *125*, 1731-1737.

(66)   Kool, J.; Ramautar, R.; van Liempd, S. M.; Beckman, J.; de Kanter, F. J. et al. Rapid On-line Profiling of Estrogen Receptor Binding Metabolites of Tamoxifen. *J Med Chem* **2006**, *49*, 3287-3292.

(67)   van Liempd, S. M.; Kool, J.; Reinen, J.; Schenk, T.; Meerman, J. H. et al. Development and validation of a microsomal online cytochrome P450 bioreactor coupled to solid-phase extraction and reversed-phase liquid chromatography. *J Chromatogr A* **2005**, *1075*, 205-212.

(68)   Rowley, M.; Hallett, D. J.; Goodacre, S.; Moyes, C.; Crawforth, J. et al. 3-(4-Fluoropiperidin-3-yl)-2-phenylindoles as high affinity, selective, and orally bioavailable h5-HT(2A) receptor antagonists. *J Med Chem* **2001**, *44*, 1603-1614.

(69)   Ioannides, C.; Lewis, D. F. Cytochromes P450 in the bioactivation of chemicals. *Curr Top Med Chem* **2004**, *4*, 1767-1788.

(70)   Gan, T. J. Selective serotonin 5-HT3 receptor antagonists for postoperative nausea and vomiting: are they all the same? *CNS Drugs* **2005**, *19*, 225-238.

(71)   Surgand, J. S.; Rodrigo, J.; Kellenberger, E.; Rognan, D. A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors. *Proteins* **2006**, *62*, 509-538.

(72)   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 Metabolism and Disposition* **2001**, *29*, 936-944.

(73)   Jongejan, A.; Leurs, R. Delineation of receptor-ligand interactions at the human histamine H1 receptor by a combined approach of site-directed mutagenesis and computational techniques - or - how to bind the H1 receptor. *Arch Pharm (Weinheim)* **2005**, *338*, 248-259.

(74)   Sippl, W. Receptor-based 3D QSAR analysis of estrogen receptor ligands--merging the accuracy of receptor-based alignments with the computational efficiency of ligand-based methods. *J Comput Aided Mol Des* **2000**, *14*, 559-572.

(75)   Ridderstrom, M.; Zamora, I.; Fjellstrom, O.; Andersson, T. B. Analysis of selective regions in the active sites of human cytochromes P450, 2C8, 2C9, 2C18, and 2C19 homology models using GRID/CPCA. *J Med Chem* **2001**, *44*, 4072-4081.

(76)   Zhu, B. T.; Han, G. Z.; Shim, J. Y.; Wen, Y.; Jiang, X. R. Quantitative Structure-Activity Relationship (QSAR) of Various Endogenous Estrogen Metabolites for Human Estrogen Receptor {alpha} and {beta} Subtypes: Insights into the Structural Determinants Favoring a Differential Subtype Binding. *Endocrinology* **2006**.

(77)   De Rienzo, F.; Fanelli, F.; Menziani, M. C.; De Benedetti, P. G. Theoretical investigation of substrate specificity for cytochromes P450 IA2, P450 IID6 and P450 IIIA4. *J Comput Aided Mol Des* **2000**, *14*, 93-116.

(78)   Paul, N.; Kellenberger, E.; Bret, G.; Muller, P.; Rognan, D. Recovering the true targets of specific ligands by virtual screening of the protein data bank. *Proteins* **2004**, *54*, 671-680.

Interacties tussen cytochromen P450 en medicijnen:
Het voorspellen van bindingsoriëntaties en bindings-affiniteiten in CYP2D6 met behulp van computer-simulaties.

Arsenicum, strychnine, curare, botuline, koolmonoxide en cyaankali, bekende gifstoffen die in lage doses tot de dood kunnen lijden. Parcelsus stelde halverwege de zestiende eeuw echter al dat in feite "*al het voedsel en alle drank, als men er te veel van gebruikt, vergift is*". Uitlaatgassen, sigarettenrook, asbest, weekmakers in plastic, voortdurend worden we blootgesteld aan giftige stoffen. Het lichaam is gelukkig uitgerust met mechanismen om de schade van deze stoffen zoveel mogelijk te beperken. Het lichaam kan zich van gifstoffen ontdoen door deze verbindingen met behulp van enzymen om te zetten in stoffen die minder giftig zijn. Deze omzettingsproducten zijn beter in water oplosbaar en kunnen daardoor gemakkelijker via de nieren worden uitgescheiden. Enzymen zijn eiwitten die chemische reacties sneller kunnen laten verlopen en de belangrijkste enzymen die kunnen ontgiften zijn de zogenaamde Cytochroom P450 (CYP) enzymen. Door de omzetting van lichaamsvreemde stoffen kunnen echter ook verbindingen met een hogere giftigheid ontstaan. Deze actieve stoffen kunnen binden aan DNA, vetten of eiwitten en hierdoor de normale celgroei ernstig verstoren. Remming van CYP enzymen door lichaamsvreemde stoffen kan ook ongewenste gevolgen hebben. Via dit mechanisme wordt het lichaamseigen metabolisme van bijvoorbeeld hormonen verstoord en kunnen ongewenste effecten optreden, zoals leververgiftiging of kanker. Daarnaast zetten CYPs ook stoffen om die een heilzame werking kunnen hebben, zoals medicijnen.

Er zijn zo'n 60 verschillende menselijke CYP-enzymen bekend en de omzettingscapaciteit van CYPs verschilt per persoon. Dit leidt ertoe dat niet iedereen hetzelfde reageert op een bepaalde dosis van een geneesmiddel. Het tempo waarin een geneesmiddel uit het lichaam wordt uitgescheiden is immers bepalend voor de medicijn-concentratie in het bloed, hetgeen weer van invloed kan zijn op het effect het geneesmiddel. De variatie in omzettingssnelheid is grotendeels erfelijk bepaald. Sommige CYP-enzymen zijn genetisch polymorf, dat wil zeggen dat er mutante vormen van dit specifieke CYP enzym voorkomen die door mutatie een andere enzymactiviteit hebben. Daarnaast zijn ook andere factoren van invloed op de activiteit van CYP-enzymen. Stoffen in tabaksrook kunnen bijvoorbeeld de aanmaak van bepaalde CYPs verhogen, terwijl medicijnen die interacties aangaan met CYPs de omzetting van andere medicijnen kunnen blokkeren. Het is van belang om bij de ontwikkeling van nieuwe medicijnen in te schatten in welke mate deze verbindingen worden omgezet door de CYP-enzymen in individuen met verschillende CYP-activiteit, en welke producten hierbij worden gevormd. Hiervoor zijn verschillende methoden beschikbaar, waaronder zogenaamde 'high throughput screening' technieken waarbij op een snelle manier stoffen in de reageerbuis (*in vitro*) worden getest op affiniteit voor CYPs. Tegenwoordig kan men dit echter ook met computertechnieken (*in silico*) uitvoeren. Het katalytisch centrum van CYP-enzymen, de plek in het eiwit waar de daadwerkelijke chemische reactie plaatsvindt, is een heem-molecuul met een ijzeratoom. Wanneer geneesmiddelen worden opgenomen in de bindingsholte van CYP-enzymen worden ze op die plaatsen geoxideerd die het meest reactief zijn en die zich het dichtst bij het ijzeratoom bevinden. Computermodellen kunnen worden ingezet om de oriëntaties en reactiviteit van geneesmiddelen *in* CYP-enzymen en de interactie krachten (affiniteiten) *met* CYP-enzymen te voorspellen.

Ongeveer een kwart van de geneesmiddelen die op de markt zijn, wordt door het unieke CYP2D6-enzym gemetaboliseerd. Vele geneesmiddelen zoals antidepressiva, neuroleptica, bèta-blokkers en antiarrhythmica zijn substraten van CYP2D6. Tevens is CYP2D6 genetisch polymorf: veel mensen hebben een sterk verlaagde of sterk verhoogde CYP2D6-activiteit.

Wetenschappelijke studies naar de wijze waarop medicijnen interacties aangaan met het CYP2D6-enzym zijn daarom van belang voor de ontwikkeling van veilige en efficiënte medicijnen, en computermodellen kunnen in een vroeg stadium bij dit ontwikkelingsproces worden ingezet. De centrale vraag van dit proefschrift luidt dan ook: op welke manieren kunnen we met computersimulaties voorspellen hoe medicijnen zich in de bindingsholte van CYP2D6 oriënteren en wat hun affiniteit is met de CYP2D6 bindingsholte?

In **Hoofdstuk 2** wordt een overzicht gegeven van de belangrijkste computersimulatie studies ter beschrijving van CYP-enzymen. De in deze studies beschreven computermodellen zijn ofwel gebaseerd op de liganden (remmers, substraten) van een bepaalde CYP, gebaseerd op de CYP-eiwit structuur, ofwel gebaseerd op de beschrijving van interacties tussen CYPs en hun liganden. Om de functie en structuur van CYPs zo nauwkeurig mogelijk te kunnen beschrijven wordt tegenwoordig steeds meer gebruik gemaakt van combinaties van verschillende modellen en benaderingen.

Computermodellen kunnen niet alleen gebruikt worden om verklaringen te geven voor de resultaten van experimentele studies in het laboratorium, maar ook om hypotheses op te stellen en voorspellingen te doen ten aanzien van de uitkomsten van experimenten. Op deze manier kan de computer worden gebruikt om laboratoriumexperimenten te ontwerpen, terwijl de uitkomsten van deze experimenten op hun beurt weer kunnen worden gebruikt om het computermodel te testen en te verfijnen. We hebben de kristalstructuren van substraatgebonden vormen van een verwant CYP-enzym gebruikt als mal voor de vervaardiging van een computer(homologie)model van CYP2D6. In **Hoofdstuk 3** laten we zien hoe we dit homologie model hebben gebruikt om aminozuren in de bindingsholte van CYP2D6 te identificeren die invloed kunnen hebben op de manier waarop medicijnen kunnen binden aan het CYP2D6-enzym. In het laboratorium bleken deze voorspellingen inderdaad te kloppen en de effecten van de mutaties van deze aminozuren bleken zelfs substraatafhankelijk. Op deze manier konden we enerzijds aantonen dat belangrijke details van ons drie-dimensionale computermodel overeenkwamen met de werkelijkheid. Anderzijds leverden deze experimentele validatiestudies weer nieuwe inzichten op ten aanzien van CYP2D6-substraat interacties die we konden proberen te verklaren en voorspellen met onze computermodellen. Sterker nog, aan het einde van het in dit proefschrift beschreven promotie-onderzoek verscheen er een kristalstructuur van CYP2D6 die in grote mate overeenkomt met ons computermodel.

In **Hoofdstukken 4** en **5** hebben we zogenaamde "automated docking" technieken gebruikt om te voorspellen in welke oriëntaties medicijnen binden aan CYP-enzymen en hoe sterk de interactiekrachten zijn tussen medicijnen en CYPs. De bindingsoriëntatie van het medicijn ten opzichte van het katalytisch centrum van de CYP, het ijzeratoom van de heemgroep, bepaalt of het medicijn wordt omgezet en welke omzettingsproducten hierbij worden gevormd. De sterkte van de interactie, ofwel de bindingsaffiniteit van een medicijn voor een bepaalde CYP, is van invloed op de mate waarin het medicijn wordt omgezet. Tevens bepaalt de bindingsaffiniteit in welke mate een medicijn A een ander medicijn B uit de CYP kan verdringen en op deze manier de omzetting van medicijn B kan remmen. Docking technieken maken gebruik van een docking algoritme om mogelijke bindingsoriëntaties te genereren en scoring functies om de energetisch meest gunstige oriëntatie te selecteren. Er zijn veel verschillende docking algoritmes en scoring functies ontwikkeld in de afgelopen jaren. Het is daarom van belang om te weten met welke benaderingen de meest nauwkeurige voorspellingen kunnen worden gedaan ten aanzien van CYPs (Hoofdstuk 4) en CYP2D6 in het bijzonder (Hoofdstuk 5). Normaal gesproken worden watermoleculen buiten beschouwing gelaten tijdens docking simulaties, maar in

veel CYPs die wij hebben bestudeerd bleek het meenemen van watermoleculen in de berekeningen juist een positieve invloed te hebben op onze voorspellingen. Uiteindelijk hebben we met de optimale docking strategiën de relatieve bindingsaffiniteiten van 20.000 moleculen met CYP2D6 voorspeld (*virtual screening*) en verbindingen geselecteerd met de hoogste docking scores. Deze verbindingen bleken na experimentele tests in het laboratorium ook daadwerkelijk een hoge affiniteit voor CYP2D6 te bezitten.

Subtiele verschillen in de chemische structuur van medicijnen en enzymen kunnen significante effecten hebben op hun onderlinge interacties (substraat bindingsoriëntaties en bindingsaffiniteiten). Om deze verschillen in interacties nauwkeurig te kunnen voorspellen is een tijdsafhankelijke beschrijving van het enzym-substraat complex vereist. Moleculaire dynamica (MD) simulaties zijn bij uitstek een methode om de bewegingen van enzym-substraat complexen in de tijd te beschrijven. Tijdens deze simulaties laten we de atomen van het systeem bewegen volgens de wetten van Newton en op die manier nemen we meerdere conformaties van het enzym-substraat complex mee in onze berekeningen. **Hoofdstuk 6** laat zien dat we door een combinatie van automated docking en veel langdurige MD-simulaties in staat waren om experimenteel bepaalde bindingsoriëntaties van MDMA en MDMA-analogen te voorspellen. MDMA, beter bekend als partydrug XTC, werd vroeger gebruikt als afslankmiddel. De experimentele resultaten waren gebaseerd op spectroscopische metingen van kernspinresonantie (NMR) van protonen van XTC en op metingen van de kinetiek van de vorming van specifieke metabolieten door CYP2D6. Aan de hand van de spectroscopische metingen kan de gemiddelde afstand van de protonen tot het ijzeratoom van de heemgroep worden bepaald, terwijl de enzymologische metingen aangeven welke delen van het substraat zich het dichtst bij het ijzeratoom bevinden. Onze MD-simulaties voldeden zowel aan de spectroscopische als aan de enzymologische metingen. Op deze manier konden we zelfs verklaren waarom XTC stereo-isomeren zich verschillend oriënteren in normale (wild-type) en gemuteerde vormen (fenylalanine 120 naar alanine) van CYP2D6.

In **Hoofdstuk 7** hebben we MD-simulaties en de zogenaamde Thermodynamische Integratie (TI) methode gebruikt om de affiniteit te voorspellen van propranolol stereo-isomeren (een bèta-blokker medicijn) met wild-type en gemuteerde (fenylalanine 483 naar alanine) CYP2D6. De TI-methode maakt gebruik van alchemie: in kleine stapjes wordt molecuul A in molecuul B veranderd en aan de hand van de verschillende simulaties met deze fysisch onmogelijke verbindingen ("een beetje van A en een beetje van B") kan het verschil in vrije energie tussen moleculen A en B worden afgeleid. Het omzetten van de ene stereo-isomeer van propranolol (*R*) in de andere stereo-isomeer (*S*) in zowel wildtype als F483A-mutante CYP2D6 leverde verschillen in vrije energie op die overeen kwamen met experimenteel bepaalde verschillen in affiniteit. Tevens konden de MD-simulaties een verklaring geven voor het feit dat *R*-propranolol minder goed bindt aan de F483A-mutant dan aan wild-type, terwijl *S*-propranolol aan beide vormen van CYP2D6 even goed bindt. *R*- en *S*-propranolol vormen minder waterstofbruggen in de F483A-mutant, en alleen *S*-propranolol kan dit compenseren door meer hydrofobe interacties aan te gaan met de F483A-mutant.

Niet alleen de interactiekrachten tussen het medicijn en de CYP in de bindingsholte kunnen bepalend zijn voor de mate van omzetting van het medicijn. Ook de manier waarop het medicijn deze bindingsholte kan bereiken is een belangrijke factor bij de voorspelling van interacties tussen medicijnen en CYPs. We hebben daarom in **Hoofdstuk 8** de mogelijke kanalen onderzocht waarlangs medicijnen kunnen binnenkomen en omzettingsproducten het CYP2D6 eiwit weer kunnen verlaten. Deze studies suggereren dat slechts bepaalde delen van het CYP2D6 van structuur moeten veranderen om verbindingen uit het enzym te laten ontsnappen of juist toegang te verschaffen tot de

bindingsholte. Een belangrijke rol schijnt te zijn weggelegd voor de negatief geladen glutamaat (E) 216 bij dit proces. Dit negatief geladen aminozuurresidu werkt als een anker- en herkenningspunt voor positief geladen CYP2D6-substraten. Ook andere aminozuurresiduen die zich langs deze toegangskanalen bevinden zijn potentiële kandidaten voor toekomstige experimentele mutatiestudies.

Door het in dit proefschrift beschreven onderzoek hebben we nu een betere kijk op de factoren die bepalend zijn voor de wijze waarop medicijnen zich in de bindingsholte van CYP2D6 oriënteren en hoe sterk hun affiniteit is voor deze bindingsholte. De in deze studies ontwikkelde methodes kunnen worden ingezet tijdens de ontwikkeling van nieuwe medicijnen om te voorspellen of deze verbindingen worden omgezet door CYP2D6 en welke omzettingsproducten zich hierbij vormen. Op deze manier kan worden vermeden dat het medicijn een slechte therapeutische werking heeft in individuen met een verhoogde CYP2D6 activiteit (snelle afbraak van het medicijn) of giftige bijverschijnselen vertoont als gevolg van verlaagde (langdurig hoge medicijnconcentraties in het bloed) dan wel verhoogde (vorming van giftige omzettingsproducten) CYP2D6 activiteit. Tot slot kan de geïntegreerde benadering van computersimulatie en experimentele validatie zoals beschreven in dit proefschrift dienen als een blauwdruk voor toekomstig onderzoek aan andere enzymen.

List of publications

This thesis led to the following publications:

Chapter 2:
Chris de Graaf, Nico P.E. Vermeulen and K. Anton Feenstra
"Cytochrome p450 in silico: an integrative modeling approach"
J Med Chem **2005***, 48*, 2725-2755.

Chapter 3:
Chris de Graaf, Chris Oostenbrink, Peter H.J. Keizers, Barbara M.A. van Vugt-Lussenburg, Robert A.B. van Waterschoot, Richard A. Tschirret-Guth, Jan N.M. Commandeur and Nico P.E. Vermeulen
"Molecular modeling guided site-directed mutagenesis of cytochrome P450 2D6"
Curr Drug Metab **2006**, *in press*

Chapter 4:
Chris de Graaf, Pavel Pospisil, Wouter Pos, Gerd Folkers and Nico P.E. Vermeulen
"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.

Chapter 5:
Chris de Graaf, Chris Oostenbrink, Peter H.J. Keizers, Tushar van der Wijst, Aldo Jongejan and Nico P.E. Vermeulen
"Catalytic site prediction of and virtual screening accuracy of cytochrome P450 2D6 substrates by consideration of water and rescoring in automated docking"
J Med Chem **2006**, *49*, 2417-2430.

Chapter 6:
Peter H.J. Keizers, Chris de Graaf, Frans J. de Kanter, Chris Oostenbrink, K. Anton Feenstra, Jan N.M. Commandeur and Nico P.E. Vermeulen
"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

Chapter 7:
Chris de Graaf, Chris Oostenbrink, Peter H.J. Keizers, Barbara M.A. Lussenburg, Nico P.E. Vermeulen
"Free energies of binding of R- and S- propranolol to wildtype and mutant Cytochrome P450 2D6 from molecular dynamics simulations"
Submitted to *Eur J Biophys* **2006**

Related publications:

Peter H.J. Keizers, Barbara M.A. Lussenburg, Chris de Graaf, Lenny M. Mentink, Nico P.E. Vermeulen and Jan N.M. Commandeur
"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.

Barbara M.A. Lussenburg, Peter H.J. Keizers, Chris de Graaf, Mats Hidestrand, Magnus Ingelman-Sundberg
"The role of phenylalanine 483 in cytochrome P450 2D6 is strongly substrate dependent"
Biochem Pharmacol **2005**, *70*, 1253-1261.

Aldo Jongejan, Chris de Graaf, Nico P.E. Vermeulen, Rob Leurs and Iwan J.P. de Esch
"The role and application of in silico docking in chemical genomics research"
In: *Methods in Molecular Biology*: Chemical Genomics: Reviews and Protocols, 2005. Ed. E.D. Zanders, The Humana Press Inc. **2005**.

Peter H.J. Keizers, Loek H. Schraven, Chris de Graaf, Mats Hidestrand, Magnus Ingelman-Sundberg, Nico P.E. Vermeulen and Jan N.M. Commandeur
"Role of the conserved threonine 309 in mechanism of oxidation by cytochrome P450 2D6"
Biochem Biophys Res Commun **2005**, *338*, 1065-1074.

Aike Stortelder, Peter H.J. Keizers, Chris Oostenbrink, Chris de Graaf, Petra de Kruijf, Nico P.E. Vermeulen, Cees Gooijer, Jan N.M. Commandeur and Gert van der Zwan
"Binding of 7-methoxy-4-(aminomethyl)-coumarin to wild-type and W128F mutant cytochrome P450 2D6 studied by time-resolved fluorescence spectroscopy"
Biochem J **2006**, 393, 635-643.

Robert A.B. van Waterschoot, Peter H.J. Keizers, Chris de Graaf, Nico P.E. Vermeulen and Richard A. Tschirret-Guth, Topological role of cytochrome P450 2D6 active site residues.
Arch Biochem Biophys **2006**, *447*, 53-58.

Peter H.J. Keizers, Ben R. van Dijk, Chris de Graaf, Barbara M.A. van Vugt-Lussenburg, Nico P.E. Vermeulen and Jan N.M. Commandeur
"Metabolism of N-substituted 7-methoxy-4-(aminomethyl)-coumarins by cytochrome P450 2D6 mutants indicates additional substrate interaction points"
Xenobiotica **2006**, *36*, 763-771.

K. Anton Feenstra, Chris de Graaf and Nico P.E. Vermeulen
"Cytochrome P450 protein modeling and ligand docking"
In: *Drug Metabolism / Drug Discovery & Development*. Ed. Paul W. Erhardt, Univ. of Toledo, in press **2006**.

Eric-Wubbo Lameijer, Chris de Graaf, Daniël Acohen, Chris Oostenbrink and Ad IJzerman
"Evolutionary algorithms in *de novo* molecule design: comparing atom-based and fragment-based optimization"
To be submitted to *J Chem Inf Comput Sci* **2006**

Alois Bonifacio, Peter H.J. Keizers, André R. Groenhof, Chris de Graaf,
Jan N.M. Commandeur, Nico P.E. Vermeulen, Andreas W. Ehlers, Koop Lammertsma,
Cees Gooijer and Gerd van der Zwan
"Altered spin state equilibrium in the T309V mutant of cytochrome P450 2D6: a spectroscopic and computational study"
In preparation **2006**

Publications not related to the work in this thesis:

René van Herwijnen, Chris de Graaf, Harrie A.J. Govers and John R. Parsons
"Estimation of kinetic parameter for the biotransformation of three-ring azaarenes by the phenanthrene-degrading strain Sphingomonas sp. LH128"
Environ Toxicol Chem **2004**, *23*, 331-338.

Niels Jonkers, Remi W.P.M. Laane, Chris de Graaf, Pim de Voogt
"Fate modeling of nonylphenol ethoxylates and their metabolites in the Dutch Scheldt and Rhine estuaries: validation with new field data"
Estuarine Coastal and Shelf Science **2005**, *62*, 141-160.

Curriculum Vitae

Chris de Graaf was born on November 12th 1978 in Leiderdorp, the Netherlands. In 1997 he graduated from high school (Gymnasium) at the Eerste Christelijk Lyceum in Haarlem. In the same year he started his study in Chemistry at the Universiteit van Amsterdam (UvA). He did his major internship in Environmental Chemistry at the Section Environmental Chemistry and Toxicology of the Institute of Biodiversity and Ecosystem Dynamics (IBED), resulting in three Master's theses: on the bacterial degradation of azaarenes (under the supervision of Dr. John R. Parsons), on computational molecular modeling of NDO-substrate interactions (Prof. Dr. Harrie A.J. Govers), and on the construction of a fate model to describe the behaviour of nonylphenol ethoxylates (Dr. Pim de Voogt). In 1998 and 1999 he received grants from the Association of the Dutch Chemical Industry (Vereniging Nederlandse Chemische Industrie (VNCI)) for his study results as a Dutch Chemistry student, and he was awarded by the Division of Environmental Chemistry of the Royal Dutch Chemical Society (Koninklijke Nederlandse Chemische Vereniging (KNCV)) for the best Dutch Master's thesis in environmental chemistry in the year 2002. Both for his B.Sc. ('propedeuse') and for his M.Sc. he graduated with honours (*cum laude*). He performed his PhD research described in this thesis from January 2002 to January 2006 at the Division Molecular Toxicology of the Leiden/Amsterdam Centre for Drug Research (LACDR) at the department of Chemistry and Pharmaceutical Sciences of the Vrije Universiteit (VU) Amsterdam, under the supervision of Prof. Dr. Nico P.E. Vermeulen and Dr. Chris Oostenbrink. During this period he did research internships at the ETH Zürich to learn new automated docking approaches with Prof. Dr. Gerd Folkers, and at the EML Research Institute in Heidelberg to learn new molecular dynamics simulation approaches with Dr. Rebecca C. Wade. Since February 2006, he is a post-doctoral fellow in the Bioinformatics of the Drug group of the CNRS 7175 Unit at the Universitée Louis Pasteur (ULP) in Strasbourg (France), performing G-protein coupled receptor (GPCR) modeling studies under the supervision of Dr. Didier Rognan and in collaboration with AstraZeneca.

Nawoord

```
 G              C    F    G
```
How does it feel?
With no direction at home
Like a complete unknown
Like a rolling stone

Singer-songwriting poetical wizzard Bob Dylan verwoordt de verwarring, desoriëntatie, isolatie, en ontheemding. Geen grip op het eigen lot, schijnbaar stuurloos als 'rolling stones'. De naam van andere favoriete muzikale meesters overigens. Het zijn gevoelens die de wetenschapper overvallen, die de promovendus in hun greep houden, alleen en opgesloten in zijn eigen wetenschappelijke puzzels, gebundeld in wat uiteindelijk een opus magnum behoort te worden, behoort te zijn. Het proefschrift, hopelijk heeft het boekwerk u kunnen bekoren, u kunnen doen verwonderen. Ik heb in ieder geval erg veel plezier beleefd aan en tijdens het werk dat leidde tot dit alles. Ik bedank daarom allen die mij hebben geholpen en geïnspireerd, allen die mij in mijn weg van een rolling stone hebben bijgestaan. Een dankwoord middels een dichtrijm, geleid door het ritme van Bob Dylan's eigen opus magnum.

```
C                    Dm7         Em                              F       G
```
Nico is de man of MolTox clan, met promotor gen, en ook een fan van zijn pupillen
Je stond me bij, en steunde mij, een stuurde mij naar buiten en de weide wereld in
```
F              G        F                    G
```
Je vertrouwen in in silico, en geloof in een jonge AiO
```
F                    Em          Dm7                        C
```
Een maecenas en tuh-tuh-tuh, so-so, struinend door de zalen ter KTP-info
```
C                              F        G
```
Lettend op de grote lijnen en de puntjes op de i

Chris, ik ben aan jou, en je hersens en je trouw zoveel verschuldigd
Dertig maanden geen, maar vervolgens wat voor één! Super-visor
Je smeedde een gezellig team, vol intelligentie en met self esteem
Vliegensvlug correcties, manuscripten grime, fysische precisie, vrije energie
Het was een eer om jouw eerste promovendus te zijn

Peter, teamgenoot, amigo, makker, paranymf, mijn ware CYP brother in arms
Veel gelachen, veel geschreven, een onderzoekers web geweven tussen onze labs
Vol respect keek ik af hoe jij, steeds een AiO-stapje voor was op mij
Vele congressen bezochten wij, Praag, Mainz, Dallas, en daarbij blijven wij
In Holland, Italië, en Frankrijk altijd een biertje doen

Alois, il mio caro amico, grazie di tutto il tuo supporto
A social life sound-bord, scientific sparring sport, Raman maestro
Italiaanse meester in Hollandse gezelligheid, kokende-drinkende-entertainende-gastvrijheid
Mountain man and sweater fan, muziek maak partner, with Diego, jam
Trieste e bella, i ricordi sono belli

Saskia, jij promootte mij en het project, dat mijn positie heeft gedekt, het CCM
Dus ook dank aan de CCM'ers, scheikunde discipline mengers, voor hun cross-over inzichten
Aike, André, Andreas, Cees, Gert, Harm
Johnny, Koop, Marcel, Mirka, Patrick, Stéphane
Laat de onderzoeksbanden voortduren en voortgaan

Pavel, dock around the clock, the Czech-n'-Zurich rock, l'importance de l'eau
Thank you as a scientist and a friend, in Zurich, Praha, and wherever, man
Gerd, thank you, for charismetic support, inviting me to your lab, powerful positive reference
Rebecca, thank you for inspiration and intelligence, randomly accellerating from Neckarthal to EML with the bike you lent
Ainsi que mes collegues et amis Strasbourgeois, je vous remercie pour tout

Anton, roomy, review nijver, MD meester, scriptjes schrijver, in P275
Mp3-pc-dj, muziek schalde door de silly ruimte van CMCT
Aldo, jij toonde mij, de flair en lef om wetenschapper te zijn
Autodidact, allround, en leergierigheid, een voorbeeld voor het overwinnen van onzekerheid
Marola, gezellig collega klankbord op de tonen van de Counting Crows

Jeroen, oh halfgod Herc, Kreta, kermis in Heemskerck, tl-buis tent
Eierbakker, commando sluiper, een goede vriend om mee te zuip ..eh, na noeste arbeid
Sebastiaan, Parijs-kompaan, goudgeel ouwehoeren na onze onderzoekers baan
Achter Noordwijkse & Lunterse, (be)togen staan, essentiële AiO-excercitie ter contact en faam
Zoals ook Robert en Jelle is onderwezen

Micaela, quoi, p&t%in, tellement gentille, sympa, oranje staat je goed
Eva, blondy, we had a laugh, "snel-op-de-kast", gezelligheid, Lamme Goedzak & scientific vibe
Leading to a superb next CMCT laudatio, en dank aan Eric-Wubbo, de evolutione fragmento
Barbara, mutanten-maakster, van F naar A, moleculair biologe, met ratio in silico
Jan, wetenschaps scherp rechter, muzikale fan

Studenten Wouter, Tushar, Ahmed, Maya, onderwezen mij, smeedden data, daarom mijn dank
Naar al mijn Moltox collega's alle lof en dank, kom strofen tekort, maar de lijst is nog lang
Atilla, Ed, Francine, John, Jozef, Hande, Hilmi, Jennifer,
Laura, Martijn, Peter, Regina, Rolien, Sara
En alle andere gezellige Nederlandse & buitenlandse vakgenoten, in laboratoria en op symposia

Marjo, vriendschap, gelijk van geest, cultuur en kunst, herkennings feest, als mentale steun
Wetenschaps interesse overstegen, maar via wetenschap contact gekregen, science klikt
Niels, charmante globetrotter, inspiratio voor mijn eigen post-doc avontuur, amicum respicio
Arjen, Bourgondisch gastvrij genieter, makker, gulle lach, Daniël, trouwe vriend, corps drinkgelag
Moleculaire dynamica regina, werkbezoekjes en etentjes met Jocelyne

Chiel, vriend, met interesse in mijn vreemde wetenschapspuzzels, tot morele steun gediend
Evenals Bas, Bob, Sanne, Irene, Joost, Rijk, Pim, Vilija, Tim, Johan, Wiek, Esther, Enny, Kim, André, Daniël, Erik, Frank, Gabriëlle, Lennart, Mattijs, Martijn, Michel, Monique, Ron, Rogier, Steven, Sudi, Suzanne, Thyra, Yu Ling, Ronald, Cora, Mirjam, Alice, Jeroen, Florine, Sandrijn, Thijs, en een nog langere lijst van vrienden & vriendinnen, Kersthuisjevierders, ACD'ers, Haarlemmers, Amsterdammers, waar ook ter wereld
En ook mijn voetbal makkers van Alliance, HFC, van Headroom, urban funk rock in tha house
Willem, Ciro, Brian, Jules, muziek zit in het hart, zal altijd inspireren in leven en wetenschap
Marie, I thank you for all the wunderful memories

Pieter, als eerste doceerde jij mij, motiveerde jij mij voor scheikunde en wetenschap
Ook mijn UvA-MTC leermeesters John, Harrie, Pim, bedankt voor jullie milieuchemische lessen
Omi, Oma, Nel, Monique, familieliefde, interesse en steun, banden waar ik op leun, liefs & dank
De laatste strofes zijn voor mijn grootste supporters, inspirators, top-ouders, top-managers, met gevoel voor mijn wetenschaps-sport
Papa en mama, ik kom woorden te kort, maar dit boekje is voor jullie, trots in het hart
Pluim op mijn werk, mijn gift voor jullie onvoorwaardelijke steun en liefde

Iedereen, bedankt voor alles! Door jullie voelde ik mij juist niet verloren en ontheemd. Maar de vervreemdende verwondering, het mysterie van de wetenschap, moet blijven bestaan, zal altijd blijven bestaan. Dus tot slot, als parafrase op Dylans poëtische woorden:

That's how it feels
To be with a home
To be loved and known
Still like a rolling Stone