Pharmacogenetic modeling of human cytochrome P450 2D6

On the force of variation in inducing toxicity

Inauguraldissertation

Zur Erlangung der Würde eines Doktors der Philosophie
vorgelegt der Philosophisch-Naturwissenschaflichen Fakultät der Universität Basel

von

Charleen Georgette Don

Basel, 2020

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Daniel Ricklin
PD Dr. Martin Smieško
Prof. Dr. Daan P. Geerke

Basel, den 18. Februar 2020

Prof. Dr. Martin Spiess
(Dekan)

Contents

"Meden agan" ~ 'nothing in excess'

Chilon of Sparta (6th century BC)

Chapter 1

General introduction

In part based on the review article:

Out-compute drug side effects: Focus on cytochrome P450 2D6 modeling

Charleen G. Don & Martin Smieško.

WIREs Comput Mol Sci 2018;8:e:1366

1.1 Historical perspective

"*What is there that is not poison? All things are poison and nothing is without poison. Solely the dose determines that a thing is not poison.*"

This famous dictum of the Swiss scientist Paracelsus (1493-1541) lies at the core of the research described in this thesis. Even though he only held a position for less than a year at the university of Basel due to his unpredictable temperament, during his life he lay the foundation for modern safety/toxicity research and assessment [1]. His most prominent studies showed that the body's response to a foreign substance could be beneficial or harmless when given in low concentrations whereas toxic at high concentrations.

One of the earliest application of this concept can be traced back to the Olympic Games in Greece during the third century BC. Eager to win the games, athletes relied on natural performance enhancing "drugs" such as for example hallucinogenic mushrooms to overcome fatigue and injuries or also "analgesic bread" which was prepared with opium from poppies and relieved pain. The emergency for doctors and cooks to prescribe the exact dosage to the athletes was paramount in winning and avoiding any deleterious effects during the games which nevertheless did occur from time to time [2]. More than 500 years later Paracelsus concept has become one of the most fundamental principles in toxicology and it is daily applied to determine the toxicity of a substance by identifying the dose-response correlation.

Over time we became aware that apart from the crucial function of dose in toxicity assessments several other essential factors contribute to determining the fate of a substance to be harmful to our body or not. For instance, the type of dose (e.g. inhaled, ingested, intravenous, dermal), the intrinsic physicochemical properties of the substance (e.g. solubility, molecular weight, flexibility), or exposure frequency (e.g. duration and single/multiple times, circadian rhythms) can impact its toxic potential. Also the organism itself can have an impact on the substance; gender (e.g. differences in body weight), age (some chemicals are more toxic to children), genetic predisposition (the genetic make up of a person can increase their susceptibility towards a chemical), the microbial enzyme

composition in our gut or also the type of species (human or animal) can again influence the degree of toxicity [3,4]. Regulations and guidelines have been tailored by governmental institutions such as the US Federal Drug Administration (FDA) and European Medicine Agency (EMA) that should prevent or limit adverse effects of chemicals (e.g. drugs, food additives, or cosmetics) we daily interact with and might form a potential health risk. Foreign substances (xenobiotics) that found a way into our body usually encounter one of our first lines of defence at the liver, the cytochrome P450 superheme family. The enzymes selectively recognise xenobiotics and metabolize them into more hydrophilic compounds which facilitates their secretion mainly by the kidneys.

We have gradually unravelled parts of these truly extraordinary and complex functioning enzymes. The wealth of information we gathered helped to solve parts of the giant P450 puzzle and answer questions such as what their function is, where they are located or how the catalysis cycle is organized. The first experimental piece of evidence that has proven the existence of this unique superheme family dates back to the late 1940s [5]. Millers and co-workers studied the metabolism of the aminoazo dye N-N-dimethyl-4-aminoazobenzene and its correlation with the formation of tumours. During the experiments, they discovered that a particulate fraction of rat liver cells was able to catalyze the metabolism of the aminoazo dye studied [6]. Around the same time, similar *in vitro* metabolism experiments were performed at the Brodie Laboratory. Julius Axelrod, who was awarded the medicine Nobel prize in 1970, uncovered the presence of an enzyme system found in rabbit liver microsomes which could oxidize psychoactive drugs (e.g. amphetamines and ephedrine). The metabolism process required a soluble factor, which he identified to be NADPH [5,7]. The official "P450" part of the family name came from spectrophotometric studies performed in the Chance Laboratory. The prime focus of the lab was to understand the reactions of oxidative phosphorylation catalyzed in the liver and heart mitochondria. After identifying the existence of a heme-binding protein in the supernatant of the rat liver mitochondria (now known as cytochrome b) by spectrophotometric studies in 1954, their discovery four years later was determining for the P450 research field. The German physical chemist Klingenberg observed that when carbon monoxide was gassed through a cuvette

containing rat liver microsomes and a reducing agent (e.g. NADPH), a spectrum with maximum absorbance at 450 nm was recorded [8]. This initiated several studies aimed to characterize the found pigments in the rat liver microsomes. Sato and Omura determined the pigments as heme-proteins in 1962 [9]. They designated the pigments "cytochromes".

1.2 CYPs in drug discovery and medical treatment

"*It is more important to know what sort of person has a disease than to know what sort of disease a person has.*" As the Greek physician Hippocrates already pointed out 2500 years ago, determining the genetic profile (genotype) of a person to personalize the medication accordingly is crucial for effective treatment and avoid side effects. Metabolism of drugs, the process in which a non-polar compound (parent drug) is chemically modified into a polar compound (metabolite) occurs mainly in the liver [10]. The reaction is catalysed by various enzymes and divided into phase I (oxidation, reduction and hydrolysis) and phase II (conjugation). During phase I the parent drug is chemically modified into a more polar metabolite by introduction of a polar group (e.g. OH or SH). The produced phase I metabolite(s) can be pharmacologically inactive, active (parent drug is inactive, called prodrug), or some of the metabolites are active but less than the original drug. During phase II, the phase I produced metabolites are conjugated with a charged group (e.g. glutathionide or glycine) which further increases their solubility and promotes the secretion mainly by the kidneys. CYPs are the major drug-metabolizing phase I enzymes and their activity is genetically determined. Inter-individual CYP-genotype variability can alter drug metabolism (faster, slower or none) after administering the same drug dose to different patients [13]. Slower drug metabolism results in increased plasma concentrations and overexposure of the drug thereby increasing toxicity risks. On the other hand, faster elimination of the drug prevents it from reaching sufficient bioavailability and efficacy [14]. Therefore, drug discovery, clinical trials and personalized therapy include CYP-related assessments to determine the drug metabolism profile (which isoforms are involved) and potentially identify the the CYP genomic profile of patients to adjust the drug therapy accordingly and minimize the chance on unexpected adverse reactions.

1.2.1 The famous P450 family member CYP2D6

The human P450 cytochrome superfamily is encoded by 57 functional genes of which primarily members of family 1, 2 and 3 (particularly CYP1A2, CYP2E1, CYP3A4, CYP2C9, CYP2C19 and CYP2D6) account for the metabolism of therapeutic drugs (>75%) whereas other family members are involved in metabolism of endogenous substrates such as fatty acids, eicosanoids and vitamins [10]. CYP2D6 is responsible for about 25% of overall drug clearance despite its low share of 2% of the liver cytochrome pool (concentration) [11]. In addition, CYP2D6 is highly polymorphic with over 100 variant alleles known (http://www.PharmVar.org). This number is still growing with the ongoing P450 sequencing projects worldwide [12].

Pharmacogenetic variation of the P450 encoding genes (duplication, base pair insertions/deletions, copy number variations and single nucleotide polymorphisms (SNPs) [13] has often a dramatic impact on the expression and functional activity of the enzyme [14] which will be discussed in section 1.2.2. The CYP2D6 variants CYP2D6*4, CYP2D6*5, CYP2D6*10, CYP2D6*17 and CYP2D6*41 are among the most frequently observed allelic variants throughout populations worldwide [11,15,16]. The overall P450 fold is conserved among the different CYP families [17,18]. The structure is composed out of twelve α-helices (A-L) and four small β-sheet domains (**Figure 1.1A**). The P450 family carefully conserved the region around the heme in order to perform the generally uniform P450 catalytic cycle. The conserved cavity is formed by helices D, E, I, J, K, L, β-sheet 1 and 2, and the meander loop [18]. The most variable regions are the domains that are involved in the formation of the channels and control the passage of ligands and co-factors through the cytochrome.

Currently, several channels are known based on X-ray studies and Molecular Dynamics (MD) simulations [19,20]. Cojocaru et al. performed a channel study on 26 P450 (mammalian, bacterial and archaea) X-ray structures and provided a general nomenclature. The naming is based on the lining of the secondary structure elements involved. The five most commonly observed channels (solvent, 2a, 2ac, 2c, and 2f) are located within the region of helices F and G [21,22]. Six general substrate recognition sites (SRS) are known (**Figure 1.1B, 1.1C**) [23]. Important CYP2D6 residues directly involved in the binding include (i) Glu216 and

Asp301 associated with recognition and positioning of the ligand, and (ii) Phe120, Phe381 and Phe483, suggested to control orientation of the aromatic rings found in most substrates. From the two essential recognition residues, Glu216, not found in any other CYP2 subfamily, is thought to be foremost responsible for initial ligand binding, which is supported by experimental and modeling studies [24,25].

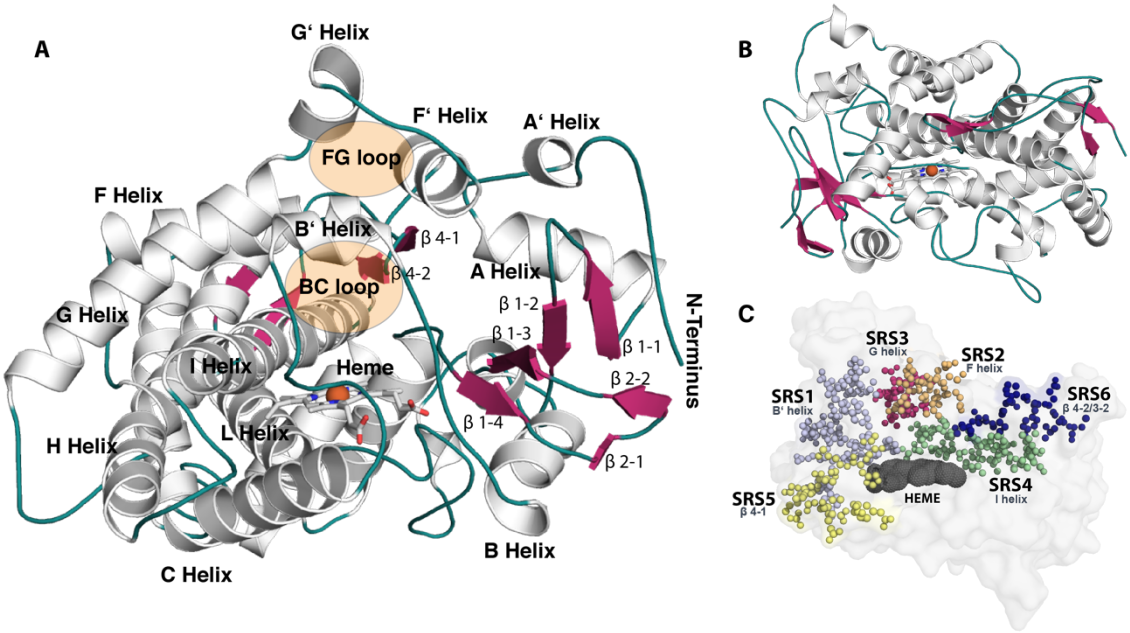
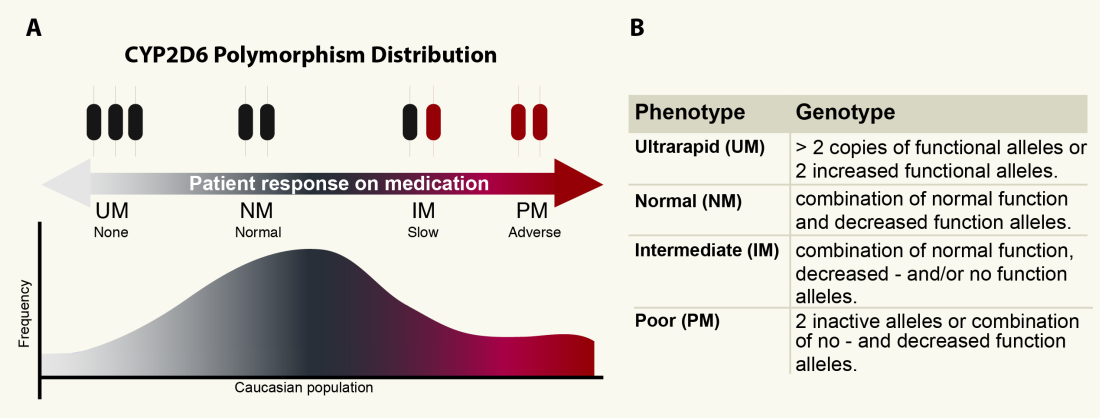


Figure 1.1 **Cytochrome P450 2D6 (CYP2D6) structural characteristics. A)** CYP2D6 (PDB ID 3QM4) is displayed with annotated secondary structural elements. The α-helices are shown in white, β-sheets in pink, and loops in cyan. The FG and BC loop, important for regulation of the binding cavity access, are highlighted in orange. The CYP2D6 core of the active site cavity is formed by amino acids located on helices I, F, G, loop KL and FG, and the heme. **B)** Same CYP2D6 structure as shown in (a) but rotated to match the orientation shown in c. **C)** The six major substrate recognition sites (SRS) are displayed and located around: The B' helix (SRS1, gray), the F helix (SRS2, orange), the G helix (SRS3, pink), the I helix (SRS4, green), β 4-1 (SRS5, yellow), and β 4-2/3-2 (SRS6, blue)

1.2.2 Clinical relevance of CYP2D6 pharmacogenetics

The amount of clinical evidence regarding which *CYP2D6* allelic variants may predispose an individual to a higher risk of developing a disease or guide drug choice and/or dose is continuously growing [26]. Therapeutic areas where *CYP2D6* polymorphism has shown a

major impact are psychiatry (ADHD, depression) and pain treatment [27]. Other areas where CYP2D6's role is still controversial include oncology (e.g. breast cancer), cardiology (e.g. cardiac arrhythmias) and neurological disorders (e.g. Parkinson disease) [28–30]. The presence of any other phenotype (ultrarapid (UM), intermediate (IM) or poor (PM) metabolizers) than the normal (NM) one can lead to treatment failure, life-threatening adverse reactions and even fatality in some cases (**Figures 1.2A and 2B**) [31].



Phenotype	Genotype
Ultrarapid (UM)	> 2 copies of functional alleles or 2 increased functional alleles.
Normal (NM)	combination of normal function and decreased function alleles.
Intermediate (IM)	combination of normal function, decreased - and/or no function alleles.
Poor (PM)	2 inactive alleles or combination of no - and decreased function alleles.

Figure 1.2 **CYP2D6 polymorphism distribution among Caucasians. A)** the genetic modification is shown (normal = black; impaired = red gene; explained in table B) with the corresponding phenotype, enzyme activity, and its relative frequency. Depending on the genetic differences among patients, the phenotypes are categorized as ultrarapid (UM), normal (NM), intermediate (IM), or poor (PM) metabolizers. **B)** Overview of different phenotypes and the corresponding genotypic description.

In case of Schizophrenia and Parkinson disease, *CYP2D6* polymorphism has been suggested to increase susceptibility towards the development of these diseases [29,32]. Improving our general understanding of *CYP2D6* polymorphism and the mechanisms underlying CYP2D6 functioning can contribute to the development of improved tools that support the currently changing practice of physicians to more frequently incorporate the patient pharmacogenetic profiling for determining the appropriate drug and its dose in order to avoid aforementioned risks [33]. A predictive *in silico* CYP2D6 tool would ideally provide information on the likelihood of a drug-CYP2D6 interaction and give an indication (score) on the potential toxicity. SNPs based amino acid changes could be incorporated into the CYP2D6 model as well to evaluate the impact of *CYP2D6* polymorphism-induced structural changes on drug-cytochrome interactions.

However, even if the CYP2D6 phenotype is unambiguously identified, there are additional relevant factors which need to be assessed for establishing an accurate clinical phenotype: (i) defining the overall contribution of the CYP2D6-mediated pathway and of other possible pathways in the clearance of the drug from the body, (ii) the activity of the parent molecule and the metabolite(s), (iii) co-administered drugs (or herbs) and (iv) the therapeutic index of the drug [34]. All of these factors would require assessment by the pharmacokinetic modelling experts during the drug design process.

For clinical cases where a drug depends on CYP2D6 metabolism, polymorphisms can have various consequences for the metabolism and activity of the drug [30,35]. The drug can be either administered as an active compound (e.g. *S*-metoprolol), which frequently will be metabolized into an inactive metabolite for easier clearance from the body or as a prodrug (e.g. codeine) requiring CYP2D6 metabolism to form the active compound needed for reaching its pharmacological effect (e.g. pain relief). If the patient's phenotype deviates from normal CYP2D6 activity, the enzyme activity can either be higher, lower or absent (**Figure 1.2A,B**). Subjects with PM and IM phenotypes are characterized by little to no and decreased CYP2D6 activity respectively and compromise drug metabolism. In contrast, subjects with an UM phenotype (usually three or more functional copies of *CYP2D6*) show increased enzyme activity and metabolize drugs at a higher than the normal rate [35].

For a prodrug, poor metabolism can potentially lead to poor drug efficacy with an increased risk of a therapeutic failure. Moreover, the prodrug might accumulate in the body increasing the chance of drug-induced side effects. On the other hand, ultrarapid metabolism of the prodrug can lead to sudden elevated (toxic) quantities of the active metabolite that can result in death as for instance was the case with codeine [36]. For an active drug, the potential consequences are slightly different. Good drug efficacy might be achieved for slow metabolizers, nevertheless, gradual accumulation of the active drug increases again the risk of drug-induced side effects and may require dose lowering. Ultrarapid metabolizers usually suffer from poor drug efficacy leading to an increased risk of treatment failure [31]. In some cases, increasing the dose might be an option depending on the drug dosage index. If ample evidence is found that proves a clear link between *CYP2D6* allelic variation

and the clinical treatment outcome, a standard genotype test might be implemented that would help to determine the optimal patient dose and enhance the chance on a successful treatment outcome.

To address this issue, the Clinical Pharmacogenetics Implementation Consortium (CPIC), established in 2009, procure dosing guidelines for over 90 drugs for which genetic variants (and the corresponding phenotype) showed to have an impact on drug response [37]. The guidelines (https://cpicpgx.org, https://www.pharmgkb.org) are based on different types and levels of evidence and help physicians to justify the use of pharmacogenetic tests in the clinic. Currently, 12 drugs are listed with CYP2D6 guidelines and more convincing proofs need to be collected for over 40 drugs as the number of cases with adverse responses is still insufficient. A clinical example of 'codeine' (one of the CPIC listed CYP2D6 drugs) which illustrates the complexity of determining the correlation between *CYP2D6* allelic variation and drug metabolism unambiguously and also the potential added value that recommendation of *CYP2D6* testing can have to lower adverse responses in patients and optimize the choice and strategy of the clinical treatment can be found in reference 38.

1.3 Enzymology of CYP2D6

The word "enzyme" was derived from the greek word "ἔνζυμον" which means "leavened" or "in yeast". An enzyme (a protein) enables catalysis of a compound (substrate) by transforming it into a new product. The speed-up of the reaction requires the presence of the enzyme, also called the catalyst. Without the enzyme the activation barrier of the reaction would usually be too high and the chance on a spontaneous reaction would be very low. In addition, the enzyme is not consumed during the reaction, hence can be smartly recycled all times. Thus enzymology studies focus on understanding the overall chemical reaction process and its environmental setting.

1.3.1 P450 catalytic cycle

The P450 catalytic reaction depends on the presence of several co-factors including the ligand(s), molecular oxygen, water and the cytochrome P450 reductase (CPR, also known as

POR). The latter is responsible for the electron delivery by interaction of its FMN domain with the enzyme (**Figure 1.3**) [39].

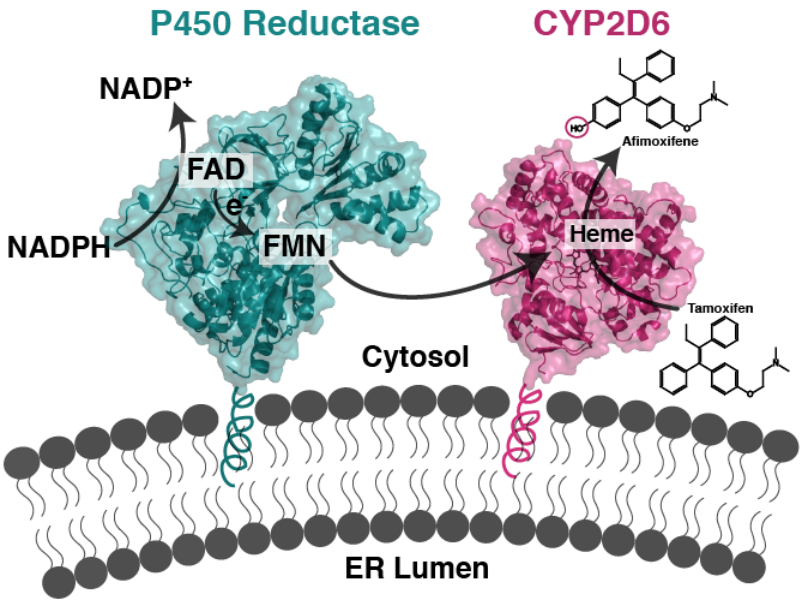


Figure 1.3 **CYP2D6 Microenvironment displaying the three major components required for the catalytic cycle.** Molecular environment of cytochrome P450: CYP2D6 (497 aa) anchored (drawn part) to the membrane (endoplasmic reticulum) at its N-terminus side (pink, PDB ID 4WNU[i]) together with its redox partner NADPH-Cytochrome P450 Reductase (CPR, 676 aa) (cyan, PDB ID 1AMO[i]). CPR shuttles the electrons from nicotinamide adenine dinucleotide phosphate (NADPH) to the CYP where they are used for the catalytic oxidation of tamoxifen into 4-hydroxtamoxifen (afimoxifene). It is proposed that lipophilic substrates can access the active site directly through the membrane (pathway II) in addition to the cytosolic entry (pathway I) [40].

The consensus P450 catalytic cycle can be seen in **Figure 1.4**; at the beginning of the cycle, CYP2D6 is in a resting state in which a water molecule is covalently bound to the iron (III) **(A)**. Upon binding of a substrate to the ferric enzyme, the distal water is displaced as the sixth heme ligand **(B)**. The first reduction of the iron (III) complex reduces the iron to the (deoxy)iron (II) state, also known as "the ferrous state" **(C)**. This enables the binding of a molecular oxygen, forming a ferrous oxyiron(II) complex **(D)**. The second reduction results in an activated oxygen species **(E)**. Being a very strong base, it abstracts a proton and forms the iron(II)-peroxide intermediate **(F)**. A second protonation follows, which usually occurs at the distal OH group, enabling heterolytic oxygen-oxygen bond scission and a water

molecule leaves (one oxygen atom with two protons and electrons) resulting in the iron (IV) coupled porphyrin radical cation **(G)**. This highly reactive oxyferryl intermediate incorporates an oxygen atom into the substrate, producing the oxidized product (R-OH) **(H)**. Dissociation of the product turns the cytochrome back to its initial resting state **(A)**. The second reduction can however also sometimes occur on the proximal oxygen and a so-called 'uncoupling' reaction occurs. During the catalytic disruption, a hydrogen peroxide is formed which leaves the active site and the ferri-heme intermediate is regenerated **(B)**. Whenever the delivery of the second proton is inhibited (e.g. by mutation of an involved active side residue) the uncoupling reaction will occur [41].

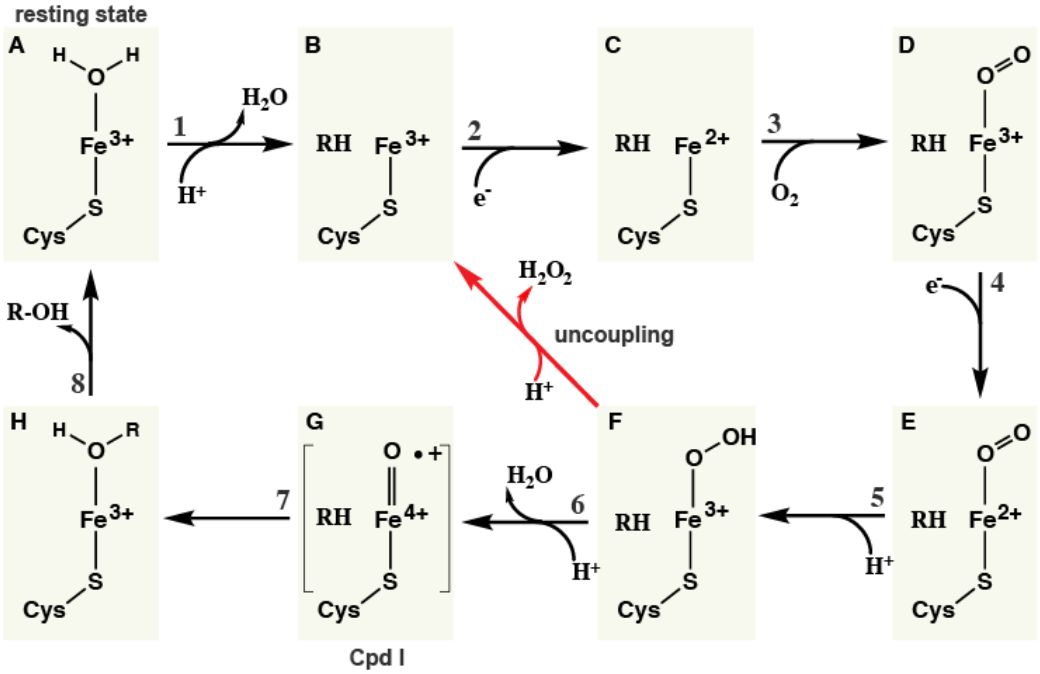


Figure 1.4 **Consensus cytochrome P450 catalytic cycle.**

1.3.2 Hydrogen abstraction

The majority of drugs metabolized by CYP2D6 usually have a rather hydrophobic character and contain one or more aromatic rings (e.g. warfarin or codeine). Understanding the mechanism(s) underlying the oxidation reaction is of importance to avoid undesired metabolites causing toxicity and improve our ability to develop computational tools that are able to predict the most likely metabolites.

Hydroxylation of a strong aliphatic or aromatic C-H bond by CYPs is mediated through its high-valent oxo-ferryl species, known as compound I (Cpd I). In order to understand the

catalytic mechanism, characterization of the major reactive intermediate(s) is crucial to know before studying the reactive pathways. Due to the highly reactive nature of Cpd I, it took several years before Rittle and Green successfully characterized Cpd I of CYP119 in 2010 [42]. Their results confirmed that Cpd I exists as an oxoiron(IV) porphyrin cation radical and is the key intermediate species for hydrogen abstraction. An important other intermediate is the hydroxoiron(IV) complex, known as compound II (Cpd II) or rebound intermediate. It has been shown that the reactivity of the metal-oxo system towards C-H bond scission is correlated to the reaction driving force; the free energy difference between the broken C-H bond and the formed O-H bond in the ferryl oxygen. Hence, the more basic the ferryl in Cpd II (pK_a ~12), the stronger the ferryl-hydroxy bond, and the more effective the oxidation reaction [43,44]. Within this context it has been suggested that the proximal Cysteine ligand plays an essential role in increasing the basicity of Cpd II (the ferryl-oxygen) by pushing electrons towards it. Thus this tight basicity regulation enables the enzyme to achieve a feasible reduction potential for C-H bond scission without causing oxidative damage to itself.

Aliphatic or aromatic C-H hydroxylation is generally thought to be primary catalyzed by the high-valent oxo-ferryl species, compound I (Cpd I). The now generally accepted rebound mechanism for aliphatic hydrogen abstracted was proposed by Groves et al in 1976 [45]. The mechanism in shown in **Figure 1.5**. In the first step, the oxoiron (IV) porphyrin cation radical intermediate abstracts an aliphatic hydrogen from the substrate and a Fe^{IV} – OH (Cpd II) species and a radical substrate is formed.

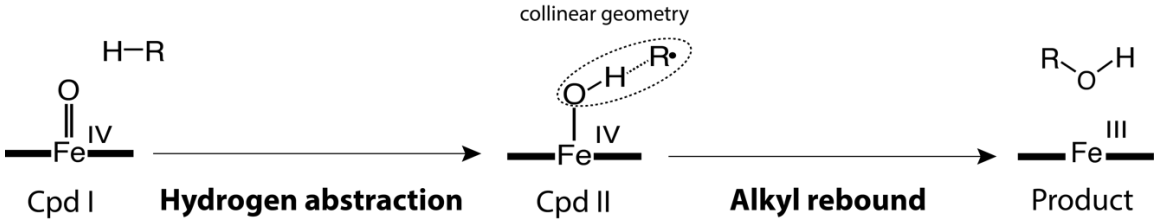


Figure 1.5 **Rebound mechanism for aliphatic hydrogen abstraction.**

In the second step, the "rebound" of the substrate radical to the iron hydroxyl forming the product. During the reaction, the transition state reaches a collinear Fe-O ---- H-C geometry [45]. Camphor hydroxylation by $P450_{cam}$ was among the first QM/MM studies that supported

the proposed rebound mechanism [45]. Values for the activation barrier are generally found to be in a range of 11 kcal/mol to 25 kcal/mol. Depending on the used QM method (e.g. DFT or B3LYP) force field and system set-up (environmental residues included in the QM region and water) the activation barrier can be slightly higher or lower.

Aromatic oxidation by CYPs has shown to produce both phenolic and benzene epoxide products (**Figure 1.6**) [47,48]. The latter product can induce toxic effects in our body as it likely will react with nucleophiles such as DNA or proteins and should be avoided if possible. Initially it was suggested that the epoxide (**Figure 1.6, pathway C**) was a necessary intermediate based on the observed shift of the hydrogen from the hydroxylation side to the closest carbon (**Figure 1.6, pathway B**) [49]. The product of this reaction will be the epoxide but when non-enzymatic ring-opening occurs of this group, a phenol can also be formed. Later experimental and computational studies provided evidence for an additional pathway without epoxide intermediate, which is now known as the addition-rearrangement mechanism (**Figure 1.6, pathway A**) [50,51]. This pathway is generally found to be lowest energy pathway.

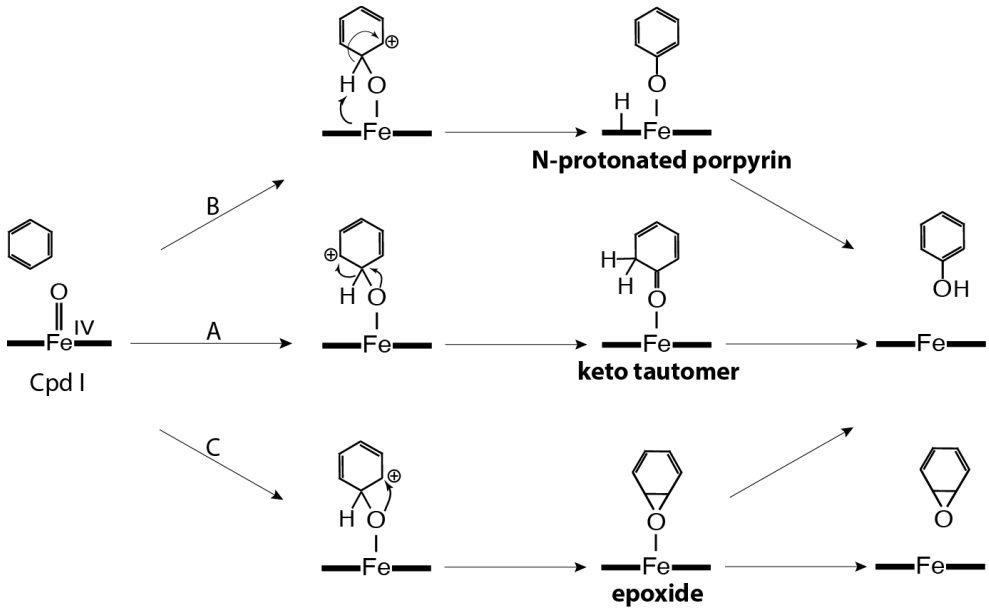


Figure 1.6 **Three proposed rearrangement pathways for aromatic hydrogen abstraction [49]**.

Moreover, the step with the highest energy barrier is found to be the addition of Cpd I to a substrate carbon that leads to the tetrahedral intermediate s-complex [49]. From this state,

different products can be formed depending on the rearrangement of the intermediate. Another possible pathway based on DFT studies proceeds through an N-protonated porphyrin intermediate. From this intermediate, the ketone and/or direct phenol complex formation is possible [52,53]. Which products will be formed will be the outcome of the interplay between active site geometry and/or the substrate reactivity which is well illustrated by a study of Oláh et al [54]. Experimental studies on the oxidation of dextromethorphan by CYP2D6 did not lead to any aromatic hydroxylation of the compound. Though MD simulation and docking studies demonstrated that the putative reaction sites all were viable since they were energetically accessible for the heme-oxo, QM/MM studies provided the key to the experimentally only observed O-demethylation. The activation energy for the aromatic carbon oxidation (around 14 kcal/mol – 17 kcal/mol) had a much higher activation energy compared to the aliphatic hydrogen abstraction (around 10 kcal/mol – 12 kcal/mol).

1.4 State-of-the art CYP2D6 modeling

The journey that a potential drug has to take before entering the market is tough and takes on average between 10 and 15 years and costs millions [55]. The computer-aided drug design (CADD) unit of a pharmaceutical company has become indispensable to assist in the drug design and development process and largely contributes to lower R&D costs [56,57]. Typically, the initial discovery stage involves the identification of a therapeutic target against which a drug needs to be developed. Subsequently, if lucky, several lead compounds are identified and multiple cycles of optimization follow. Here bioinformatics approaches as well as structure-based - and/or ligand-based modeling approaches are applied (**Figure 1.7**) [58]. The best lead candidates continue to a preclinical *in vivo* and *in vitro* testing phase and if positive results return clinical trials (phase 1-3) follow. During the three clinical phases, 79% of the clinical failures can be attributed to safety or efficacy issues as an clinical drug study found over the time period 2016-2018 [59]. Accurate drug metabolism prediction, especially of CYP2D6, is one of the crucial aspects during the drug design process as this may contribute to lower the attrition rate as explained in the clinical relevance section 1.2.2.

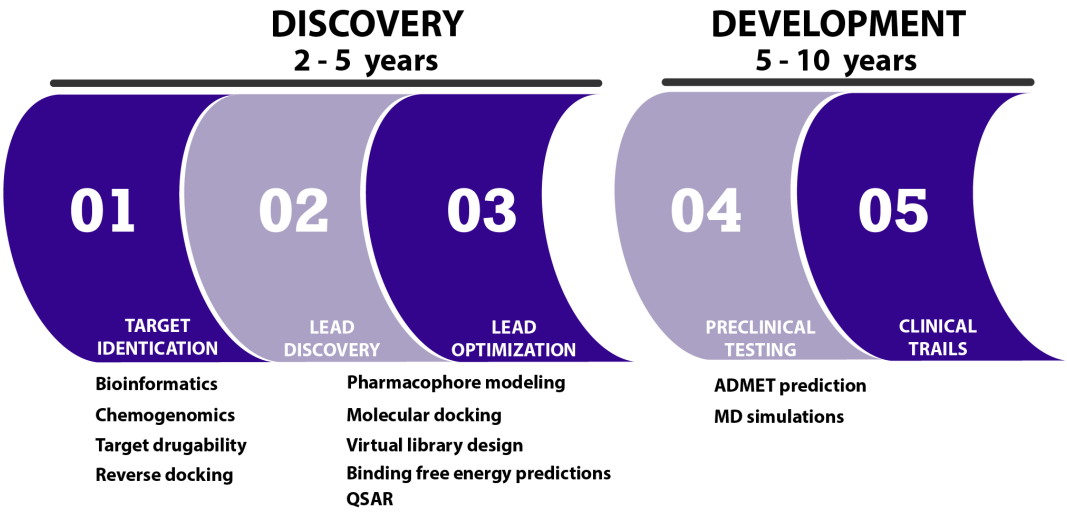


Figure 1.7 **Overview of the drug discovery and development stages and the major computational methods that are applied in the different pipeline phases.**

In silico modeling of drug metabolism can provide the most comprehensive atomic-level data that can assist medical chemists early on in the drug development process. The potential toxicity attributable to CYPs can be predicted by assessing the probability and degree of compound interaction with the enzyme [60–62]. Currently, the modeling approaches in predictive toxicology typically rely on information about the wild type protein, but in case of highly polymorphic targets such as CYP2D6 an accomplished study should simultaneously consider all important allelic variants (SNPs). The approach can employ a ligand- or structure-based strategy (or both) depending on the available structural data, computational resources and the toxicology endpoint(s) to investigate (**Figure 1.8**) [60,61]. The next sections will discuss both CADD CYP2D6 approaches (sections 1.4.1 and 1.4.2) and describe the basic concepts of MD simulations and QM/MM methods (sections 1.4.3 and 1.4.4) as they were the major techniques used for studying CYP2D6 in this thesis.

1.4.1 Ligand-based strategies

These are commonly applied to predict for instance the site of metabolism (SoM), biotransformation(s), binding affinity or the produced metabolite(s) using physicochemical and biological data **(Figure 1.8**). Pharmacophore modeling methods require a set of compounds with known activity which are aligned on a selected template by matching their molecular properties [63]. A hypothetical structure-activity model is then built by

mathematically extracting the critical structural and electronic features from the molecules, which are required for their interaction with the target and correlation with their biological activity. Such a model can cover predictions regarding the SoM or the CYP2D6 ligand character (inhibitor, substrate). In drug discovery, the pharmacophore approach can be conveniently used as a filter to screen large compound databases for ligands with particular desired or undesired properties (High Throughput Screening (HTS)). The main drawback of this otherwise fast and powerful approach is the uncertainty about the bioactive conformation and the assumption that all compounds interact with the target following a similar binding mode without inducing any substantial rearrangement of the target (induced fit), which is not always the case [64,65]. An advanced version of the pharmacophore abstraction, the three-dimensional (3D) Quantitative Structure-Activity Relationship (QSAR) modeling, relies on the ligand 3D structures. The main advantage of this approach is its ability to predict, in addition to the qualitative properties, also quantitative properties, especially the ligand binding affinity [64]. Nevertheless, inherent lack of any 3D structural information regarding the highly flexible target renders pharmacophore methods less suited for studying the impact of SNPs on drug-enzyme interactions.

1.4.2 Structure-based strategies

If (along with known ligands) also structural data of atomic-resolution is available for the cytochrome, the ligand and the structure-based approach can be advantageously combined to search and subsequently obtain a consensus outcome. If two methodologically different approaches provide the same result (e.g. suggest favourable binding interaction) then such a result can be considered more reliable (**Figure 1.8**). In case of cytochromes, information about the target can be integrated in the assessment of possible toxicologically relevant outcomes (e.g. nature of the metabolites, inhibition).

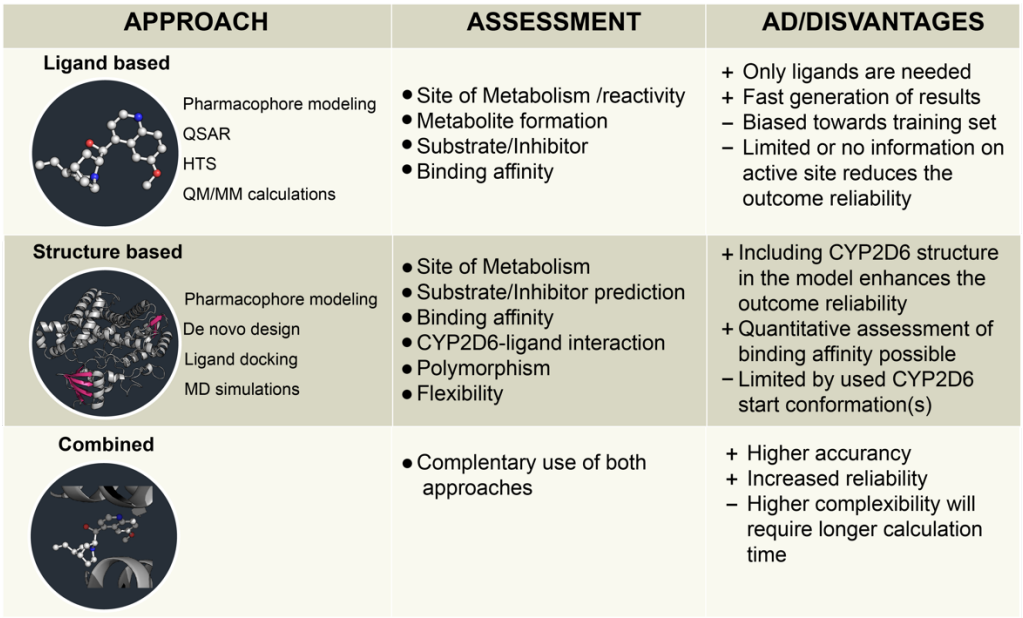


Figure 1.8 **Overview of computational strategies applied in cytochrome P450 2D6 toxicology modeling.** If experimental atomic data on the target is available, a structure-based or combined (ligand- and structure-based) approach is advantageous. This enables the modeling of the ligand in the native binding pocket with increased accuracy.

Modeling of the native CYP2D6 active site enables addressing possible issues related to different phenotypes (SNPs), identifying likely ligand binding mode(s) and exploring the dynamic features of the ligand-enzyme complex. For building such an advanced model a well-resolved structure of the complex is required. Before 2005, no experimental coordinates of CYP2D6 were available in the Protein Data Bank (PDB), hence the structure had to be predicted by homology modeling. Introducing a ligand into the cytochrome is known to trigger conformational changes of the active site [66]. Therefore, molecular docking and MD simulation studies usually rely on a X-ray or homology model based on a holo-determined CYP structure [67]. A very important step is the validation of the model. This is typically done by assessing the ability of the model to reproduce known ligand binding mode(s) and by cross-docking. The first CYP2D6 homology model was built based on the cytochrome P450 101 (from organism *Pseudomonas putida*) although it contained only a part of the active site [68]. Several more complex homology models followed based on template structures that converged with the cytochrome species X-ray release history (from easier bacterial towards more complex mammalian/human). However, studies based

on homology modeling must be always interpreted with caution: even if a high sequence similarity exists between the template and query sequences, a certain degree of uncertainty about the position of the active site residues is still unavoidable [69]. This can have a major impact on modeling results, especially if delicate information is sought for (e.g. ligand binding modes featuring complex pattern of weak intermolecular interactions). A more comprehensive explanation of pitfalls in homology modeling can be found in the literature [70–72]. In general, the homology models were used to investigate the enzyme fold, active site composition, ligand-CYP2D6 interactions and to predict ligand SoM specificity [24,73–76].

Molecular Docking. With a CYP2D6 3D model in hand, energetically favourable binding conformations of the ligand can be explored towards wild-type and variants using molecular docking. This procedure can be divided into two phases. In the first phase, an algorithm generates a pool of possible binding poses. In the second phase, the generated poses are scored and ranked based on the contribution from various energy components (e.g. H-bonding, electrostatics, van der Waals, entropy, desolvation, etc.). When applying this method to CYP2D6, several critical points need to be addressed as they can significantly impact the reliability of the results and the overall docking performance:
(i) flexibility of the active site: since the majority of docking algorithms still does not feature full protein flexibility (with a few exceptions e.g. HADDOCK [77]) popular strategies applied to CYP2D6 that can deal with this challenge to some extent are ensemble docking [78] or induced-fit docking [79]. Ensemble docking utilizes a strategy in which multiple protein conformations are used to represent the flexibility of the protein. The ensemble can be collected for example from an MD trajectory or determined experimentally (e.g. NMR spectroscopy). The ensemble is explicitly treated during the docking and besides movement of the side chains it can also account for the protein backbone flexibility [80]. This is an advantage compared to the induced-fit approach, in which only readjustment of the side chains in the active site is allowed. Upon ligand docking, the induced-fit algorithm then

typically searches for the most stable binding site arrangement by a combinatorial scan of the side chain rotamer libraries [81,82].

(ii) importance of including structural water molecules during the docking: most water molecules present in the X-ray structure are fundamental for the enzyme function and important for ligand recognition [81,82]. However, the majority of the previous CYP2D6 docking studies removed the solvent from the active site before docking. Including water into the CYP docking procedure is still questionable since the outcome of several studies assessing the impact of in- or excluding water molecules upon docking ligands to CYP2D6 (or other isoforms) is not consistent [73,81].

(iii) conformation of the ligand: the lowest energy pose is not necessarily the most reactive one and the ligand might be able to bind in multiple orientations within the cytochrome. This fact should be kept in mind when selecting liganded CYP2D6 X-ray structures for computational studies [65]. If the docking approach heavily depends on the co-crystallized ligand template or the shape of the cavity associated with this particular ligand, the pose accuracy for dissimilar compounds might be very likely compromised. Instead of focusing too much on searching for the pose with the most favourable energy terms, a suitable clustering approach can be used to identify the representative binding pose(s) within a reasonable energy window that orients the possible SoM more closely toward the heme.

(iv) scoring function: no universal scoring function exists so far and different strategies are applied for their parameterizing (knowledge-based, empirical, force-field based or consensus). Considering the numerous approximations on which the force fields are based, obtaining directly realistic binding free energies of chemical accuracy for the docked ligands is extremely challenging. Quantum mechanical (QM) calculations or hybrid approaches offer a more accurate treatment of intermolecular interactions, however at an increased computational time cost which usually disqualifies them from a large scale application (e.g. virtual screening) [83]. With regard to the hydrophobic nature of the CYP2D6 active site, a scoring function that underestimates the lipophilic contribution would also likely suffer from loss of accuracy during the filtering and ranking process [84]. Furthermore, special parameterization of the metal-ligand interactions (the cytochrome heme group) is required

to represent the contacts between the ligand groups (lone pair donor) and the heme iron atom. The choice of the heme state (covalently bound to dioxygen or not) requires the corresponding heme parameter set to equilibrate well with all the energy terms within the scoring function. Nowadays, several functions exist that are parameterized for cytochrome studies including Chemscore, Gold and Glide (OPLS_2005 force field) [84,85]. The most popular molecular docking programs and scoring functions used for CYP2D6 studies are Autodock, Gold (Chemscore, Goldscore), FlexX, DOCK and Cscore [61,86].

Site of metabolism. In order for the catalytic reaction to occur, the ligand needs to be in close vicinity of the heme-oxo iron atom and present a matching intrinsic reactivity group towards the activated oxygen species. A cut-off value of 6 Å is recommended for use during the pose assessment. The value is based on the assumption that the oxidation reaction is initiated by the extraction of a hydrogen atom, located at the ligand SoM, forming the radical iron-oxo species (compound I). This requires a SoM-oxyferryl distance bellow 3 Å, but taking thermal motion into account the upper limit might extend to a maximum distance of 6 Å [87]. More sophisticated approaches such as MetaSite [88] or IDSite [89] use a combined strategy which takes also into account the reactivity of the ligand by means of quantum mechanics calculations (in addition to the geometric criteria). In addition, DR-Predictor combines flexible docking, reactivity calculations and machine-learning algorithms to predict CYP mediated SoMs (addressing the regioselectivity as well) [90].

Molecular dynamics. Though automated docking protocols are eagerly used due to their speed and low cost, the dynamics-related properties such as binding affinity calculations, residence time, flexibility and channel formation remain in the twilight zone or out of reach without using Monte Carlo - or molecular dynamics (MD) simulation methods. The latter allows for studying the enzyme-ligand complex's multidimensional energy landscape over a given time period and capturing its thermodynamic and kinetic fingerprint by determining the relative probability of various states and the energy barriers separating them. Thoroughly sampled time-evolved conformations can be used to predict the free energy difference between CYP2D6 ligand-bound and unbound states. This way useful insights can be gained on the dynamic character of the ligand-cytochrome interaction, thus

simultaneously addressing flexibility and binding affinity. Most commonly used methods are thermodynamic integration (TI), free energy perturbation (FEP) or linear interaction energy (LIE) [91,92]. The latter has proven to be of particular value for fast (1 ns production time) and accurate CYP2D6 binding affinity predictions (RMSD error < 1 kcal.mol^{-1}) [93].

Several studies have employed MD simulations to investigate channel formation [21], water dynamics [81], effect and importance of the membrane on CYP behaviour [94–97], binding and reaction mechanisms [76,98], refinement of docking results, and *CYP2D6* polymorphism [21,99,100]. These studies are essential for extending our current CYP2D6 knowledge and subsequently seamlessly integrate this information into advanced CYP2D6 *in silico* techniques. The most frequently used MD programs are GROMACS, GROMOS, AMBER, CHARMm, and Desmond [61,86]. The technique is limited by the trade-off between accuracy and simulation time: the more detailed and complex the model system, the more demanding the calculations and hence the longer the computational time required. The usefulness of the data generated from short time trajectories (bellow 100 ns) might be doubted considering the uncertainty of the time needed for most biological phenomena to occur (e.g. large conformational changes develop on a micro-second time scale). A detailed description of MD simulation challenges can be found in ref [101]. Nevertheless, for studying protein-ligand movements and calculation of properties that do not require electron transfer between atoms (e.g. bond formation or polarization), the method can be accurate enough to generate data of sufficient quality. In addition, the computational power available to perform a large number of calculations within an acceptable time frame and at reasonable costs is levelling the last couple of years and becomes more attractive for industrial use as well (e.g. APO A1 protein with 92'000 atoms, 118 ns/day, on 1 GeForce GTX Titan Xp GPU) [102].

QM/MM. Deciphering the dynamical aspects of P450 by use of molecular mechanistic (MM) force fields has proven to be of tremendous value. As described in the previous section, many important physiological P450 processes have been studied such as the binding of ligands, flexibility of the binding pocket and tunnel formation, but for some events MD simulations are not suitable. The classical additive force fields treat atoms as point charges

and bonds as harmonic oscillators, changes in electronic structure caused by a chemical reaction can not directly be modeled using this technique.

To study the reactivity of P450s, *ab initio* methods such as quantum mechanics (QM) are favorable. Considering the fact that increasing system size increases the computational time and resources needed, initial P450 studies were performed on a small iron-porphyrin system using primarily density functional theory (DFT) to describe the electronic system [103]. These studies contributed largely to our understanding of the electronic structure of the active species, the different intermediate states, and also its specificity. However, the necessity of including the protein environment in the active site calculations is crucial as it was found to be the defining factor for the different specificity of the CYP isoforms [104,105]. Therefore, the hybrid quantum mechanical/molecular mechanical (QM/MM) approach is more suitable to study reactivity patterns. Schöneboom performed the first QM/MM study using $P450_{cam}$ (also known as CYP101) and revealed that the oxo-heme moiety is stabilized by specific hydrogen bonding interactions between the cysteine sulfur and nearby residue thereby sustaining the Fe-S bond and preventing the sulfur from oxidation [106]. Subsequently, following QM/MM studies have been mainly focused on elucidating the electronic structures of all the species in initially $P450_{cam}$ but later on also for the human isoforms, and the reaction mechanism underlying the oxidation of several organic molecules [46,49,104,107–110]. The major metabolism reactions studied using a QM/MM approach include hydroxylation (aliphatic or aromatic) [111], double bond epoxidation [52,112] and sulfoxidation [113,114].

1.4.3 Computer simulation

Molecular Dynamics (MD) simulation is a powerful method to explore structure and dynamics of biomolecular systems complementing experimental studies. In order to calculate and predict properties of the system under study, a reliable model and accurate algorithm is indispensable. In addition, carefully post-processing the simulation trajectories is crucial as well. A protein can be viewed at from four different layers: 1) its primary structure which is defined by its amino acid sequence, 2) its secondary structure, composing the α-helices and β-sheets, 3) its tertiary structure, defining the spatial relationship of at least two

secondary structural elements and thus its overall monomeric protein fold and 4) its quaternary structure which is the arrangement of multiple folded monomeric proteins into a multi-subunit. Such an assembly enables flexibility and movement of the protein and can be rather seen as a Boltzmann-weighted ensemble of poses/configurations instead of just one rigid configuration. MD and Monte Carlo (MC) simulations are two methods to generate such a Boltzmann-weighted ensemble of e.g. CYP2D6 poses. MC generates moves of atoms based on the change in potential energies and momentum conservation is not preserved. MD simulations on the other hand, solve Newtons equations of motion:

$$\boldsymbol{F}_i = m_i \boldsymbol{r}_i \qquad (1.1)$$

where m_i is the mass -, r_i the position - and F_i the force on the particle (atom) i. This makes it a more suitable choice to study the movement of proteins over time. Depending on the computational power available and the size of the biomolecule studied, the simulation time scale can vary from ns to several μs.

Before starting MD simulation, four basic choices need to be made in order to obtain a reliable outcome [115];

i) Selecting the appropriate degrees of freedom set is essential to avoid sampling irrelevant space or simulating for prohibitive time to obtain maximal accuracy and precision of the calculations.

The most fine-grained degrees of freedom possible in MD simulations are models in which atoms are treated as particles as done in quantum-chemical (QM) calculations. This approach allows to study the interactions between nuclei and electrons which is necessary for instance to determine chemical reactivity as demonstrated in chapter 2. On the other end of the scale are coarse-grained (CG) models which largely reduce the degrees of freedom by grouping atoms into larger 'pseudo' atoms or beads. Therefore, CG MD simulations can lower the computational costs by several orders of magnitude enabling numerical simulations over longer time scales [116]. The classical atomistic model lies in between the two previous described models and treats one atom as one particle.

Finding an appropriate model to address the problem in question is thus a delicate matter that needs a suitable balance between accuracy and speed.

ii) Another related choice that will influence the outcome of the calculations strongly is the type of force field and parameter set used. The classical Hamiltonian is usually referred to as force field can be divided into a kinetic K (p) and a potential V (**r**) energy term:

$$H\ (\boldsymbol{p},\boldsymbol{r}) = K(\boldsymbol{p}) + V(\boldsymbol{r}) \tag{1.2}$$

in which $\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2,\ldots,\mathbf{p}_N)$ and $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2,\ldots,\mathbf{r}_N)$ and the 3N-dimensional vectors of the particle momenta and positions for a system of N particles. The kinetic term is independent from the particle position:

$$K(\boldsymbol{p}) = \sum_{i=1}^{N} \frac{\boldsymbol{P}}{2m_i} = \sum_{i=1}^{N} \frac{1}{2} m_i \boldsymbol{v}^2 \tag{1.3}$$

m_i is the mass of particle i and the 3N-dimentional vector of the particle velocities, $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2,\ldots,\mathbf{v}_N)$ which is the time derivate of the position vector $\mathbf{v} = \mathbf{r}$.

The potential energy term in **Equation 1.2** is the sum of several bonded (b) and nonbonded (nb) particle interactions for a given configuration:

$$V(\boldsymbol{r}) = V^b\ (\boldsymbol{r}) +\ V^{nb}\ (\boldsymbol{r}) \tag{1.4}$$

Depending on the force field, the (non)bonded terms can vary. Commonly the bonded interaction terms include bond-stretching, dihedral-angle bending and bond-angle bending terms to preserve planarity of the groups and e.g. chirality. The non-bonded terms include a Lennard-Jones (LJ) term to account for particle dispersion and repulsion plus an electrostatic term:

$$V^{nb}(\boldsymbol{r}) = \sum_{\substack{nb \\ pairs\ (i,j)}} \left\{ \left[\frac{C_{12}(i,j)}{r_{ij}^6} - C_6(i,j) \right] \frac{1}{r_{ij}^6} + \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1} \left[\frac{1}{r_{ij}} - \frac{\frac{1}{2}C_{rf} r_{ij}^2}{r_{rf}^3} - \frac{1-\frac{1}{2}C_{rf}}{R_{rf}} \right] \right\} \tag{1.5}$$

r_{ij} denotes the distance between particle i and j, the C_{12} (i,j) the repulsive and C_6 (i,j) the attractive LJ parameters, q_i and q_j the partial charges of particle i and j, ε_0 the permittivity

of the vacuum, ε_1 the relative dielectric permittivity of the system, and C_{rf} and R_{rf} the reaction field parameters.

The force acting on particle i is defined by the negative of the gradient of the potential energy term V (**r**). Combined with **Equation 1.1** this results in:

$$m_i \boldsymbol{r}_i = \boldsymbol{F}_i = -\frac{\partial}{\partial r_i} V(\boldsymbol{r}) = -\frac{\partial}{\partial \boldsymbol{r}_i} \left(V^b(\boldsymbol{r}) + V^{nb}(\boldsymbol{r}) \right) \quad (1.6)$$

The parameters for the terms in the force field are derived from different sources. For an atomistic model, the bond-length and bond-angle potential energy terms parameters can be obtained from quantum chemical calculations, spectroscopic or X-ray diffraction data. Parameters for the non-bonded torsional-angle interaction term and the partial charges are usually derived from quantum-chemical calculations on small molecules in the gas phase. The fitting of the parameters is also a factor which can be done in different ways. For example, OPLS2005 (largely based on OPLS-AA) or GROMOS fit the parameters of the non-bonded terms (Lennard-Jones parameters and charges) in such a way that they reproduce experimental thermodynamic properties (heats of vaporization, densities) [109,117]. The fitting of the charges can also be done by reproducing the electrostatic potential obtained from QM calculations as done by CHARMM11 or AMBER [118,119].

iii) Most real-life biomolecular systems are a mix of proteins and/or other molecules in an aqueous solution which would be far too large to sample all the degrees of freedom from with the current computational power. However, treatment of the spatial boundary conditions is crucial to avoid a negative impact of the surface on the calculated properties. This can be circumvented by using periodic boundary conditions in which the box with the molecular system is surrounded by an infinite number of copies of itself, mimicking the presence of an infinite bulk environment. The infinite copies introduce at the same time periodic contributions of long-range interactions which is an artefact of the method. The introduced error is negligible small for the Lennard-Jones interactions but larger for the electrostatic Coulomb interactions and needs to be corrected. The reaction-field method truncates the non-bonded interactions beyond a certain cut-off radius after which a homogeneous medium of constant dielectric permittivity is presumed.

iv) Ensuring adequate sampling of the relevant configurational space is a nontrivial task but crucial for accurate calculation of properties such as the free energy of the system or its enthalpy. For a biomolecular system the degrees of freedom can be as large as 10^4 to 10^6 or even larger and the corresponding movements can evolve on a femtosecond to millisecond time scale [115]. The diversity of movements is captured by the potential energy function and reflected in the free energy surface of the system containing energy hills and basins with varying heights and depths. Identifying the global minimum and sampling the regions that are most relevant in defining the free energy of the system can therefore become a difficult task. The phase space probability distribution function P(r,p) defines the probability of occurrence for a certain state of the system and for a canonical ensemble (NVT) can be expressed as:

$$P_{NTV}\,(\boldsymbol{r},\boldsymbol{p}) = \frac{e^{-\beta H(\boldsymbol{r},\boldsymbol{p})}}{\iint e^{-\beta H(\boldsymbol{r},\boldsymbol{p})} d\boldsymbol{p}d\boldsymbol{r}} = \frac{e^{-\beta H(\boldsymbol{r},\boldsymbol{p})}}{h^{3N} N! Z(N,V,T)} \tag{1.7}$$

where β is the inverse of the Boltzmann constant, k_B, multiplied by the temperature, T, N is the number of particles with volume V, h Planck's constant, and Z(N,V,T) the canonical partitioning function:

$$Z(N,V,T) = \frac{1}{h^{3N} N!} \iint e^{-\beta H(\boldsymbol{r},\boldsymbol{p})} d\boldsymbol{p}d\boldsymbol{r} \tag{1.8}$$

the factor N! is used in case of indistinguishable particles. At constant temperature, T, and pressure, p, (referred to as the isothermal-isobaric ensemble) the phase-space probability becomes volume dependent;

$$P(\boldsymbol{r},\boldsymbol{p},V) = \frac{e^{-\beta H(\boldsymbol{r},\boldsymbol{p})+pV}}{h^{3N} N! Z(N,\boldsymbol{p},T)} \tag{1.9}$$

and the partitioning function becomes:

$$Z(N,p,T) = \frac{1}{h^{3N} N!} \iiint e^{-\beta H(\boldsymbol{r},\boldsymbol{p})+pV} d\boldsymbol{p}d\boldsymbol{r}\, dV \tag{1.10}$$

Subsequently, the desired quantity (Q) can be calculated as the expected value (average indicated by the angular brackets) of the property over the whole phase space for the isothermal-isobaric or the canonical ensemble by:

$$\langle Q \rangle_{NVT} = \iint Q(\boldsymbol{r}, \boldsymbol{p})\, P(\boldsymbol{r}, \boldsymbol{p})\, d\boldsymbol{p} d\boldsymbol{r} \tag{1.11}$$

Which proves the fact that an efficient algorithm can generate an ensemble by sampling enough configurations within an acceptable amount of time. In order to accomplish this, the time step used in the integration of the equations of motion is key. Increasing the time step size enables to visit more configurations within the same number of integration steps. However, too large time steps can lead to inaccuracies, therefore this parameter needs to be selected carefully. In this thesis, the Desmond package implemented in Schrödinger suite was used for all MD simulations which applies the Velocity Verlet algorithm for integration:

$$\boldsymbol{r_i}\,(t + \Delta t) = \boldsymbol{r_i}(t) + \Delta t v_i\,(t) + \frac{\Delta t^2}{m_i} f_i(t) + O\Delta t^3 \tag{1.12}$$

$$\boldsymbol{v_i}\,(t + \Delta t) = \boldsymbol{v_i}(t) + \frac{\Delta t}{2m_i}(f_i(t) + f_i(t + \Delta t) + O\Delta t^3 \tag{1.13}$$

A time step of 0.002 ps has been used as recommended. In addition, the isothermal-isobaric (NPT, T=298, p = 1 atm) ensemble was used for simulations since most biological system function under these conditions.

1.4.4 QM/MM

The biochemical systems that can be described at any level of *ab initio* theory are limited because most of the quantum mechanics (QM) methods scale linear (or worse) with system size. However, to model processes in which chemical bonds are formed or broken a purely molecular mechanics (MM) approach will neither be successful since the force fields do not treat electrons explicitly but together with the nuclei as effective atoms. Focussing only on the active site to model e.g. enzymatic reactions is also not satisfying because neglecting

the environment and solvent effects will reduce the accuracy of the calculations [120]. One popular strategy is the combination of multiple resolutions within one simulation thereby creating a powerful synergy that overcomes the limitation of the individual methods. In such a hybrid approach, the system is usually divided into a smaller part (inner subsystem), where the chemical reactivity occurs and a larger part (outer subsystem) containing the rest of the system. Subsequently, the inner subsystem (e.g. the reactive heme and molecule) is modeled at the level of QM, whereas the larger subsystem (e.g. the protein environment) by MM, which is computational much cheaper in use. This hybrid QM/MM concept was first introduced by Warshel and Levitt in 1976 [121]. Studying biochemical systems using this approach gained popularity after the later performed studies by Karplus and coworkers [122].

Before preparing and starting a QM/MM simulation, five key aspects need to be considered: (i) the size of the QM region, (ii) the choice of the QM Hamiltonian, (iii) the choice of force field used for the MM region, iv) boundary and coupling of the QM/MM regions, and (v) boundary conditions for the MM region.

Defining the optimal QM region is necessary as different system sizes can considerably affect the energies and potentially cause convergence issues [123]. For QM/MM geometry optimization it has been shown that a QM system size of 6 to 13 residues can reach convergence, larger systems have frequently convergence problems and large varying energies (up to 70 kJ/mol) [124].

The inner QM region can be described at various levels. The basic concept behind QM calculations is that by providing the coordinates, the net charge and total spin of a biochemical system, the total energy and wave-function can be determined by solving the Schrödinger equations. From this result measurable properties (e.g. heat of vaporization, activation barrier energies) can be calculated. However, the Schrödinger equation can solely be analytically solved for single electron systems. For larger systems only approximate numerical solutions can be calculated. The Hartree-Fock (HF) is the most basic approach but due to its rather approximate result more accurate approaches have been developed using it as a starting point such as the perturbation theory (e.g. Møller-Plesset second-order

perturbation (MP2)) or the series expansions (e.g. coupled cluster calculations with single, double and perturbatively treated triple excitations). A very frequently used approximation is the density functional theory (DFT) which is not based on the wave-function but on the electron density. Hence the calculations are a function of only the three coordinates of Cartesian space, and not of all involved particles as the case is for the wavefunction. A great variety of DFT approximations exists such as PBE [125], TPSS [126] or BP86 [127]. In addition, hybrid DFT methods exist that involve a part of the exchange from the HF method; B3LYP or M06-2x [128,129] both hybrid meta-generalized gradient approximations (GGA's). The quality of the DFT calculations are usually comparable to MP2 results while computational much more inexpensive [130]. The electronic wave function is represented by a set of functions, the basis sets. In general, the more extensive the basis set, the more accurate the electrons of the involved atoms in the system are described and the more reliable the results. Semi-empirical QM approaches can speed up the calculations largely by using a minimal basis set and substituting the integrals by empirical parameters. Examples are AM1 [131], PM3 [132], OM1-3, or SCC-DFTB [133]. One prominent shortcoming of both HF and DFT approximations is the lack of a proper treatment of the London dispersion interactions. Several methods have been developed to account for this short-coming including the DFT-D2 or B3LYP-D3 [134–136].

For the outer MM region, the most popular force fields applied are: the Chemistry at Chemistry at HARvardMolecular Mechanics (CHARMM) [137,138] force field, the Optimized Potentials for Liquid Simulations – All Atoms (OPLS-AA/OPLS3) force field [139,140], the Amber force field [119,141], and the Groningen Molecular Simulations (GROMOS) force field [142,143].The coupling of the inner and outer region is the most critical aspect of the QM/MM approach. To calculate the total energy of the complete system (S), two coupling schemes exists that incorporate the inner (I), outer (O) and the linking (L) region (also referred to as boundary region) [144].

The energy for the subtractive (Sub) coupling scheme is expressed by:

$$E_{QM/MM}^{\boldsymbol{Sub}} = E_{MM}(S) + E_{QM}(I + L) - E_{MM}(I + L) \tag{1.14}$$

The energy for the most commonly used additive (Add) coupling scheme is expressed by:

$$E_{QM/MM}^{\boldsymbol{Add}} = E_{MM}(O) + E_{QM}(I+L) + E_{QM-MM}(I,O) \tag{1.15}$$

in which

$$E_{QM-MM}(I,O) = E_{QM-MM}^{bonded} - E_{QM-MM}^{vdW} - E_{QM-MM}^{elec} \tag{1.16}$$

This is the so-called 'coupling term' which describes the bonded, van der Waals (vdW) and electrostatic (elec) interactions between the QM and MM region. The electrostatics is the most dominant interaction type between the regions and is handled by different levels of approximations: Mechanical embedding treats the interactions at the MM level and is therefore less accurate and often used compared to electrostatic embedding. Here the electrostatics interactions are fully treated at the QM level and the two other types of interactions (vdW and bonded) both at the MM level. The point charges on the MM atoms are included in the QM calculations. Therefore, polarization of the QM region by the MM region is explicitly accounted for. However, to obtain polarization both ways (MM on QM and QM on MM) a polarizable force field is required for the MM part and a QM program which can deal with polarizabilities which are both not yet too common. One remaining challenge is the treatment of covalent bonds in the linking/boundary region which need to be correctly truncated. A very specific treatment would be by placing hybrid orbitals [121,145,146] or using frozen localized orbitals [121,147,148] at the boundary to saturate the bond but more commonly applied is the hydrogen-link-atom scheme. In this approach a link atom (hydrogen) is placed to saturate the bond. Positional restraining can be applied to keep the geometrical constraints of the former bond. One important artefact of the link approach is over-polarization caused by the partial charges close to the QM region. Several methods have been developed to deal with this artefact [149,150]. For example, the charge shift scheme shifts the MM charges that are located at the boundary region of the covalent

bonds to the next layer of MM atoms, thereby preserving the total charge and dipole moment [151,152].

Both mechanical and electrostatical embedding can be used for the additive or subtractive scheme. The advantage of using an additive scheme is that the MM calculations do not require parameters for the QM region and link atoms as they are not part of the MM calculations. The subtractive scheme does require accurate parameters for the full system because MM calculations also cover the QM region. It is therefore also considered to be more difficult in preparing the system for the QM/MM calculations [153].

Different QM/MM programs are nowadays available in several QM and or MM software packages including Q-site [154], Gaussian (ONIOM) [155] or CHARMM [156]. In addition, ChemShell [157] or ComQum [158] are independent interfaces that combine different QM and MM software.

References

1. Grandjean P. Paracelsus Revisited: The Dose Concept in a Complex World. Basic Clin Pharmacol Toxicol. 2016;119: 126–132. doi:10.1111/bcpt.12622

2. Tsatsakis AM, Vassilopoulou L, Kovatsi L, Tsitsimpikou C, Karamanou M, Leon G, et al. The dose response principle from philosophy to modern toxicology: The impact of ancient philosophy and medicine in modern toxicology science. Toxicol Rep. 2018;5: 1107–1113. doi:10.1016/j.toxrep.2018.10.001

3. Horii I. The principle of safety evaluation in medicinal drug - how can toxicology contribute to drug discovery and development as a multidisciplinary science? J Toxicol Sci. 2016;41: SP49–SP67. doi:10.2131/jts.41.SP49

4. Koontz JM, Dancy BCR, Horton CL, Stallings JD, DiVito VT, Lewis JA. The Role of the Human Microbiome in Chemical Toxicity. Int J Toxicol. 2019;38: 251–264. doi:10.1177/1091581819849833

5. Brodie BB, Axelrod J, Cooper JR, Gaudette L, La Du BN, Mitoma C, et al. Detoxication of drugs and other foreign compounds by liver microsomes. Science. 1955;121: 603–604. doi:10.1126/science.121.3147.603

6. Attia SM. Deleterious effects of reactive metabolites. Oxidative medicine and cellular longevity. 2010. doi:10.4161/oxim.3.4.13246

7. Axelrod J. The enzymatic demethylation of ephedrine. J Pharmacol Exp Ther. 1955;114: 430–438.

8. Klingenberg M. Pigments of rat liver microsomes. Arch Biochem Biophys. 1958;75: 376–386. doi:10.1016/0003-9861(58)90436-3

9. Omura T, Sato R. A new cytochrome in liver microsomes. J Biol Chem. 1962;237: 1375–1376.

10. Guengerich FP, Waterman MR, Egli M. Recent Structural Insights into Cytochrome P450 Function. Trends Pharmacol Sci. 2016;37: 625–640. doi:10.1016/j.tips.2016.05.006

11. Ingelman-Sundberg M. Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity. Pharmacogenomics J. 2005;5: 6–13. doi:10.1038/sj.tpj.6500285

12. Zhou Y, Ingelman-Sundberg M, Lauschke VM. Worldwide distribution of cytochrome P450 alleles: A meta-analysis of population-scale sequencing projects. Clin Pharmacol Ther. 2017; n/a-n/a. doi:10.1002/cpt.690

13. Probst-Schendzielorz K, Viviani R, Stingl JC. Effect of Cytochrome P450 polymorphism on the action and metabolism of selective serotonin reuptake inhibitors. Expert Opin Drug Metab Toxicol. 2015;11: 1219–1232. doi:10.1517/17425255.2015.1052791

14. Johansson I, Ingelman-Sundberg M. Genetic polymorphism and toxicology--with emphasis on cytochrome p450. Toxicol Sci. 2011;120: 1–13. doi:10.1093/toxsci/kfq374

15. Gaedigk A. Complexities of CYP2D6 gene analysis and interpretation. Int Rev Psychiatry. 2013;25: 534–553. doi:10.3109/09540261.2013.825581

16. Peñas-LLedó EM, LLerena A. CYP2D6 variation, behaviour and psychopathology: implications for pharmacogenomics-guided clinical trials. British Journal of Clinical Pharmacology. 2014;77: 673–683. doi:10.1111/bcp.12227

17. Poulos TL. Cytochrome P450 flexibility. PNAS. 2003;100: 13121–13122. doi:10.1073/pnas.2336095100

18. Werck-Reichhart D, Feyereisen R. Cytochromes P450: a success story. Genome Biol. 2000;1: reviews3003.1-reviews3003.9.

19. Cojocaru V, Winn PJ, Wade RC. The ins and outs of cytochrome P450s. Biochim Biophys Acta. 2007;1770: 390–401. doi:10.1016/j.bbagen.2006.07.005

20. Wade RC, Winn PJ, Schlichting I, Sudarko null. A survey of active site access channels in cytochromes P450. J Inorg Biochem. 2004;98: 1175–1182. doi:10.1016/j.jinorgbio.2004.02.007

21. de Waal PW, Sunden KF, Furge LL. Molecular Dynamics of CYP2D6 Polymorphisms in the Absence and Presence of a Mechanism-Based Inactivator Reveals Changes in Local Flexibility and Dominant Substrate Access Channels. PLoS One. 2014;9. doi:10.1371/journal.pone.0108607

22. Waring MJ, Arrowsmith J, Leach AR, Leeson PD, Mandrell S, Owen RM, et al. An analysis of the attrition of drug candidates from four major pharmaceutical companies. Nat Rev Drug Discov. 2015;14: 475–486. doi:10.1038/nrd4609

23. 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.

24. de G, Oostenbrink C, Keizers PHJ, van V-L, van W, Tschirret-Guth RA, et al. Molecular modeling-guided site-directed mutagenesis of cytochrome P450 2D6. Current Drug Metabolism. 2007;8: 59–77. doi:10.2174/138920007779315062

25. Guengerich FP. Cytochromes P450, Drugs, and Diseases. Molecular Interventions. 2003;3: 194. doi:10.1124/mi.3.4.194

26. Crews KR, Hicks JK, Pui C-H, Relling MV, Evans WE. Pharmacogenomics and individualized medicine: translating science into practice. Clin Pharmacol Ther. 2012;92: 467–475. doi:10.1038/clpt.2012.120

27. Hicks JK, Sangkuhl K, Swen JJ, Ellingrod VL, Müller DJ, Shimoda K, et al. Clinical pharmacogenetics implementation consortium guideline (CPIC) for CYP2D6 and CYP2C19 genotypes and dosing of tricyclic antidepressants: 2016 update. Clin Pharmacol Ther. 2016. doi:10.1002/cpt.597

28. Brauch H, Schroth W, Goetz MP, Mürdter TE, Winter S, Ingle JN, et al. Tamoxifen Use in Postmenopausal Breast Cancer: CYP2D6 Matters. J Clin Oncol. 2013;31: 176–180. doi:10.1200/JCO.2012.44.6625

29. Lu Y, Peng Q, Zeng Z, Wang J, Deng Y, Xie L, et al. CYP2D6 phenotypes and Parkinson's disease risk: a meta-analysis. J Neurol Sci. 2014;336: 161–168. doi:10.1016/j.jns.2013.10.030

30. Wang B, Yang L-P, Zhang X-Z, Huang S-Q, Bartlam M, Zhou S-F. New insights into the structural characteristics and functional relevance of the human cytochrome P450 2D6 enzyme. Drug Metab Rev. 2009;41: 573–643. doi:10.1080/03602530903118729

31. Zackrisson AL, Lindblom B, Ahlner J. High Frequency of Occurrence of CYP2D6 Gene Duplication/Multiduplication Indicating Ultrarapid Metabolism Among Suicide Cases. Clinical Pharmacology & Therapeutics. 2010;88: 354–359. doi:10.1038/clpt.2009.216

32. Gassó P, Mas S, Alvarez S, Trias G, Bioque M, Oliveira C, et al. Xenobiotic metabolizing and transporter genes: gene-gene interactions in schizophrenia and related disorders. Pharmacogenomics. 2010;11: 1725–1731. doi:10.2217/pgs.10.158

33. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. Nature. 2015;526: 68–74. doi:10.1038/nature15393

34. Hicks JK, Swen JJ, Gaedigk A. Challenges in CYP2D6 phenotype assignment from genotype data: a critical assessment and call for standardization. Curr Drug Metab. 2014;15: 218–232.

35. Gaedigk A, Sangkuhl K, Whirl-Carrillo M, Klein T, Leeder JS. Prediction of CYP2D6 phenotype from genotype across world populations. Genet Med. 2017;19: 69–76. doi:10.1038/gim.2016.80

36. Kelly LE, Rieder M, van den Anker J, Malkin B, Ross C, Neely MN, et al. More codeine fatalities after tonsillectomy in North American children. Pediatrics. 2012;129: e1343-1347. doi:10.1542/peds.2011-2538

37. Relling M, Klein T. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. Clin Pharmacol Ther. 2011;89: 464–467. doi:10.1038/clpt.2010.279

38. Don CG, Smieško M. Out-compute drug side effects: Focus on cytochrome P450 2D6 modeling. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2018;8: e1366. doi:10.1002/wcms.1366

39. Estrada DF, Laurence JS, Scott EE. Cytochrome P450 17A1 Interactions with the FMN Domain of Its Reductase as Characterized by NMR. J Biol Chem. 2016;291: 3990–4003. doi:10.1074/jbc.M115.677294

40. Scott EE, Wolf CR, Otyepka M, Humphreys SC, Reed JR, Henderson CJ, et al. The Role of Protein-Protein and Protein-Membrane Interactions on P450 Function. Drug Metab Dispos. 2016;44: 576–590. doi:10.1124/dmd.115.068569

41. Guengerich FP. Mechanisms of Cytochrome P450-Catalyzed Oxidations. ACS Catalysis. 2018;8: 10964–10976. doi:10.1021/acscatal.8b03401

42. Rittle J, Green MT. Cytochrome P450 Compound I: Capture, Characterization, and C-H Bond Activation Kinetics. Science. 2010;330: 933–937. doi:10.1126/science.1193478

43. Mayer JM. Proton-coupled electron transfer: a reaction chemist's view. Annu Rev Phys Chem. 2004;55: 363–390. doi:10.1146/annurev.physchem.55.091602.094446

44. Huang X, Groves JT. Beyond ferryl-mediated hydroxylation: 40 years of the rebound mechanism and C–H activation. J Biol Inorg Chem. 2017;22: 185–207. doi:10.1007/s00775-016-1414-3

45. Groves JT, McClusky GA. Aliphatic Hydroxylation via Oxygen Rebound. Oxygen Transfer Catalyzed by Iron. American Chemical Society. 1976;98: 859–861. doi:https://doi.org/10.1021/ja00419a049

46. Lin H, Schöneboom JC, Cohen S, Shaik S, Thiel W. QM/MM Study of the Product−Enzyme Complex in P450 $_{cam}$ Catalysis. The Journal of Physical Chemistry B. 2004;108: 10083–10088. doi:10.1021/jp0493632

47. Lovern MR, Turner MJ, Meyer M, Kedderis GL, Bechtold WE, Schlosser PM. Identification of benzene oxide as a product of benzene metabolism by mouse, rat, and human liver microsomes. Carcinogenesis. 1997;18: 1695–1700. doi:10.1093/carcin/18.9.1695

48. Jerina DM, Daly JW, Witkop B, Zaltzman-Nirenberg P, Udenfriend S. 1,2-naphthalene oxide as an intermediate in the microsomal hydroxylation of naphthalene. Biochemistry. 1970;9: 147–156. doi:10.1021/bi00803a019

49. Bathelt CM, Mulholland AJ, Harvey JN. QM/MM modeling of benzene hydroxylation in human cytochrome P450 2C9. J Phys Chem A. 2008;112: 13149–13156. doi:10.1021/jp8016908

50. Benzene Oxidation by 'Bare' FeO+ in the Gas Phase - Schröder - 1992 - Helvetica Chimica Acta - Wiley Online Library. [cited 18 Jul 2019]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/hlca.19920750429

51. Shiota Y, Suzuki K, Yoshizawa K. Mechanism for the Direct Oxidation of Benzene to Phenol by FeO+. Organometallics. 2005;24: 3532–3538. doi:10.1021/om050136b

52. Bathelt CM, Ridder L, Mulholland AJ, Harvey JN. Aromatic Hydroxylation by Cytochrome P450: Model Calculations of Mechanism and Substituent Effects. J Am Chem Soc. 2003;125: 15004–15005. doi:10.1021/ja035590q

53. Bathelt CM, Ridder L, Mulholland AJ, Harvey JN. Mechanism and structure-reactivity relationships for aromatic hydroxylation by cytochrome P450. Org Biomol Chem. 2004;2: 2998–3005. doi:10.1039/B410729B

54. Olah J, Mulholland AJ, Harvey JN. Understanding the determinants of selectivity in drug metabolism through modeling of dextromethorphan oxidation by cytochrome P450. Proceedings of the National Academy of Sciences. 2011;108: 6050–6055. doi:10.1073/pnas.1010194108

55. Morgan S, Grootendorst P, Lexchin J, Cunningham C, Greyson D. The cost of drug development: a systematic review. Health Policy. 2011;100: 4–17. doi:10.1016/j.healthpol.2010.12.002

56. Yu W, MacKerell AD. Computer-Aided Drug Design Methods. Methods Mol Biol. 2017;1520: 85–106. doi:10.1007/978-1-4939-6634-9_5

57. Kore PP, Mutha MM, Antre RV, Oswal RJ, Kshirsagar SS. Computer-Aided Drug Design: An Innovative Tool for Modeling. 2012;2012. doi:10.4236/ojmc.2012.24017

58. Sliwoski G, Kothiwale S, Meiler J, Lowe EW. Computational Methods in Drug Discovery. Pharmacol Rev. 2014;66: 334–395. doi:10.1124/pr.112.007336

59. Dowden H, Munro J. Trends in clinical success rates and therapeutic focus. Nature Reviews Drug Discovery. 2019 [cited 15 May 2019]. doi:10.1038/d41573-019-00074-z

60. Kirchmair J, Göller AH, Lang D, Kunze J, Testa B, Wilson ID, et al. Predicting drug metabolism: experiment and/or computation? Nat Rev Drug Discov. 2015;14: 387–404. doi:10.1038/nrd4581

61. Kirchmair J, Williamson MJ, Tyzack JD, Tan L, Bond PJ, Bender A, et al. Computational Prediction of Metabolism: Sites, Products, SAR, P450 Enzyme Dynamics, and Mechanisms. J Chem Inf Model. 2012;52: 617–648. doi:10.1021/ci200542m

62. Raies AB, Bajic VB. In silico toxicology: computational methods for the prediction of chemical toxicity. WIREs Comput Mol Sci. 2016;6: 147–172. doi:10.1002/wcms.1240

63. Ehrlich P. Über den jetzigen Stand der Chemotherapie. Ber Dtsch Chem Ges. 1909;42: 17–47. doi:10.1002/cber.19090420105

64. Acharya C, Coop A, Polli JE, MacKerell AD. Recent Advances in Ligand-Based Drug Design: Relevance and Utility of the Conformationally Sampled Pharmacophore Approach. Curr Comput Aided Drug Des. 2011;7: 10–22.

65. Gay SC, Roberts AG, Halpert JR. Structural Features of Cytochromes P450 and Ligands that Affect Drug Metabolism as Revealed by X-ray Crystallography and NMR. Future Med Chem. 2010;2: 1451–1468.

66. Rowland P, Blaney FE, Smyth MG, Jones JJ, Leydon VR, Oxbrow AK, et al. Crystal structure of human cytochrome P450 2D6. J Biol Chem. 2006;281: 7614–7622. doi:10.1074/jbc.M511232200

67. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov. 2004;3: 935–949. doi:10.1038/nrd1549

68. Koymans L, Kelder GMD-OD, Koppele JM te, Vermeulen NPE. Cytochromes P450: Their Active-Site Structure and Mechanism of Oxidation. Drug Metabolism Reviews. 1993;25: 325–387. doi:10.3109/03602539308993979

69. Kemppainen LM, Kemppainen TT, Reippainen JA, Salmenniemi ST, Vuolanto PH. Use of complementary and alternative medicine in Europe: Health-related and sociodemographic determinants. Scand J Public Health. 2018;46: 448–455. doi:10.1177/1403494817733869

70. BORDOGNA A, PANDINI A, BONATI L. Predicting the Accuracy of Protein–Ligand Docking on Homology Models. J Comput Chem. 2011;32: 81–98. doi:10.1002/jcc.21601

71. Rodrigues JPGLM, Melquiond ASJ, Karaca E, Trellet M, van Dijk M, van Zundert GCP, et al. Defining the limits of homology modeling in information-driven protein docking. Proteins. 2013;81: 2119–2128. doi:10.1002/prot.24382

72. Vyas VK, Ukawala RD, Ghate M, Chintha C. Homology Modeling a Fast Tool for Drug Discovery: Current Perspectives. Indian J Pharm Sci. 2012;74: 1–17. doi:10.4103/0250-474X.102537

73. de Graaf C, Oostenbrink C, Keizers PHJ, van der Wijst T, Jongejan A, Vermeulen NPE. Catalytic Site Prediction and Virtual Screening of Cytochrome P450 2D6 Substrates by Consideration of Water and Rescoring in Automated Docking. J Med Chem. 2006;49: 2417–2430. doi:10.1021/jm0508538

74. Ito Y, Kondo H, Goldfarb PS, Lewis DFV. Analysis of CYP2D6 substrate interactions by computational methods. JOURNAL OF MOLECULAR GRAPHICS & MODELLING. 2008;26: 947–956.

75. Kjellander B, Masimirembwa CM, Zamora I. Exploration of Enzyme−Ligand Interactions in CYP2D6 & 3A4 Homology Models and Crystal Structures Using a Novel Computational Approach. J Chem Inf Model. 2007;47: 1234–1247. doi:10.1021/ci600561v

76. Maréchal J-D, Kemp CA, Roberts GCK, Paine MJI, Wolf CR, Sutcliffe MJ. Insights into drug metabolism by cytochromes P450 from modelling studies of CYP2D6-drug interactions. Br J Pharmacol. 2008;153: S82–S89. doi:10.1038/sj.bjp.0707570

77. van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastritis PL, Karaca E, et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol. 2016;428: 720–725. doi:10.1016/j.jmb.2015.09.014

78. Martiny VY, Carbonell P, Chevillard F, Moroy G, Nicot AB, Vayer P, et al. Integrated structure- and ligand-based in silico approach to predict inhibition of cytochrome P450 2D6. Bioinformatics. 2015;31: 3930–3937. doi:10.1093/bioinformatics/btv486

79. Rossato G, Ernst B, Smiesko M, Spreafico M, Vedani A. Probing Small-Molecule Binding to Cytochrome P450 2D6 and 2C9: An In Silico Protocol for Generating Toxicity Alerts. ChemMedChem. 2010;5: 2088–2101. doi:10.1002/cmdc.201000358

80. Chaudhury S, Gray JJ. Conformer selection and induced fit in flexible backbone protein-protein docking using computational and NMR ensembles. J Mol Biol. 2008;381: 1068–1087. doi:10.1016/j.jmb.2008.05.042

81. Santos R, Hritz J, Oostenbrink C. Role of water in molecular docking simulations of cytochrome P450 2D6. J Chem Inf Model. 2010;50: 146–154. doi:10.1021/ci900293e

82. Wong SE, Lightstone FC. Accounting for water molecules in drug design. Expert Opinion on Drug Discovery. 2011;6: 65–74. doi:10.1517/17460441.2011.534452

83. Lonsdale R, Mulholland AJ. QM/MM modelling of drug-metabolizing enzymes. Current Topics in Medicinal Chemistry. 2014;14: 1339–1347. doi:10.2174/1568026614666140506114859

84. Kirton SB, Murray CW, Verdonk ML, Taylor RD. Prediction of binding modes for ligands in the cytochromes P450 and other heme-containing proteins. Proteins. 2005;58: 836–844. doi:10.1002/prot.20389

85. Oda A, Yamaotsu N, Hirono S. New AMBER force field parameters of heme iron for cytochrome P450s determined by quantum chemical calculations of simplified models. J Comput Chem. 2005;26: 818–826. doi:10.1002/jcc.20221

86. de Graaf C, Vermeulen NPE, Feenstra KA. Cytochrome P450 in Silico: An Integrative Modeling Approach. J Med Chem. 2005;48: 2725–2755. doi:10.1021/jm040180d

87. Olsen L, Oostenbrink C, Jørgensen FS. Prediction of cytochrome P450 mediated metabolism. Adv Drug Deliv Rev. 2015;86: 61–71. doi:10.1016/j.addr.2015.04.020

88. Cruciani G, Carosati E, De Boeck B, Ethirajulu K, Mackie C, Howe T, et al. MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. J Med Chem. 2005;48: 6970–6979. doi:10.1021/jm050529c

89. Li J, Schneebeli ST, Bylund J, Farid R, Friesner RA. IDSite: An accurate approach to predict P450-mediated drug metabolism. J Chem Theory Comput. 2011;7: 3829–3845. doi:10.1021/ct200462q

90. Huang T, Zaretzki J, Bergeron C, Bennett KP, Breneman CM. DR-Predictor: Incorporating Flexible Docking with Specialized Electronic Reactivity and Machine Learning Techniques to Predict CYP-Mediated Sites of Metabolism. J Chem Inf Model. 2013;53: 3352–3366. doi:10.1021/ci4004688

91. van Dijk M, ter Laak AM, Wichard JD, Capoferri L, Vermeulen NPE, Geerke DP. Comprehensive and Automated Linear Interaction Energy Based Binding-Affinity Prediction for Multifarious Cytochrome P450 Aromatase Inhibitors. J Chem Inf Model. 2017;57: 2294–2308. doi:10.1021/acs.jcim.7b00222

92. Vosmeer CR, Kooi DP, Capoferri L, Terpstra MM, Vermeulen NPE, Geerke DP. Improving the iterative Linear Interaction Energy approach using automated recognition of configurational transitions. J Mol Model. 2016;22: 31. doi:10.1007/s00894-015-2883-y

93. Perić-Hassler L, Stjernschantz E, Oostenbrink C, Geerke DP. CYP 2D6 Binding Affinity Predictions Using Multiple Ligand and Protein Conformations. Int J Mol Sci. 2013;14: 24514–24530. doi:10.3390/ijms141224514

94. Barnaba C, Gentry K, Sumangala N, Ramamoorthy A. The catalytic function of cytochrome P450 is entwined with its membrane-bound nature. F1000Res. 2017;6. doi:10.12688/f1000research.11015.1

95. Berka K, Paloncýová M, Anzenbacher P, Otyepka M. Behavior of Human Cytochromes P450 on Lipid Membranes. J Phys Chem B. 2013;117: 11556–11564. doi:10.1021/jp4059559

96. Jeřábek P, Florián J, Martínek V. Membrane-Anchored Cytochrome P450 1A2–Cytochrome b5 Complex Features an X-Shaped Contact between Antiparallel Transmembrane Helices. Chem Res Toxicol. 2016;29: 626–636. doi:10.1021/acs.chemrestox.5b00349

97. Navrátilová V, Paloncýová M, Berka K, Otyepka M. Effect of Lipid Charge on Membrane Immersion of Cytochrome P450 3A4. J Phys Chem B. 2016;120: 11205–11213. doi:10.1021/acs.jpcb.6b10108

98. Sasahara K, Mashima A, Yoshida T, Chuman H. Molecular dynamics and density functional studies on the metabolic selectivity of antipsychotic thioridazine by cytochrome P450 2D6: Connection with crystallographic and metabolic results. Bioorg Med Chem. 2015;23: 5459–5465. doi:10.1016/j.bmc.2015.07.043

99. Don CG, Smieško M. Microsecond MD simulations of human CYP2D6 wild-type and five allelic variants reveal mechanistic insights on the function. PLoS ONE. 2018;13: e0202534. doi:10.1371/journal.pone.0202534

100. Fischer A, Don CG, Smieško M. Molecular Dynamics Simulations Reveal Structural Differences among Allelic Variants of Membrane-Anchored Cytochrome P450 2D6. J Chem Inf Model. 2018;58: 1962–1975. doi:10.1021/acs.jcim.8b00080

101. van Gunsteren WF, Daura X, Hansen N, Mark AE, Oostenbrink C, Riniker S, et al. Validation of Molecular Simulation: An Overview of Issues. Angew Chem Int Ed Engl. 2018;57: 884–902. doi:10.1002/anie.201702945

102. Shaw DE, Dror RO, Salmon JK, Grossman JP, Mackenzie KM, Bank JA, et al. Millisecond-scale Molecular Dynamics Simulations on Anton. Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis. New York, NY, USA: ACM; 2009. pp. 39:1–39:11. doi:10.1145/1654059.1654099

103. Rydberg P, Lonsdale R, Harvey JN, Mulholland AJ, Olsen L. Trends in predicted chemoselectivity of cytochrome P450 oxidation: B3LYP barrier heights for epoxidation and hydroxylation reactions. J Mol Graph Model. 2014;52: 30–35. doi:10.1016/j.jmgm.2014.06.002

104. Lonsdale R, Oláh J, Mulholland AJ, Harvey JN. Does Compound I Vary Significantly between Isoforms of Cytochrome P450? J Am Chem Soc. 2011;133: 15464–15474. doi:10.1021/ja203157u

105. Shaik S, Cohen S, Wang Y, Chen H, Kumar D, Thiel W. P450 Enzymes: Their Structure, Reactivity, and Selectivity—Modeled by QM/MM Calculations. Chem Rev. 2010;110: 949–1017. doi:10.1021/cr900121s

106. Schöneboom JC, Lin H, Reuter N, Thiel W, Cohen S, Ogliaro F, et al. The elusive oxidant species of cytochrome P450 enzymes: characterization by combined quantum mechanical/molecular mechanical (QM/MM) calculations. J Am Chem Soc. 2002;124: 8142–8151. doi:10.1021/ja026279w

107. Postils V, Saint-André M, Timmins A, Li X-X, Wang Y, Luis JM, et al. Quantum Mechanics/Molecular Mechanics Studies on the Relative Reactivities of Compound I and II in Cytochrome P450 Enzymes. Int J Mol Sci. 2018;19. doi:10.3390/ijms19071974

108. van der Kamp MW, Mulholland AJ. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. Biochemistry. 2013;52: 2708–2728. doi:10.1021/bi400215w

109. BANKS JL, BEARD HS, CAO Y, CHO AE, DAMM W, FARID R, et al. Integrated Modeling Program, Applied Chemical Theory (IMPACT). J Comput Chem. 2005;26: 1752–1780. doi:10.1002/jcc.20292

110. Cohen S, Kozuch S, Hazan C, Shaik S. Does substrate oxidation determine the regioselectivity of cyclohexene and propene oxidation by cytochrome p450? J Am Chem Soc. 2006;128: 11028–11029. doi:10.1021/ja063269c

111. Lai W, Shaik S. Can Ferric-Superoxide Act as a Potential Oxidant in P450cam? QM/MM Investigation of Hydroxylation, Epoxidation, and Sulfoxidation. J Am Chem Soc. 2011;133: 5444–5452. doi:10.1021/ja111376n

112. Hou QQ, Sheng X, Wang JH, Liu YJ, Liu CB. QM/MM study of the mechanism of enzymatic limonene 1,2-epoxide hydrolysis. Biochim Biophys Acta. 2012;1824: 263–268. doi:10.1016/j.bbapap.2011.08.014

113. Dubey KD, Wang B, Vajpai M, Shaik S. MD simulations and QM/MM calculations show that single-site mutations of cytochrome P450BM3 alter the active site's complexity and the chemoselectivity of oxidation without changing the active species. Chem Sci. 2017;8: 5335–5344. doi:10.1039/C7SC01932G

114. Sharma PK, De Visser SP, Shaik S. Can a single oxidant with two spin states masquerade as two different oxidants? A study of the sulfoxidation mechanism by cytochrome p450. J Am Chem Soc. 2003;125: 8698–8699. doi:10.1021/ja035135u

115. Gunsteren WF van, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, et al. Biomolecular Modeling: Goals, Problems, Perspectives. Angewandte Chemie International Edition. 2006;45: 4064–4092. doi:10.1002/anie.200502655

116. Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A. Coarse-Grained Protein Models and Their Applications. Chem Rev. 2016;116: 7898–7936. doi:10.1021/acs.chemrev.6b00163

117. Scott WRP, Hünenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, et al. The GROMOS Biomolecular Simulation Program Package. The Journal of Physical Chemistry A. 1999;103: 3596–3607. doi:10.1021/jp984217f

118. Foloppe N, Jr ADM. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. Journal of Computational Chemistry. 2000;21: 86–104. doi:10.1002/(SICI)1096-987X(20000130)21:2<86::AID-JCC2>3.0.CO;2-G

119. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, et al. The Amber biomolecular simulation programs. J Comput Chem. 2005;26: 1668–1688. doi:10.1002/jcc.20290

120. Groenhof G. Introduction to QM/MM Simulations. In: Monticelli L, Salonen E, editors. Biomolecular Simulations. Totowa, NJ: Humana Press; 2013. pp. 43–66. doi:10.1007/978-1-62703-017-5_3

121. Warshel A, Levitt M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J Mol Biol. 1976;103: 227–249. doi:10.1016/0022-2836(76)90311-9

122. Field MJ, Bash PA, Karplus M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. Journal of Computational Chemistry. 1990;11: 700–733. doi:10.1002/jcc.540110605

123. Tian B, Strid Å, Eriksson LA. Catalytic roles of active-site residues in 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase: an ONIOM/DFT study. J Phys Chem B. 2011;115: 1918–1926. doi:10.1021/jp111525p

124. Sumner S, Söderhjelm P, Ryde U. Effect of Geometry Optimizations on QM-Cluster and QM/MM Studies of Reaction Energies in Proteins. J Chem Theory Comput. 2013;9: 4205–4214. doi:10.1021/ct400339c

125. Perdew null, Burke null, Ernzerhof null. Generalized Gradient Approximation Made Simple. Phys Rev Lett. 1996;77: 3865–3868. doi:10.1103/PhysRevLett.77.3865

126. Tao J, Perdew JP, Staroverov VN, Scuseria GE. Climbing the density functional ladder: nonempirical meta-generalized gradient approximation designed for molecules and solids. Phys Rev Lett. 2003;91: 146401. doi:10.1103/PhysRevLett.91.146401

127. Becke AD. Density-functional exchange-energy approximation with correct asymptotic behavior. Phys Rev A. 1988;38: 3098–3100. doi:10.1103/PhysRevA.38.3098

128. Becke AD. A new mixing of Hartree–Fock and local density-functional theories. J Chem Phys. 1993;98: 1372–1377. doi:10.1063/1.464304

129. Lee C, Yang W, Parr RG. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys Rev B. 1988;37: 785–789. doi:10.1103/PhysRevB.37.785

130. Zawada A, Kaczmarek-Kędziera A, Bartkowiak W. On the potential application of DFT methods in predicting the interaction-induced electric properties of molecular complexes. Molecular H-bonded chains as a case of study. J Mol Model. 2012;18: 3073–3086. doi:10.1007/s00894-011-1312-0

131. Applications of quantum mechanical molecular models to chemical problems. Part 70. Quantum mechanical molecular models | The Journal of Physical Chemistry. [cited 2 Oct 2019]. Available: https://pubs.acs.org/doi/10.1021/j100257a004

132. Optimization of parameters for semiempirical methods I. Method - Stewart - 1989 - Journal of Computational Chemistry - Wiley Online Library. [cited 2 Oct 2019]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.540100208

133. Thiel W. Semiempirical quantum–chemical methods. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2014;4: 145–157. doi:10.1002/wcms.1161

134. Grimme S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. J Comput Chem. 2006;27: 1787–1799. doi:10.1002/jcc.20495

135. Grimme S, Antony J, Ehrlich S, Krieg H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J Chem Phys. 2010;132: 154104. doi:10.1063/1.3382344

136. Burns LA, Vázquez-Mayagoitia Á, Sumpter BG, Sherrill CD. Density-functional approaches to noncovalent interactions: a comparison of dispersion corrections (DFT-D), exchange-hole dipole moment (XDM) theory, and specialized functionals. The Journal of chemical physics. 2011;134: 084107. doi:10.1063/1.3545971

137. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †. The Journal of Physical Chemistry B. 1998;102: 3586–3616. doi:10.1021/jp973084f

138. Mackerell AD. Empirical force fields for biological macromolecules: overview and issues. J Comput Chem. 2004;25: 1584–1604. doi:10.1002/jcc.20082

139. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. J Phys Chem B. 2001;105: 6474–6487. doi:10.1021/jp003919d

140. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. - PubMed - NCBI. [cited 2 Oct 2019]. Available: https://www.ncbi.nlm.nih.gov/pubmed/26584231

141. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins. 2006;65: 712–725. doi:10.1002/prot.21123

142. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase - Schuler - 2001 - Journal of Computational Chemistry - Wiley Online Library. [cited 21 Aug 2019]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/jcc.1078

143. Schmid N, Eichenberger AP, Choutko A, Riniker S, Winger M, Mark AE, et al. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. Eur Biophys J. 2011;40: 843–856. doi:10.1007/s00249-011-0700-9

144. Cao L, Ryde U. On the Difference Between Additive and Subtractive QM/MM Calculations. Front Chem. 2018;6. doi:10.3389/fchem.2018.00089

145. Gao J, Amara P, Alhambra C, Field MJ. A Generalized Hybrid Orbital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations. J Phys Chem A. 1998;102: 4714–4721. doi:10.1021/jp9809890

146. Pu J, Gao J, Truhlar DG. Generalized Hybrid-Orbital Method for Combining Density Functional Theory with Molecular Mechanicals. Chemphyschem. 2005;6: 1853–1865. doi:10.1002/cphc.200400602

147. Philipp DM, Friesner RA. Mixed ab initio QM/MM modeling using frozen orbitals and tests with alanine dipeptide and tetrapeptide. Journal of Computational Chemistry. 1999;20: 1468–1494. doi:10.1002/(SICI)1096-987X(19991115)20:14<1468::AID-JCC2>3.0.CO;2-0

148. Ferré N, Assfeld X, Rivail J-L. Specific force field parameters determination for the hybrid ab initio QM/MM LSCF method. J Comput Chem. 2002;23: 610–624. doi:10.1002/jcc.10058

149. Lin H, Truhlar DG. QM/MM: what have we learned, where are we, and where do we go from here? Theor Chem Acc. 2006;117: 185. doi:10.1007/s00214-006-0143-z

150. Senn HM, Thiel W. QM/MM methods for biomolecular systems. Angew Chem Int Ed Engl. 2009;48: 1198–1229. doi:10.1002/anie.200802019

151. Sherwood P, Vries AH de, Collins SJ, Greatbanks SP, Burton NA, Vincent MA, et al. Computer simulation of zeolite structure and reactivity using embedded cluster methods. Faraday Discuss. 1997;106: 79–92. doi:10.1039/A701790A

152. König PH, Hoffmann M, Frauenheim Th, Cui Q. A Critical Evaluation of Different QM/MM Frontier Treatments with SCC-DFTB as the QM Method. J Phys Chem B. 2005;109: 9082–9095. doi:10.1021/jp0442347

153. Roßbach S, Ochsenfeld C. Influence of Coupling and Embedding Schemes on QM Size Convergence in QM/MM Approaches for the Example of a Proton Transfer in DNA. J Chem Theory Comput. 2017;13: 1102–1107. doi:10.1021/acs.jctc.6b00727

154. Murphy RB, Philipp DM, Friesner RA. A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. Journal of Computational Chemistry. 2000;21: 1442–1457. doi:10.1002/1096-987X(200012)21:16<1442::AID-JCC3>3.0.CO;2-O

155. Svensson M, Humbel S, Froese RDJ, Matsubara T, Sieber S, Morokuma K. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels−Alder Reactions and Pt(P(t-Bu)3)2 + H2 Oxidative Addition. J Phys Chem. 1996;100: 19357–19363. doi:10.1021/jp962071j

156. Riahi S, Rowley CN. The CHARMM-TURBOMOLE interface for efficient and accurate QM/MM molecular dynamics, free energies, and excited state properties. J Comput Chem. 2014;35: 2076–2086. doi:10.1002/jcc.23716

157. Sherwood P, de Vries AH, Guest MF, Schreckenbach G, Catlow CRA, French SA, et al. QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. Journal of Molecular Structure: THEOCHEM. 2003;632: 1–28. doi:10.1016/S0166-1280(03)00285-9

158. Ryde U, Olsson MHM. Structure, strain, and reorganization energy of blue copper models in the protein. International Journal of Quantum Chemistry. 2001;81: 335–347. doi:10.1002/1097-461X(2001)81:5<335::AID-QUA1003>3.0.CO;2-Q

Chapter 2

Deciphering reaction determinants of altered-activity CYP2D6 variants by well-tempered metadynamics simulation and QM/MM calculations

Manuscript submitted.

Abstract

The xenobiotic metabolizing enzyme CYP2D6 is the P450 cytochrome family member with the highest rate of polymorphism. This causes changes in the enzyme activity and specificity, which can ultimately lead to adverse reactions during drug treatment. To avoid or lower CYP-related toxicity risks, prediction of the most likely positions within a molecule where a metabolic reaction might occur is paramount. In order to obtain accurate predictions, it is crucial to understand all phenomena within the active site of the enzyme that contribute to an efficient substrate recognition and the subsequent catalytic reaction together with their relative weight within the overall thermodynamic context. This study aims to define the weight of the driving forces upon the C-H bond activation within CYP2D6 wild-type and a clinically relevant allelic variant with increased activity (*CYP2D6*53*) featuring two amino acid mutations in close vicinity of the heme. First, we investigated the steric and electrostatic complementarity of the substrate bufuralol using well-tempered metadynamics simulations with the aim to obtain the free energy profiles for each site of metabolism within the different active sites. Secondly, the stereoelectronic complementarity was determined for each SoM within the two different active site environments. Relying on the well-tempered metadynamics simulation energy profiles of each SoM, we identified the binding mode that was closest to the preferred transition state geometry for efficient C-H bond activation. The binding modes were then used as starting structures for the quantum mechanics/molecular mechanics calculations performed to quantify the corresponding activation barriers. Our results show the relevance of the steric component in orienting the SoM in an energetically accessible position towards the heme. However, the corresponding intrinsic reactivity and electronic complementarity within the active site must be accurately evaluated in order to obtain a meaningful reaction prediction from which the predominant sites of metabolism can be determined. The F120I mutation lowered the activation barrier for the major site and one of the minor sites of metabolism. However, it had neither impact on the CYP2D6 enantioselectivity preference of the oxidation reaction nor on the stereoselectivity from the substrate point of view.

Introduction

Cytochrome (CYP) P450 monooxygenases are thiolate-ligated heme enzymes that have a crucial function in the detoxification of xenobiotics in our body [1–3]. Catalytic reactions, including aliphatic and aromatic hydroxylation, epoxidation, and heteroatom oxidation, introduce polar functional groups to the substrate molecule increasing the solubility of potentially harmful compounds [4–6]. The biotransformation of such compounds is essential to facilitate their secretion in kidneys and to avoid bioaccumulation and toxicity. Moreover, numerous marketed drugs rely on this oxidation mechanism to achieve the bioactivation [5,7]. However, during drug therapy, the fact that drugs can act as substrates, inhibitors or both towards CYPs requires careful monitoring of the substance fate within the organism as otherwise e.g. unexpected drug-drug interactions may lead to manifestation of toxicity, possibly even with a fatal outcome [8]. Hence, from the early drug development phase on, potential CYP-related toxicity issues need to be addressed.

Computational techniques which are able to provide information regarding the predominant binding modes, binding affinities and therefrom site(s) of metabolism (SoMs) at CYPs can be of high value to improve the therapeutic and safety profile of lead compound and avoid late-stage undesired CYP-mediated adverse reactions [9]. The better we understand the forces that contribute to driving the catalytic reaction and their weight in the overall balance, the more accurate and reliable we can develop CYP-related predictive tools. However, the complete framework in which CYPs operate is complex. To start, the lead compound (ligand) can be metabolized by one or several P450 subtypes (isoforms) [10]. Depending on the selectivity of the isoform, a preference for the oxidation at a particular functional group (chemoselectivity), a stereoisomer preference (stereoselectivity) or particular part of the compound (regioselectivity) will apply. State-of-the-art computational metabolism prediction tools commonly used include FAME [11,12], GLORY [13], SOMP [14] or CYPreact [15]. A comprehensive overview on computational CYPP450 metabolism prediction tools can be found in the recent review of Tyzack et al. [16]. Moreover, it is postulated that the substrate oxidation occurs when CYP is in the so called compound I (cpd I) intermediate state [4,17]. Reaction of the iron(IV)oxo species with the functional group of the ligand requires a

particular geometry to reach an optimal orbital overlap for an efficient biotransformation [18,19]. For example, for aliphatic or aromatic carbon hydrogen abstraction (also referred to as C-H bond activation) the ideal Fe=O·····H-C distance is around 2 Å and the Fe=O·····H angle around 120° [18,20]. Docking methods commonly predict the binding modes and the corresponding binding affinity based on the interaction and conformational energies together with a distance criterion between the SoM and the heme-iron [21]. Challenges such as the large flexibility of the binding pocket and multiple binding modes of the substrate increase the complexity of this task. Molecular dynamics (MD) simulations can partially aid to validate the stability of the docking poses, introduce flexibility and confirm which of the substrate's sites can approach the heme-bound oxygen in a favorable orientation enabling its biotransformation [22–26]. If the lowest energy pose of the substrate also corresponds to the conformation, in which the SoM is positioned such that it is energetically most accessible for its biotransformation, remains an open question. Nevertheless, even if binding mode(s) and corresponding one or more SoMs are identified using a docking and/or MD simulations approach, the other essential question that needs to be answered is reactivity, which is defined as the sum of the substrate's intrinsic reactivity and the protein environment (enzyme cavity) effects. In order to determine this, a quantum mechanics/molecular mechanics (QM/MM) modeling approach is frequently used [27]. Previous QM/MM studies have provided valuable insights on CYP450 selectivity by investigating activation energies for e.g. CYP2D6 dopamine formation [28] or dextromethorphan CYP2D6 metabolism [20]. Both studies confirmed the importance of including the protein environment for successful metabolite predictions as it stabilizes the ligand-heme complex [29]. Hence, for an accurate isoform specific metabolism prediction the contribution of both the binding affinity and the reactivity needs to be properly assessed. In this study we have investigated (i) *the interplay between the enzyme binding pocket and ligand pose (shape complementarity)* by performing well-tempered metadynamics (WT-MTD) simulations to obtain the free energy surfaces (FES) for each of the SoMs of a well characterized classical CYP2D6 substrate (bufuralol) in CYP2D6 wild-type (WT) (**Figure 2.1**). In addition, an increased activity (*CYP2D6*53*, F120I/A122S) [30] and diminished-to-non-functional (E216Q/F483G) [31]

CYP2D6 variant have been included to examine the impact of amino acid mutations on the binding mode(s) and subsequent SoM positioning with regard to the oxyferryl species; the altered activity variants have both two mutations within close vicinity of the heme which will have an impact on the steric and electronic component of the catalytic reaction. However, from the two variants, only *CYP2D6*53* is clinically relevant. Until today, no clinically relevant variant with diminished-to-no activity has been identified that harbours mutations only within the active site. Therefore, the non-functional CYP2D6 variant was selected from a combinatorial mutation study aimed at identifying improved activity variants [31]. Determining the FES profiles of bufuralol SoMs within the three different active site variants will provide new insight on the reliability of using solely the lowest ranked energy pose(s) from docking for the prediction of the most likely SoMs and its capability in distinguishing major SoM(s) from minor ones. In a second step, we wanted to explore (ii) *the electronic complementarity of bufuralol and the binding pocket* by determining to what extent the point mutations influence the protein-ligand interaction. For each SoM in CYP2D6 WT and *CYP2D6*53* we deduced from its earlier determined WT-MTD energy profile the binding mode that was closest to the preferred reaction state geometry and calculated the corresponding activation barrier by performing QM/MM calculations on reactant, products and their transition states. To our knowledge, this is the first *in silico* study performed probing the driving forces within both CYP2D6 WT and the clinical allelic variant *CYP2D6*53* upon C-H bond activation. The results of the combined WT-MTD and QM/MM approach will contribute to improved understanding of CYP2D6 selectivity and guide the further development of the state-of-the-art techniques used for metabolism predictions.

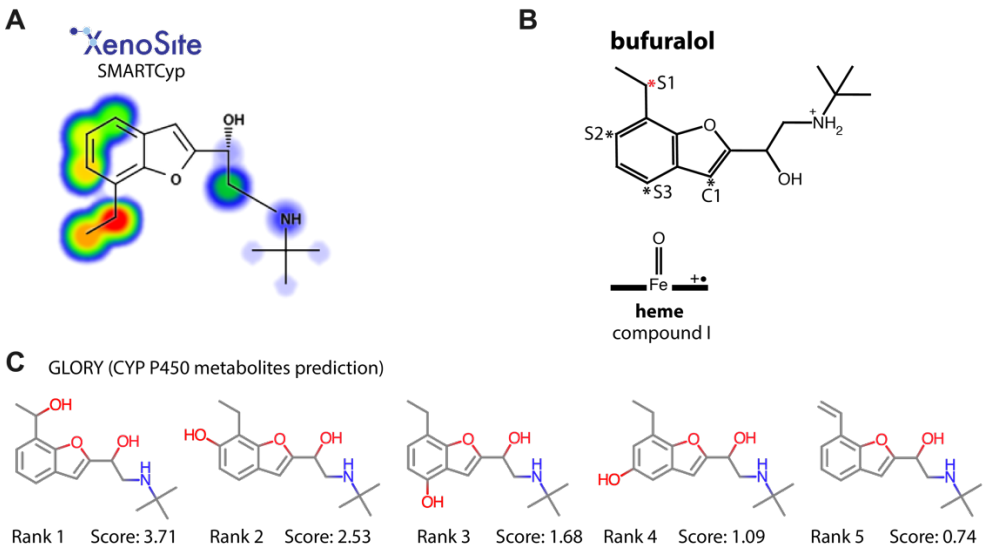


Figure 2.1 **Overview of the CYP2D6 substrate bufuralol and its experimentally confirmed sites of metabolism (SoMs S1-S3)** [10]. **A**) CYP2D6 regio-selectivity predictions (SoMs) by SMARTCyp [32,33]. The red/yellow areas correspond well with the experimentally determined SoMs. Some deviation is found for the yellow-green areas predicted by SMARTCyp. **B**) The SoMs (S1-S3) and non-reactive control site (C1) investigated in this study. **C**) The CYP P450 metabolites of bufuralol predicted by GLORY [13].

Results & Discussion

Free energy potentials of the wild-type SoMs and altered activity variants

In order to explore the free energy profile of each SoM within close distance of the oxyferryl group, WT-MTD was used as an enhanced sampling technique. Two different sets of collective variables (CVs) were chosen initially to determine the CV set representing most accurately the pathway of the oxidation reaction studied. The first CV set (referred to as CV-set 1) used as CVs the distance between the SoM carbon and the oxyferryl (its O atom) (CV1) and the angle between the Fe=O and the hydrogen for the aliphatic carbon positioned SoM (Fe=O·····H) or the carbon in case of an aromatic positioned SoM (Fe=O·····C) (CV2) (**Figure 2.2**).

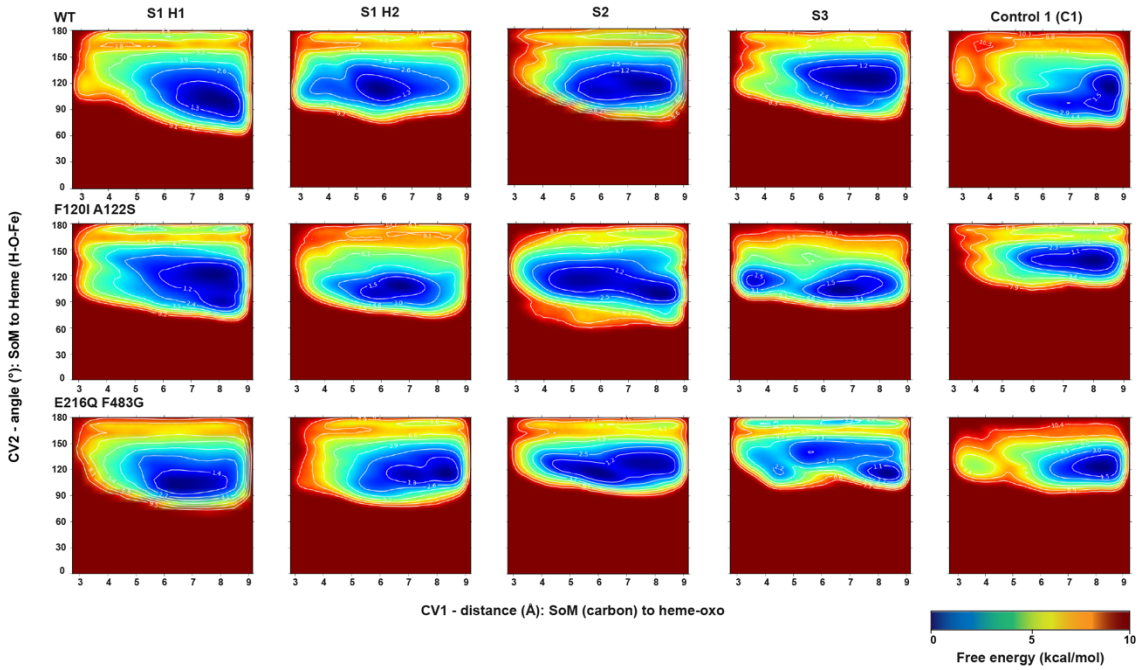


Figure 2.2. **Free energy surface (FES) graphs displayed for the sites of metabolism (SoMs, S1-S3) of bufuralol as well as an non-reactive control site (C1) using CV-set 1 calculated over 200 ns WT-MTD simulation.**

The second CV set tested (referred to as CV-set 2) included the same CV1 as in the CV-set 1 but as CV2 the angle between the oxyferryl and the hydrogen and carbon of the SoM (O·····H-C) (CV2) (**Figure S2.1**). The motivation for selecting these CVs is that the positioning of the substrate's hydrogen towards the oxyferryl is critical for hydrogen abstraction reaction efficient proceeding. As explained before, an optimal orbital overlap between the substrate's hydrogen and the oxyferryl (Fe=O·····H) is achieved with a distance of around 2 Å, an angle close to 120° and a linear angle between the oxyferryl and the hydrogen-carbon bond of the substrate (O·····H-C) of 180° [18,20,34]. Monitoring the distance and one of the two angles will show how energetic favorable it is for the substrate to adopt the geometry required to initiate the oxidation reaction in which the hydrogen atom is abstracted. For selecting the CV-set that enables to distinguish well between the two different states (reactant and transition state geometry during the C-H bond activation reaction) the system needs to diffuse in the entire phase space to be able to reach convergence. Most efficient sampling of space for each CV over time was found for CV-set 1 (**Figures S2.2-S2.3**) and the system showed to diffuse efficiently in the CV space (not being stuck in a local minimum (**Figure S2.4**). Additional convergence evaluations performed for CV-set 1 ensuring efficient sampling can be found at the Materials and Methods section (**Figures S2.5-S2.11**).

For all CYP variants studied, an unfavourable FES area (red) for each SoM is observed bellow an Fe-O·····H angle of 60°, which is as expected since steric hindrance prevents it from reaching a lower value (**Figure 2.2**). The lowest energy area for the monitored angle (CV2) varies between 70° and 160° (**Figure 2.2**, dark blue area). The FES profiles of the CYP2D6 wild-type (WT) indicates that based on the distance between the SoM and the oxyferryl group, the major SoM can be distinguished from minor; site S1_H2 is most likely to get in close vicinity of the heme with a preferred Fe-O·····H angle of about 120°. Sites S2 and S3 reach CV1 values of approximately 5 Å which is similar to the control (C1) CV1 distance meaning that the minor SoMs cannot be assigned any ranking priority. The FES profiles of the increased activity variant (F120I A122S) show that there is no difference between major or minor site; S2 and S3 reach a SoM to oxyferryl group distance of approximately 3 Å (dark blue area, **Figure 2.2**) whereas S1_H1 and S1_H2 reach only a distance of approximately 4 Å. The closest SoM to oxyferryl group distance for the variant with no activity (E216Q F483G) was found for its S2 SoM (~ 4 Å) all other SoMs remain further away (~ 5 Å). The experimental setup for this variant only monitored and quantified the metabolism of the major metabolite (S1: 1'hydroxybufuralol), therefore production of S2 or S3 metabolites cannot be excluded [31]. Overall, the analysis of the FES profiles suggests the following: (i) the lowest energy poses (**Figure 2.2**, dark blue areas) are for all SoMs located further than 3 Å away from the heme oxygen. This may be interpreted as an indication that using only the lowest energy (top-ranked) poses from docking would give a non-conclusive SoM prediction. (ii) using solely the free energy profile of a substrate pose in the active site to classify the SoM as a major or a minor site can be possible as shown for the wild-type enzyme, however for the two altered activity variants it gave an ambiguous answer. To improve the reliability of such a classification prediction, including the intrinsic reactivity of the ligand and its electronic complementarity with the active site will enhance the accuracy of the SoM predictions as shown by other studies [18,29,34]. To access this second aspect, the frames closest to the favored transition state geometry for an efficient oxidation reaction (deduced from each SoM FES profile) were selected from each SoM WT-MTD simulation for CYP2D6 WT and the

clinically relevant variant *CYP2D6*53*. Subsequently, the frames were used as input (reactant structure) for the QM/MM calculations to determine the activation barrier of each.

QM/MM calculations

Typically CYP2D6 substrate recognition in the active site occurs through a salt-bridge interaction between the substrate's protonated nitrogen atom and the carboxyl group of the active site residue(s) D301 and/or E216. This allows orienting the substrate SoM(s) for a more or less favorable orbital overlap with the reactive oxyferryl group which thus lowers or increases the activation barrier and subsequently the efficiency of the hydrogen abstraction [10,35,36].

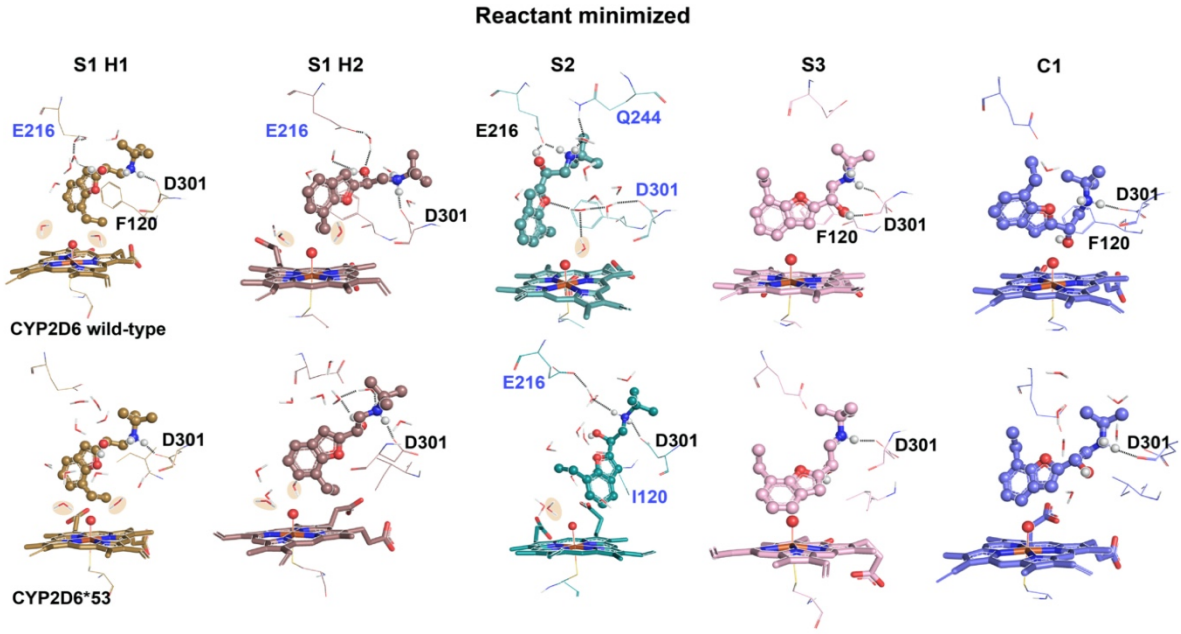


Figure 2.3. **QM/MM minimized reactant structures of all SoMs within CYP2D6-WT (top) and *CYP2D6*53* (bottom).** Important CYP2D6 active site residues interacting with bufuralol are displayed; black labelled residues indicate direct contact, whereas blue labelled residues indicate water-mediated contact. In addition, water molecules within 3 Å of the heme-iron are highlighted in orange.

This recognition mechanism is observed in all the starting structures; a direct salt bridge between the protonated aliphatic nitrogen atom of bufuralol and the carboxyl group of the active site residue D301 could be identified in all structures, except for the S2 SoM CYP2D6 WT in which the D301 interaction was bridged by two water molecules (**Figure 2.3**). For the anchoring residue E216, a direct salt bridge between bufuralol's basic nitrogen and the E216

carboxyl group was found in S2 CYP2D6 WT. Water-mediated interactions between bufuralol's hydroxyl group and the carboxyl group of active site residue E216 were identified in S1_H1, S1_H2 CYP2D6 WT, and S2 *CYP2D6*53*. It has been observed before that the substrate's benzene ring can be positioned such that its plane approaches a perpendicular angle with the plane of the heme group, called "side-on" or such that both planes are parallel, called "face-on" [18,20,37]. All the reactant structures are "side-on" positioned. Previous QM/MM studies for benzene oxidation within CYP2C9 found similar barrier heights for both orientations of the benzene ring (18.1-21.7 kcal/mol) [37]. For SoMs S1_H1, S1_H2, and S2 in CYP2D6 WT and *CYP2D6*53* one or two water molecules were found to be within 3 Å of the oxyferryl group oxygen atom (**Figure 2.3**, highlighted orange). It has been earlier determined that water molecules can lower the activation barrier (on average by 3-4 kcal/mol) if they stabilize the transition state structure [38]. However, all the transition state structures were inspected and this was not the case. A more detailed overview of the type of interactions between bufuralol and the active site residues including the distances can be found in **Table S2.3**. Furthermore, the (dis-) similarity of the reactant SoM binding poses in CYP2D6 WT and *CYP2D6*53* was determined by calculating the ΔRMSD of bufuralol in place after superposition of the QM region (**Table 2.1**). The most similar binding pose was found for the major S1_H1 SoM, ΔRMSD of 1.8 Å. For the minor SoMs S2 and S3 more dissimilar binding poses were identified with ΔRMSD values of 6.4 Å and 4.0 Å respectively. In addition, the C1 SoM binding poses were quite dissimilar as well with a ΔRMSD of 4.7 Å. A focused view on the binding modes of the reactants displaying the QM region without the heme moiety can be found in **Figure 2.4**.

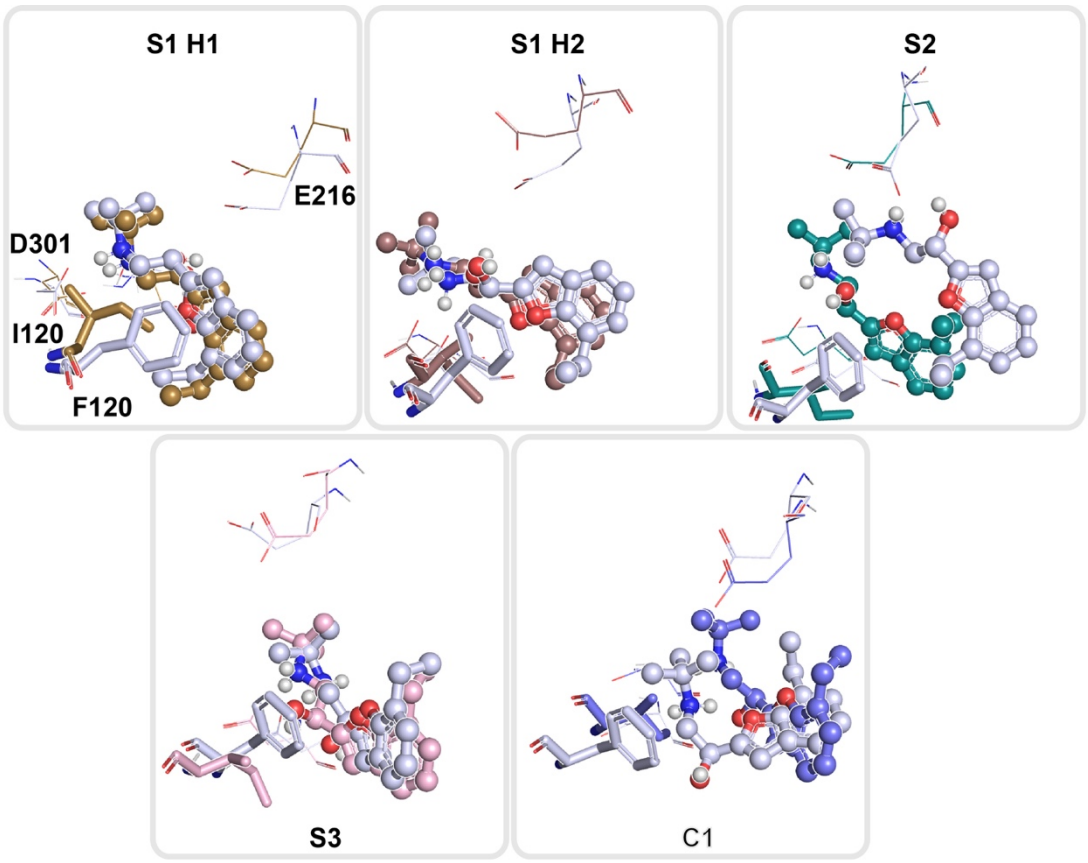


Figure 2.4. **Important Focused view on the interaction of F120 CYP2D6 WT (colored light-blue) or I120 *CYP2D6*53* (differently colored for each SoM) within the minimized reactant structure**. CYP2D6 WT, F120: π-π stacking with the aromatic group of bufuralol is observed for S1_H1 and S1_H2, π-cation interaction for S3 and C1, in S2 bufuralol is positioned more distant and unfavourable. *CYP2D6*53*: S1_H1 is oriented most optimal and closest to I120, S1_H2 and S2 are oriented less optimal and S3 and C1 are positioned furthest away from I120.

QM/MM energy barriers. An overview of the determined activation barriers can be found in **Table 2.1,** the most important geometric values (indicative of a good orbital overlap) of the input (reactant) and minimized reactant and TS structures can be found in **Table 2.2**. A detailed overview on the interaction types between bufuralol and the residues within the two different active sites environments determined in the starting structure (reactant) used for the QM/MM calculations can be found in **Table S2.1**.

Table 2.1. **QM/MM activation barriers (kcal/mol, M06-2X/LACV3P*) of the hydrogen abstraction of bufuralol's SoMS studied in CYP2D6 WT and *CYP2D6*53***. The activation barrier was calculated using the quartet spin state. The last column displays the RMSD difference between each SoM binding pose of CYP2D6 WT and *CYP2D6*53*.

	Activation Energy ΔE [kcal/mol]		ΔRMSD (Å)
SoM	CYP2D6 WT	CYP2D6*V53	bufuralol
S1_H1	**13.3**	**12.4**	**1.8**
S1_H2	**22.8**	**31.9**	**2.7**
S2	**22.9**	**21.2**	**6.4**
S3	**28.8**	**61.8**	**4.0**
C1	**57.2**	**42.4**	**4.7**

For both CYP2D6 WT and *CYP2D6*53* the aliphatic hydrogen abstraction led to the lowest activation barriers, 13.3 kcal/mol and 12.4 kcal/mol respectively, with a regioselectivity for S1_H1. The reaction enantioselectivity at S1 remained the same upon the F120I mutation (*R*-hydroxylation). This is in contrast with the F120A mutation, which reverses the enantioselectivity from *R* to S [39]. A possible explanation is that isoleucine is similar in size (volume) compared to phenylalanine, whereas alanine is much smaller thus offering space for different binding modes. The activation barriers of the aromatic (minor) SoMs displayed a trend consistent with experiment; S2 was found to be less reactive than the major SoM with an activation barrier of 22.9 kcal/mol in the CYP2D6 WT and 21.2 kcal/mol in *CYP2D6*53*, but more reactive than S3; the activation for the CYP2D6 WT was 5.9 kcal/mol higher and 40 kcal/mol higher for *CYP2D6*53*.

As a note, the transition state search for SoM S3 in *CYP2D6*53* has been performed twice, using two slightly different intermediate structures as an input for the search. However, the single point calculations performed on the resulting transition state structure led to the same activation barrier. The large difference of 33 kcal/mol between the S3 SoMs may be caused by the dissimilar orientation of the aromatic moiety of bufuralol; in CYP2D6 WT it is placed more perpendicular towards the heme plane whereas in *CYP2D6*53* it is more tilted (**Figure 2.3**). The less favorable conformation likely contributes to the considerably larger distance of 2.59 Å between the cysteine thiolate group and the heme iron atom (Fe-S) (**Table 2.2**).

Table 2.2. **Overview of the geometry parameters of the minimized reactant and transition state structures**. The selected angles and distances are indicative for reaching optimal orbital overlap and subsequently an efficient catalysis reaction. The minimized value is indicated first, in the brackets the values of the input structure (before the minimization) are given in bold.

structure		Fe=O·····H/C[A] (Å)	C-H (Å)	S-Fe (Å)	Fe=O (Å)	Fe=O·····H/C[A] (°)	O·····H-C (°)
wt S1_H1							
	R	2.66 **(2.41)**	1.10 **(1.09)**	2.37 **(2.28)**	1.77 **(1.86)**	116.6 **(121.5)**	114.1 **(100.7)**
	T	1.31 **(1.26)**	1.27 **(1.32)**	2.40 **(2.36)**	1.82 **(1.85)**	**124.2 (120.6)**	159.3 **(166.3)**
wt S1_H2							
	R	2.23 **(2.09)**	1.10 **(1.24)**	2.50 **(2.94)**	1.86 **(2.03)**	117.2 **(120.6)**	144.0 **(158.8)**
	T	1.30 **(1.24)**	1.26 **(1.33)**	2.65 **(2.55)**	1.76 **(1.8)**	117.3 **(123.6)**	170.1 **(158.1)**
wt S2							
	R	3.3 **(2.61)**	1.09 **(1.08)**	2.39 **(2.39)**	1.75 **(1.78)**	125.6 **(126.4)**	-
	T	1.41 **(1.9)**	1.11 **(1.10)**	2.40 **(2.50)**	1.91 **(1.7)**	125.4 **(125.6)**	-
wt S3							
	R	3.18 **(2.33)**	**1**.08 **(1.08)**	2.43 **(2.44)**	1.74 **(1.64)**	118.2 **(122.7)**	-
	T	1.92 **(1.88)**	1.07 **(1.09)**	2.51 **(2.43)**	1.8 **(1.78)**	120.1 **(127.3)**	
wt C1							
	R	3.03 **(3.9)**	1.08 **(1.08)**	2.38 **(2.38)**	1.77 **(1.72)**	151.8 **(134.5)**	-
	T	1.89 **(1.97)**	1.08 **(1.11)**	2.50 **(2.38)**	1.81 **(1.87)**	131.0 **(132.8)**	-
V53 S1_H1							
	R	2.67 **(2.55)**	1.10 **(1.09)**	2.37 **(2.34)**	1.78 **(1.87)**	110.9 **(116.5)**	105.6 **(104.1)**
	T	1.28 **(1.29)**	1.30 **(1.32)**	2.42 **(2.50)**	1.84 **(1.74)**	117.1 **(123.1)**	152.6 **(159.9)**
V53 S1_H2							
	R	2.71 **(2.29)**	1.10 **(1.11)**	2.35 **(2.38)**	1.77 **(1.75)**	119.1 **(114.2)**	130.8 **(124.2)**
	T	0.97 **(0.97)**	**2**.09 **(1.40)**	2.37 **(2.38)**	1.82 **(1.75)**	106.9 **(125.6)**	153.0 **(123.1)**
V53 S2							
	R	2.94 **(2.67)**	1.09 **(1.08)**	2.35 **(2.35)**	1.79 **(1.68)**	145.3 **(139.5)**	-
	T	1.89 **(1.95)**	1.09 **(1.04)**	2.40 **(2.36)**	1.89 **(1.76)**	130.9 **(130.5)**	
V53 S2							
	R	3.20 **(2.94)**	1.08 **(1.06)**	2**.38 (2.49)**	1.76 (**1.78)**	121. 0 **(112.3)**	-
	T	1.87 **(1.92)**	1.08 **(1.10)**	2.59 **(2.39)**	1.80 **(1.76)**	120.2 **(124.4)**	-
V53 C1							
	R	3.74 **(3.32)**	1.08 **(1.08)**	2.43 **(2.44)**	1.74 **(1.76)**	136.7 **(125.9)**	-
	T	1.40 **(1.89)**	1.09 **(1.10)**	2.44 **(2.45)**	1.93 **(1.77)**	126.0 **(125.9)**	-

[A] The distance or angle displayed is between Fe=O····H for aliphatic hydroxylation (S1_H1, S1_H2) and between Fe=O·····C for aromatic hydroxylation (S2, S3, C1).

The distance between the cysteine thiolate group and oxyferryl is one of the major driving forces in the cleavage of the C-H bond. Its electron pushing effect towards the oxyferryl lowers the redox potential and at the same time increases the basicity of the oxyferryl. Therefore, a larger Fe-S distance is unfavorable as it decreases the electron pushing effect

towards the oxyferryl and diminishes its capability to cleave the C-H bond [36] (see the materials and methods section for a more detailed description of the pull-push C-H cleavage mechanism). Also, it is generally known that a longer Fe=O and C-H bond distance whereas a shorter S-Fe and Fe=O·····H bond distance point towards a more product-like character of the TS structure [18]. Compared to the Fe-S bond distances determined for the other TS-heme complexes, a larger distance (above 2.5 Å) appears to be one of the factors that will lead to a less reactive SoM (CYP2D6 WT: S1_H2, S3 and C1, CYP2D6, **Table 2.2**). Moreover, the impact of the F120I on the S3 SoM in *CYP2D6*53* seems to give the aromatic moiety more freedom to adjust its orientation which in this case leads to a less optimal geometry for the oxidation reaction to proceed. This assumption is supported by the outcome from the single point calculations (M06-2X/LACV3P**) performed on the minimized reactant structure of only bufuralol and F120 or I120 (**Table 2.3)**.

Table 2.3. **The QM interaction energy between ligand and the residue 120 calculated for each minimized reactant geometry.**

	Energetic contribution [kcal/mol]	
SoM	ΔE F120	ΔE I120
S1_H1	-6.82	-5.21
S1_H2	-5.56	-3.14
S2	-2.79	-3.10
S3	-11.25	-2.82
C1	-8.71	-0.53

All of the single point calculation values used to determine the energetic contribution of F120 or I120 can be found in **Table S2.2**. The results show that F120 has a larger energetic contribution than I120 within the binding poses (except for S2). Though both amino acids are hydrophobic and can stabilize bufuralol by van der Waals interactions, only the aromatic ring of phenylalanine can form stabilizing non-covalent interactions with bufuralol through π-π stacking and/or cation-π interactions; particularly for S1_H1 and S1_H2 a face-to-face π stacking positioning can be seen (**Figure 2.4**). The less restraining influence of I120 is expected to give the major S1 and minor S2 SoM of bufuralol increased conformational freedom to position the SoM in a more favorable orientation (improved orbital overlap)

towards the oxyferryl, thereby lowering the activation barrier. However, for S3 the result was the opposite.

Comparing the determined S3 *CYP2D6*53* activation barrier to the non-reactive C1 control SoMs activation barrier values which reached values above 40 kcal/mol, it is very unlikely that S3 oxidation within the *CYP2D6*53* variant will occur. As mentioned in the method section, the experimental set-up for the functional characterization of *CYP2D6*53* only monitored the metabolism of the major SoM S1, not S2 or S3. It is therefore not possible to support the QM/MM results with experimental data and make any unambiguous statements regarding the impact on the F120I on orienting bufuralol more favorably. Nevertheless, based on our QM/MM study we predict S2 to be metabolized (though to a lesser extent than the major SoM) and S3 to be non-reactive in *CYP2D6*53*.

In support of this outcome are the results found in other QM/MM studies focusing on aliphatic and aromatic C-H bond activation; Olah et al. studied dextromethorphan oxidation in CYP2D6 and determined an activation barrier of 12.4 kcal/mol for aliphatic carbon hydrogen abstraction and a value of 17.2 kcal/mol for the side-on aromatic carbon oxidation (quartet, B3LYP/LACVP**) [20]. In addition, several other QM/MM studies investigating C-H bond activation within different P450 systems found similar activation barrier values (**Table S2.3**). It should be kept in mind that the several factors will contribute to the variation found in the determined activation values from other QM/MM studies including (i) the (dis-)similarity of the substrates studied (ii) different isoforms will have a different active site and steric effects, (iii) the method and set-up used (functional, basis set, inclusion of dispersion effects, size QM region, etc.). Moreover, the landmarking study of Mulholland et al. studying aliphatic and aromatic hydrogen abstraction within CYP2C9 clearly demonstrated that the activation barrier can span a considerable range (about 6.5 kcal/mol for aliphatic and 4.5 kcal/mol for aromatic hydroxylation) depending on the binding mode used as an input for the QM/MM calculations. The lower energy barriers determined are suggested to be most representative of the reaction pathway taking place in the enzyme [34]. Analysis of the geometric values of the determined TS-heme complexes showed to be in accordance with the values found in other C-H bond activation QM/MM studies [20,40]. For QM gas-phase

calculations, the O·····H-C angle can reach the ideal linear angle (180°) [34]. However, for the QM/MM calculations the active site residues restrain the O·····H-C angle of the transition state and the angle ranges from 159° to 170°.

Conclusion

This study focused on improving our current understanding regarding the driving forces affecting the substrate C-H bond activation within different CYP2D6 active site environments known to induce altered enzyme activities. Learning to what extent a change in active site composition influences the kinetic behaviour of a ligand and its ability to adopt an orientation that favors oxidation, is essential in order to develop accurate computational metabolism prediction tools. Computational metabolite predictions are valuable for example within the early-stage drug design process where this information can be cost-efficiently obtained and applied in the lead compound optimization. The first part of this study, utilizing well-tempered metadynamics simulations, demonstrated that the lowest energy poses of the substrate studied (bufuralol) within different CYP2D6 active sites environments did not correspond to the energetically most accessible conformation for oxidation. Also, determining the significance of the binding pose in order to assign it major or minor SoM using solely the free energy profile of the substrate within the active site can lead to an incomplete answer as demonstrated by the altered-activity variants. In the second part of the study, QM/MM calculations for each SoM of bufuralol within CYP2D6 WT and the clinical relevant allelic variant *CYP2D6*53* (F120I/A122S) were performed in order to determine the corresponding aliphatic and aromatic C-H bond activation barriers. The results show that the stereoelectronic complementarity of the substrate in the active site is an indispensable factor that needs to be quantified for obtaining a meaningful and reliable metabolism prediction. In particular, we observed and quantified that the F120I mutation in *CYP2D6*53* had a less restraining influence on the kinetics of bufuralol, giving it more freedom to orient the dominant SoMs in a favorable geometry which contributed to a more efficient oxidation reaction (lower activation barrier). Furthermore, we found the mutation had neither impact on the CYP2D6 enantioselectivity preference of the oxidation reaction at bufuralol, nor on its stereoselectivity. We therefore postulate that depending on the type(s) of mutation(s)

within the active site, the reaction selectivity and efficiency can be modulated by enabling the substrate to reach a more or less favorable active site stabilized and energetically accessible conformation. The placement of the substrate regarding the oxyferryl group can in turn promote or impede the desired orbital overlap for transferring the electrons, thereby the probability of reaching the TS which subsequently increases or decreases the activation barrier. Within the CYP2D6 reaction balance studied the stereoelectronic complementarity seems to have predominant weight in determining its selectivity though the spatial binding mode is an essential driving force as well.

Materials and methods

Substrate selection. Bufuralol, a beta-adrenoceptor antagonist, was selected for studying the energetic accessibility of its three experimentally well-characterized SoMs (denoted as "S1-S3" in this study). It is a typical CYP2D6 substrate having a basic nitrogen around 5-7 Å away from its site(s) of oxidation and an aromatic planar system [10]. The U.S. Food and drug Administration (FDA) recommends the substrate as *in vitro* marker for CYP2D6 studies (http://www.fda.gov/) [41]. The major metabolites formed are 1'-hydroxybufuralol (referred to as S1_H1 and S1_H2 for strereoselective reason), 1'-oxobufuralol and 1'2'-ethenylbufuralol [10]. Minor metabolites are 6-hydroxybufuralol (referred to as S2) and 4'-hydroxybufuralol (referred to as S3) [42,43]. The regio-selectivity predictor SMARTCyp confirmed 1-'OH-bufuralol (S1) to be the major site of metabolism. The intrinsic reactivity of the two experimentally verified minor sites (S2 and S3) were also indicated (**Figure 2.1A**). An overview of the SoMs investigated in this study can be found in **Figure 2.1B** and the predicted metabolites by GLORY [13] in **Figure 2.1C** which are in agreement with experiment. As a negative control, the site referred to as C1 at the aromatic carbon located in the furan ring was selected. For the *CYP2D6*53* variant, increased activity has been determined by measuring the activity of bufuralol 1'-hydroxylation (K_M and V_{max}) using a substrate concentration of 80 µM. Results indicated a 4-fold increase in intrinsic clearance (CL_{int}) compared to CYP2D6 WT [30]. The minor SoMs (S2 and S3) were not monitored in this study.

Protein preparation. The CYP2D6 crystal structure with PDB ID 4WNW was retrieved from the Protein Data Bank (PDB: https://www.rcsb.org/) and processed using the protein preparation wizard of Maestro small-molecule drug discovery suite (v. 2017-2). Missing residues and hydrogen atoms were added assuming the pH of 7.4, bond orders were assigned, and waters, the co-crystalized ligand and co-factors were removed. Compound I state was modelled (Fe^{3+} bound to O^{2-}, zero-order bond). The distance between the iron and the oxygen was set to 1.97 Å. Subsequently, a short heavy-atom restrained minimization was performed using the OPLS_2005 force field [44].

Ligand preparation and docking. The three-dimensional (3D) conformation of bufuralol was obtained from DrugBank (accessed: October 2018). The structure was prepared using LigPrep wizard of Maestro small-molecule drug discovery suite (OPLS_2005 force field, Schrödinger v. 2017-2). The preparation procedure included bond order assignment, protonation evaluation (at pH 7.4), and a standard LigPrep minimization of the conformation. Bufuralol was subsequently docked to CYP2D6 wild-type using default docking protocol using Glide [45]. The best scored docking pose was used for subsequent simulations. After docking, the two amino acid mutations for each variant were introduced into the CYP2D6 wild-type. This way the WT-MTD as well as were started from the same ligand pose/orientation.

Well-tempered metadynamics simulations. The systems used for WT-MTD were prepared following the standard procedure for running MD simulations; the CYP2D6-ligand systems were solvated using the TIP3P water model in orthorhombic boxes of the periodic boundary system with 10.0 Å cutoff from the protein in each of the three dimensions. Net charge of the system was neutralized by addition of sodium ions. For the WT-MTD simulations, the OPLS_2005 force field was applied as implemented in Desmond (version 2016-4). Box dimensions were 68 x 85 X 83 $Å^3$ with 39711 water atoms. Before starting the WTD-MD production runs of 200 ns in a NPT ensemble and standard conditions (T= 300 K, p= 101.325 kPa), a relaxation phase was applied with the Nose-Hoover thermostat coupled with the

Martyna-Tobias-Klein barostat using a relaxation time of 1.0 ps and a time step of 2.0 fs at 300 K. Long-range coulombic interactions were handled with the smooth particle mesh Ewald (PME) method with a cutoff for short-range interactions of 9 Å. The M-SHAKE algorithm was used to constrain bonds to hydrogens, no hydrogen mass partitioning was applied. Simulations were performed on a GeForce GTX Titan GPU. The compound I state of the heme-complex was modeled for the simulations as this enabled to analyse the FES profile associated with the orientation of the substrate SoMs with respect to the oxyferryl group which is critical for an efficient hydrogen abstraction reaction.

Well-tempered metadynamics parameters. Two collective variables were chosen: (i) the distance between the oxyferryl and the SoM carbon atoms, (ii) the angle between the iron, the oxyferryl and the aliphatic hydrogen (Fe=O·····H) or the aromatic carbon (Fe=O·····C) atoms. In addition, as an alternative set (CV-set 2) the angle between the oxyferryl oxygen atom, the aliphatic/aromatic carbon hydrogen and the aliphatic/aromatic carbon (O·····H-C) atoms was monitored. Both angles need to reach an optimal geometry (around 120° and 180° respectively, **(Figure 2.5**) for a most efficient reaction process [20].

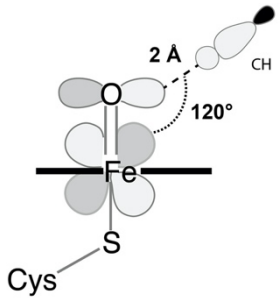


Figure 2.5. **CYP2D6 The orientation of the substrate and cpd I for C-H bond activation is displayed which facilitates optimal orbital overlap and subsequently leads to low activation barriers [18,20,34].**

The width of the Gaussian was set to 0.05 Å for the distance CV1, and 2.5° for the angle CV2. The width was chosen by one-fourth of the average fluctuation of the CV which is needed to get out of the free energy minimum. The initial height of the Gaussians was set to 0.03 kcal/mol and the deposition time was 0.09 ps. For CV1 the exploration of CV space

was limited by adding an upper wall bias of 10 Å in order to keep the ligand within close vicinity of the heme and sampling the relevant region of the free energy surface. All simulations were run for 200 ns. The negative of the sum of the Gaussians gives the FES. Convergence of the WT-MTD simulations was evaluated by a) performing two replicas in which the seed was changed to assign different starting velocities for the major site of metabolism (S1) and the negative control (C1). The range of the FES profile minima (blue area) should be superposable. b) Verification of efficient sampling of space for each CV over time (**Figures S2.2-S2.3**). c) The height of the Gaussian deposited during the simulation time was evaluated to ensure that the simulation was not stuck in a local minimum and a smooth FES was achieved (**Figure S2.4**). d) Evaluation of the last 50 ns at 25 ns interval, to verify that the FES profile did not change significantly (**Figures S2.6-S2.11**). In addition, a visual inspection of the distribution of the frames over the FES was done, which supported that the simulation sampled the CV space efficiently and a smooth FES was reached and thus are an accurate representation of the free energies.

QM/MM modelled hydrogen abstraction. A remarkable fact of CYP450 enzymes is that nature designed them as such that the heme-thiolate complex can oxidize a thermodynamically very stable C-H bond but it will not oxidize nearby residues (e.g. tyrosine, tryptophan or histidine) which would also destroy the functionality of the enzyme. In order to break a strong C-H bond (typically around 100 kcal/mol) and form a stronger O-H bond, both the basicity and redox potential of the Fe=O bond are important driving forces [46]. The cysteine thiolate group has a strong electron pushing effect towards the oxyferryl thereby lowering the redox potential but increasing the basicity of the oxyferryl oxygen (pK_a= ~12) which in turn leads to a greater pull for C-H cleavage [47]. Aliphatic hydrogen abstraction occurs through a rebound-mechanism shown in **Figure 2.6**. The hydrogen abstraction step is the rate-limiting step and is therefore the step modelled in this study for determining the activation barrier of the aliphatic hydrogen abstraction [18,34,48,49].

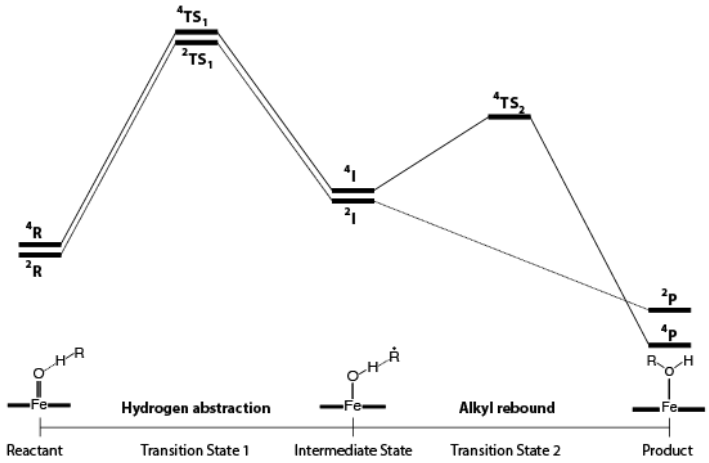


Figure 2.6. **Aliphatic carbon hydrogen abstraction energy diagram. The different energy levels are shown for the doublet and quartet spin state [18].** For the quartet spin state, the reactant, intermediate and product states were modelled and geometrically optimized with QM/MM calculations. A transition state search was performed to find the first intermediate states (TS_1) in order to calculate the activation barriers for all the different SoMs.

For aromatic hydrogen abstraction, the reaction proceeds differently. Both a phenolic and benzene epoxide product can be formed through three major pathways; with a keto tautomer (**Figure 2.7**, pathway A), an N-protonated porphyrin (**Figure 2.7** pathway B) or an epoxide as intermediate product (**Figure 2.7**, pathway C). The addition-rearrangement pathway (**Figure 2.7,** pathway A) has been found to be the lowest energy pathway [50,51]. The step with the highest energy barrier is found to be the addition of the Cpd I to the substrate carbon which leads to the tetrahedral intermediate s-complex [37]. Therefore, this step was modelled to determine the activation barrier of the aromatic oxidation. To quantify the impact of the mutations it is assumed that neither the reaction pathway is changed, nor other factors (e.g. substrate pathway, large-scale movements, etc.). The relative changes in activation barrier are thus the result of the differences in stabilizing inter- and intramolecular forces (electrostatic multipole, electrostatic penetration, exchange, induction and dispersion contributions) within the active site [52]. From the two point mutations, the F120I mutation (hydrophobic aromatic to hydrophobic aliphatic) is most relevant for the reaction barrier calculations considering its vicinity to the reactive heme center. Previous studies demonstrated the stereo- and regioselectivtity role of F120 and that the type of amino acid substitution and substrate influence the outcome [35,36,53,54].

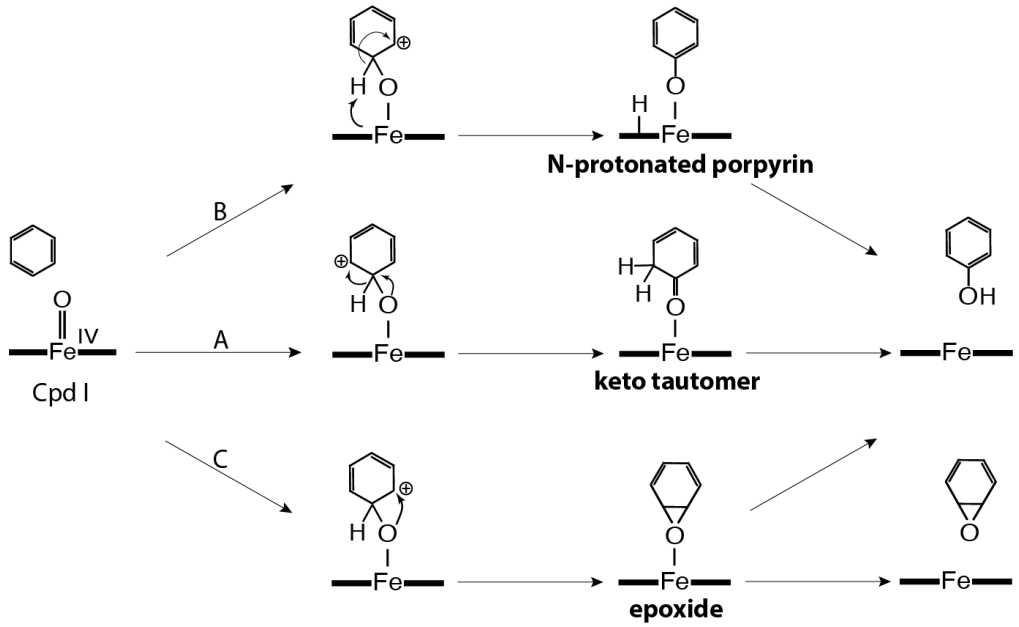


Figure 2.7. **Three proposed rearrangement pathways for aromatic hydrogen abstraction [37].** The step with the highest energy barrier is found to be the addition of the Cpd I to the substrate carbon which leads to the tetrahedral intermediate s-complex. Modeling of this state has been therefore the focus of this study in order to define the corresponding height of the activation barrier.

For 1'hydroxylation of bufuralol (major SoM) the F120A mutation reversed the enantioselectivity from R to S whereas the metabolite stereoselectivity remained the same [39]. In addition, other studies using different substrates showed similar outcome (changed regio- and stereoselectivity) and highlighted the determining role of the intrinsic reactivity, in particular for molecules which have large mobility within the active site [35,36,53,54]. To determine the activation barrier which accurately represents the reactivity of the enzyme, selecting a starting structure with an energetically accessible geometry resembling closely the transition state is crucial [18]. A very precise study would use multiple starting structures for each SoM to obtain a realistic average barrier value through Boltzmann-weighted calculations [55]. However, the aim of this study was not to determine absolute activation barrier heights, but rather to perform a relative comparison study in order to identify how the ligand's intrinsic reactivity is influenced by the mutation of surrounding residues.

QM/MM procedure. QM/MM analysis was performed using the Qsite module implemented in the Schrödinger Suite (2016-4) [56,57]. It combines Jaguar (version 9.3, release 15) and Impact modules (OPLS force field) for the QM and MM calculations respectively [58] (Impact,

version 6.4, Schrödinger, LLC, New York, NY, 2014). The starting structures (reactant structure) for QM/MM optimization were obtained from the WT-MTD simulations. Structures which have a SoM that is in close vicinity of the oxyferryl oxygen atom (< 3 Å) and have an Fe=O·····H angle between 110° and 130° are considered energetically accessible for oxidation (**Figure 2.8)** [18,20,34]. Therefore, for each SoM, the frame that fulfilled these criteria was selected from the corresponding WT-MTD. The selected reactant structures for each SoM in CYP2D6 wild-type and *CYP2D6*53* used for the QM/MM calculations can be found in **Figure 2.3**.

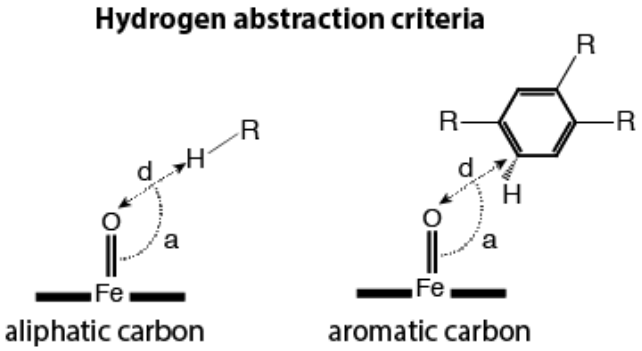


Figure 2.8. **Overview selection criteria applied to choose the frames from the well-tempered MTD for QM/MM calculations.** The frames closest to the heme oxygen (d < 3 Å) and with an Fe=O·····H angle between 110° and 130° were selected [18,20,34,59].

The criteria are based on the previously performed QM studies which determined the TS geometries for each oxidation pathway (aliphatic/aromatic) [18,20,60]. The QM/MM geometry optimization was performed with the M06-2X functional and the polarized basis set LACVP*, diffuse basis functions (LACVP*+) were added only for the negative charged atoms [18,20,61]. Single point calculations were performed at the M06-2X/LACV3P*+ level. The basis set selection is frequently a compromise between accuracy of the results and the feasibility of the computational time required for the calculations. The Minnesota effective core potential functionals have been parameterized such that they implicitly account for dispersion effects [62–64] and have previously shown to perform well for similar P450 systems [20,34,38,61]. The CYP2D6 WT QM region included the ligand, the heme and the binding pocket residue side chains (hydrogen caps between C_α and C_β bond) of PHE120 (or ILE 120 in one of the variants), GLU216, ASP301 and CYS443. Residues beyond 6 Å from the selected

QM region were treated as frozen during the MM optimization. All waters were treated within the MM optimization also the water molecules within the active site. Our main motivation to exclude the active site water molecules from the QM region is that the number of water molecules varied among the starting structures which, depending on their position, may impact the results and diminish the impact of the mutated residue. Also, we inspected the positions of the water molecules to evaluated if they might be able to stabilize the transition state geometry by interaction with the oxyferryl or SoMs which was not the case. Previous studies showed that this can lower the activation barrier in the order of 3-4 kcal/mol on average [38]. The conjugate gradient algorithm was applied and tight MM optimization convergence criteria were used; maximum number of cycles 9999, initial step size of the minimization cycle 0.05, maximum step size 1, energy and gradient as convergence criterion with an energy change of 10^{-5} kcal / mol, and gradient criterion of 0.01 kcal/(mol Å). It has been shown before that for hydrogen abstraction the low (doublet) and intermediate (quartet) are very close in energy [18,65]. For computational reason we studied all structures using the quartet spin state.

RMSD calculation. The RMSD difference between the minimized SoM reactant binding pose of the CYP2D6 WT and *CYP2D6*53* variant was calculated by superposition of the C_α atoms of the QM region and calculating the ΔRMSD of bufuralol in place.

References

1. Werck-Reichhart D, Feyereisen R. Cytochromes P450: a success story. Genome Biol.:2000;1(6):reviews3003.

2. Hodges RE, Minich DM. Modulation of Metabolic Detoxification Pathways Using Foods and Food-Derived Components: A Scientific Review with Clinical Application. J Nutr Metab. 2015 [cited 2018 Oct 29];2015.

3. Dubey KD, Shaik S. Cytochrome P450—The Wonderful Nanomachine Revealed through Dynamic Simulations of the Catalytic Cycle. Acc Chem Res. 2019 Feb 19;52(2):389–99.

4. Guengerich FP. Mechanisms of Cytochrome P450-Catalyzed Oxidations. ACS Catalysis. 2018 Dec 7;8(12):10964–76.

5. Rendic S, Guengerich FP. Survey of Human Oxidoreductases and Cytochrome P450 Enzymes Involved in the Metabolism of Xenobiotic and Natural Chemicals. Chem Res Toxicol. 2015 Jan 20;28(1):38–42.

6. Mansuy D. The great diversity of reactions catalyzed by cytochromes P450. Comp Biochem Physiol C, Pharmacol Toxicol Endocrinol. 1998 Nov;121(1–3):5–14.

7. Guengerich FP. Intersection of the Roles of Cytochrome P450 Enzymes with Xenobiotic and Endogenous Substrates: Relevance to Toxicity and Drug Interactions. Chem Res Toxicol. 2017 17;30(1):2–12.

8. Zanger UM, Schwab M. Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. Pharmacology & Therapeutics. 2013 Apr;138(1):103–41.

9. Don CG, Smieško M. Out-compute drug side effects: Focus on cytochrome P450 2D6 modeling. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2018;8(5):e1366.

10. Wang B, Yang L-P, Zhang X-Z, Huang S-Q, Bartlam M, Zhou S-F. New insights into the structural characteristics and functional relevance of the human cytochrome P450 2D6 enzyme. Drug Metab Rev. 2009;41(4):573–643.

11. Šícho M, de Bruyn Kops C, Stork C, Svozil D, Kirchmair J. FAME 2: Simple and Effective Machine Learning Model of Cytochrome P450 Regioselectivity. Journal of Chemical Information and Modeling. 2017 Aug 28;57(8):1832–46.

12. Šícho M, Stork C, Mazzolari A, de Bruyn Kops C, Pedretti A, Testa B, et al. FAME 3: Predicting the Sites of Metabolism in Synthetic Compounds and Natural Products for Phase 1 and Phase 2 Metabolic Enzymes. J Chem Inf Model. 2019 Aug 26;59(8):3400–12.

13. de Bruyn Kops C, Stork C, Šícho M, Kochev N, Svozil D, Jeliazkova N, et al. GLORY: Generator of the Structures of Likely Cytochrome P450 Metabolites Based on Predicted Sites of Metabolism. Front Chem. 2019;7:402.

14. Rudik A, Dmitriev A, Lagunin A, Filimonov D, Poroikov V. SOMP: web server for in silico prediction of sites of metabolism for drug-like compounds. Bioinformatics. 2015 Jun 15;31(12):2046–8.

15. Tian S, Djoumbou-Feunang Y, Greiner R, Wishart DS. CypReact: A Software Tool for in Silico Reactant Prediction for Human Cytochrome P450 Enzymes. J Chem Inf Model. 2018 25;58(6):1282–91.

16. Tyzack JD, Kirchmair J. Computational methods and tools to predict cytochrome P450 metabolism for drug discovery. Chem Biol Drug Des. 2019 Apr;93(4):377–86.

17. Rittle J, Green MT. Cytochrome P450 Compound I: Capture, Characterization, and C-H Bond Activation Kinetics. Science. 2010 Nov 12;330(6006):933–7.

18. Shaik S, Cohen S, Wang Y, Chen H, Kumar D, Thiel W. P450 Enzymes: Their Structure, Reactivity, and Selectivity—Modeled by QM/MM Calculations. Chem Rev. 2010 Feb 10;110(2):949–1017.

19. Podgorski MN, Coleman T, Chao RR, De Voss JJ, Bruning JB, Bell SG. Investigation of the requirements for efficient and selective cytochrome P450 monooxygenase catalysis across different reactions. Journal of Inorganic Biochemistry. 2020 Feb 1;203:110913.

20. Olah J, Mulholland AJ, Harvey JN. Understanding the determinants of selectivity in drug metabolism through modeling of dextromethorphan oxidation by cytochrome P450. Proceedings of the National Academy of Sciences. 2011 Apr 12;108(15):6050–5.

21. Olsen L, Oostenbrink C, Jørgensen FS. Prediction of cytochrome P450 mediated metabolism. Adv Drug Deliv Rev. 2015 Jun 23;86:61–71.

22. Don CG, Smieško M. Microsecond MD simulations of human CYP2D6 wild-type and five allelic variants reveal mechanistic insights on the function. PLoS ONE. 2018;13(8):e0202534.

23. Fischer A, Smieško M. Spontaneous Ligand Access Events to Membrane-Bound Cytochrome P450 2D6 Sampled at Atomic Resolution. Sci Rep. 2019 Nov 11;9(1):16411.

24. Nair PC, McKinnon RA, Miners JO. Cytochrome P450 structure-function: insights from molecular dynamics simulations. Drug Metab Rev. 2016 Aug;48(3):434–52.

25. Louet M, Labbé CM, Fagnen C, Aono CM, Homem-de-Mello P, Villoutreix BO, et al. Insights into molecular mechanisms of drug metabolism dysfunction of human CYP2C9*30. PLOS ONE. 2018 May 10;13(5):e0197249.

26. Li J, Zhang H, Liu G, Tang Y, Tu Y, Li W. Computational Insight Into Vitamin K1 **ω**-Hydroxylation by Cytochrome P450 4F2. Front Pharmacol. 2018;9:1065.

27. Groenhof G. Introduction to QM/MM Simulations. In: Monticelli L, Salonen E, editors. Biomolecular Simulations [Internet]. Totowa, NJ: Humana Press; 2013 [cited 2019 Sep 26]. p. 43–66.

28. Schyman P, Lai W, Chen H, Wang Y, Shaik S. The Directive of the Protein: How Does Cytochrome P450 Select the Mechanism of Dopamine Formation? J Am Chem Soc. 2011 May 25;133(20):7977–84.

29. Schöneboom JC, Lin H, Reuter N, Thiel W, Cohen S, Ogliaro F, et al. The elusive oxidant species of cytochrome P450 enzymes: characterization by combined quantum mechanical/molecular mechanical (QM/MM) calculations. J Am Chem Soc. 2002 Jul 10;124(27):8142–51.

30. Sakuyama K, Sasaki T, Ujiie S, Obata K, Mizugaki M, Ishikawa M, et al. Functional Characterization of 17 CYP2D6 Allelic Variants (CYP2D6.2, 10, 14A–B, 18, 27, 36, 39, 47–51, 53–55, and 57). Drug Metab Dispos. 2008 Dec 1;36(12):2460–7.

31. Geier M, Braun A, Fladischer P, Stepniak P, Rudroff F, Hametner C, et al. Double site saturation mutagenesis of the human cytochrome P450 2D6 results in regioselective steroid hydroxylation. FEBS J. 2013 Jul 1;280(13):3094–108.

32. Rydberg P, Gloriam DE, Zaretzki J, Breneman C, Olsen L. SMARTCyp: A 2D Method for Prediction of Cytochrome P450-Mediated Drug Metabolism. ACS Med Chem Lett. 2010 Jun 10;1(3):96–100.

33. Matlock MK, Hughes TB, Swamidass SJ. XenoSite server: a web-available site of metabolism prediction tool. Bioinformatics. 2015 Apr 1;31(7):1136–7.

34. Lonsdale R, Houghton KT, Żurek J, Bathelt CM, Foloppe N, de Groot MJ, et al. Quantum Mechanics/Molecular Mechanics Modeling of Regioselectivity of Drug Metabolism in Cytochrome P450 2C9. J Am Chem Soc. 2013 May 29;135(21):8001–15.

35. de Graaf C, Vermeulen NPE, Feenstra KA. Cytochrome P450 in Silico: An Integrative Modeling Approach. J Med Chem. 2005 Apr 1;48(8):2725–55.

36. de G, Oostenbrink C, Keizers PHJ, van V-L, van W, Tschirret-Guth RA, et al. Molecular modeling-guided site-directed mutagenesis of cytochrome P450 2D6. Current Drug Metabolism. 2007;8(1):59–77.

37. Bathelt CM, Mulholland AJ, Harvey JN. QM/MM modeling of benzene hydroxylation in human cytochrome P450 2C9. J Phys Chem A. 2008 Dec 18;112(50):13149–56.

38. Tian L, Friesner RA. QM/MM Simulation on P450 BM3 Enzyme Catalysis Mechanism. J Chem Theory Comput. 2009;5(5):1421–31.

39. Masuda K, Tamagake K, Katsu T, Torigoe F, Saito K, Hanioka N, et al. Roles of phenylalanine at position 120 and glutamic acid at position 222 in the oxidation of chiral substrates by cytochrome P450 2D6. Chirality. 2006;18(3):167–76.

40. Lonsdale R, Harvey JN, Mulholland AJ. Compound I Reactivity Defines Alkene Oxidation Selectivity in Cytochrome P450cam. J Phys Chem B. 2010 Jan 21;114(2):1156–62.

41. Research C for DE and. Drug Development and Drug Interactions: Table of Substrates, Inhibitors and Inducers. FDA [Internet]. 2019 Feb 9 [cited 2019 Sep 26]; Available from: http://www.fda.gov/drugs/drug-interactions-labeling/drug-development-and-drug-interactions-table-substrates-inhibitors-and-inducers

42. Kim J, Lim Y-R, Han S, Han J-S, Chun Y-J, Yun C-H, et al. Functional influence of human CYP2D6 allelic variations: P34S, E418K, S486T, and R296C. Arch Pharm Res. 2013 Dec;36(12):1500–6.

43. Hanna IH, Kim M-S, Guengerich FP. Heterologous Expression of Cytochrome P450 2D6 Mutants, Electron Transfer, and Catalysis of Bufuralol Hydroxylation: The Role of Aspartate 301 in Structural Integrity. Archives of Biochemistry and Biophysics. 2001 Sep 15;393(2):255–61.

44. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des. 2013 Mar;27(3):221–34.

45. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. J Med Chem. 2004 Mar 1;47(7):1750–9.

46. Groves JT. Cytochrome P450 enzymes: understanding the biochemical hieroglyphs. F1000Research. 2015; 4;178.

47. Groves JT. Enzymatic C-H bond activation. Nat Chem. 2014 Feb;6(2):89–91.

48. Ogliaro F, Harris N, Cohen S, Filatov M, de Visser SP, Shaik S. A Model "Rebound" Mechanism of Hydroxylation by Cytochrome P450: Stepwise and Effectively Concerted Pathways, and Their Reactivity Patterns. J Am Chem Soc. 2000 Sep 1;122(37):8977–89.

49. Kumar D, de Visser SP, Shaik S. How Does Product Isotope Effect Prove the Operation of a Two-State "Rebound" Mechanism in C−H Hydroxylation by Cytochrome P450? J Am Chem Soc. 2003 Oct 1;125(43):13024–5.

50. Schröder D, Schwarz H. Benzene Oxidation by 'Bare' FeO+ in the Gas Phase. Helvetica Chimica Acta. 1992;75(4):1281–7.

51. Shiota Y, Suzuki K, Yoshizawa K. Mechanism for the Direct Oxidation of Benzene to Phenol by FeO+. Organometallics. 2005 Jul 1;24(14):3532–8.

52. Sokalski WA. The physical nature of catalytic activity due to the molecular environment in terms of intermolecular interaction theory: derivation of simplified models. Journal of Molecular Catalysis. 1985 Jun 1;30(3):395–410.

53. Flanagan JU, Maréchal J-D, Ward R, Kemp CA, McLaughlin LA, Sutcliffe MJ, et al. Phe120 contributes to the regiospecificity of cytochrome P450 2D6: mutation leads to the formation of a novel dextromethorphan metabolite. The Biochemical journal. 2004;380(Pt):353–60.

54. Maréchal J-D, Kemp CA, Roberts GCK, Paine MJI, Wolf CR, Sutcliffe MJ. Insights into drug metabolism by cytochromes P450 from modelling studies of CYP2D6-drug interactions. Br J Pharmacol. 2008 Mar;153(Suppl 1):S82–9.

55. Lonsdale R, Oláh J, Mulholland AJ, Harvey JN. Does Compound I Vary Significantly between Isoforms of Cytochrome P450? J Am Chem Soc. 2011 Oct 5;133(39):15464–74.

56. Murphy RB, Philipp DM, Friesner RA. A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. Journal of Computational Chemistry. 2000;21(16):1442–57.

57. Philipp DM, Friesner RA. Mixed ab initio QM/MM modeling using frozen orbitals and tests with alanine dipeptide and tetrapeptide. Journal of Computational Chemistry. 1999 Nov 15;20(14):1468–94.

58. Bochevarov AD, Harder E, Hughes TF, Greenwood JR, Braden DA, Philipp DM, et al. Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. International Journal of Quantum Chemistry. 2013;113(18):2110–42.

59. Reinhard FGC, Visser SP de. Biodegradation of cosmetics products: A computational study of Cytochrome P450 metabolism of phthalates. In 2017.

60. de Visser SP, Ogliaro F, Sharma PK, Shaik S. What factors affect the regioselectivity of oxidation by cytochrome p450? A DFT study of allylic hydroxylation and double bond epoxidation in a model reaction. J Am Chem Soc. 2002 Oct 2;124(39):11809–26.

61. Jerome SV, Hughes TF, Friesner RA. Successful application of the DBLOC method to the hydroxylation of camphor by cytochrome p450. Protein Sci. 2016 Jan;25(1):277–85.

62. Zhao Y, Truhlar DG. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. Theoretical Chemistry Accounts. 2008 May;120(1–3):215–41.

63. Walker M, Harvey AJA, Sen A, Dessent CEH. Performance of M06, M06-2X, and M06-HF density functionals for conformationally flexible anionic clusters: M06 functionals perform better than B3LYP for a model system with dispersion and ionic hydrogen-bonding interactions. J Phys Chem A. 2013 Nov 27;117(47):12590–600.

64. Attia AAA, Lupan A, Silaghi-Dumitrescu R. Spin state preference and bond formation/cleavage barriers in ferrous-dioxygen heme adducts: remarkable dependence on methodology. RSC Adv. 2013 Nov 15;3(48):26194–204.

65. de Visser SP, Kumar D, Cohen S, Shacham R, Shaik S. A predictive pattern of computed barriers for C-h hydroxylation by compound I of cytochrome p450. J Am Chem Soc. 2004 Jul 14;126(27):8362–3.

Supporting Information

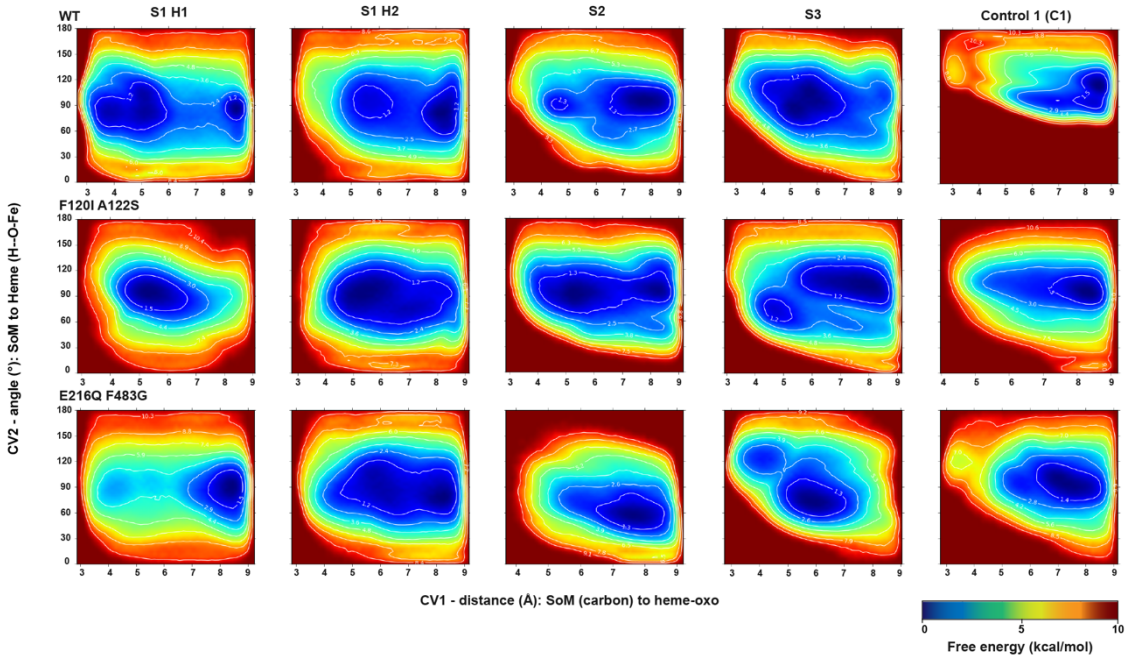


Figure S2.1 **Free energy surface (FES) graphs is displayed for the experimentally known sites of metabolism (SoMs, S1-S3) and control site C1 of bufuralol and control using CV-set 2 calculated over 200ns WT MTD simulations The distance (Å) between the heme-oxo group and the SoM carbon.**

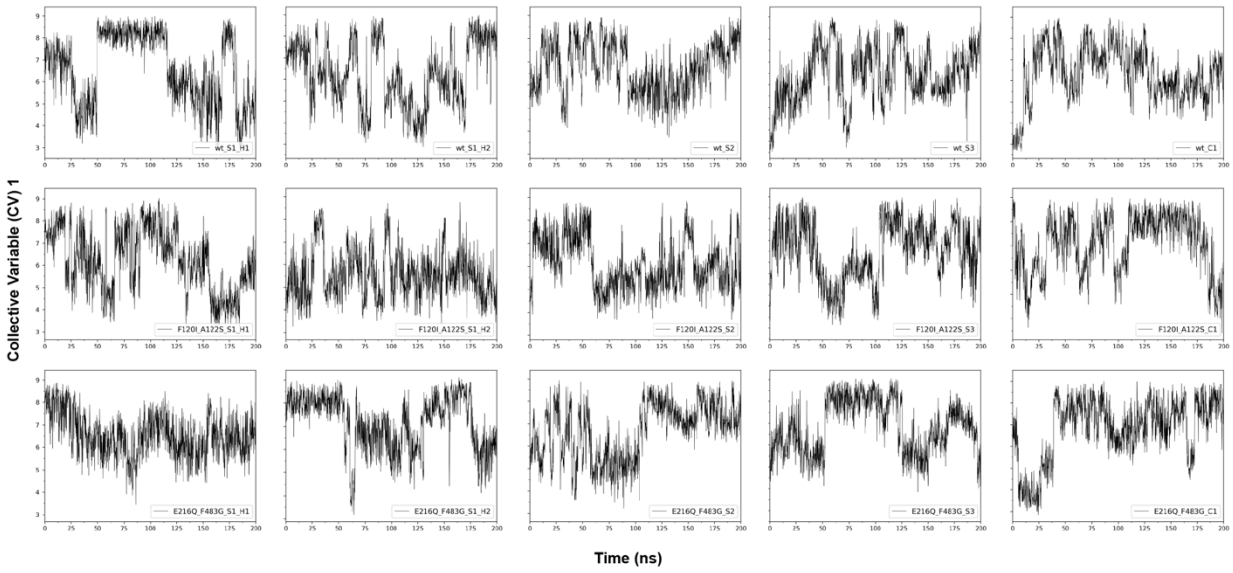


Figure S2.2. **Diffusion assessment of CV1 (Å) during WT-MTD.** Convergence assessment of metadynamics simulation for the five monitored bufuralol SoMs in all CYP2D6 variants (wild-type (top row), F120IA122S (middle row) and E216QF483G (bottom row) by assessing the time-evolution behaviour of CV1 (distance SoM to heme-oxo group).

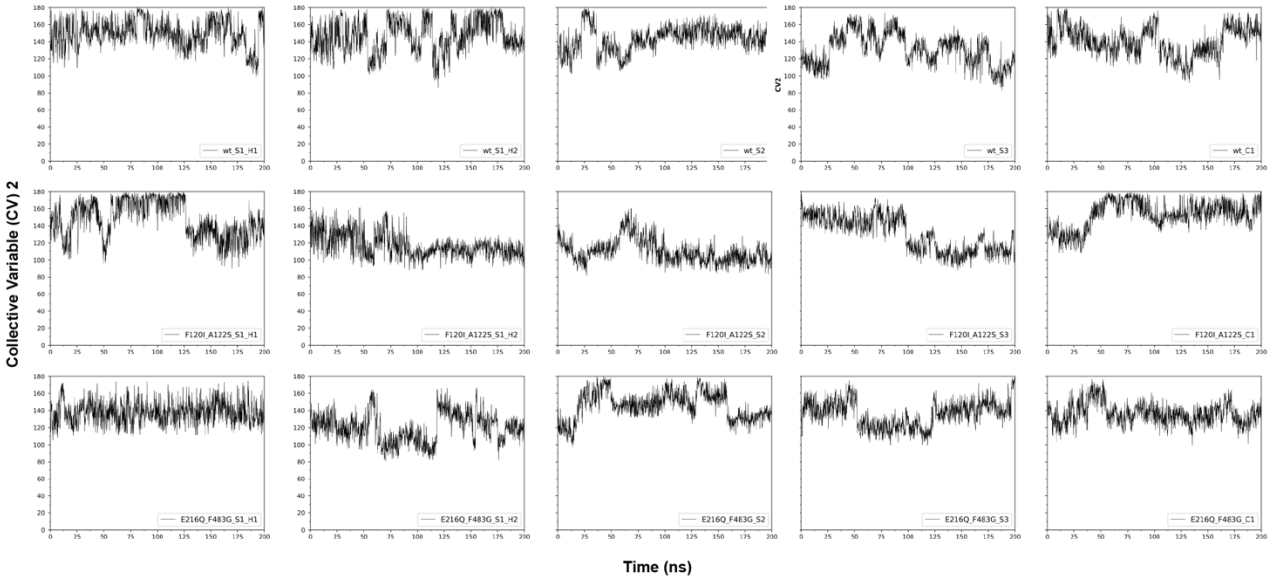


Figure S2.3 **Diffusion assessment of CV2 (degrees) during WT-MTD.** Convergence assessment of metadynamics simulation for the five monitored bufuralol SoMs in all CYP2D6 variants (wild-type (top row), F120IA122S (middle row) and E216QF483G (bottom row) by assessing the time-evolution behaviour of CV2 (the angle Fe-O·····H).

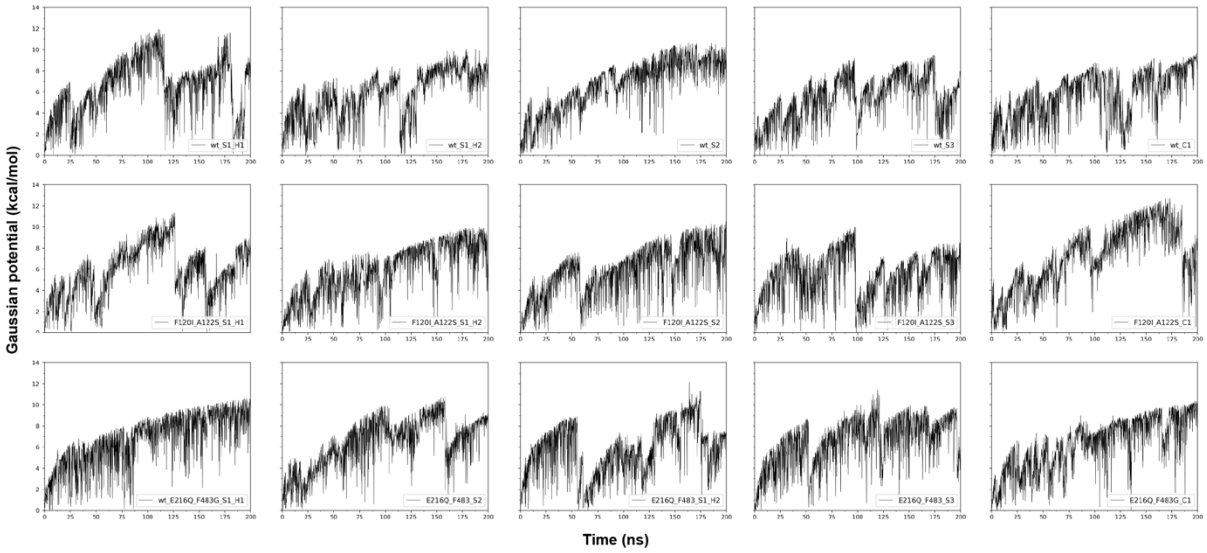


Figure S2.4. **Gaussian potential development over time (Fe·····H angle).** Convergence assessment of metadynamics simulation for the five monitored bufuralol SoMs in all CYP2D6 variants (wild-type (top row), F120IA122S (middle row) and E216QF483G (bottom row)) revealing that no simulation was trapped in a local minimum.

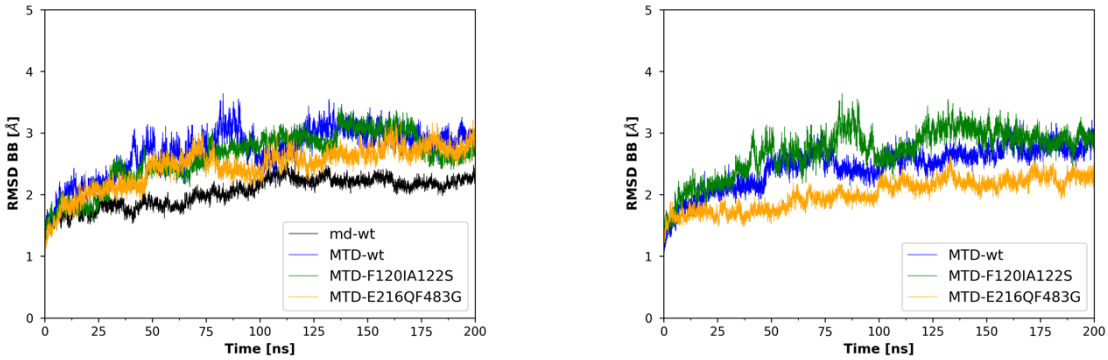


Figure S2.5. **The RMSD (backbone) for the different simulations.** In black the unbiased CYP2D6 wild-type (WT), the well-tempered metadynamics (MTD) of S1_H1 is shown for WT CYP2D6 (blue), F120IA122S (green) and E216QF483G (orange) on the left for the (O·····H-C angle) and the right the (Fe-O·····H angle).

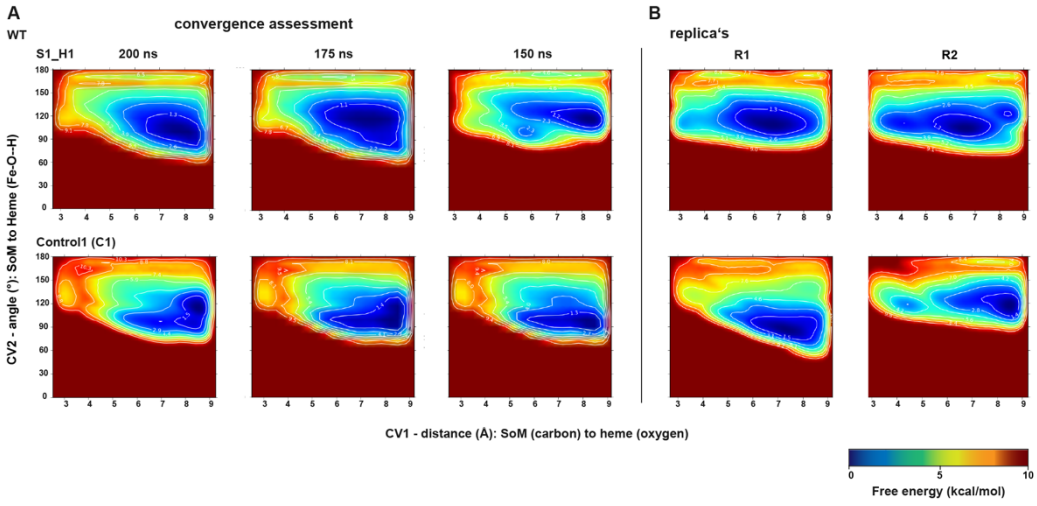


Figure S2.6. **Reproducibility assessment of FES simulations (FeO -- H angle) for wild-type (WT) CYP2D6 with bufuralol using the major site of metabolism (S1_H1) and control (C1) SoM. A)** Convergence assessment of metadynamics simulation for bufuralol S1_H1 (top) and C1 (bottom) by FES analysis at 150 ns, 175 ns and 200 ns**. B)** Replica simulations (R1 and R2) ran for bufuralol S1_H1 (top) and C1 (bottom) using different start velocities.

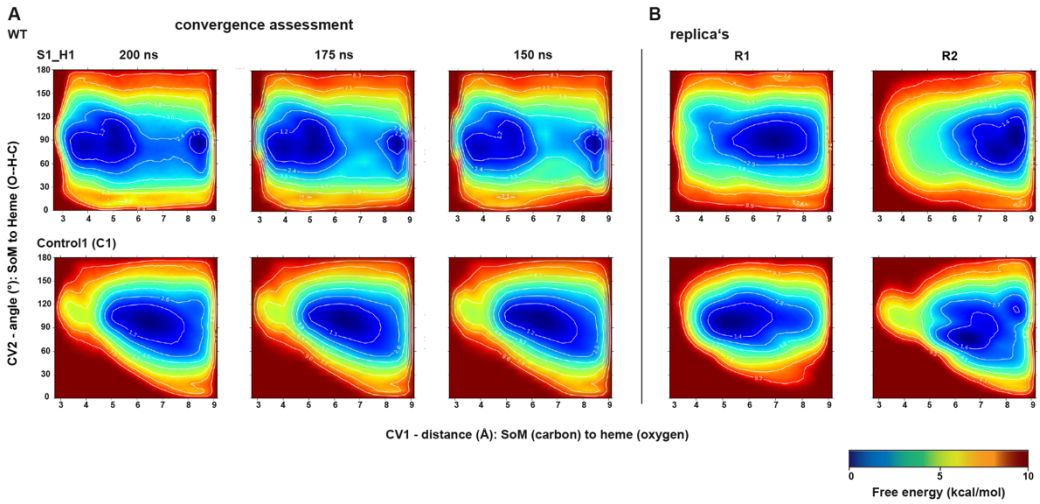


Figure S2.7. **Reproducibility assessment of FES simulations (O·····H-C angle) for wild-type (WT) CYP2D6 with bufuralol using the major site of metabolism (S1_1) and control (C1) SoM. A)** Convergence assessment of metadynamics simulation for bufuralol S1_H1 (top) and C1 (bottom) by FES analysis at 150 ns, 175 ns and 200 ns. **B**) Replica simulations (R1 and R2) ran for bufuralol S1_H1 (top) and C1 (bottom) using different start velocities.

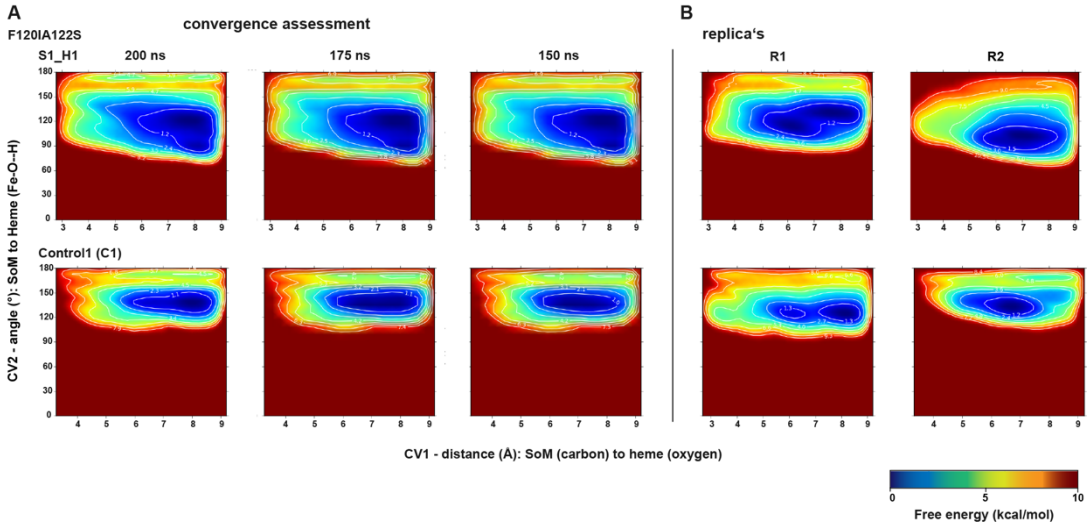


Figure S2.8. **Reproducibility assessment of FES simulations (FeO·····H angle) for the increased activity CYP2D6 variant (F120IA122S) with bufuralol using the major site of metabolism (S1_H1) and control (C1) SoM. A)** Convergence assessment of metadynamics simulation for bufuralol S1_H1 (top) and C1 (bottom) by FES analysis at 150 ns, 175 ns and 200 ns**. B)** Replica simulations (R1 and R2) ran for bufuralol S1_H1 (top) and C1 (bottom) using different start velocities.

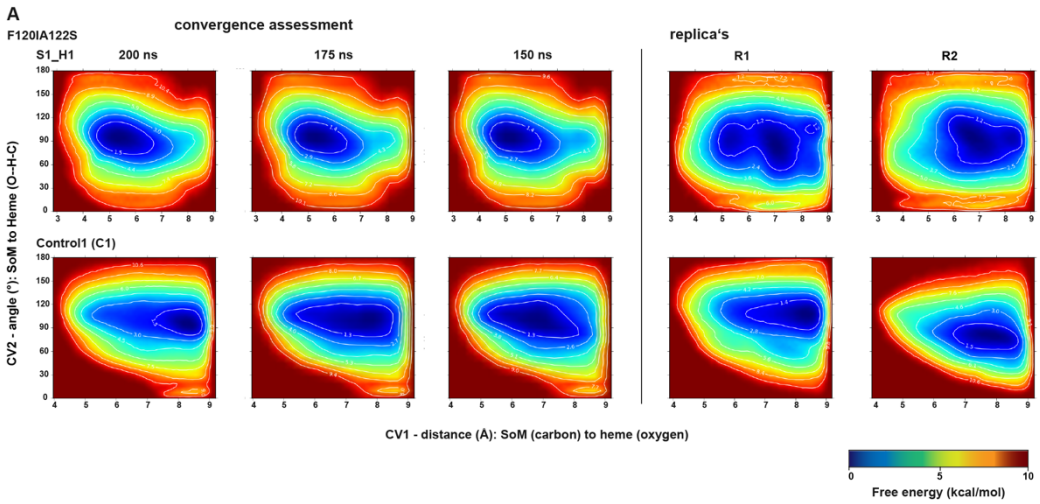


Figure S2.9. **Reproducibility assessment of FES simulations (O·····H-C angle) for the increased activity CYP2D6 variant (F120IA122S) with bufuralol using the major site of metabolism (S1_1) and control (C1) SoM. A)** Convergence assessment of metadynamics simulation for bufuralol S1_H1 (top) and C1 (bottom) by FES analysis at 150 ns, 175 ns and 200 ns. **B)** Replica simulations (R1 and R2) ran for bufuralol S1_H1 (top) and C1 (bottom) using different start velocities.

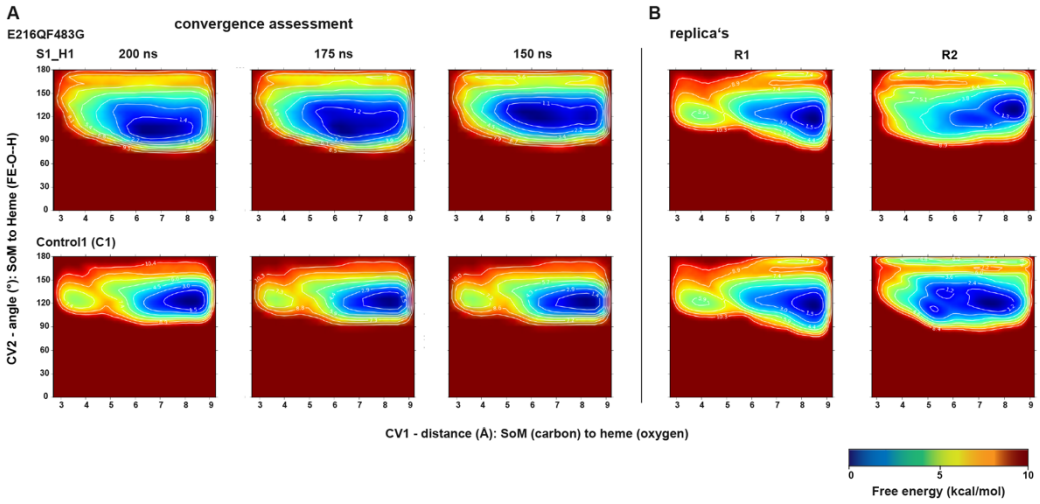


Figure S2.10. **Reproducibility assessment of FES simulations (FeO·····H angle) for the no-activity CYP2D6 variant (E216QF483G) with bufuralol using the major site of metabolism (S1_H1) and control (C1) SoM. A)** Convergence assessment of metadynamics simulation for bufuralol S1_H1 (top) and C1 (bottom) by FES analysis at 150 ns, 175 ns and 200 ns**. B)** Replica simulations (R1 and R2) ran for bufuralol S1_H1 (top) and C1 (bottom) using different start velocities.

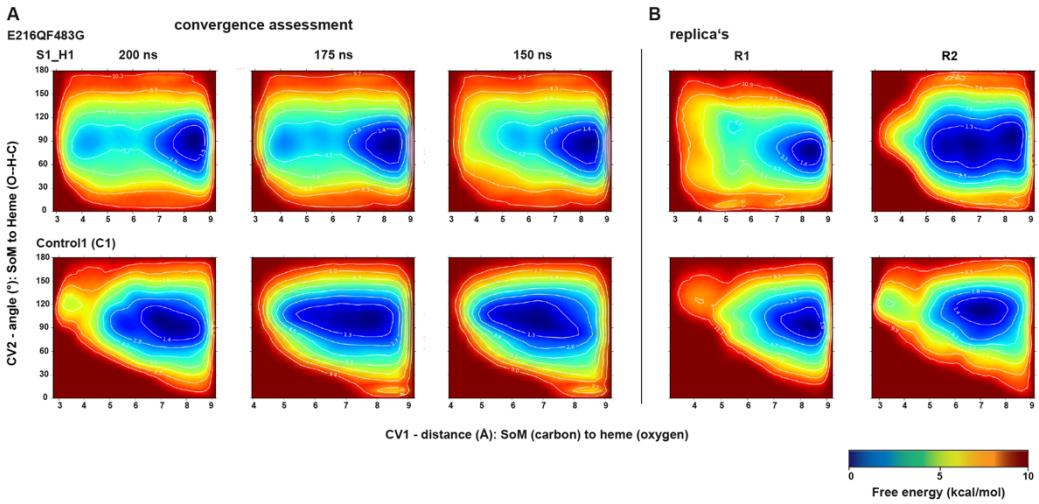


Figure S2.11. **Reproducibility assessment of FES simulations (O·····H-C angle) for the no-activity CYP2D6 variant (E216QF483G) with bufuralol using the major site of metabolism (S1_1) and control (C1) SoM.** **A)** Convergence assessment of metadynamics simulation for bufuralol S1_H1 (top) and C1 (bottom) by FES analysis at 150 ns, 175 ns and 200 ns. **B)** Replica simulations (R1 and R2) ran for bufuralol S1_H1 (top) and C1 (bottom) using different start velocities.

Table S2.1. **Overview of CYP2D6 - bufuralol interactions within each starting structure (reactant) used for QM/MM calculations.** Consensus amino acid (aa) interactions (direct) for CYP2D6 WT and the allelic variant (V53) are in bold. For water-mediated interactions the number of molecules is indicated in brackets and subsequently the distance.

Reactant	direct ligand-enzyme interaction	mode of interaction	distance (Å)	water-mediated ligand-enzyme interaction
Major site				
wt S1_H1	**D301**	**ionic**	**1.6**	**E216 (2)**
	F120	**ππ stacking**	**~3.5**	
	S304	**H-bond**	**1.9**	
V53 S1_H1	**D301**	**ionic**	**1.6**	
	S304	**H-bond**	**2.0**	
wt S1_H2	**D301**	**ionic**	**1.5**	**E216 (1)**
				D301 (2)
V53 S1_H2	**D301**	**ionic**	**1.7**	
Minor Sites				
wt S2	**E216**	**ionic**	**1.7**	**D301 (2)**
	E216	**ionic**	**1.9**	**Q244 (1)**
V53 S2	**D301**	**ionic**	**1.9**	**E216 (1)** **I120 (1)**
wt S3	**D301**	**ionic**	**1.6**	
	D301	**H-bond (bb)**	**1.9**	
	F120	**ππ stacking**	**~3.8**	
V53 S3	**D301**	**ionic**	**1.7**	**E216 (1)**
Control				
wt C1	**D301**	**ionic**	**1.7**	
	S304	**H-bond**	**2.2**	
	F120	**with N..**	**~4.5**	
	F483	**ππ stacking**	**~4.0**	
V53 C1	**D301**	**ionic**	**3.9**	

Table S2.2 **Single point calculation (SPC) results performed on the reactant minimized structure of only bufuralol and F120 in CYP2D6 WT or I120 in CYP2D6*53 (M06-2X/LACV3P**).** The ΔE value was determined after performing SPC on reactant binding pose of the bufuralol-F120 or bufuralol-I120 residues (hartrees) (A), only bufuralol (hartrees) (B), or only F120 or I120 (hartrees) (C) and subtracting A and B from C (kcal/mol).

CYP2D6 WT	A: buf-F120	B: buf	C: F120	ΔE= (A - B - C)
S1_H1	-1100.50233	-828.98965	-271.50182	-6.8
S1_H2	-1100.49505	-828.98425	-271.50194	-5.6
S2	-1100.48193	-828.97554	-271.50195	-2.8
S3	-1100.50373	-828.98378	-271.50202	-11.2
C1	-1100.50004	-828.98536	-271.50080	-8.7
CYP2D6*53	**A: buf-I120**	**B: buf**	**C: I120**	**ΔE= (A - B - C)**
S1_H1	-987.39456	-828.98647	-158.39979	-5.2
S1_H2	-987.38892	-828.98348	-158.40044	-3.2
S2	-987.38807	-828.98361	-158.39952	-3.1
S3	-987.39394	-828.98894	-158.40051	-2.8
C1	-987.29627	-828.89776	-158.39767	-0.5

Table S2.3. **Overview of activation barriers determined by previous QM/MM P450 oxidation studies.** The activation barriers (kcal/mol) displayed are for the quartet spin state (S=3/2) and side-on unless otherwise indicated.

isoform	Substrate	Aliph HA	Arom HA	F-BS*	Reference
CYP2D6	dextromethorphan	12.4	17.2	B3LYP/LACVP**	[9]
CYP2C9	benzene	-	18.1/21.7	B3LYP/LACV3P/6-31G	[37]
CYP101	camphor			B3LYP/LACV3P**	[52]
CYP3A4	Phtalate	19.9	22.8	B3LYP/LACV3P+	[56]
CYP102A1	N-Palmitoylglycine	18.2	-	B3LYP/ LACV3P+/6-31G	[53]
CYP101	N-Palmitoylglycine	21.8	-	B3LYP/ LACV3P+/6-31G	[53]

* F-BS: functional and basis set combination used for most of the QM atoms in the single point calculation.

Chapter 3

Microsecond MD simulations of human CYP2D6 wild-type and five allelic variants reveal mechanistic insights on the function

Based on the article:

Microsecond MD simulations of human CYP2D6 wild-type and five allelic variants reveal mechanistic insights on the function

Charleen G. Don & Martin Smieško

PLoS ONE 13(8): e0202534

Abstract

Characterization of cytochrome P450 2D6 (CYP2D6) and the impact of the major identified allelic variants on the activity of one of the most dominating drug-metabolising enzymes is essential to increase drug safety and avoid adverse reactions. Microsecond molecular dynamics simulations approach has been performed to capture the dynamic signatures of this complex enzyme and five allelic variants with diverse enzymatic activity. In addition to the apo simulations, three substrates (bufuralol, veliparib and tamoxifen) and two inhibitors (prinomastat and quinidine) were included to explore their influence on the structure and dynamical features of the enzyme. Our results indicate that the altered enzyme activity can be attributed to changes in the hydrogen bonding network within the active site, and local structural differences in flexibility, position and shape of the binding pocket. In particular, the increased (CYP2D6*53) or the decreased (CYP2D6*17) activity seems to be related to a change in dynamics of mainly the BC loop due to a modified hydrogen bonding network around this region. In addition, the smallest active site volume was found for CYP2D6*4 (no activity). CYP2D6*2 (normal activity) showed no major differences in dynamic behaviour compared to the wild-type.

Introduction

Genetic polymorphism in CYP2D6, a monooxygenase enzyme metabolizing around 25% of the therapeutic drugs [1], frequently leads to altered enzyme activity (increased, decreased or none) which in turn has an impact on the drug efficacy and the occurrence of adverse reactions [2]. Depending on the genetic variant, four phenotypes can be assigned: (i) ultrarapid metabolizer (UR), (ii) normal metabolizer (NM), (iii) intermediate metabolizer (IM) and (iv) poor metabolizer (PM) [3]. During the last decade, increased awareness concerning the risks that CYP2D6 polymorphism can bear on treatment outcome has lead to relabeling of several CYP2D6 metabolized drugs with additional guidelines on drug dosage in case of polymorphism by the FDA [4]. In addition, the clinical pharmacogenetics implementation consortium (CPIC) has been procuring therapeutic guidelines for several drugs that have a

high risk of adverse reactions when administered to patients with a phenotype deviating from the normal CYP2D6 activity (wild-type) [5].

The enzymatic activity can be altered in various ways. Intrinsic properties of CYP2D6 such as a localized hydrophobic region in the binding pocket (V119, F120, L121, F219, L220, R221, V370, P371, L372, V374) or the strong electrostatic field of the two carboxylates (E216 and D301) enabling binding and orienting of the ligand, are essential for its unique substrate specificity [6–8]. In addition, the ligand flux between the exterior and the buried active side is tightly regulated through the formation of tunnels [6]. Environmental conditions have been shown to contribute to the metabolic capacity of the enzyme, too. Its redox partner the cytochrome P450 reductase (CPR/POR) provides the electrons needed for the redox reaction [7]. The presence of other isoforms is also important as it has been observed that P450s (including CYP2D6) are able to form homomeric and heteromeric complexes [8]. In addition, the polarity and organization of the membrane is essential for proper anchoring and enzymatic function [9]. How all these factors exactly influence and modify CYP2D6 function on the molecular level is still poorly understood.

Using in silico methods for extending our understanding of the enzyme's essential bond-forming capabilities, channel formation and overall plasticity is burgeoning [10–17]. This trend is driven by the growing amount of structural data (the majority of available mammalian CYP2D6 x-ray structures have been released after 2014) as well as ever increasing computational power allowing to reach the millisecond time scale of the molecular dynamics (MD) simulations [18]. Such simulations are the method of choice for studying the biomolecular structural and dynamical aspects on atomic level (based on its thermodynamics and kinetics) and conveniently complement experimental investigations [19,20]. In general, MD simulation studies (varying from 5 ns to 250 ns) that focused on CYP2D6 and different variants (CYP2D6*34, CYP2D6*17-2, CYP2D6*17-3, CYP2D6*53, CYP3D6*2, CYP2D6*10, CYP2D6*14A, CYP2D6*51, CYP2D6*62) showed that the global structural fold remains similar for all [12,15,17]. However, local changes mainly found at the loop regions were demonstrated to alter the flexibility (increased/decreased) of one particular variant and they are thought to correlated with the enzyme activity [12,15,21]. Moreover, the CD-, GH-, FG-,

and BC loops displayed increased or decreased flexibility compared to the wild-type. Especially the last two loops are positioned close to several important tunnels (2a/2ac/4 and 2b/2e/2c respectively) which allows the regulation of ligand flux between the outer environment and the buried active site. If the loops become more rigid or flexible upon a particular amino acid mutation this will translate into different thermodynamics and altered enzyme activity [17]. It is known that enzymatic reactions occur on the millisecond-to-second time scale, hence the need to prolong the MD simulation studies to a longer time scale in order to capture an improved overall framework of the links between enzyme structure, movement and its catalytic action [22]. Improved information and any novel insights regarding CYP2D6 polymorphs could also potentially contribute to minimizing the interindividual differences in pharmacological and toxicological responses to a drug (e.g. altered binding mode of a drug in the binding pocket of an allelic variant) during drug discovery and translate into more focused pharmacovigilance [23–25].

Our pioneering study focuses on exploring such dynamic phenomena contributing to enzyme activity on a larger time scale (1 μs) for CYP2D6 wild-type (WT) and five allelic variants (CYP2D6*2, CYP2D6*10, CYP2D6*17, CYP2D6*4 and CYP2D6*53) (**Table 3.1**). The selection criteria of the CYP2D6 variants were procured based on (i) the functional activity and clinical relevance, for each phenotype at least one variant was selected (**Table 1.1**). At the moment, only one CYP2D6 variant with increased in vitro activity is identified. A list containing all currently identified allelic CYP2D6 variants (> 100) can be found at www.pharmvar.org. (ii) the location and overlap of the mutations within the CYP2D6 structure (**Figure 3.1**): except for CYP2D6*53 (increased activity) all the other selected variants have at least one mutation in common. Furthermore, three substrates and two inhibitors were selected in order to investigate the way they are accommodated in the active site and influence the enzyme flexibility.

The 15 MD simulations performed in this study extend our existing knowledge on the structural and functional relationship of CYP2D6 wild-type and five variants. Our results suggest that the altered enzyme activity can be attributed to both changes in hydrogen bonding network within the active site as well as local structural differences in flexibility,

position and shape of the binding pocket – in particular the loop regions (FG and BC) essential for the regulation of the ligand access to the heme.

Table 3.1 **Overview of all MD simulations performed (1 µs).**

Run ID	Description
wt_a	CYP2D6*1 (wild-type) apo enzyme (normal activity)
V2_a	CYP2D6*2 apo enzyme (normal activity)
V10_a	CYP2D6*10 apo enzyme (decreased activity)
V17_a	CYP2D6*17 apo enzyme (decreased activity)
V4_a	CYP2D6*4 apo enzyme (no activity)
V53_a	CYP2D6*53 apo enzyme (increased activity)
wt_buf	CYP2D6*1 (wild-type) enzyme run with bufuralol (substrate)
wt_vel	CYP2D6*1 (wild-type) enzyme run with veliparib (substrate)
wt_tam	CYP2D6*1 (wild-type) enzyme run with tamoxifen (substrate)
wt_pri	CYP2D6*1 (wild-type) enzyme run with prinomastat (inhibitor)
wt_qui	CYP2D6*1 (wild-type) enzyme run with quinidine (inhibitor)
V17_qui	CYP2D6*17 enzyme run with quinidine (inhibitor)
V17_pri	CYP2D6*17 enzyme run with prinomastat (inhibitor)
V53_qui	CYP2D6*53 enzyme run with quinidine (inhibitor)
V53_pri	CYP2D6*53 enzyme run with prinomastat (inhibitor)

Single Nucleotide Polymorphisms (SNPs) of the five CYP2D6 variants

Binding specificity and enzyme activity is controlled by a diversity of factors as mentioned in the introduction. The amino acid constitution of the enzyme contributes largely to its stability (e.g. intra-molecular hydrogen bonding network) and function (e.g. more hydrophobic lining channel residues regulating typical lipophilic substrate flux to and from the heme) [1]. Depending on the location of mutation (e.g. surface, substrate recognition site or active site) and its nature (e.g. hydrophobic into hydrophilic) the impact on the enzyme activity and stability will be more or less pronounced. For the eight SNPs among the CYP2D6 variants in this study (**Figures 3.1 and 3.2**), half of them caused a polarity change and three reversed hydrophobicity or hydrophilicity (**Table S3.1**). The SNP P34S in CYP2D6*4 and CYP2D6*10 is known to perturb the proline-rich PPGP motif near the N-terminus site crucial for proper folding and membrane anchoring of the enzyme [26]. It has been observed that P34S is solely responsible for decreased to almost abolished enzymatic activity [27]. Both F120I and A122S SNPs in CYP2D6*53 (increased activity) are located at the BC-loop, in close vicinity of the CYP2D6 heme-iron. Several site-directed - and molecular modeling studies have proven the relevance of Phe^{120} to CYP2D6 substrate binding, orientation and regiospecificity of CYP2D6

[28–30]. Substitution of Phe^{120} by Ile is expected to reduce the local stacking interactions (increase BC loop flexibility) and to give substrates easier access towards the iron-heme. CYP2D6*2 with R296C located at the I-helix (N-terminus side) and S486T located at the β 4-2 loop, has a similar activity compared to the wild-type [31]. Both the more conserved Ser^{486} substitution with Thr and positively charged Arg^{296} substitution with Cys seem not to have a major impact on CYP2D6 [17,32].

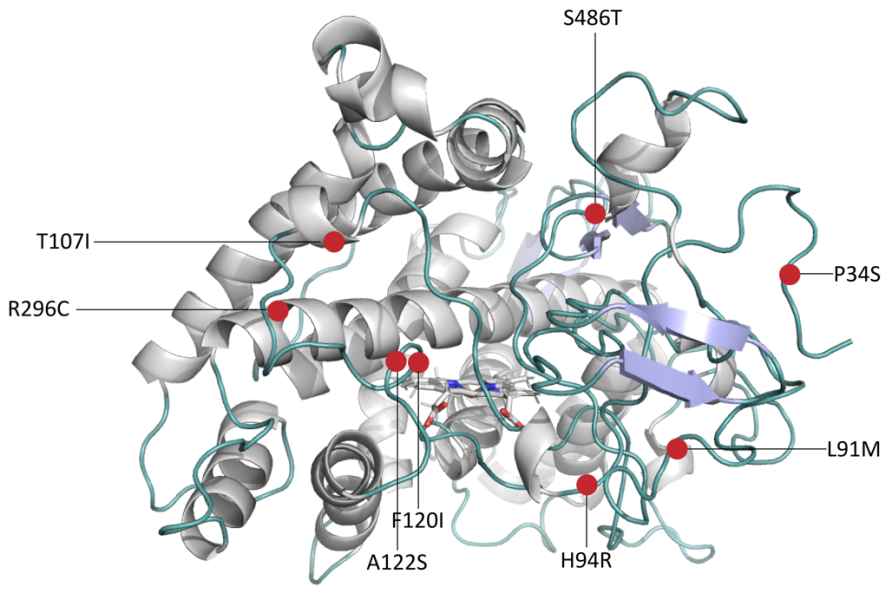


Figure 3.1 **CYP2D6 structure showing the positions of the single nucleotide polymorphisms (SNPs).** The SNPs are highlighted in red. SNPs in CYP2D6*2 are R296C and S486T, in CYP2D6*10 are P34S, S486T in CYP2D6*17 are T107I, R296C and S486T, in CYP2D6*4 are P34S, L91M, H94R and S486T, in CYP2D6*53 F120I and A122S (see overview **Table 3.2 and Figure 3.2**).

These two SNPs are also found in CYP2D6*17 (decreased activity) in addition to the SNP T107I. The significant role of the latter residue has been demonstrated by experimental research in which decreased enzymatic activity was observed with only the SNP T107I [33]. The hydrogen bond forming residue Thr^{107} is located at the B' helix in the center of the BC loop. It can be assumed that substitution with Ile will increase local hydrophobicity and interactions, which in turn will stabilize the structure and reduce the flexibility of the BC loop,

potentially resulting in a decreased enzyme activity. Among Caucasians, the average identified allelic frequencies are 38%, 18%, 4%, 2%, 0.3%, for the wild-type (CYP2D6*1), CYP2D6*4, CYP2D6*2, CYP2D6*10, and CYP2D6*17 respectively [34]. The global distribution regarding the allelic frequency of CYP2D6*53 is not yet available.

Table 3.2 **Overview of the CYP2D6 allelic variants together with information regarding their relevance and impact on the enzyme.**

CYP2D6 Allele	Sequence modification [A]	CYP2D6 activity	Nearest SRS[B]	Allelic frequency [34] (%)[C]:
Normal Metabolizer (NM)				
*1 (wild-type)	None	Normal		Americans (50%), Caucasians (38%)
*2	R296C, S486T	Normal	6	Asians (central, 30%), Caucasians (4%)
Intermediate Metabolizer (IM)				
*10	P34S, S486T	Decreased	5/6	Asians (east, 43%), Caucasians (2%)
*17	T107I, R296C, S486T	Decreased	1/3/4/6	Africans (19%), Caucasians (0.3%)
Poor Metabolizer (PM)				
*4	P34S, L91M, H94R, S486T	Inactive	5/6	Caucasians (18%)
Ultrarapid Metabolizer (UM)				
*53	F120I, A122S	Increased	1	Global frequency data incomplete

[A] data obtained www.pharmvar.org accessed on 20.12.2017, [B] Substrate Recognition Site (SRS), [c] the population identified to have the highest occurrence regarding the CYP2D6 allele is indicated first, followed by the occurrence found for Caucasians.

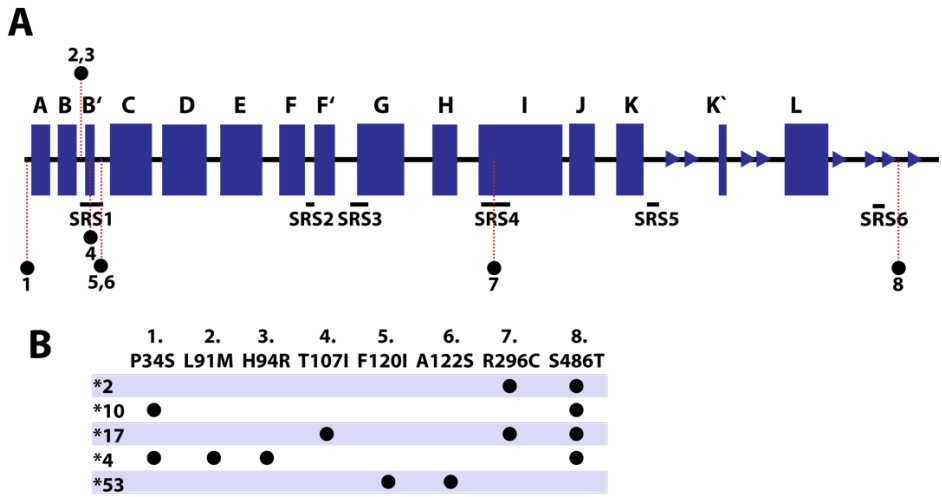


Figure 3.2. **Overview of the CYP2D6 sequence, the secondary elements and the amino acid mutations for the five variants.** (**A**) The CYP2D6 sequence is shown together with the location of the substrate recognition site (SRS1-SRS6) and the location of the mutations. (B) The table displays the CYP2D6 allelic variants and their mutations.

Materials and methods

PDB selection

It has been observed that the presence of a ligand in the binding pocket induces conformational changes mainly at helices A, B, F, G, the first β–sheet, and loops BC, AB, and FG. The distance between the later three loops and the position of their connecting helices determine a more closed/open protein conformation (**Figure 3.3**) [21,35].

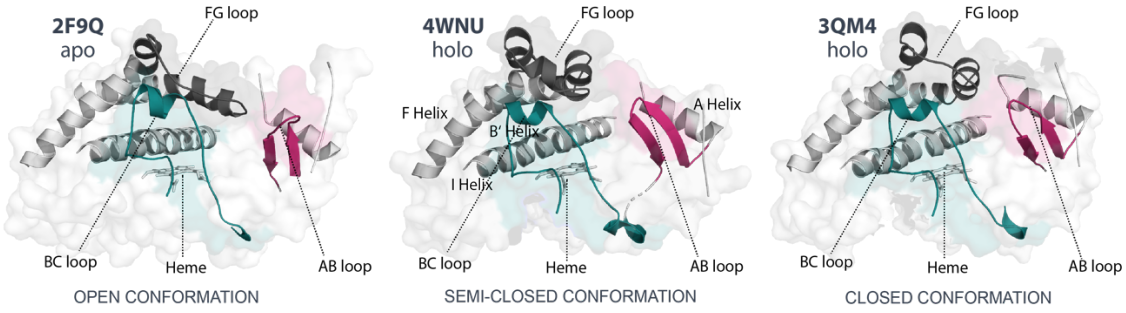


Figure 3.3 **CYP2D6 x-ray structures illustrating the conformational diversity (open/closed).** The apo structure 2F9Q (left) shows a more open conformation as defined by the distances between the FG- (grey), BC- (cyan) and AB- (pink) loops. The holo 3QM4 complex (with prinomastat, right) represents one of the most closed CYP2D6 conformations, whereas the holo 4WNU complex (with quinidine, middle) a semi-closed conformation.

A start conformation with an overall fold that is semi-closed was preferred for this study and the CYP2D6 quinidine complex 4WNU was selected [36]. As with all available mammalian CYP2D6 x-ray structures, part of the N-terminus (till G31) was truncated to increase its solubility and facilitate the crystal growing process. Furthermore, residues 38-52, 145-147 and 498-501 all located outside the active side, were missing in chain A. In our model the missing residues were filled in using Prime [37]. The iron-heme was modelled as Fe^{3+}. This oxidation state corresponds with the active catalytic ferrous state [38]. The Schrodinger Protein Preparation Wizard was used for pre-processing of the x-ray structure [39]. Cofactors (5 molecules) that were within the binding pocket were replaced (at their oxygen atom positions) with water molecules: glycerol (replaced with 3 waters) and DMSO (replaced with 2 waters). To optimize the geometry of residues P41 and G42 in the x-ray structure, an additional loop refinement for residues 31-58 was performed using Prime. The resulting

structure was minimized using a hybrid method of the steepest decent and the limited memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithms with a convergence threshold of 0.3 kcal mol^{-1} $Å^{-1}$. Subsequently, this structure was used for generation of all the CYP2D6 variants. After the applied mutations to the structure, each new protein structure was minimized again with the same parameter settings as for the wild-type.

Ligands

The ligand selection was based on several criteria: (i) classical CYP2D6 binder; (ii) if there was data available for the ligand from previous computational/experimental studies; (iii) clinical relevance (**Figure 3.4**). Bufuralol (β–blocker) and quinidine (antiarrhythmic drug) were selected as classical CYP2D6 substrate and inhibitor respectively. Next, substrates, tamoxifen (estrogen receptor antagonist) and veliparib (PARP inhibitor) were included, as any novel (polymorphism) information regarding their binding behaviour would be relevant from a clinical perspective and potentially contribute to the future of personalized medicine/anti-cancer drug development. In addition, for quinidine and prinomastat (matrix metalloproteinase inhibitor) a crystal structure is available which allows comparing of the dynamical and docking simulations with their observed native pose (**Table S3.3**) [40]. CYP2D6 metabolizes bufuralol into 1'-hydroxybufuralol [41], tamoxifen into 4-hydroxyltamoxifen [42], and veliparib into a lactam containing metabolite [43]. Although quinidine contains the usual CYP2D6 substrate characteristics (basic nitrogen, and flat hydrophobic moiety) it is first metabolized into 3-hydroxy- and O-demethylated quinidine when Phe^{120} is mutated to alanine (or also by E216Q/D301Q) [1,44]. Therefore, the F120I mutation can be expected to allow quinidine to be metabolized. The sites of metabolism (SoMs) for studied ligands are indicated in **Figure 3.4** and their distance to the heme in Supplementary Figure **S3.5**. As a note, the enzyme kinetics is a complex problematic where one needs to consider numerous possibilities on how a molecule can bind the enzyme. For example, the metabolite of a substrate can act as an inhibitor or, if a substrate has a slow off-rate, one could argue, if it thereby also acts as inhibitor. However, these effects are outside of the scope of this study. A detailed discussion on enzyme binding, the associated terms

and kinetics can be found in the literature [41,42]. The binding pose of the ligands tamoxifen, veliparib and bufuralol within the CYP2D6 pocket used for MD simulations was generated using our in-house docking software (DOLINA) [45]. The generation of the ligand poses is based on pharmacophore matching and allows for local induced-fit changes by combinatorial rearrangement of the binding site side chains. The highest ranked pose was selected as a start conformation for MD simulations. For both prinomastat and quinidine the conformation was adapted from the corresponding CYP2D6 x-ray complex (PDB IDs 3QM4, 4WNU).

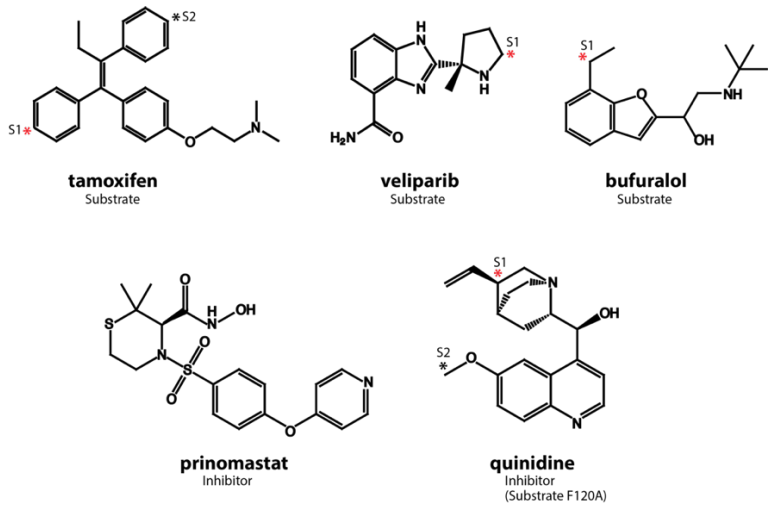


Figure 3.4 **Overview of the CYP2D6 ligands. Three substrates (tamoxifen, veliparib, and bufuralol) and two inhibitors (prinomastat and quinidine) were used in this study**. The red asterisk indicates the primary site of metabolism (SoM, **Figure S4**), and the black asterisk a secondary SoM. Quinidine might be metabolized in CYP2D6*53, hence the known SoMs are indicated as well.

Molecular Dynamics simulations

Each prepared CYP2D6 variant was solvated using TIP3P water model in an orthorhombic box with 10.0 Å cut-offs from the protein in each dimension, and the net system charge was neutralized by adding counterions (sodium ions). The OPLS_2005 force field as implemented in Desmond (version 2016-4) was used [46]. The system was minimized using a steepest decent algorithm until a gradient threshold of 0.1 kcal mol^{-1} $Å^{-1}$. The dimension of the box for each prepared CYP2D6 variant system (after minimization) were on average 86 x 82 x 87 $Å^3$. The production simulations with the total duration of 1.0 µs (NPT ensemble and standard

conditions T = 300 K, p = 101.325 kPa) and with frames sampled every 100 ps (in total 10 000 frames were saved per simulation) were performed. The simulations were performed under NPT ensemble, and the Nose-Hoover thermostat at a relaxation time of 1.0 ps with the Martyna-Tobias-Klein barostat were combined with a relaxation time of 2.0 ps at 300 K. The Particle Mesh Ewald (PME) method was used to treat the long-range interactions, the cutoff for short-range interactions was set to 9 Å. Bonds to hydrogen atoms were constrained with the M-SHAKE algorithm and no hydrogen mass partitioning was applied. The two replica simulations were started with different initial velocities. On average one microsecond simulation took 10 days to finish (61195 atoms) on one GeForce GTX Titan GPU.

Results and Discussion

Overall fold and system equilibration

The root mean square deviation (RMSD) measured over the whole trajectory (1 μs) shows that the wild-type (CYP2D6*1) and all other CYP2D6 variant simulations reached equilibrium at or below the RMSD of 6 Å from the starting structure (**Figure 3.5**). **Figure S3.2** displays only the last 100 ns and **Figure S3.3** an additional 400 ns of simulation time (1,4 μs simulation in total, only ran for the wild-type apo structure) confirming that all systems were properly equilibrated. In addition, two replica simulations were performed with the apo wild-type using different starting velocities to assure the integrity of the system (**Table S3.4**). The RMSD and RMSF values of the two replica were calculated and compared to the wild-type showing that the intrinsic properties of the system remained conserved in the replicas. Time to convergence varied between 20 ns ≤ t ≤ 250 ns. Fast convergence (t ≤ 50 ns) is observed for most of the wild-type simulations (wt_apo, wt_pri, wt_vel, wt_buf) as well as for CYP2D6*2 (V2_a). The dynamic global and local fluctuations of CYP2D6*2 (mutations R296C and S486T) are expected to be similar to the wild-type as they share the same enzyme activity. All other apo CYP2D6 variants (V10_a, V17_a, V4_a and V53_a) needed longer time (100 ns ≤ t ≤ 250 ns) before reaching equilibrium. This is likely related to the mutations applied, causing local structural instabilities requiring longer times to converge (equilibrate). In general, simulations

run with a ligand reached plateau in a short time (t ≤ 50 ns). Wild-type simulations run with quinidine (wt_qui) and tamoxifen (wt_tam) needed longer time (100 ns and 250 ns respectively) to reach equilibrium.

From the apo state simulations, CYP2D6*53 showed the largest conformational change followed by CYP2D6*17, whereas CYP2D6*2, CYP2D6*10 and CYP2D6*4 resembled the dynamics of the wild-type. Furthermore, except for CYP2D6*53 with quinidine (V53_qui), the simulations including the inhibitors prinomastat and quinidine (wt_qui, wt_pri, V17_qui, V17_pri and V53_pri), and the substrate veliparib (wt_vel) led to a more consistent protein conformation (RMSD curves 1 Å to 2 Å bellow the wild-type apo). CYP2D6*53 with quinidine (V53_qui) shows more conformational transitions (higher RMSD values) compared to prinomastat (V53_pri). In the CYP2D6*17 simulations, the presence of both inhibitors (V17_qui, V17_pri) showed a similar conformational stability.

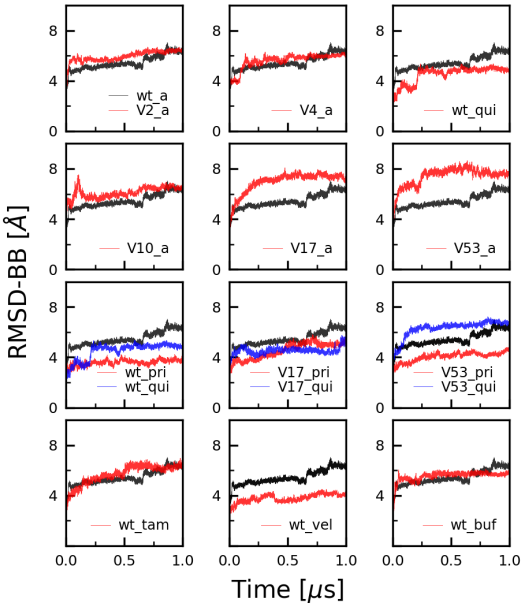


Figure 3.5 **Backbone root mean square deviation (RMSD) graphs for all CYP2D6 simulations.** The wild-type simulation is shown black in all graphs. The convergence time varied between 20 ns ≤ t ≤ 250 ns. Compared to the wild-type apo simulations; low degree of fluctuation was observed for wild-type veliparib (wt_vel), quinidine (wt_qui), and prinomastat (wt_pri) and the most flexible variant was CYP2D6*53 apo (V53_a) followed by CYP2D6*17 apo (V17_a).

For all apo simulations the start conformation and the most populated (representative) cluster conformation during the 1 μs trajectory can be found in **Figure 3.6**.

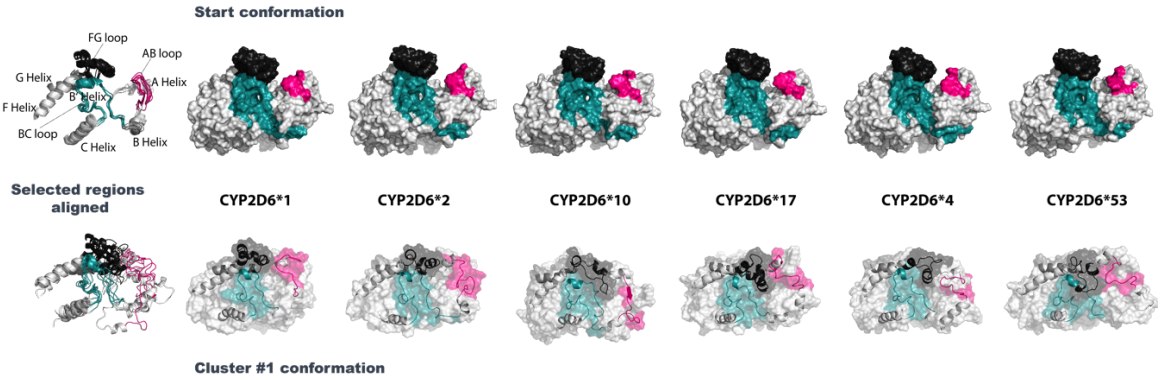


Figure 3.6 **The CYP2D6 starting - and most representative conformation during MD simulations for CYP2D6*1 (wild-type) and the five variants.** Structural elements relevant for defining a more open or closed conformation are the regions of the BC loop colored cyan, AB loop colored pink, and FG loop colored black. (**Top**) For all apo CYP2D6 variants, the surface of the start conformations is shown at the beginning (t = 0 µs) of the simulation. All CYP2D6 variants displayed a similar start conformation with an average distance of 13 Å between the FG loop region and AB loop region. (**Bottom**) For all apo CYP2D6 variants, the most populated conformation (obtained by the RMSD clustering) during the simulation are displayed, revealing a semi-closed fold compared to the start conformation. The surface (transparent) is overlaid with the cartoon representation of the selected regions. The alignment in the first image shows that the FG, BC and AB loops are at different positions in the most prevalent conformation for all the CYP2D6 variants.

Overall RMSD values between the CYP2D6 variant apo start conformation and the most populated cluster conformation (cluster #1, occurrence > 60%) calculated for all CYP2D6 variant simulations (backbone carbon atom alignment) varied between 2.2 Å and 3.3 Å. CYP2D6*17 featured the highest ΔRMSD (3.3 Å) whereas CYP2D6*17 with quinidine the lowest (2.2 Å) ΔRMSD was seen. Though the start conformations are very similar to each other (ΔRMSD < 1.3 Å) the clustered conformations are locally more diverse and display a semi-closed fold (ΔRMSD > 3.3 Å). The active site volume increased upon reaching equilibrium, mainly attributed to a rearrangement of the FG-, BC-, and AB- loops and the connecting helices (F, G, I A, and B) as determined in other studies as well (**Table S3.2**) [14]. Compared to the apo wild-type, CYP2D6*2, CYP2D6*10, CYP2D6*17 and CYP2D6*53 were found to display a more confined fold of the active site, whereas CYP2D6*4 showed a different arrangement mainly of the FG loop and AB loop (more towards the outside) leading to a less confined cavity (more open substrate access state) and a lower volume (8% decrease compared to wild-type apo). The volumes for the wild-type holo conformations varied largely

compared to the apo wild-type (from -11% to 26%) suggesting not all ligands possess equal capabilities to favorably engage with the active site and stabilize the protein structure upon binding. The F120I mutation in the active site of CYP2D6*53 contributed to larger observed volumes compared to all others (on average +35%).

Protein structure dynamics (e.g. its folding and overall stability) is known to have an impact on its function [47]. The enzymatic catalysis of CYP2D6 can be indirectly influenced by the distribution of the charged residues at the solvent accessible surface (by for example influencing the overall stability of the protein or the ligand access or egress through long-range electrostatics) and the non-polar solvent accessible surface (also known as buried solvent area, BSA). Therefore, the differences in solvent accessible surface area (SASA) (polar/apolar) and BSA were investigated for the wild-type and allelic variants to explore if there were any patterns observed that could be correlated with the altered activity of the variants (**Table S3.2**). The ratios within each variant between total surface area, SASA (polar/apolar) and BSA seemed to be consistent within the variants. In general, the SASA was found to vary between -2% and +5%, with one outlier: CYP2D6*53 apo had an increased SASA of 8% compared to the wild-type. An explanation for this larger value might be related to the fact that increased enzyme activity is also achieved by an enlargement of the access tunnel(s) connecting the surface with the active site. At the same time, the catalytic efficiency is retained by optimal alignment of the active site with reduced steric hindrance of F120I, allowing to offer larger active site volumes and a higher through-put of substrates compared to the wild-type. Within the apo variants, only CYP2D6*4 showed a remarkable decrease (13%). The BSA analysis of the holo conformations showed very large volume changes, varying between -38% (wt_tam) and +70% (V17_pri), which makes it difficult to assign any meaning to the observed BSA values other than that enzyme active site is very plastic and can accommodate various ligand shapes and adjust for an optimal interaction. However, considering the known differences among CYP families in charged surface areas which is linked to the binding of CPR, further studies focused on surface properties might prove valuable for deciphering enzyme activity and allostery [48–50].

Local flexibility differences

To gain insight in the local structural differences among the CYP2D6 variant simulations, the root mean square fluctuation (RMSF) calculations were performed excluding first 250 ns of the overall trajectory, (the average determined equilibration time). The values were normalized and plotted by subtracting the wild-type simulation values from the CYP2D6 variant simulation. Fluctuations around the N-terminus should be interpreted with caution, since this is the site where CYP2D6 is normally anchored to the membrane which would normally stabilize and reduce the local flexibility. Since the simulations are performed without the membrane, higher RMSF fluctuations are expected for this region.

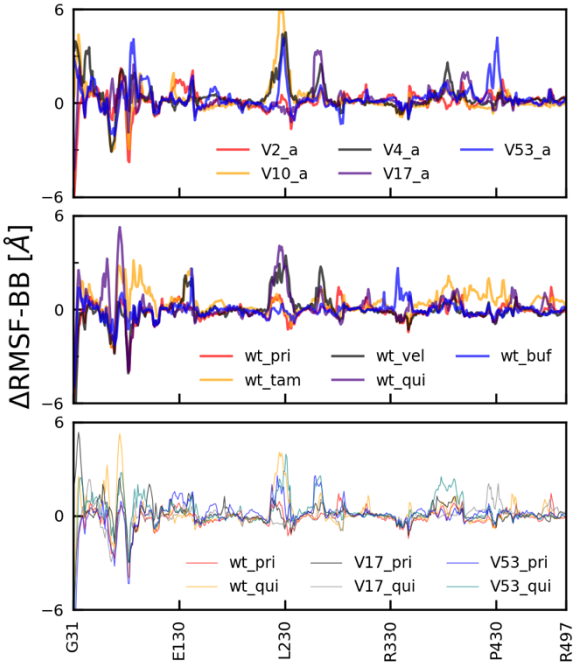


Figure 3.7 **Root mean square fluctuation (RMSF) graphs for all CYP2D6 simulations, excluding the average equilibration time (first 250 ns)**. In the RMSF graphs, the curves were normalized by subtracting the wild-type RMSF values from the CYP2D6 variant simulation. The grey areas indicate the position of the loop regions. The Largest RMSF fluctuations were found at the B helix and around the loop regions, especially AB, FG, GH and KL loops.

For the apo variant simulations (**Figure 3.7**, first row), all loop regions showed larger fluctuations compared to the wild-type, in particular AB, FG, GH and KL loops. In addition, CYP2D6*4, CYP2D6*17 and CYP2D6*53 displayed a larger peak (≥ 4 Å) at the β–2-1, β–2-2

(within the KL loop). CYP2D6*53 displayed the highest flexibility for the KL loop region (P430, E431 and A432). The position of the B helix fluctuated the most for all the CYP2D6 variants, which is likely too high due to the absence of the membrane normally surrounding and stabilizing this site.

Similar flexibility patterns were found for the wild-type simulations ran with ligands (**Figure 3.7**, second row). However, the ΔRMSF peaks were in general less intensive which suggests that the presence of a ligand leads to a more stable complex which is in agreement with other studies [12,15]. From all the wild-type holo simulations, quinindine (inhibitor, wt_qui) and veliparib (substrate, qui_vel) induced the largest ΔRMSF peaks (± 3 Å) around the FG and GF loop region, and only bufuralol (wt_buf) showed a pronounced peak (3 Å, D349) located at the J' helix.

The inhibitor CYP2D6 simulation series (**Figure 3.7**, third row) showed similar flexibility RMSF patterns and intensities compared to the wild-type holo simulations. CYP2D6*17 with quinidine (v17_qui) showed a peak close to the N-terminus side of the L-helix, as observed for wild-type with tamoxifen (wt_tam). In addition, CYP2D6*53 with quinidine displayed higher flexibility around the β 2-1, β 2-2 region (within the KL loop). Considering the important function of the FG- and BC loop in determining the fold and access towards the iron-heme, in-depth analysis for these two loops were performed separately (**Figure S3.1**) and are commented on bellow.

Intra-molecular hydrogen bonding network analysis

Hydrogen bonding analysis for the mutated amino acids was performed to deduce any significant structural changes that could affect the enzyme stability for the CYP2D6 variants compared to the wild-type (**Table 3.3**). Default Maestro hydrogen bonding criteria were applied (maximal distance from the hydrogen atom to the acceptor atom: 2.8 Å, minimum donor hydrogen bonding angle: 120°, and minimum acceptor Hydrogen bonding angle: 90°). For each variant the hydrogen bonding partners of the mutated residues were identified and compared with the wild-type. The two SNPs found in CYP2D6*2 (R296C, S486T) showed no significant impact on the hydrogen bonding interactions.

Table 3.3 **Hydrogen bonding network analysis of CYP2D6 variants, focus on mutated amino acids.**

CYP2D6 Variant activity (mutations)	Interaction(s) Occurrence (%)		
CYP2D6*2 similar (R296C, S486T)			
wt	R296	+ D252	65%
		+ A300	71%
CYP2D6*2	C296	+ A300	93%
wt	S486	+ V480	100%
		+ L484	29%
CYP2D6*2	T486	+ V480	100%
CYP2D6*4 none (P34S, L91R, H94R, S486T)			
wt	P34	+ none	
CYP2D6*4	S34	+ G36	31%
		+ W75	27%
wt	L91	+ R88	36%
CYP2D6*4	M91	+V87	74%
wt	H94	+ G44	34%
		+ R440	72%
CYP2D6*4	R94	+ A90	30%
		+ E383	33%
wt	S486	+ V480	100%
		+ L484	29%
CYP2D6*4	T486	+ V480	100%
CYP2D6*53 increased (F120I, A122S)			
wt	F120	+ R101	92%
		+ D301	100%
CYP2D6*53	I120	+ D301	98%
wt	A122	+ R441	100%
CYP2D6*53	S122	+ none	

CYP2D6 Variant activity (mutations)	Interaction(s) Occurrence (%)		
CYP2D6*10 decreased (P34S, S486T)			
wt	P34	+ none	
CYP2D6*10	S34	+ none	
wt	S486	+ V480	100%
		+ L484	29%
CYP2D6*10	T486	+ V480	100%
		+ I312	62%
CYP2D6*17 decreased (T107I, R296C, S486T)			
wt	T107	+ N255	58%
CYP2D6*17	I107	+ F122	86%
		+ L110	38%
wt	R296	+ D252	65%
		+ A300	71%
CYYP2D6*17	C296	+ A300	83%
wt	S486	+ V480	100%
		+ L484	29%
CYP2D6*17	T486	+ V480	100%

The two major backbone – backbone (BB – BB) interactions observed most of the time (> 70%) for the wild-type (Arg^{296} with Ala^{300} and Ser^{486} with Val^{480}) were also observed for CYP2D6*2. Indeed, Cys^{296} with Ala^{300} and Thr^{486} with Val^{480} were also identified for CYP2D6*17 (R296C, S486T, T107I) the majority of the time (> 70%). Our observation is that the latter interaction facilitates in stabilization of the β–sheet (β 4-1, β 4-2). The wild-type hydrogen bonding interaction of Thr^{107} with Asn^{255} (G-helix, side-chain(SC)-SC, 58%) was altered for CYP2D6*17. The mutated Ile^{107}, located at the BC loop, interacted most of the time with Phe^{112} (86%) and for a smaller time period with Leu^{110} (38%). This could be an indication that in this variant the BC loop is less flexible due to the extended hydrophobic

network compared to the wild-type and indirectly contributes to a more flexible FG loop due to the missing N255 bond (as supported by the FG – and BC loop RMSD analysis, see **Figure S3.1**). Based on this data, we propose that the substrate access through the closest tunnels (2c/2e) is hindered in this variant, which might translate to the decreased activity of the enzyme. Handa et al. performed similar analysis for wild-type and CYP2D6*17, though only over a 5 ns MD trajectory [17]. For the T107I mutation different hydrogen bonding interactions were observed for the wild-type (L110, G11, F112 and V104) and CYP2D6*17 (G111 and F112). For the R296C mutation the wild-type interacted with D252 which is in agreement with our observation, though the A300 interaction is missing in both cases. S486T formed for both wild-type and CYP2D6*17 a hydrogen bond with V480 as seen in our simulations. The difference in results can be attributed to a rather short simulation time of the previous modeling study.

The decreased activity of CYP2D6*10 (P34S, S486T) is mainly caused by the SNP P34S as explained in our SNP section. Though as in the wild-type no hydrogen bonding interactions were observed, it is known that this proline-rich motif (PPGP) is crucial for protein folding [26]. In addition, the hydrogen bonding results for this residue should be treated with caution since our simulations lack the membrane, which would normally stabilize the N-terminal part of the enzyme (including Pro^{34}). This is also valid for the P34S mutation found in CYP2D6*4 (P34S, L91M, H94R, S486T).

An interesting observation for CYP2D6*4 was found for the SNP H94R located at the BC loop. Earlier studies indicate that a conserved arginine found in P450s is acting as a gatekeeper of the water tunnel; if interacting with the heme propionates, water molecules are prevented from accessing the active site from the surface [51,52]. Proper regulation of water molecules supply at the active site is crucial for efficient enzyme functioning. In CYP2D6 there are two conserved arginines (R440, R441). In the wild-type His^{94} interacts the majority of the time with Arg^{440} (72%, BB-SC), whereas the mutated Arg^{94} interacted with Glu^{383} and Ala^{90} for around 30% of the time. No compensating hydrogen bonding partners were identified for Arg^{440} in this variant, meaning the side-chain is free to move around and might thereby interfere with the normal functioning of this water channel. Another MD

simulation study of 250 ns duration suggested that the hydrogen bonding interactions of the protein with the heme also largely contribute to proper embedding of the heme group within the enzyme [12]. The loss of interactions between the heme and the enzyme (including Arg^{441}) likely worsens the heme recognition and incorporation which in turn might lead to no enzyme activity as observed by experiment for CYP2D6*62 [53]. However, analysis of Arg^{441} of both wild-type and CYP2D6*4 showed similar hydrogen bonding interaction profiles (with V119, A122, and the heme porphyrin ring of the heme, > 60%). In addition, no weakened interactions between the heme and the protein were found for CYP2D6*4 and the wild-type. Therefore, we suggest that the absence of activity for this variant could be related to both (i) the the disruption of the proline rich motif by the SNP P34S preventing optimal enzyme folding and interaction with the membrane, and in addition (ii) the SNP H94R might decrease efficient regulation of the water tunnel by its large positively charged side-chain in such a way hat it decreases enzyme activity.

For the SNP A122S in CYP2D6*53 (F120I, A122S) the Ala^{122} interaction with Arg^{441} found in the wild-type (100%) was lacking for Ser^{122}. No compensating protein hydrogen bonding interactions were identified for Arg^{441}. Phe^{120} interacted with Asp^{301} and Arg^{101} in the wild-type (> 92%). Though the latter was lacking for Ile^{120}, the more important interaction with Asp^{301}, known to be essential for the specificity of CYP2D6, was maintained [54]. RMSD analysis of the BC loop showed a 1.5 Å increase in flexibility on average for the BC loop compared to the wild-type (**Figure S3.1**). We would therefore propose that the increase in enzyme activity for CYP2D6*53 is potentially related to the loss of the otherwise BC-loop-stabilizing hydrogen bonding interactions observed in the wild-type.

CYP2D6 Ligand Analysis

RMSD calculations were performed for each ligand with the starting coordinates being used as the reference. Most stable dynamics were found for the wild-type with veliparib, quinidine and bufuralol (wt_vel, wt_qui, wt_buf) (**Figure S3.4**). Prinomastat showed the lowest stability among all wild-type holo simulations. Comparison of variants CYP2D6*17 and CYP2D6*53 revealed that the latter induced the largest change from the initial conformation (V53_pri,

V53_qui). The spike for prinomastat in CYP2D6*53 around 0.3 µs is due to a shift of the pyridyl ring, which later on is followed by a movement (after 0.5 µs) of the inhibitor closer towards the heme (6 Å). In the wild-type and CYP2D6*17 this displacement of prinomastat did not occur. Instead, during the whole simulation time it remained close to the initial pose with small fluctuations of the pyridyl ring. Despite the reduced steric hindrance by the F120I substitution, quinidine was not found to move closer towards the heme during simulation with CYP2D6*53. In all three simulations (wt_qui, V53_qui, V17_qui) it kept its position constant. In the following sections, the functional groups or atoms described in the parenthesis always refer to the ligand.

Prinomastat

The time-averaged binding mode of prinomastat (inhibitor) in a similar binding mode as observed in the x-ray structure for the wild-type, CYP2D6*17 (3.6 Å) and CYP2D6*53 (**Figure 3.8, A-C**). Dominant hydrogen bonding interactions with D301 (hydroxyl group of the hydroxamic acid), E216 and Q244 (with one of the sulfonyl oxygens) were most of the time (> 70%) present in all three cases (wt_pri, V17_pri, V53_pri). Other hydrogen bonding interactions observed for the wild-type for a shorter period of time (25%-50%) included G212, S217, I369, and A482. For CYP2D6*17 and CYP2D6*53 an additional hydrogen bonding interaction was observed with R221 (with oxygen of the hydroxamic acid). During the simulation, the pyridyl nitrogen which is known to be coordinated towards the heme iron for its inhibitor effect [55], was found more often for CYP2D6*17 and CYP2D6*53 in close vicinity (6 Å or 7 Å on average) of the heme than for the wild-type (11 Å on average). Stabilizing hydrophobic interactions included L110, F112, F120, L213, F247, A300, V308, F483, and L484 observed during all three simulations.

Quinidine

The quinuclidine moiety of quinidine was pointing in all three cases (wt, V17, V53) towards the heme and the quinolone ring into the direction of the F- and G helices (**Figure 3.8, D-F**). Hydrogen bonding interactions were observed in all three cases with E216 (with the

protonated nitrogen of the quinuclidine moiety and the hydroxyl hydrogen), Q244 (with the hydroxyl oxygen and the ether oxygen), S304 (with the protonated nitrogen of the quinuclidine moiety) and G212 (with the hydrogen of the hydroxyl group). Only for the wild-type and CYP2D6*17 a hydrogen bonding interaction was observed between D301 and the protonated nitrogen. Hydrophobic ligand protein contacts were observed for all three cases with I106, L110, F112, F12, L121, L213, F247, L248, I297, A300, V308, and F483. The SNP F120I did not force the SoM of quinidine to move closer to the heme in order to form the 3-hydroxyquinidine metabolite as known for the F120A mutation (**Figures 3.8F and S3.5**) [44]. Substitution of Phe^{120} with Ile^{120} is likely to form a similar steric hindrance energy barrier considering the similarity in size and hydrophobicity of the two.

Bufuralol, tamoxifen and veliparib

The three substrate wild-type simulations all pointed the SoM in direction of the heme during the whole simulation time (**Figure 3.8, G-I**). Bufuralol and tamoxifen had an average SoM to heme distance of 6.0 Å and 4.4 Å respectively, whereas veliparib was kept more distant (10.4 Å) (**Figure S3.5**). Hydrogen bonds between E216 (protonated nitrogen and the hydrogen of the hydroxyl group) and between the oxygen of the hydroxyl group and Q244, D301 and S304 were observed for bufuralol.

For tamoxifen only two hydrogen bonds were observed: D301 (protonated nitrogen) and S304 (with the oxygen). Veliparib formed hydrogen bonds with G212 (hydrogens of the amine group near the oxygen), E216 (hydrogens of the protonated nitrogen), R221 (nitrogen of the imidazole), Q244 (hydrogen of the amine group near the oxygen), D301 (hydrogens of the protonated nitrogen) S304 and E215 (both with the hydrogens of the nitrogen in the imidazole). Furthermore, L110, F112, F120, L121, L213, F247, A305, V308, and F483 were the major stabilizing hydrophobic interactions observed for all three substrates.

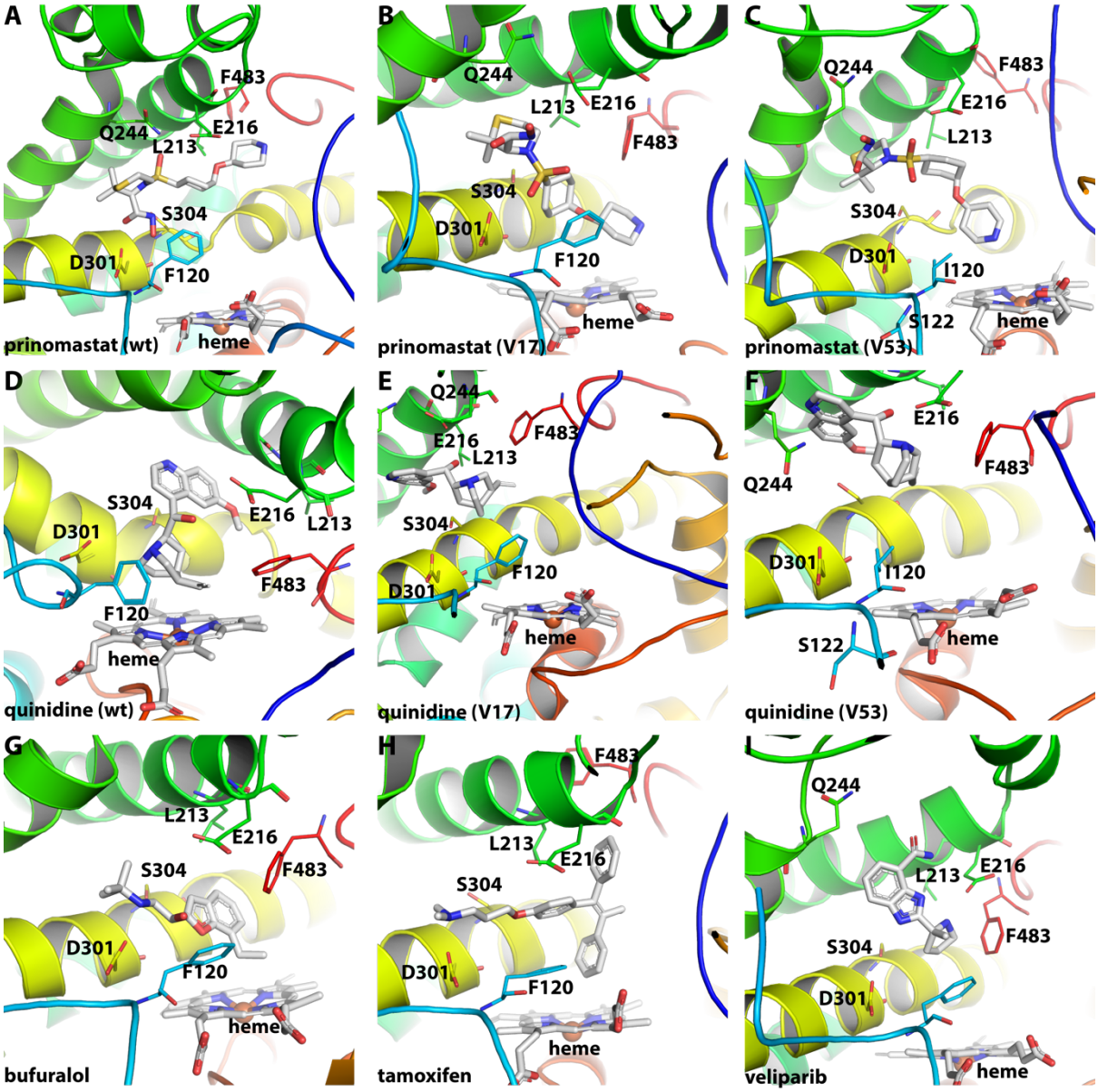


Figure 3.8 **Dominant ligand conformations during 1 µs MD simulation.** (**A-C**) Prinomastat (inhibitor) binding mode is displayed for wild-type (**A**), CYP2D6*17 (**B**) and CYP2D6*53 (**C**). (**D-F**) quinidine (inhibitor) binding mode is displayed for wild-type (**D**), CYP2D6*17 (**E**) and CYP2D6*53 (**F**). (**G-I**) The wild-type binding mode is shown for the substrates bufuralol (**G**), tamoxifen (**H**) and veliparib (**I**).

Crystal structure comparison

To assess the dynamic behaviour of six key active site residues (F120, E216, Q244, D301, S304 and F483) in the presence of different ligands and among the different variants, the most prevalent conformation was aligned with the x-ray structure (PDB ids: 4WNU or 3QM4) (**Figure S3.5**).

The side-chain of F120 and S304 did show variation in the torsion angles. A reversed positioning of the S304 side chain was observed compared to the x-ray structure for 5 of the 15 conformations (wt_apo, V17_apo, wt_pri, wt_tam, V53_qui). E216 showed a small displacement in most conformations. Largest displacements among all conformations were found for F483 and Q244. Both residues are located at more flexible regions of the CYP2D6. As previous research indicated, F483 (located at the β 4-2 sheet and SRS 6) is known to fulfil an important role in the binding of the ligand, and also to be close to the solvent channel [10,15,56]. Experiment showed that the F483I mutant was able to metabolize testosterone, whereas wild-type was not [56]. Prinomastat (wt_pri, V53_pri) and veliparib (wt_vel) both show F483 to be moved further away compared to the x-ray position. These two ligands are considerably large and contain several bulky groups. We propose that the binding energetics (unfavourable interactions) will induce F483 to move away and the ligand size will have a determining role in the location of F483 within the active site. The flexibility and displacement of Q244 (located at the FG loop or extended G helix and SRS 3) is coupled to the flexibility of the FG loop. For instances, the wild-type simulation with bufuralol and prinomastat showed overall a more stable behaviour of the enzyme with no extreme fluctuations of the FG loop (**Figure 3.7**), which is reflected in the similar position compared to the x-ray structure. On the other hand, the wild-type with veliparib showed larger fluctuations of the FG loop which in turn led to larger displacement of Q244.

Conclusion

This long time-scale (1 microsecond) MD simulation study confirmed earlier observed CYP2D6 dynamics and provided additional insights on the structure-function relationship of the enzyme. The wild-type MD simulations showed a stabilizing effect of the ligand on the structure. The folding of primarily the FG loop and secondary the AB- and BC loop around the active site is reshaped upon the ligand binding, which contributed also to the larger identified volumes and a more closed (semi-closed) state. Differences in flexibility and arrangement of the key residues lining the active site, together with the intra- and intermolecular forces among the variants with changed enzymatic activity, suggest the need

for a precise positioning of these factors to control optimal proceeding of the catalytic reaction, which is tightly coupled to the kinetics of the enzyme. The location (e.g. at the active site, or SRS) and type (conservation) of amino acid mutation appears to be relevant for maintaining a functional structural fold as well as for the regulation mechanisms (e.g. hydrogen binding network) the enzyme employs to steer the binding of ligands and cofactors. Hence, simulating the enzyme dynamics on a long time scale in the presence of explicit solvent is important for a proper understanding of the activity of the enzyme under various conditions (e.g. substrate, inhibitor, polymorphs, etc.). Such mechanistic information is of particular relevance for the drug development process as it can be directly utilized within the design of drugs in order to rationally avoid or at least limit the cytochrome liability. The observed differences among wild-type and clinically relevant allelic variants justify the need of their detailed screening using in silico approaches based on docking and MD simulations. In addition, considering the importance of the thermodynamics of the catalytic reaction, additional polymorphism studies focused on determining the free-energy barrier changes would be valuable to improve the link between differences in observed dynamical behaviour and enzyme activity. Undoubtedly, an important limitation of this study is the missing anchor part. Since CYP2D6 is anchored to the membrane at the N-terminus site, we expect less fluctuations and possibly slightly altered arrangement of the domains in the close vicinity of the membrane bilayer. We assume that this could have impact also on the tunnel structure and function. Therefore, we aim at performing additional CYP2D6 studies with protein natively anchored to the membrane.

References

1. Wang B, Yang L-P, Zhang X-Z, Huang S-Q, Bartlam M, Zhou S-F. New insights into the structural characteristics and functional relevance of the human cytochrome P450 2D6 enzyme. Drug Metab Rev. 2009;41: 573–643. doi:10.1080/03602530903118729

2. Teh LK, Bertilsson L. Pharmacogenomics of CYP2D6: molecular genetics, interethnic differences and clinical importance. Drug Metab Pharmacokinet. 2012;27: 55–67.

3. Caudle KE, Dunnenberger HM, Freimuth RR, Peterson JF, Burlison JD, Whirl-Carrillo M, et al. Standardizing terms for clinical pharmacogenetic test results: consensus terms from the Clinical Pharmacogenetics Implementation Consortium (CPIC). Genet Med Off J Am Coll Med Genet. 2017;19: 215–223. doi:10.1038/gim.2016.87

4. Stingl J, Viviani R. Polymorphism in CYP2D6 and CYP2C19, members of the cytochrome P450 mixed-function oxidase system, in the metabolism of psychotropic drugs. J Intern Med. 2015;277: 167–177. doi:10.1111/joim.12317

5. Hicks J, Bishop J, Sangkuhl K, Müller D, Ji Y, Leckband S, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 and CYP2C19 Genotypes and Dosing of Selective Serotonin Reuptake Inhibitors. Clin Pharmacol Ther. 2015;98: 127–134. doi:10.1002/cpt.147

6. Cojocaru V, Winn PJ, Wade RC. The ins and outs of cytochrome P450s. Biochim Biophys Acta. 2007;1770: 390–401. doi:10.1016/j.bbagen.2006.07.005

7. Pandey AV, Flück CE. NADPH P450 oxidoreductase: Structure, function, and pathology of diseases. Pharmacol Ther. 2013;138: 229–254. doi:10.1016/j.pharmthera.2013.01.010

8. Reed JR, Backes WL. Formation of P450•P450 Complexes and Their Effect on P450 Function. Pharmacol Ther. 2012;133: 299–310. doi:10.1016/j.pharmthera.2011.11.009

9. Barnaba C, Gentry K, Sumangala N, Ramamoorthy A. The catalytic function of cytochrome P450 is entwined with its membrane-bound nature. F1000Research. 2017;6. doi:10.12688/f1000research.11015.1

10. Ito Y, Kondo H, Goldfarb PS, Lewis DFV. Analysis of CYP2D6 substrate interactions by computational methods. J Mol Graph Model. 2008;26: 947–956.

11. Hochleitner J, Akram M, Ueberall M, Davis RA, Waltenberger B, Stuppner H, et al. A combinatorial approach for the discovery of cytochrome P450 2D6 inhibitors from nature. Sci Rep. 2017;7. doi:10.1038/s41598-017-08404-0

12. Fukuyoshi S, Kometani M, Watanabe Y, Hiratsuka M, Yamaotsu N, Hirono S, et al. Molecular Dynamics Simulations to Investigate the Influences of Amino Acid Mutations on Protein Three-Dimensional Structures of Cytochrome P450 2D6.1, 2, 10, 14A, 51, and 62. PloS One. 2016;11: e0152946. doi:10.1371/journal.pone.0152946

13. Sasahara K, Mashima A, Yoshida T, Chuman H. Molecular dynamics and density functional studies on the metabolic selectivity of antipsychotic thioridazine by cytochrome P450 2D6: Connection with crystallographic and metabolic results. Bioorg Med Chem. 2015;23: 5459–5465. doi:10.1016/j.bmc.2015.07.043

14. Martiny VY, Carbonell P, Chevillard F, Moroy G, Nicot AB, Vayer P, et al. Integrated structure- and ligand-based in silico approach to predict inhibition of cytochrome P450 2D6. Bioinformatics. 2015;31: 3930–3937. doi:10.1093/bioinformatics/btv486

15. de Waal PW, Sunden KF, Furge LL. Molecular Dynamics of CYP2D6 Polymorphisms in the Absence and Presence of a Mechanism-Based Inactivator Reveals Changes in Local Flexibility and Dominant Substrate Access Channels. PLoS ONE. 2014;9. doi:10.1371/journal.pone.0108607

16. Santos R, Hritz J, Oostenbrink C. Role of water in molecular docking simulations of cytochrome P450 2D6. J Chem Inf Model. 2010;50: 146–154. doi:10.1021/ci900293e

17. Handa K, Nakagome I, Yamaotsu N, Gouda H, Hirono S. In Silieo Study on the Inhibitory Interaction of Drugs with Wild-type CYP2D6.1 and the Natural Variant CYP2D6.17. Drug Metab Pharmacokinet. 2014;29: 52–60. doi:10.2133/dmpk.DMPK-13-RG-044

18. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. Nature. 2015;526: 68–74. doi:10.1038/nature15393

19. Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE. Biomolecular Simulation: A Computational Microscope for Molecular Biology. Annu Rev Biophys. 2012;41: 429–452. doi:10.1146/annurev-biophys-042910-155245

20. Bermudez M, Mortier J, Rakers C, Sydow D, Wolber G. More than a look into a crystal ball: protein structure elucidation guided by molecular dynamics simulations. Drug Discov Today. 2016;21: 1799–1805. doi:10.1016/j.drudis.2016.07.001

21. Poulos TL. Cytochrome P450 flexibility. Proc Natl Acad Sci. 2003;100: 13121–13122. doi:10.1073/pnas.2336095100

22. Benkovic SJ, Hammes-Schiffer S. A perspective on enzyme catalysis. Science. 2003;301: 1196–1202. doi:10.1126/science.1085515

23. Suvarna BS. Role of pharmacogenomics in the drug development. Kathmandu Univ Med J. 2009;7: 172–176. doi:10.3126/kumj.v7i2.2716

24. Wooller SK, Benstead-Hume G, Chen X, Ali Y, Pearl FMG. Bioinformatics in translational drug discovery. Biosci Rep. 2017;37: BSR20160180. doi:10.1042/BSR20160180

25. He Z-X, Chen X-W, Zhou Z-W, Zhou S-F. Impact of physiological, pathological and environmental factors on the expression and activity of human cytochrome P450 2D6 and implications in precision medicine. Drug Metab Rev. 2015;47: 470–519. doi:10.3109/03602532.2015.1101131

26. Han K-M, Chang HS, Choi I-K, Ham B-J, Lee M-S. CYP2D6 P34S Polymorphism and Outcomes of Escitalopram Treatment in Koreans with Major Depression. Psychiatry Investig. 2013;10: 286–293. doi:10.4306/pi.2013.10.3.286

27. Shiraishi T, Hosokawa M, Kobayashi K, Tainaka H, Yamaura Y, Taguchi M, et al. Effects of G169R and P34S substitutions produced by mutations of CYP2D6*14 on the functional properties of CYP2D6 expressed in V79 cells. Drug Metab Dispos Biol Fate Chem. 2002;30: 1201–1205.

28. Flanagan JU, Maréchal J-D, Ward R, Kemp CA, McLaughlin LA, Sutcliffe MJ, 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. doi:10.1042/BJ20040062

29. Venhorst J, Onderwater RC, Meerman JH, Commandeur JN, Vermeulen NP. Influence of N-substitution of 7-methoxy-4-(aminomethyl)-coumarin on cytochrome P450 metabolism and selectivity. Drug Metab Dispos Biol Fate Chem. 2000;28: 1524–1532.

30. de Graaf C, Oostenbrink C, Keizers PHJ, van der Wijst T, Jongejan A, Vermeulen NPE. Catalytic Site Prediction and Virtual Screening of Cytochrome P450 2D6 Substrates by Consideration of Water and Rescoring in Automated Docking. J Med Chem. 2006;49: 2417–2430. doi:10.1021/jm0508538

31. Abduljalil K, Frank D, Gaedigk A, Klaassen T, Tomalik-Scharte D, Jetter A, et al. Assessment of Activity Levels for CYP2D6*1, CYP2D6*2, and CYP2D6*41 Genes by Population Pharmacokinetics of Dextromethorphan. Clin Pharmacol Ther. 2010;88: 643–651. doi:10.1038/clpt.2010.137

32. Sakuyama K, Sasaki T, Ujiie S, Obata K, Mizugaki M, Ishikawa M, et al. Functional Characterization of 17 CYP2D6 Allelic Variants (CYP2D6.2, 10, 14A–B, 18, 27, 36, 39, 47–51, 53–55, and 57). Drug Metab Dispos. 2008;36: 2460–2467. doi:10.1124/dmd.108.023242

33. Yu A, Kneller BM, Rettie AE, Haining RL. Expression, Purification, Biochemical Characterization, and Comparative Function of Human Cytochrome P450 2D6.1, 2D6.2, 2D6.10, and 2D6.17 Allelic Isoforms. J Pharmacol Exp Ther. 2002;303: 1291–1300. doi:10.1124/jpet.102.039891

34. Gaedigk A, Sangkuhl K, Whirl-Carrillo M, Klein T, Leeder JS. Prediction of CYP2D6 phenotype from genotype across world populations. Genet Med Off J Am Coll Med Genet. 2017;19: 69–76. doi:10.1038/gim.2016.80

35. Yu X, Cojocaru V, Wade RC. Conformational diversity and ligand tunnels of mammalian cytochrome P450s. Biotechnol Appl Biochem. 2013;60: 134–145. doi:10.1002/bab.1074

36. Wang A, Stout CD, Zhang Q, Johnson EF. Contributions of Ionic Interactions and Protein Dynamics to Cytochrome P450 2D6 (CYP2D6) Substrate and Inhibitor Binding. J Biol Chem. 2015;290: 5092–5104. doi:10.1074/jbc.M114.627661

37. Bell JA, Cao Y, Gunn JR, Day T, Gallicchio E, Zhou Z, et al. PrimeX and the Schrödinger computational chemistry suite of programs. In: urn:isbn:978-0-470-66078-2 [Internet]. 2012 [cited 21 Mar 2018]. doi:10.1107/97809553602060000864

38. Conner KP, Woods C, Atkins WM. Interactions of Cytochrome P450s with their Ligands. Arch Biochem Biophys. 2011;507: 56–65. doi:10.1016/j.abb.2010.10.006

39. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des. 2013;27: 221–234. doi:10.1007/s10822-013-9644-8

40. Berka K, Anzenbacherová E, Hendrychová T, Lange R, Mašek V, Anzenbacher P, et al. Binding of quinidine radically increases the stability and decreases the flexibility of the cytochrome P450 2D6 active site. J Inorg Biochem. 2012;110: 46–50. doi:10.1016/j.jinorgbio.2012.02.010

41. Panicco P, Dodhia VR, Fantuzzi A, Gilardi G. Enzyme-based amperometric platform to determine the polymorphic response in drug metabolism by cytochromes P450. Anal Chem. 2011;83: 2179–2186. doi:10.1021/ac200119b

42. Mürdter T, Schroth W, Bacchus-Gerybadze L, Winter S, Heinkele G, Simon W, et al. Activity Levels of Tamoxifen Metabolites at the Estrogen Receptor and the Impact of Genetic Polymorphisms of Phase I and II Enzymes on Their Concentration Levels in Plasma. Clin Pharmacol Ther. 2011;89: 708–717. doi:10.1038/clpt.2011.27

43. Li X, Delzer J, Voorman R, Morais SM de, Lao Y. Disposition and Drug-Drug Interaction Potential of Veliparib (ABT-888), a Novel and Potent Inhibitor of Poly(ADP-ribose) Polymerase. Drug Metab Dispos. 2011;39: 1161–1169. doi:10.1124/dmd.110.037820

44. McLaughlin LA, Paine MJI, Kemp CA, Maréchal J-D, Flanagan JU, Ward CJ, 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. doi:10.1074/jbc.M505974200

45. Smieško M. DOLINA--docking based on a local induced-fit algorithm: application toward small-molecule binding to nuclear receptors. J Chem Inf Model. 2013;53: 1415–1423. doi:10.1021/ci400098y

46. Bowers KJ, Chow DE, Xu H, Dror RO, Eastwood MP, Gregersen BA, et al. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. Proceedings of the ACM/IEEE SC 2006 Conference. 2006. pp. 43–43. doi:10.1109/SC.2006.54

47. Visscher KM, Kastritis PL, Bonvin AMJJ. Non-interacting surface solvation and dynamics in protein–protein interactions. Proteins Struct Funct Bioinforma. 2015;83: 445–458. doi:10.1002/prot.24741

48. Navrátilová V, Paloncýová M, Berka K, Otyepka M. Effect of Lipid Charge on Membrane Immersion of Cytochrome P450 3A4. J Phys Chem B. 2016;120: 11205–11213. doi:10.1021/acs.jpcb.6b10108

49. Sündermann A, Oostenbrink C. Molecular dynamics simulations give insight into the conformational change, complex formation, and electron transfer pathway for cytochrome P450 reductase. Protein Sci Publ Protein Soc. 2013;22: 1183–1195. doi:10.1002/pro.2307

50. McMasters DR, Torres RA, Crathern SJ, Dooney DL, Nachbar RB, Sheridan RP, et al. Inhibition of Recombinant Cytochrome P450 Isoforms 2D6 and 2C9 by Diverse Drug-like Molecules. J Med Chem. 2007;50: 3205–3213. doi:10.1021/jm0700060

51. Oprea TI, Hummer G, García AE. Identification of a functional water channel in cytochrome P450 enzymes. Proc Natl Acad Sci U S A. 1997;94: 2133–2138.

52. Otyepka M, Skopalík J, Anzenbacherová E, Anzenbacher P. What common structural features and variations of mammalian P450s are known to date? Biochim Biophys Acta. 2007;1770: 376–389. doi:10.1016/j.bbagen.2006.09.013

53. Klein K, Tatzel S, Raimundo S, Saussele T, Hustert E, Pleiss J, et al. A natural variant of the heme-binding signature (R441C) resulting in complete loss of function of CYP2D6. Drug Metab Dispos Biol Fate Chem. 2007;35: 1247–1250. doi:10.1124/dmd.107.015149

56. Paine MJI, McLaughlin LA, Flanagan JU, Kemp CA, Sutcliffe MJ, Roberts GCK, 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. doi:10.1074/jbc.M209519200

54. Wang A, Savas U, Hsu M-H, Stout CD, Johnson EF. Crystal structure of human cytochrome P450 2D6 with prinomastat bound. J Biol Chem. 2012;287: 10834–10843. doi:10.1074/jbc.M111.307918

55. de G, Oostenbrink C, Keizers PHJ, van V-L, van W, Tschirret-Guth RA, et al. Molecular modeling-guided site-directed mutagenesis of cytochrome P450 2D6. Curr Drug Metab. 2007;8: 59–77. doi:10.2174/138920007779315062

56. Smith G, Modi S, Pillai I, Lian LY, Sutcliffe MJ, Pritchard MP, 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.

Supporting Information

Table S3.1 **Overview of amino acid property change by each mutation**

CYP2D6 Mutation	Polarity	change	Hydrophobicity/hydrophilicity	change	Charge change	Lipophilic character change
1. P34S	nonpolar to polar	√	hydrophobic to hydrophilic	√	neutral to neutral	--
2. L91M	nonpolar to nonpolar	x	hydrophobic to hydrophobic	x	neutral to neutral	-
3. H94R	neutral/basic polar to basic polar	x	hydrophilic to hydrophilic	x	neutral to positive	none
4. T107I	polar to neutral nonpolar	√	hydrophilic to hydrophobic	√	neutral to neutral	++
5. F120I	nonpolar to nonpolar	x	hydrophobic to hydrophobic	x	neutral to neutral	+
6. A122S	nonpolar to polar	√	hydrophobic to hydrophilic	√	neutral to neutral	-
7. R296C	basic polar to nonpolar	√	hydrophilic to hydrophilic	x	positive to neutral	++
8. S486T	polar to polar	x	hydrophilic to hydrophilic	x	neutral to neutral	none

Table S3.2 **MD simulations parameter calculations for all CYP2D6 variants. The Solvent Accessible Surface Area (SASA) (FreeSASA v2.0.2), buried solvent area (PISA webserver) and overall volume (POVME v2.0) were calculated for the CYP2D6 dominant cluster conformation for each CYP2D6 variant**

#1 Cluster Conformation CYP Variant	Total SASA (Å^2)	% of total area	Polar SASA (Å^2)	% of total area	Apolar SASA (Å^2)	% of total area	BSA (Å^2)	% of total surface area	Volume (Å^3)	SOM Fe distance (Å)	ΔRMSD (vs. 4wnu)
wt_a	22195	93.8	9494	40.1	12701	53.7	1463	6.2	635	na	2.2
V2_a	21862	93.9	9248	39.7	12614	54.2	1408	6.1	716	na	2.6
V10_a	22283	93.7	9588	40.3	12694	53.4	1487	6.3	797	na	1.9
V17_a	22603	93.7	9783	40.6	12819	53.1	1522	6.3	663	na	2.8
V4_a	22072	94.6	9333	40.0	12738	54.6	1271	5.4	582	na	2.2
V53_a	24022	94.2	10132	39.7	13890	54.5	1471	5.8	836	na	2.8
wt_tam	22477	96.1	9274	39.7	13203	56.5	902	3.9	734	4.4	2.4
wt_vel	22398	94.0	9134	38.3	13264	55.7	1427	6.0	567	10.4	2.3
wt_buf	23246	94.6	10025	40.8	13221	53.8	1319	5.4	707	6	1.8
wt_pri	21745	90.3	8998	37.3	12748	52.9	2348	9.7	671	14	2.6
V17_pri	22779	90.1	9385	37.1	13394	53.0	2496	9.9	790	7	1.8
V53_pri	22584	90.2	9651	38.6	12933	51.7	2440	9.8	905	6	2.8
wt_qui	22579	91.5	9450	38.3	13128	53.2	2088	8.5	801	6	2.6
V17_qui	22018	90.5	9324	38.3	12694	52.2	2300	9.5	721	11.2	1.8
V53_qui	22600	91.3	9348	37.7	13252	53.5	2163	8.7	892	10.8	2.4
4wnu (qui)	20373	89.0	8612	37.6	11761	51.4	2509	11.0	856	na	na
2f9q (apo)	20762	86.0	8885	36.8	11878	49.2	3369	14.0	814	na	0.6
3qm4 (pri)	20306	90.6	8247	36.8	12060	53.8	2096	9.4	872	na	1.0

Table S3.3 **Overview ligand properties**[1]

Ligand	PSA (Å^2)	LogP	LogS	H_{Don}	H_{Acc}	#$Bond_{rotatable}$	Molar volume (cm^3)
tamoxifen	12.5	7.1	-5.6	0	2	8	356
veliparib	83.8	1.1	-3.0	3	3	2	192
bufuralol	45.4	3.2	-3.9	2	3	5	245
prinomastat	112.3	1.7	-4	2	6	4	307
quinidine	45.6	2.8	-3.0	1	4	4	266

1 Information obtained from DrugBank and Chemspider

Table S3.4 **Replica runs (2) of apo wild-type, RMSD and RMSF values compared**

Simulation	Mean RMSD (Å)	SD	Mean RMSF (Å)	SD
wt_a	5.4	0.6	1.53	1.19
wt_a_R1	5.6	0.6	1.59	1.21
wt_a_R2	6.5	0.7	1.88	1.7

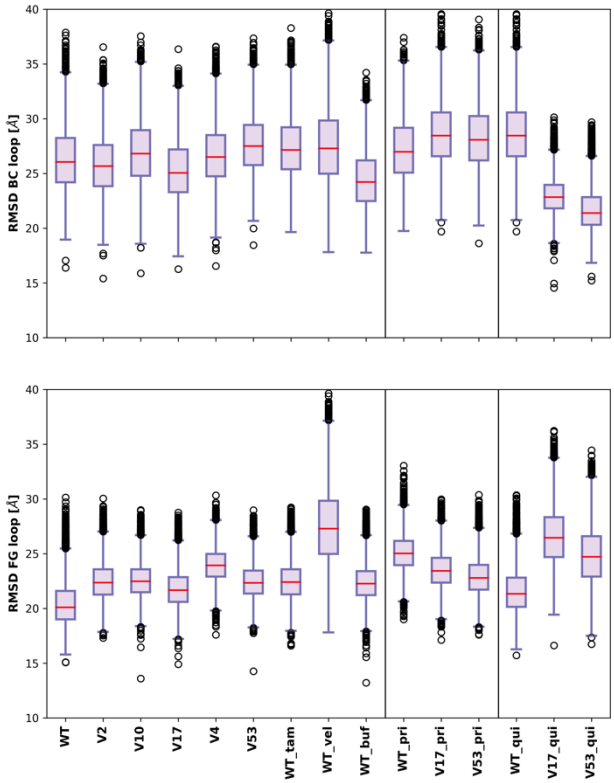


Figure S3.1 **Box-and-whisker plots for the BC and FG loops calculated over the whole trajectory for all CYP2D6 variants.** The distribution of the data points: the pink boxes indicate the upper - and lower quartile (25% of the data is greater than or less than this value), the caps indicate the greatest and the smallest value excluding outliners. The outliners represent more than or less than 3/2 times the upper or lower quartile, and the red line indicate the median value.

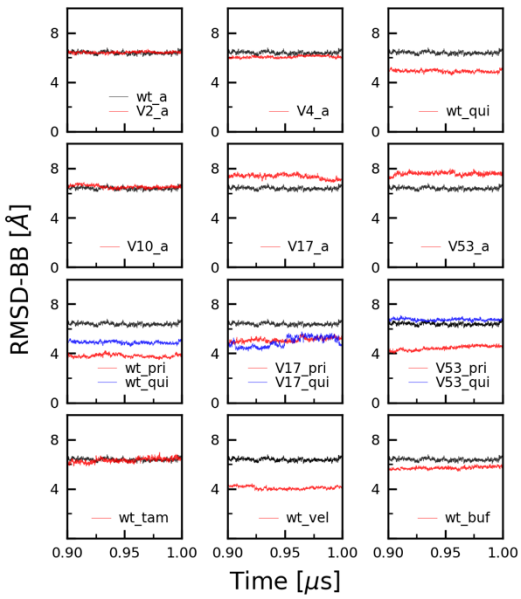


Figure S3.2 **Backbone root mean square deviation (RMSD) graphs for all CYP2D6 simulations – last 100 ns.** The wild-type simulation is shown black in all graphs. The RMSD graphs of the last 100 ns that confirm that the simulations were equilibrated.

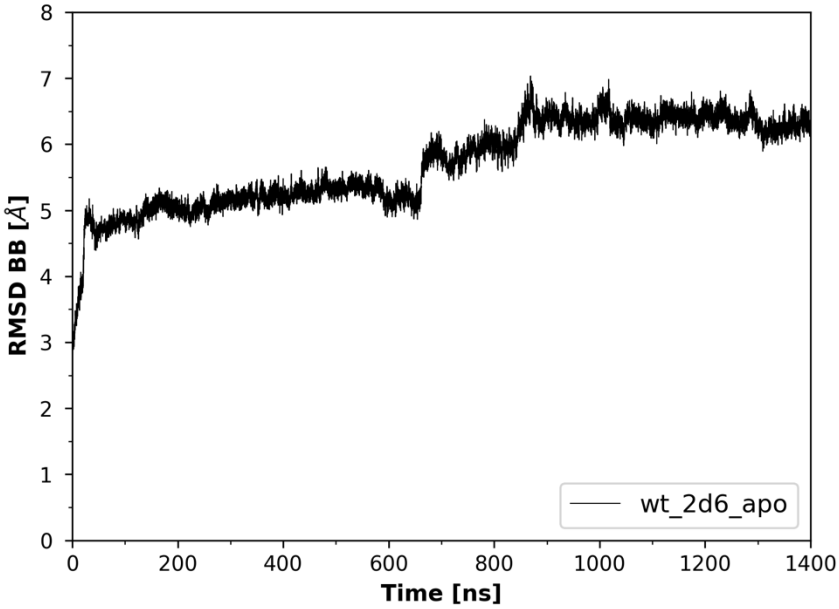


Figure S3.3 **Backbone root mean square deviation (RMSD) graph for the apo CYP2D6 wild-type calculated over 1.4 µs.** The additional 400 ns simulation ran for the wild-type simulation (black) confirms that the simulation reached plateau at the end.

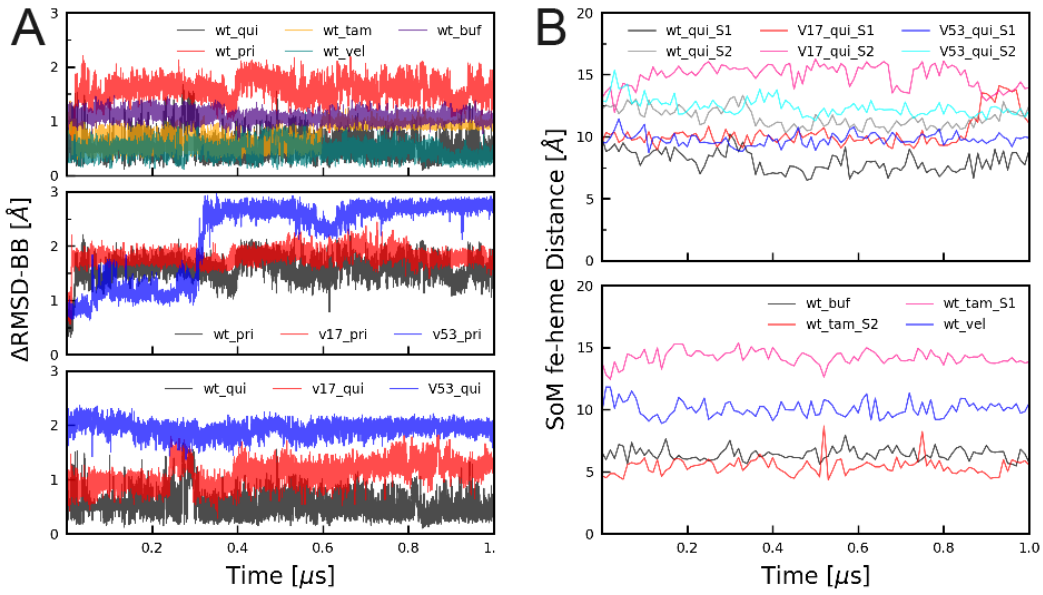


Figure S3.4 **RMSD calculated over 1.0 μs for the holo CYP2D6 variants (A) and SoM-heme distance (B). (A, top)** RMSD curves for the wild-type CYP2D6 simulations with quinidine (wt_qui, black), prinomastat (wt_pri, red), tamoxifen (wt_tam, yellow), veliparib (wt_vel, teal), and bufuralol (wt_buf, indigo). **(A, center)** RMSD curves for prinomastat with wild-type (wt_pri, black), CYP2D6*17 (v17_pri, red), and CYP2D6*53 (v53_pri, black). **(A, bottom)** RMSD curves for quinidine with wild-type (wt_qui, black), CYP2D6*17 (v17_qui, red), and CYP2D6*53 (v53_qui, blue). For the wild-type, prinomastat (inhibitor) demonstrates the highest flexibility, whereas veliparib (substrate) and quinidine (inhibitor) the lowest. The CYP2D6*53 variant with increased enzyme activity shows increased flexibility (±2Å) for both prinomastat and quinidine compared to the wild-type. (**B, top**) Distances between the heme and the site of metabolism (measured as nearest carbon atom) of quinidine with with wild-type (wt_qui_S1, black, wt_qui_S2, grey), CYP2D6*17 S1 (V17_qui_S1, red, V17_qui_S2, pink), and CYP2D6*53 S1 (V53_qui_S1, blue, V53_qui_S2, cyan). (**B, bottom**) Site of metabolism shown for wild-type with bufuralol (wt_buf, black) tamoxifen (wt_tam_S1, pink, wt_tam_S2, red) and veliparib (wt_vel, blue).

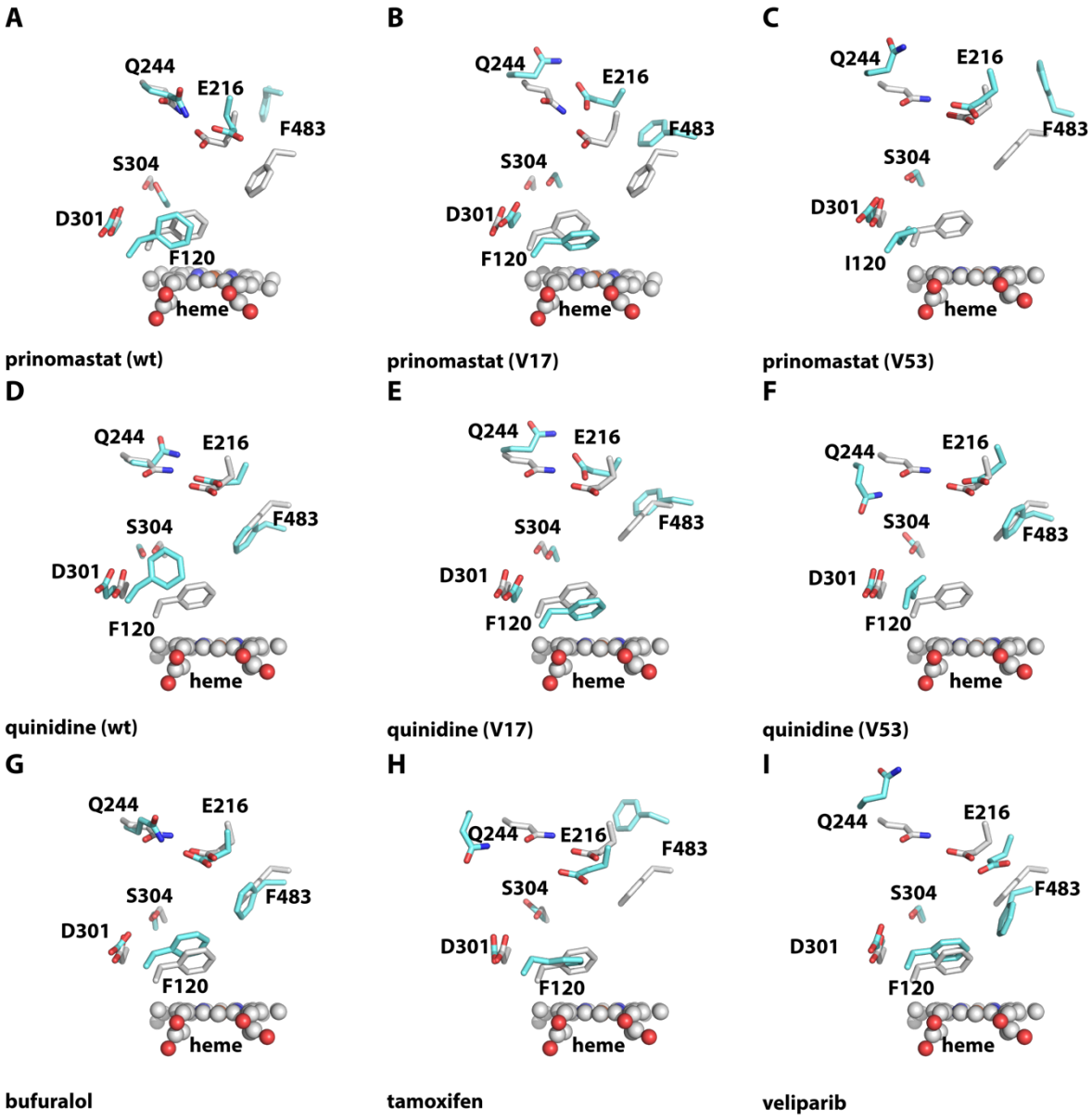


Figure S3.5 **Key residues (E216, D301, F120, F483, S304 and Q244) located in the active site of each CYP2D6 most prevalent conformation compared to the x-ray structures (ligands are left out for clarity).** The active site is shown for the most prevalent conformation in cyan, and the x-ray structure in white (A-C: 3QM4, D-I: 4WNU). **(A-C)** Prinomastat (inhibitor) with wild-type (wt) (**A**), CYP2D6*17 (V17) (**B**) and CYP2D6*53 (V53) (**C**), **(D-F)** Quinidine (inhibitor) with wild-type (wt) (**D**), CYP2D6*17 (V17) (**E**) and CYP2D6*53 (V53) (**F**), **(G-I)** wild-type together with substrates: bufuralol (**G**), tamoxifen (**H**), and veliparib (**I**).

Chapter 4

Molecular dynamics simulations reveal structural differences between allelic variants of membrane-anchored cytochrome P450 2D6

Based on the article:

Molecular dynamics simulations reveal structural differences between allelic variants of membrane-anchored cytochrome P450 2D6

André Fischer, Charleen G. Don & Martin Smieško.

J. Chem. Inf. Model.20185891962-1975

Abstract

Cytochrome P450 2D6 (CYP2D6) is an enzyme that is involved in the metabolism of roughly 25% of all marketed drugs and therefore belongs to the most important enzymes in drug metabolism. CYP2D6 features a high degree of genetic polymorphism that can significantly alter the individual metabolic activity. This can either lead to abolishment of the therapeutic effect of a drug or to enhanced plasma levels potentially causing adverse effects. In this study, starting from the crystal structure, we built a full-length membrane-anchored model of wild-type CYP2D6 as well as five variants with different enzymatic activity. We validated our models with available experimental data and compared their structural properties. The main focus of this study was to identify differences that could mechanistically explain the altered activity of the variants and improve our understanding of their functionality. We observed differences in the opening frequencies and minimal diameters of tunnels that connect the buried active site to the surrounding solvent environment. The variants CYP2D6*4 and CYP2D6*10 associated with missing or decreased activity showed lower opening frequencies of the tunnels compared to the wild-type. Furthermore, the altered fold in the N-terminal anchor region and the decreased volume of the active site cavity of CYP2D6*4 offer an explanation for the absence of metabolic activity. On the other hand, the mutations in CYP2D6*53 contributed to a significant enlargement of an important ligand tunnel and an extension of the active site cavity. This could explain the altered metabolic profile of this particular variant.

Introduction

The clearance of many orally administered drugs is largely influenced by the metabolic capacity of the individual [1]. A variability of drug metabolism among individuals can lead to a change in the plasma concentration which could either obliterate the effect of the drug or potentially lead to adverse effects. Cytochrome P450 2D6 (CYP2D6) metabolizes around 25% of all marketed drugs and therefore belongs to the most important enzymes involved in the phase I biotransformation [2]. Besides that, CYP2D6 is of interest to the field of enzyme engineering, where its site-directed mutagenesis resulted in a significantly increased

metabolic rate for several substrates promising, optimized enzymatic production of designated metabolite [3]. A special hallmark of CYP2D6 among cytochromes (CYPS) involved in drug metabolism is that its function can be altered due to genetic polymorphism leading to substantial interindividual differences in metabolism.

Four general phenotypes are known for CYP2D6: poor metabolizer (PM), intermediate metabolizer (IM), normal metabolizer (NM), and ultrarapid metabolizer (UM) [4]. These phenotypes arise from point mutations, single base-pair deletions or additions, gene rearrangements, gene duplications, and gene deletions [2,5].

The impact of the amino acid mutations within allelic variants on the structure and function of the enzyme is not clearly understood. Overlapping amino acid mutations among the four phenotype categories increase the complexity of correlating structural particularities with enzyme activity. Examples for such allelic variants are CYP2D6*2, CYP2D6*4, CYP2D6*10, CYP2D6*17, and CYP2D6*53 (**Table 4.1**). The positions of amino acid mutations present in these allelic variants are given in **Figure 4.1** together with their related secondary structural elements.

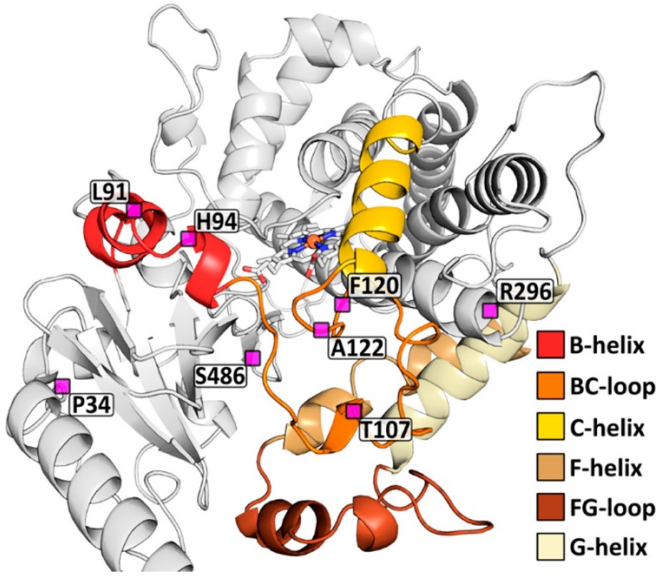


Figure 4.1 **Structural overview of CYP2D6.** The position of amino acid mutations, that CYP2D6 is prone to, are shown on its structure. Secondary structural elements of importance are indicated by different colors. We used the wild-type enzyme to compute the figure.

CYP2D6*2 is an allelic variant that shows a comparable activity to the wild-type suggesting a neglectable influence on the enzyme structure. The comparable activity results in the

declaration as NM phenotype, while CYP2D6*10 and CYP2D6*17 with decreased activity fall into the IM phenotype [6,7]. CYP2D6*10 harbors the P34S amino acid mutation causing the disruption of a conserved proline-rich segment near the N-terminus [6,8]. It is the most prevalent allele in the Asian population with frequencies up to 50% depending on the region, whereas CYP2D6*17 is more prevalent in Africans. The T107I mutation in CYP2D6*17 might lead to an altered active site structure [2]. CYP2D6*4 belongs to the most important alleles causing PM phenotype in Caucasians and also exhibits the P34S mutation [6]. Even though the absence of metabolic activity is believed to arise from effects on the level of RNA splicing three out of four amino acid mutations in this variant were shown to be degrading enzymatic activity independent of effects on the level of protein biosynthesis [9]. Furthermore, the four amino acid mutations also occur in other variants such as CYP2D6*74 and CYP2D6*82 [10,11]. CYP2D6*53 features the F120I mutation located directly in the active site, which is suspected to change the regiospecificity of metabolic reactions. Further on, CYP2D6*53 shows increased metabolic rates depending on the substrate suggesting a possible UM phenotype [6,12].

Table 4.1 **Overview on the allelic variants studied.** For each variant, the amino acid mutations and their impact on the enzymatic activity are displayed.

Variant	amino acid mutations[A]	enzyme activity [A]
CYP2D6 WT	none	normal
CYP2D6*2	R296C, S486T	normal
CYP2D6*4	P34S, L91M, H94R, S486T	none
CYP2D6*10	P34S, S486T	decreased
CYP2D6*17	T107I, R296C, S486T	decreased
CYP2D6*53	F120I, A122S	increased

[A] This data is accessible on www.pharmvar.org (accessed on January 31, 2018).

Along with polymorphism studies, CPY2D6 also started attracting attention of the molecular modeling research community [13]. This can be partly attributed to the increasing number of crystal structures of human CYP2D6 solved in the past decade [14-18]. However, available structures lack at least 30 N-terminal residues. The reason for this lies in the crystallization

procedure. The N-terminal truncation is required to increase enzyme solubility, because this α-helical segment anchors the enzyme to the membrane of the endoplasmic reticulum [14,19,20]. As the active site of CYP2D6 is buried in a cavity inside the core of the enzyme, it is suggested that tunnels leading to the active site actually govern the kinetics and specificity of the enzyme. This makes them a suitable target for selective modifications aiming on modifying either of these properties [21]. Since most drugs are rather lipophilic in nature and microsomal CYPs are membrane-anchored, an established theory exists that ligands might access the enzyme from the membrane [22-24]. Additionally, the structural motifs lining various tunnels are known to be in contact with the membrane and have been shown to modify the diversity of enzyme tunnels [20]. This highlights the need of investigating the access and egress pathways in the presence of a membrane as its natural environment. Since enzyme tunnels are influenced by the dynamic nature of the protein, it is crucial not only to investigate a single structure but also to use a time-evolved ensemble of structures representing the dynamic motion of the protein. This motion then translates into the opening or closing of tunnels that cannot be observed in a static crystal structure [21,25]. Molecular Dynamics (MD) simulations are a suitable and widely employed computational technique to produce such dynamic ensembles [25-27]. Furthermore, the growing number of CYP2D6 alleles being discovered, already exceeding 100, urges for more practical methods to support experiments with atomistic details. The activity of CYPs is further influenced by their electron donor Cytochrome P450 oxidoreductase (POR). The electrons are needed in order to perform the wide spectrum of oxidative reactions [28]. The interaction of the POR with CYPs is also suspected to influence the opening of enzyme tunnels [29]. The residues of the POR binding site were experimentally determined for CYP2B4 [30].

Until today, membrane-bound models for CYP1A2, CYP2A6, CYP2C9, CYP2D6, CYP2E1, CYP2U1, and CYP3A4 have been published [20,22-24,31,32]. These models were built and refined by the use of molecular dynamics (MD) simulations and membrane positioning protocols. In this context, most groups employed simplified coarse-grained or united-atom MD simulations to equilibrate the position of the enzyme in the membrane. However, those

models are not available to the public making it impossible to perform additional computational studies. Results from other CYPs cannot be expected to be directly applicable to CYP2D6 since CYPs greatly vary in substrate specificity despite having a similar fold. Moreover, neither of these studies focused on effects caused by genetic polymorphism related to CYP2D6. For instance, it was shown for CYP2C9 that the superficial entry points of presumed access tunnels agree with the energetically favored position of ibuprofen in the membrane [24]. Other groups showed that tunnels in CYP2U1 as well as CYP3A4 point toward the membrane and suggested the uptake of lipophilic substrates from the membrane-enzyme interface [23,31]. De Waal et al. proposed that some changes in the enzymatic activity of CYP2D6 can be explained by the accessibility of the active site for ligands [12]. In their study, they compared the tunnel opening and protein flexibility of the variants CYP2D6*****17 and CYP2D6*****53 in the absence of a membrane. They observed different degrees of tunnel opening between the analysed variants and differences in the local protein flexibility that might be related to the enzymatic activity.

In this study, we created a membrane-bound model of wild-type human CYP2D6 and the five allelic variants CYP2D6*2, CYP2D6*4, CYP2D6*10, CYP2D6*17, and CYP2D6*53. We investigated the enzyme tunnels together with other structural properties with the help of all-atom MD simulations in a cumulative length of 5.7 μs in order to find differences that could explain the altered activity of the variants. Furthermore, we analyzed the water access to the active site and the possibility of the POR binding to the wild-type model. Our findings revealed diverse dynamic behavior among the allelic variants. Most remarkably, the CYP2D6*53 variant showed an increased active site volume and an enlargement of turns 2b, which is an important ligand tunnel, that we suspect to be caused by the F120I mutation. This supports its pending designation as a cause for the UM phenotype. To our best knowledge, this is the first study to explore enzyme tunnels combined with structural hallmarks in the presence of a membrane. The uncovered structural features related to ligand tunnels, the role of individual residues, and the fold of the protein help to better understand the function of this particular CYP and five of its most important variants. Our insights regarding the cause of substrate-dependent activity between the allelic variants may

contribute to optimize processes in the field of enzyme engineering. Furthermore, the differences between the variants may be of value for the personalized drug therapy field.

Methods

Model Building and Description. Figure 4.2 schematically depicts the procedure of model building. We retrieved the structure of human CYP2D6 from the Protein Data Bank (PDB ID: 3TDA) [18] and processed it using the Protein Preparation Wizard within the Maestro Small-Molecule Drug Discovery Suite (v.2017-2) [33]. We added hydrogen atoms, assigned bond orders, and kept the crystallographic water molecules. The two zinc ions were replaced with one water molecule each. Since the N-terminal residues 31−33 were engineered for crystallization purposes, we reconstructed this short sequence to match the wild-type enzyme using the 3D Builder in Maestro. The sequence of the wild-type CYP2D6 was retrieved from UniProt [34]. After that, we refined the hydrogen-bonding network and performed a heavy atom-restrained minimization with the OPLS_2005 force field which is recommended to be used [33] and was previously employed for CYP2D6 [35]. The initial positioning of the structure in the membrane was predicted using the Orientations of Proteins in Membranes (OPM) Web-based protocol called PPM server [36]. The PPM server accepts a structure in PDB format as input and generates a new file with pseudoatoms that describe the membrane position taking into consideration energetic contributions. The structure was then embedded in a 1-palmitoyl-2-oleoylphosphatidylcholine (POPC) membrane using the Desmond system builder. We used the Desmond (v.2016-4) simulation package from Schrödinger for MD with the OPLS_2005 force field in an NPT ensemble. In this context, we combined the Nose-Hoover thermostat at a relaxation time of 1.0 ps with the Martyna-Tobias-Klein barostat with a relaxation time of 2.0 ps at 300 K. Prior to the simulation stage we applied the default Desmond relaxation protocol consisting of seven stages (**Table S4.1**).

Long-range interactions were treated with the Particle Mesh Ewald (PME) method, while the cutoff for short-range interactions was set to 9 Å. Bonds to hydrogen atoms were constrained with the M-SHAKE algorithm, and no hydrogen mass partitioning was applied.

We chose the TIP3P water model to solvate the orthorhombic boxes of the periodic boundary systems throughout this study [37]. With the resulting system, we performed a 100 ns MD simulation to equilibrate the position of the enzyme in the membrane. This simulation will be referred to as MemAs from hereafter. We built the missing transmembrane helix from Met1 to Ala30 according to the wild-type sequence using the 3-dimensional (3D) Builder in Maestro which allows to generate a segment with a predefined topology.

As a preparation for the positioning of the membrane on the helix, we applied the Web-based protocols TMHMM (v2.0) AmphipaSeek [38], and TMSEG to get an estimate of the membrane-bound residues in CYP2D6 based on its sequence. The three mentioned methods accept a protein sequence in FASTA format as input and deliver a residue-based readout. TMHMM predicts residues that are part of transmembrane helices based on a Markov Model and incorporates hydrophobicity, helix length, and charge bias into the prediction [38]. AmphipaSeek is a method that predicts amphipathic in-plane membrane anchors of monotopic membrane proteins based on pattern recognition [39]. We chose this method to get an alternative prediction in contrast to the other methods that do not specifically predict in-plane anchors. Protein Predict Open provides the TMSEG method predicting trans- membrane helices in a segment-based neural network approach [40]. The minimal amount of residues to place a POPC membrane on a peptide in Desmond is 20. We therefore used Gly2 and Met23 as border residues for the membrane placement considering the results of the Web-based protocols shown in **Table S4.2**. After embedding the α-helix, we performed a 200 ns MD with the resulting system referred to as AnchorSim hereafter. The products of both simulations were combined by connecting the catalytic domain and the transmembrane anchor with a covalent bond. Furthermore, we removed the ligand from the catalytic domain. After minimization, the resulting system was again equilibrated for 200 ns by MD with frames being collected every 200 ps. As a control, we performed multiple simulations similar to AnchorSim and MemAs differing in the positioning of the membrane on the respective structures. Likewise, we designed three more systems consisting of the anchor linked to the catalytic domain. For these simulations, the structures with different orientations of the anchor were extracted from the trajectory and linked to the catalytic

domain. These exploratory simulations varied in duration. For all control simulations, we calculated several parameters for comparison including the location of two residues in proximity to the protein–membrane interface.

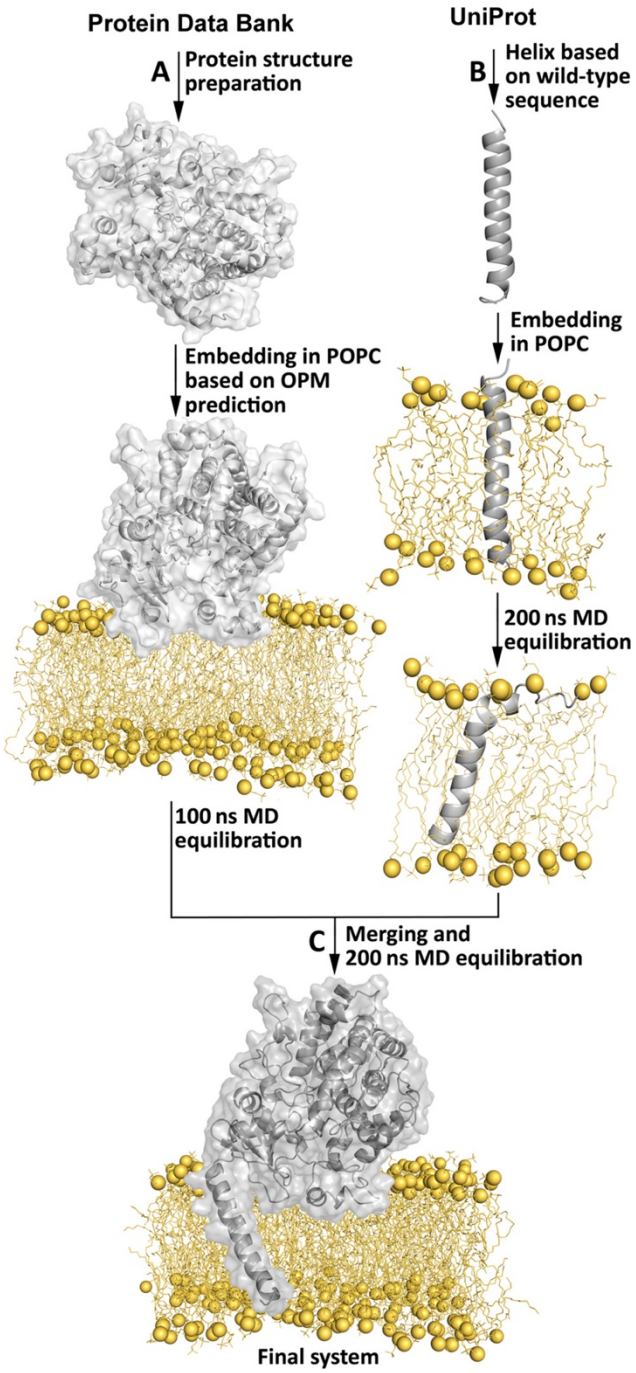


Figure 4.2. **Flowchart of the model building procedure. (A)** The catalytic domain was placed in POPC based on OPM prediction and equilibrated. **(B)** The helix was embedded in POPC and equilibrated. **(C)** Linkage of the catalytic domain to the anchor and equilibration.

Production runs. For further analysis, we continued the simulation of the previously derived system of wild-type membrane-anchored CYP2D6 from 200 to 300 ns leading to the system referred to as CYP2D6 WT. We also prepared two structures for simulations without membrane. One structure was simulated for 300 ns with the inhibitor prinomastat bound (referred to as WT inhibitor) and the other one without inhibitor respectively (referred to as WT nolig). Further on, we introduced polymorphism-related amino acid mutations to the merged system of the wild-type simulation before it was equilibrated. We focused on the five alleles CYP2D6*2, CYP2D6*4, CYP2D6*10, CYP2D6*17, and CYP2D6*53. We obtained the necessary data about the variants from the Pharmacogenetics Variation Consortium (http://www.pharmvar.org) [41]. With the resulting five systems, we performed another 200 ns equilibration run followed by a 100 ns production phase. For these simulations, we retained the same settings that we previously used for the CYP2D6 WT system. The mutations related to the allelic variants are shown in **Table 4.1**. The time step for saving frames in all the stated production simulations was 200 ps. To assess the significance of our results, we performed 12 replica simulations with a duration of 100 ns each. We started two simulations per system (wild-type and five variants) from two structures that were extracted with the trajectory_extract_frame.py script provided by Schrodinger. In this context, we selected a structure five frames before and five frames after each production simulation as starting point meaning the structures were pre-equilibrated. For MD, we used the same parameters as for the production simulations stated above. To assess the convergence of the simulations regarding structure, we used the trajectory_average_structure.py script also provided by Schrodinger to generate an average structure of the respective 100 ns simulations. Using the 3D Builder in Maestro, we then removed the membrane as well as the 31 initial residues corresponding to the flexible membrane anchor. Finally, we employed the superposition protocol in Maestro to calculate the RMSD (root-mean-square deviation) difference between the average structures from the simulations. We compared the replicas of each variant among themselves and then compared each replica to the production simulation that was included in the analysis. The resulting RMSD differences were summarized in a table, and averages for each variant were calculated.

Model validation. As previously We validated the CYP2D6 WT model using experimental data as well as data from previously published models of CYP2D6, CYP2C9, and CYP2U1 [20, 23, 24, 31, 42-46]. The heme tilt angle is often used to generally evaluate a model of a membrane-bound CYP. It is defined as the angle between the plane of the porphyrin nitrogens of the heme group and the membrane normal (z- axis). We determined the burying depth of the protein in the membrane according to Ducassou et al. They measured the distance between the mass center of the protein α-carbons and the mass center of the POPC C1-carbons in CYP2U1 [31]. The complete production phase was included into the calculation for both parameters stated above. Additionally, we determined the membrane-buried residues depending on the position of the protein α-carbons relative to the membrane after 200 ns of MD equilibration. For this purpose, we defined the headgroup region to be located between the mass centers of the C2-atoms and the N-atoms of the POPC molecules. To estimate the volume of the active site cavity, we used the Pocket Volume Measurer (POVME, v2.0) [47]. We analyzed the last 100 frames of the production phase for each system. With the aim to optimize the results from POVME, we used PyMol to visualize the results. A distance cutoff of 1.4 and a grid spacing of 1.0 were found to deliver optimal results. We equally performed these calculations with the replica simulations.
We identified the substrate-recognition site residues (referring to CYP2C9 and CYP2B4) employing sequence alignment using UGENE48 (v1.29) with the ClustalW algorithm [49]. We visualized the alignment results using JalView (v2.10.2) [50]. Finally, we determined the number of hydrogen bonds for the mutated amino acids in each variant for the complete 300 ns trajectories as well as all replica simulations.

Tunnel Analysis. For the analysis of possible access and egress tunnels in the enzyme we employed CAVER 3.0 [26]. CAVER is widely used to identify and characterize transport pathways within enzymes [12,22,25]. Prior to the tunnel analysis, we aligned 500 frames from the MD trajectories that were generated during the production runs. In order to perform the calculation, we determined the starting point using the tool CAVER Analyst 1.0 [51] by including the two active site residues Glu216 and Asp301 as well as the heme [18]. We then

performed the tunnel analysis using a probe radius of 0.9 Å and a clustering threshold of 4.5. All other parameters were left on default. We generated a heatmap, as well as a graph, for the opening frequency of the tunnels with the obtained data. We considered a tunnel to be open if its bottleneck radius was above 1.2 Å. Based on the generated data we were able to determine important bottleneck residues and the bottleneck radii. For the replica simulations, we likewise calculated the bottleneck residues, the average bottleneck radii, and the tunnel opening frequencies. The tunnel nomenclature was adapted from Wade et al. [21]. In addition, we developed a program in C-programming language to identify the water molecules that were approaching or leaving the active site through tunnels during the production phase of the CYP2D6.

WT simulation. We observed water molecules inside protein cavities to travel significantly slower than when they are free in bulk. Our program evaluated the position of each water molecule along the trajectory based on its distance to the active site, its current velocity, and how often both of these values would reach below a certain threshold in consecutive frames ("strikes"). We then manually analyzed the sorted output of the program and visually followed the respective waters in the Desmond trajectory viewer to determine their location in respect to the enzyme tunnels.

Results and Discussion

Model Building and Description. The MemAs simulation showed a stable RMSD of approximately 2 Å reaching a plateau after 40 ns of simulation time (**Figure S4.1**). On the other hand, the AnchorSim simulation showed higher fluctuations peaking at 6.3 Å due to the large movements of the non-constrained helix inside the membrane. Due to this flexibility, in the systems combined with the anchor, we calculated the RMSD only for the catalytic domain of the protein in order to get meaningful results. The simulations with the merged, membrane-bound models showed RMSD values ranging from 1.5 to 2.5 Å, and all of them readily reached equilibrium during the first 50 ns (**Figure S4.2**). According to the

root-mean-square fluctuation (RMSF) profiles the most flexible regions in all simulations were the transmembrane anchor, the FG loop, the HI loop, the BC loop, and the GH loop in decreasing order (**Figure S4.3**). Our CYP2D6 WT model reached a heme tilt angle of 43° after the 200 ns equilibration run and an average of 47 ± 5° during the production phase (**Table 4.2 and Figure S4.4**).

Table 4.2 **Overview on the variants, tunnel dynamics, and the associated structural parameters used for the validation of the respective model.**

Variant	tunnel average bottleneck radius [A, B] opening frequency [C]				*Burying depth [B]*	*Heme-tilt angle [B]*	*Active site volume [D]*
	2b	**2c**	**2e**	**4**			
CYP2D6 WT	1.4 Å 73%	1.5 Å 90%	1.1 Å 14%	1.1 Å 9%	37.7±1.2 Å	47.1±5.4°	0
CYP2D6*2	1.8 Å 94%	1.4 Å 78%	1.2 Å 23%	1.4 Å 35%	38.7±1.1 Å	59.3±4.3°	+
CYP2D6*4	1.2 Å 46%	1.2 Å 37%	1.1 Å 30%	1.2 Å 37%	38.9±1.2 Å	47.1±4.7°	-
CYP2D6*10	1.9 Å 99%	**[E]**	1.1 Å 4%	1.3 Å 39%	36.6±1.2 Å	64.5±4.8°	+
CYP2D6*17	1.8 Å 100%	1.0 Å 2%	1.0 Å 1%	1.4 Å 54%	38.8±1.5 Å	44.2±4.9°	+
CYP2D6*53	2.6 Å 99%	1.1 Å 1%	1.1 Å 15%	1.3 Å 65%	41.2±1.3 Å	48.4±4.9°	+

[A] The standard deviation for the average bottleneck radii was between 0.1 Å and 0.3 Å.
[B] Averaged values over all frames during the production phase.
[C] Percentage of frames in which the tunnel bottleneck radii were above 1.2 Å.
[D] Volume of the active site cavity estimated with CastP compared to the wild-type.
[E] Tunnel not present in the respective variant.

Hence, it was in agreement with the experimentally determined range of 38−78° for CYPs [43]. Berka et al. published the only other membrane-bound model of CYP2D6 and determined the heme tilt angle in their model to be 72° significantly differing from our value [20]. This could be due to a different membrane composition, the different crystal structure, or due to the united-atom MD approach they used. Experimental measurements in CYP3A4

bound to POPC nanodiscs revealed an average heme tilt angle of 59.7° [23]. Ducassou et al. determined the burying depth of CYP2U1 into the membrane to be decreasing from 43 to 38 Å during their simulation [31]. Similar calculations for CYP2C9 presented a value of 40 Å [24]. There is only one experimental value available regarding the height of a CYP above the membrane. It was determined for CYP2B4 by atomic force microscopy with an average of 35 ± 9 Å [44]. The burying depth of our CYP2D6 WT model calculated between the protein α-carbons and POPC C1-atoms converged from 44 Å at the beginning of the simulation to 36 Å after 300 ns (**Figure S4.5**). During the production phase, the burying depth averaged at 38 ± 1 Å. Our results therefore are in the same range as the literature values. It is accepted that the F-G loop in different CYPs is involved in the anchoring of the enzyme to the membrane [14,19, 22, 42].

Our model showed a good correlation (**Table S4.3**) to the experimental data except for a few discrepancies that may be explained by the different enzymes (CYP2B1, CYP2B4, and CYP2C2) that were used in these experiments. Membrane proteins mediate their contacts over specific protein−membrane interactions [52]. In our model Arg25, Arg115, Ser116, Lys239, and Arg242 were shown to be strongly interacting with POPC molecules via salt bridges and hydrogen bonds (**Figure S4.7**). Overall, our model is in good agreement with experimental data. All control simulations of the protein with the membrane showed a heme tilt angle that is in accordance with the experimentally determined range except for one (**Table S4.4**). The burying depth in most simulations ranged between 36 Å and 41 Å. One simulation showed a value of 34.7 Å significantly differing from the other values and was therefore not considered for additional simulations. The location of the Arg28 at the protein−membrane interface varied throughout the simulations, while the position of Gln27 remained within the headgroup region. We chose the AnchorSim, MemAs, and CYP2D6 WT simulations to be optimal after the visual inspection of the structures and careful analysis of the calculated validation parameters. The replica simulations were used to assess the convergence of the structures in individual MD simulations. This was done by superimposing average structures of the respective simulations. From the superposition, we could review the RMSD differences that are shown in **Table S4.5**. The average differences for each system

varied from 0.51 to 1.48 Å for the backbone atoms and from 0.63 to 1.55 Å for all heavy atoms of the globular domain. Only CYP2D6*2 and CYP2D6*4 showed values higher than 1 Å. The high values for CYP2D6*4 might be explained by the structural instability of the N-terminal residues. Indeed, if the RMSD difference was calculated separately from residue 32−70 and 71−497 the values were significantly higher in the N-terminal segment with 1.49 Å compared to 0.95 Å on average. For CYP2D6*2 the comparably high differences might be explained by the rather flexible character of this variant. When the RMSF profiles between the variants are compared, CYP2D6*2 displayed a high flexibility, especially around the F-G loop. Overall, the results prove that all production simulations maintained a satisfying structural convergence.

POR Accessibility and Solvent Access. As mentioned in the introduction, the POR acts as electron donor to CYP2D6 and therefore plays a crucial role in its catalytic cycle [30]. In this study, one of the main focuses was to investigate if the binding site for the POR in our models would be accessible in a similar fashion as it was observed for CYP2B4. Bridges et al. identified the POR binding site on CYP2B4 in a site-directed mutagenesis study [30]. To be able to translate their conclusions to our model, we aligned CYP2B4 to our CYP2D6 WT model as it is presented in **Figures S4.8 and S4.9**. We observed the binding site for the POR on CYP2D6 to be positively charged as it was also reported for CYP3A4 [29].

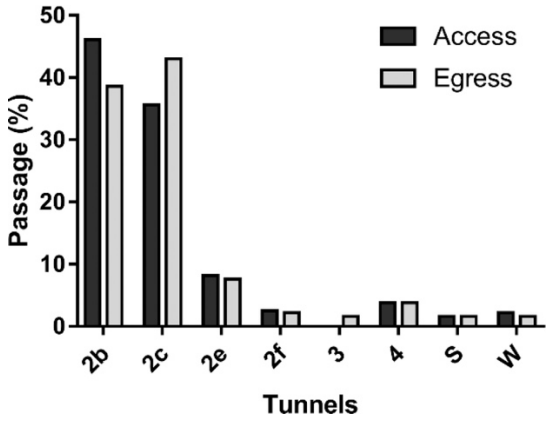


Figure 4.3 **The passage of water molecules through enzyme tunnels.** We divided the passage in access and egress considering the direction of their path. Percentages for each tunnel were calculated based on the total number of observed water molecules.

The binding site was also found to be in close vicinity to the water tunnel and faced the cytosol. This demonstrated that the POR would have the ability to bind to our model, since it is proposed to bind to a part of the enzyme that faces the cytosol [52]. It was further proposed that the POR could be involved in the opening of the water tunnel to desolvate the active site and enable ligand binding [29]. In this context, the tunnel is thought to allow passage due to a side-chain movement of a conserved arginine residue that weakens its interaction with the 7-propionate group of the heme moiety. This assumption can be supported by the fact that we observed only 12 water molecules passing through the water tunnel during the production phase (**Figure 4.3**) in the absence of the POR and without any ligand binding events. The total number of observed water molecules accessing the enzyme was 317. In our simulation, interaction between Arg441 and the heme remained stable during 99% of the production phase. As displayed in **Figure 4.4**, we did not observe any water molecules directly passing by the the heme propionates. Instead, an alternative path through a loop at the N-terminal end of the β 1-4 sheet was preferred.

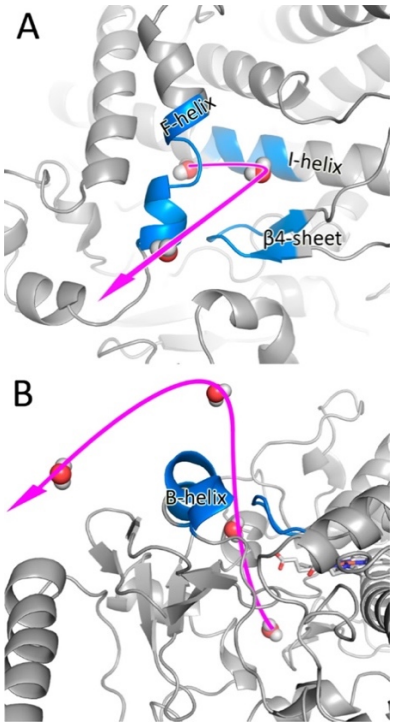


Figure 4.4 **Overview of the water egress:** through **(A)** the solvent tunnel and **(B)** the water tunnel. Snapshots of the trajectory were superimposed to represent the position of the water molecules at different time points. Secondary structure regions at the mouth of the tunnel are shown in blue. The paths of the water molecules are indicated by arrows.

We observed tunnel 2b and tunnel 2c to be the main pathways for active site hydration. In fact, we only observed 10 water molecules to pass the solvent tunnel. This indicates that the solvent tunnel shares a similarly low activity as determined for the water tunnel. This is also reflected by our results of the tunnel analysis since the solvent tunnel and the water tunnel turned out to be in a mostly closed state during the production phase (**Table S4.6**)
Taking into account the higher diffusion rates of the TIP3P water model [53], the absolute number was possibly even biased toward higher values. Hence, the water tunnel as well as the solvent tunnel turned out to be of minor importance for active site hydration in our setup. This is in agreement with results for other CYPs in the literature [12,29]. Based on the data from CAVER, we identified the lining residues of the identified tunnels and compared their hydropathic character (**Table S4.7**) according to Kyte et al. [54]. Interestingly, the average hydropathic index of the water tunnel and the solvent tunnel presented the exact same value confirming their similar roles in transport. It is also worth mentioning that the lining residues in tunnel 2b are less hydrophobic than in tunnel 2c. After considering the throughput of water molecules, one could assume that water molecules naturally prefer a less hydrophobic tunnel environment to enter the active site.

Tunnel Analysis. As previously mentioned, tunnels that connect the active site to its surrounding environment are thought to modulate substrate specificity as well as kinetic properties of the enzyme. For the analysis of enzyme tunnels, we compared the production simulations of the wild-type with and without ligand as well as the five allelic variants. In all simulations, the most prominent tunnels were tunnels 2b and 2c followed by tunnels 2e and 4 (**Table S4.6**). This is in agreement with a previous MD study on CYP2D6 without the membrane [12]. **Figure 4.5** shows that tunnels 2b, 2c, 2d, 2e, and 4 are pointing toward the membrane, whereas the other tunnels face the cytosol in our CYP2D6 WT model. A detailed overview of the spatial distribution of all tunnels in the wild- type is given in **Figure S4.10**. Tunnel 6 has only been described once in the literature to our knowledge. It is located between the helices I and K as well as the β4-sheet in CYP2A6 [55]. Even though we identified this tunnel, we consider it to be of secondary importance, since it rarely occurred

in seven of eight total simulations (**Figure 4.6**). Also, the maximal bottleneck radii were very narrow (always below 1.2 Å) during all the simulations with membrane apart from CYP2D6*17. In this variant the bottleneck radius for tunnel 6 reached values above 1.5 Å for a short period of 1.6 ns. In the WT nolig simulation without membrane, it was open in 97% of the frames with an average bottleneck radius of 1.3 Å. We considered a tunnel to be present if it occurred in at least five frames during the production phase. According to this limitation, the WT inhibitor system presented the highest number of tunnels (**Fig S4.11**).

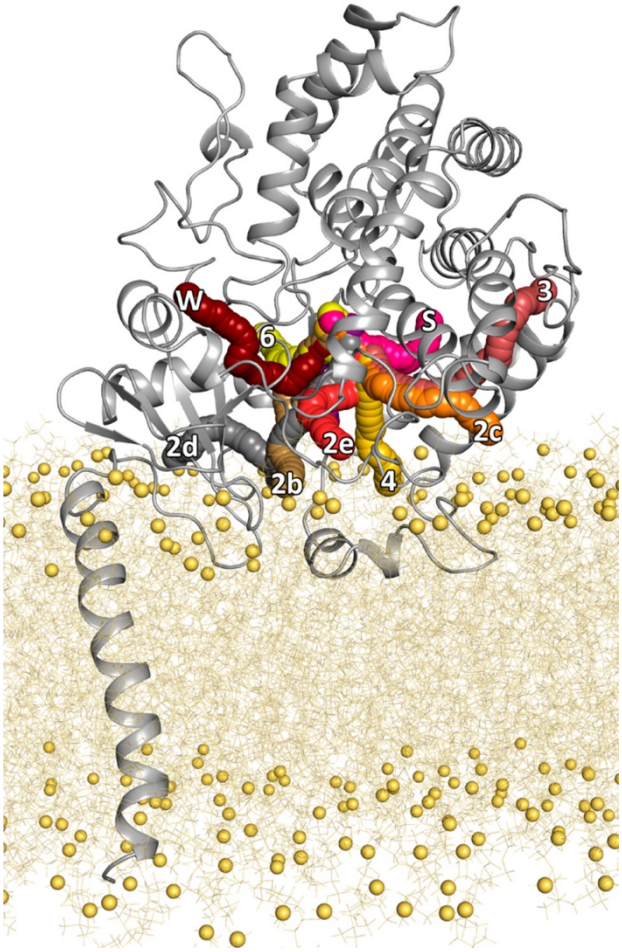


Figure 4.5 **Distribution of tunnels relative to the membrane in CYP2D6 WT.** The system is shown after 300 ns production phase. The membrane is colored in a transparent yellow, while membrane phosphorus atoms are shown in sphere representation. Water molecules are not shown for simplicity.

In total, we observed 13 of the 14 known tunnels in this system. Together with the observations on tunnel 6, this indicates an effect of the membrane on the diversity of tunnels. Berka et al. similarly observed a lower total number of tunnels in the presence of

the membrane for CYP2D6 [20]. An effect of lipids on the tunnels is likely, since the majority are in close vicinity to the membrane-binding regions of the protein. Experiments showed a stabilizing effect on the protein caused by the presence of membrane lipids [56]. This offers a possible explanation for the observations regarding the number of tunnels in the membrane-anchored CYP2D6 models. CYP2D6*2 showed a similar tunnel profile compared to the wild-type (**Figure 4.6**) approving their similar enzymatic activity. It is the only variant that presented an equal degree of opening for tunnel 2c compared to the wild-type. CYP2D6*4 exhibited a similar diversity of tunnels compared to the wild-type, even though the allele is associated with the PM phenotype. However, a closer look at the data revealed that the important tunnels overall showed a low open frequency during the production phase (**Table S4.6**). CYP2D6*4 showed the lowest degree of opening for tunnel 2b compared to all other variants. Even though tunnel 2b was one of the most important overall paths according to our data, our observations related to enzyme tunnels do not sufficiently explain the complete absence of metabolic activity in this variant. This implies the contribution of additional factors restraining the function of this variant on the level of amino acid mutations.

The complete absence of tunnel 2c in CYP2D6*10 was caused by the close positioning of the BC loop toward the G-helix which blocked this tunnel during the simulations. In fact, we observed this to be the main closing mechanism in the variants that showed a low opening frequency of this tunnel (**Figure 4.6**). Such a closing movement could also impair active site hydration, since we observed tunnel 2c to be one of the major pathways in this context. The only tunnel that showed a consistent open state in CYP2D6*10 was tunnel 2b. Indeed, this variant shows a substrate-dependent decreased activity [7] indicating the possibility that some of the tested substrates would prefer tunnel 2c as an access path. De Waal et al. observed an increased presence of tunnel 2b as well as a significantly lower occurrence of tunnel 2c in CYP2D6*17 compared to the wild-type [12]. Our results confirm their observations, since we also discovered a low occurrence of tunnel 2c as well as an increased opening frequency combined with a small enlargement of tunnel 2b compared to the wild-type (**Tables S4.6 and S4.8**). Since tunnel 2c was rarely present, the substrate-dependent

decreased activity [57,58] of this variant could again be explained by the nearly complete absence of this tunnel without any compensatory paths. Experiments measuring the affinity of different substrates related to these two variants could give more insights in this context. The mutation T107I in CYP2D6*17 is located at the center of the B-C loop, and thus lining tunnel 2e. This tunnel showed the lowest opening frequencies in CYP2D6*17, even though it occurred in over 50% of the frames (**Figure 4.6 and Table S4.6**).

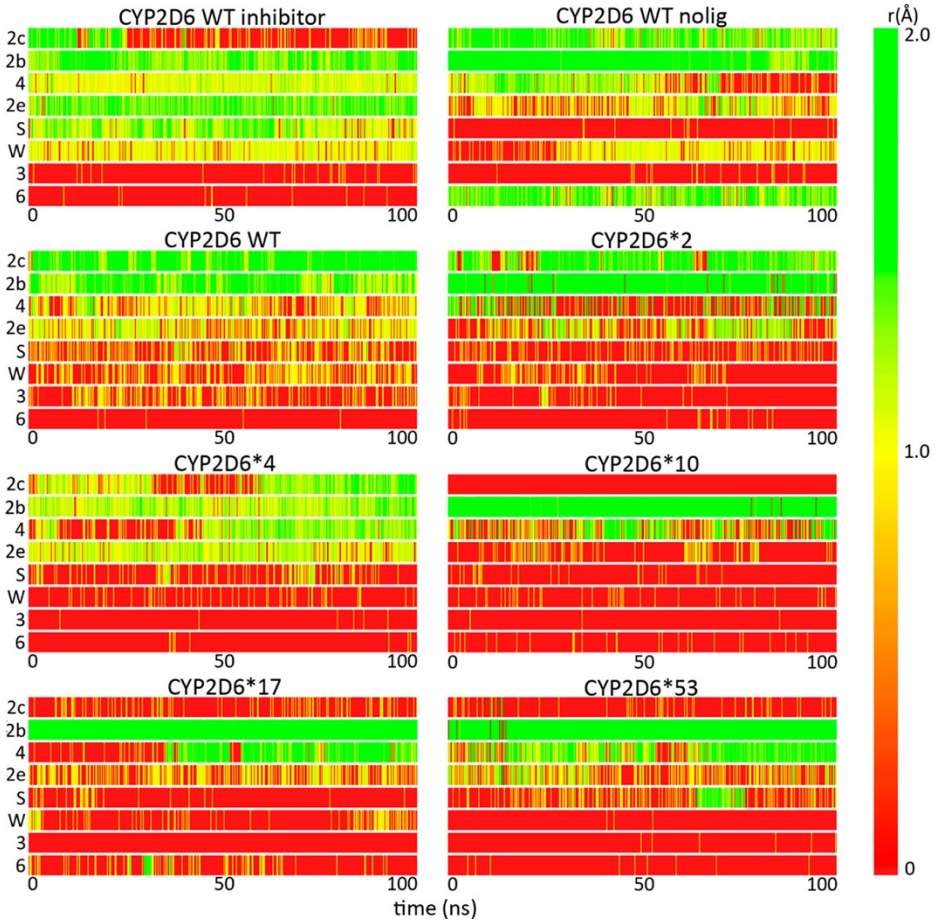


Figure 4.6 **Opening degree of enzyme tunnels in CYP2D6.** The heat maps are based on bottleneck radii showing the degree of opening for eight of the most prominent tunnels. We considered the production phases including 500 frames for the analysis. This corresponds to a simulation time of 100 ns.

This indicates that the mutation at position 107 could be locally altering in the properties of the B-C loop due to the loss of a hydroxyl group (next section). Tunnel 3 was identified as a secondary exit route for ligands by the use of random expulsion MD [59]. Both CYP2D6*10 and CYP2D6*17 did exhibit this tunnel in less than five frames during the production phase. The near absence of tunnel 3 could be contributing to the altered activity profiles of these

variants. Even though CYP2D6*53 showed only a minimal opening of tunnel 2c, it compensated this limitation by a 2-fold enlarged bottleneck radius of tunnel 2b compared to the wild- type. This significant enlargement can be explained by the fact that the F120I mutation replaces one of the major bottlenecks residues (**Table S4.9**) in tunnel 2b. The reduced steric hindrance of isoleucine compared to phenylalanine leaves more space for potential ligands to access the active site.

Based on our data, we propose that Phe120 forms a gate together with Phe483 that can be opened through a wing motion [25] regulating the opening of tunnel 2b (**Figure 4.7**).

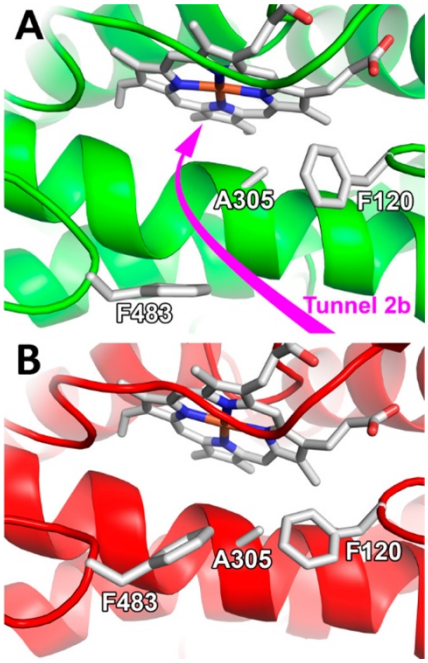


Figure 4.7 **Overview of tunnel gating.** Tunnel 2b shown in **(A)** an open state and **(B)** a closed state. The heme moiety as well as the involved residues are shown. The path of tunnel 2b in the open state is indicated by a pink arrow.

Previous studies also showed tunnel 2b to be the most open in this variant [12]. An enlargement of this tunnel could also enhance its throughput of water molecules since we measured it to be the major pathway of active site hydration in the wild-type. In our study, Phe120 not only was determined as a major bottleneck residue for tunnel 2b but was also shown to be the most prominent bottleneck residue in several other tunnels during the production phase (**Table S4.9**). It was involved in the regulation of four other tunnels including tunnel 4 suggesting similar effects on them. Indeed, our data shows that tunnel 4 presented the highest opening frequency for this variant (**Table 4.2**). Even though Phe120

was determined to be a bottleneck residue for tunnel 2c and the water tunnel, it only maintained this function in a small amount of frames during the production phase.

We already discussed different mechanisms for the regulation of tunnel 2c to take place on a globular level. Therefore, we believe this is the reason for missing effects of the F120I mutation in the gating of other tunnels. The second amino acid mutation in this variant is a replacement of alanine with serine at position 122 that is located at the B-C loop. This leads to the possibility that tunnel 2e could be affected by this mutation since its path is localized within this region. In contrast, we did not observe a significant difference in the opening of this tunnel compared to the wild-type indicating a secondary role for this mutation. As expected, the F120I mutation did not influence the opening of tunnel 2e due to its low frequency as a bottleneck residue (below 1% of the frames). Overall, our results support the fact that this variant shows an increased metabolic rate depending on the substrate. Even though, other amino acids prone to mutations at positions 91, 94, 296, and 486 are involved in lining various tunnels, the only major difference we observed was a slightly decreased opening of the solvent tunnel if the S486T mutation was present. Since this mutation is located at the mouth of the tunnel, these observations could be explained by steric effects related to the introduction of an additional methyl group at this position. For the other affected residues, we did not observe any significant differences related to the tunnels that could be correlated to the enzyme activity or active site hydration.

The replica simulations showed a good convergence regarding the average bottleneck radii (**Table S4.8**). The deviations among the production simulations and the replicas were within the standard deviation of the respective average values. The important bottleneck residues determined in the production phases remained stable except for Val308 and Val370 that showed a reduction in the number of tunnels which were bottlenecking in one replica simulation each (**Table S4.9**). This might have been caused by a rather high degree of freedom of this valine residue compared to the aromatic residues that are usually involved in tunnel gating [25]. Additionally, these two residues are located at the end of the tunnels close to the heme and therefore might be exchangeable to some degree. Despite the stable bottleneck radii of the replicas, the largest differences between the production and the

replica simulations were found in the analysis of the tunnel opening frequency based on the opening threshold of 1.2 Å (**Table S4.6**). While the variants CYP2D6*10 and CYP2D6*17 did not show any large differences, the values of some tunnels in the four other simulations showed significant differences. The replica simulations with CYP2D6 WT showed higher opening frequencies for tunnel 2b and tunnel 2c compared to the production simulations. This increase further substantiates the functionality of the wild-type enzyme structure. The system with the variant CYP2D6*2 previously showed rather high RMSD values in the assessment of the structural convergence of the replicas that were potentially caused by the increased flexibility of this variant. This was also reflected by the tunnel opening frequency of tunnel 2b and tunnel 2e in one replica simulation each. Despite the difference in these two values, the average values of the two replica simulations of CYP2D6*2 were shown to be quite similar to the production simulation. We could observe the largest deviation of tunnel opening for CYP2D6*4 in the respective replica simulations. As previously mentioned, we suspect this to be caused by the generally unstable fold of this variant. The slight deviations regarding tunnel 4 in several variants such as CYP2D6*53 might be related to the flexibility of the respective tunnel entrance, which is defined by the flexible F-G loop and might therefore open and close with an increased frequency.

Structural Comparison of Variants. The RMSD and RMSF profiles during the CYP2D6*2 simulation remained fairly comparable to the ones of the wild-type (**Figures S4.2 and S4.3**). This was to be expected, since the S486T mutation is not known to cause any significant alterations to the structure or function of CYP2D6 [57]. However, the R296C mutation also occurs in alleles with decreased function and is shown to cause structural instabilities in experiments [60]. Indeed, our analysis of the average number of hydrogen bonds showed both CYP2D6*2 and CYP2D6*17 to have lost one hydrogen bond because of the switch from arginine to cysteine in the production as well as the replica simulations (**Table S4.10**). This indicates that the fold of the enzyme could be locally destabilized. However, the combination of R296C with S486T does not influence the function in the CYP2D6*2 variant, and our previously mentioned data also indicates a function similar to the wild-type. In

CYP2D6*1, the mutations at positions 296 and 486 are combined with the mutation T107I which is suspected to be contributing to the altered activity of this variant [57]. Bapiro et al. determined that the mutation at position 107 alone cannot cause the observed reduction of activity in CYP2D6*17 [58]. It seems that the combination of all three mutations is the cause for the altered properties of this variant [7, 57]. Likewise, our results from the hydrogen bond analysis do support this data. The loss of the hydroxyl group caused by the mutation from threonine to isoleucine also resulted in the loss of a hydrogen bond over the average of the trajectory. In a different computational study with CYP2D6*17 it was shown that the mutations T107I and R296C caused a lower number or a complete loss of hydrogen bonds [35]. It is debated if residue Thr107 is involved in ligand binding and in this way alters the active site structures [2, 57, 58].

The analysis of the native crystal structure (PDB ID: 3TDA) revealed a distance of 5.6 Å to the closest heavy atom of the bound inhibitor prinomastat (**Figure 4.8**) indicating a rather weak interaction.

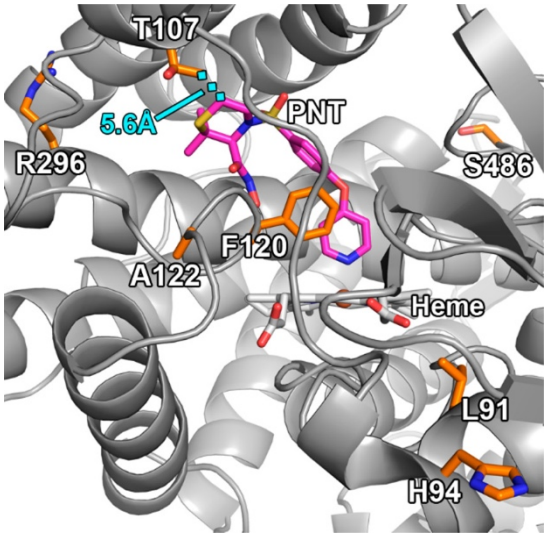


Figure 4.8 **Active site and nearby regions in the native crystal structure of CYP2D6.** The mutated amino acids are shown in orange, while the ligand prinomastat (PNT) is shown in pink. The closest distance between two heavy atoms of T107 and the ligand is shown in cyan.

This crystal structure therefore does not directly support this hypothesis. In addition, the estimated active site volume also did not indicate an altered active site structure (**Table S4.11**). The mutation at position 486 is known to have no significant impact on the enzyme activity in CYP2D6*17 [58]. Additionally, all mutations in CYP2D6*17 are located at substrate

recognition sites (SRS) providing another indication for the substrate-dependency of the observed effects [57,58]. Contrary to previous observations in the absence of the membrane [12], our data shows that CYP2D6*17 reached its RMSD equilibrium faster than any other system. Other groups previously discussed that CYPs with faster equilibration times display higher stability in vitro [12]. This points toward a more stable structure of this variant in comparison to others. The RMSF profile of this variant displayed a decreased flexibility in the region of the F-G loop essential for the formation of enzyme tunnels of the subclass 2 in our structures. This rigidity could disable substrate uptake, since it is known that residues in enzyme tunnels need to sterically adapt in order to allow ligand transport [12, 24]. On the other hand, the rigidity could result in a more stable global structure. It is worth mentioning that we observed higher flexibility on the RMSF profile in the region of helix J' contrary to the observations of a generally more rigid structure. The P34S mutation is widespread in individuals with a reduced or absent functionality of CYP2D6 [41]. As previously mentioned, it is located in a highly conserved region of the protein [8,60] and it might serve as a hinge between the transmembrane anchor and the catalytic domain [7].

Experimental data suggest that this region plays an important role for correct folding and heme incorporation [61]. This was supported by two other groups who measured lower expression levels in the presence of the P34S mutation alone [60, 62]. As previously mentioned, the complete absence of metabolic activity in CYP2D6*4 is thought be caused by the mutation from guanine to adenine at position 1934 resulting in incorrect splicing. Unfortunately, our computational setup does not allow for investigation of effects taking place on the level of protein biosynthesis but still allowed the investigation of the related amino acid changes that are sufficient to abolish enzymatic activity independent of splicing [9]. We visually compared the fold in the region of the N-terminus for both alleles containing the P34S mutation. CYP2D6*4 was shown to have an altered structure between residues Arg25 and Asp67 confirming the experimentally observed phenomena. In CYP2D6*10, on the other hand, we observed fewer alterations in this region. This suggests that the observed differences in CYP2D6*4 might not be explained by the P34S mutation alone. Indeed, experiments showed that the mutation at position 34 does not suffice to completely

abolish the enzymatic function [62]. The CYP2D6*10 allele is an impeccable example for this observation, since it harbors the P34S mutation but still shows enzymatic activity [2]. The analysis of the enzyme tunnels delivered evidence for the observed activity profile. Several tunnels presented lower opening frequencies compared to the wild-type suggesting a decreased accessibility of ligands to the active site. In addition, we calculated the average heme tilt angle during the production phases as it is presented in **Table 4.2**. We observed a value of 65 ± 5° for CYP2D6*10 which significantly differs from the values in the wild-type model, even though they are still within the experimentally determined boundaries [43]. Such a pronounced change in the heme tilt angle can impair the interaction with the POR32 and also points toward structural defects. We determined the same parameter in CYP2D6*2 to be 59 ± 4° suggesting that certain discrepancies in this range are tolerable regarding the enzymatic function. In the literature it is proposed that CYP2D6*10 shows an unstable fold, and in vitro experiments showed thermal instability for this variant at 39.5 °C [63]. The RMSD of CYP2D6*10 during the production run also presented higher values compared to other variants suggesting a certain degree of instability, even though the differences were rather small. In the N-terminal region, this variant displayed an increased flexibility in the vicinity of residue 60 further confirming our observations related to the structural abnormalities. We propose that the observed effects in this variant are mainly caused by the P34S mutation, since it is only accompanied by the mutation S486T that previously could not be associated with any aberrations. The CYP2D6*4 variant also harbors the P34S mutation combined with three others including S486T. The literature provides only little information concerning the other two mutations L91M and H94R [41]. An analysis on genetic polymorphisms in the Uygur population revealed another variant of CYP2D6 harboring exactly these two mutations [64]. The results of this study indicate that both mutations can also occur in individuals with normal enzymatic activity, although more data will be needed to confirm this observation. Upon the mutation of histidine to arginine, the amino acid at position 94 retains its basic properties. This is also reflected in predictions that were made for these mutations using two online tools, where only L91M was shown to be capable of altering the protein activity [64].

Figure 4.8 shows the positions 91 and 94 to be located far away from the active site suggesting a role in substrate recognition rather than in substrate binding. This also became apparent with the sequence alignment to CYP2C5 as we identified the SRS in CYP2D6 [58]. The results showed that both Leu91 and His94 are flanking SRS1 supporting our suggestion. Overall, we propose that the mutation P34S is the main cause for the observed differences in CYP2D6*4 with a possible secondary impact by the L91M mutation. The impact of the P34S mutation has also been shown in experiments with different DNA constructs [9]. On the RMSF profile of CYP2D6*4 we identified similar additional peaks to the ones found for CYP2D6*10 in the N-terminal region. This again confirms our visual observations related to the altered fold of CYP2D6*4. Our results regarding the active site volume estimated with POVME provide further indications toward the loss of activity in this variant (**Table 4.2 and Table S4.11**). In the last 100 frames of the MD production phase, the volume of the active site cavity of CYP2D6*4 was relevantly lower compared to the wild-type or other variants. This further points toward the hypothesis that not one single mutation but the combination of all four mutations is responsible for the alteration of the structure of the enzyme. Differences in the tunnel opening frequency (**Table S4.6**) compared to the wild-type and the similar variant CYP2D6*10 provide additional evidence for this hypothesis.

Regarding the active site volume, the replica simulations mostly showed deviations that were within the standard deviation of the average values (**Table S4.11**). One of the replica simulations of CYP2D6*2 was an exception with a comparably high value. A visual inspection of the residues in the active site in two frames with different volumes revealed Phe219, Gln244, and Phe483 to vary in position potentially contributing to the deviation between this replica and the production simulation. The differences in the hydrogen bond analysis between the replicas and the production simulations were within the error of the standard deviation (**Table S4.10**). The tunnel analysis already showed a significant enlargement of tunnel 2b in CYP2D6*53, yet additional factors could lead to the enhanced catalytic activity of this variant. The residue Phe120 is known not to be critical for the enzymatic activity of the variant, even though it directly interacts with ligands over π−π interactions. In fact, it is

proposed that the absence of this specific interaction combined with a reduced steric hindrance could lead to different poses of substrates in the active site and thus cause changes in the metabolic profile [65,66]. These assumptions are based on findings with a variant featuring the mutation F120A. This mutation replaces the bulky phenylalanine residue by a smaller, similarly hydrophobic residue making it comparable to the F120I mutation. Interestingly, the F120A variant of CYP2D6 is able to metabolize quinidine which is an inhibitor in the wild-type form [67]. Also, the regiospecificity of certain metabolic reactions with common drugs changes compared to the wild-type [66].

The pose of prinomastat in the crystal structure showed Phe120 to be oriented toward the aromatic groups of the ligand. In addition, CYP2D6*53 also showed the highest active site volume of all simulations. We therefore support the hypothesis that the mutation to a less bulky amino acid could lead to a larger conformational freedom of the ligand inside the active site cavity based on our results regarding the active site volume. Still, additional molecular docking studies and MD simulations exploring the binding modes of substrates will be needed to confirm this hypothesis. The RMSD and RMSF profiles of CYP2D6*53 suggest a rather flexible structure compared to the wild-type. There are additional peaks denoting increased mobility at residues 145, 166, and 178 and compared to the wild-type. The enhanced flexibility could contribute to its ability to perform a broader spectrum of metabolic reactions.

Conclusion

In this study, we employed all-atom molecular dynamics simulations in order to identify differences between the wild- type and five allelic variants CYP2D6*2, CYP2D6*4, CYP2D6*10, CYP2D6*17, and CYP2D6*53. We built full-length membrane-anchored models, equilibrated them, and focused on differences among them considering enzyme tunnels as well as the overall fold. We validated our membrane-bound model with available experimental data and other published models of membrane-bound CYPs. In our model, the protein mediated its contact to the membrane with the transmembrane anchor, A' anchor, A helix, and the FG loop over specific protein−membrane interactions. We showed

that the binding site for the POR in our models was accessible from the cytosol and not hidden in the membrane. The water tunnel as well as the solvent tunnel were able to transport water molecules but were shown to be only minor paths for active site hydration in our simulation conditions. To our best knowledge, we are the first group to identify tunnel 6 in CYP2D6 offering an additional possible path for ligands. As expected, the CYP2D6*2 variant demonstrated a similar tunnel profile compared to the wild-type and no major differences in the overall structure except for the loss of a hydrogen bond due to the exchange of arginine with cysteine at position 296. The overall structure in the nonfunctional CYP2D6*4 variant was shown to be altered considering our data regarding the active site volume and the change in the N- terminal fold. We propose the P34S mutation to be the major cause for the observed alterations based on our results and the available literature. The structural differences in this variant correlated with a generally decreased opening of the most prominent ligand tunnels. CYP2D6*10 is another variant that harbors the P34S mutation and showed similar changes in the N-terminal fold as CYP2D6*4 providing additional evidence for the structural impact of this mutation. Furthermore, the B-C loop in CYP2D6*10 and CYP2D6*17 shifted toward the helix G leading to the closure of tunnel 2c without any compensatory pathways. We believe this phenomenon leads to the experimental observations of substrate-dependent decreased activity in both variants. Although it is suggested in the literature, the T107I mutation in CYP2D6*17 was not shown to have an impact on ligand binding in our analysis. Instead, it led to the loss of a hydrogen bond in addition to the one it lost due to the R296C mutation. Even though we did only observe a minor opening of tunnel 2c in CYP2D6*53, this variant could compensate this restriction by a 2-fold enlargement of tunnel 2b likely due to the mutation of phenylalanine with isoleucine at position 120. Additionally, the two amino acid mutations in this variant led to an enlarged active site volume providing another possible explanation for the altered metabolic profile. To summarize, this supports the discussed and pending designation as a cause for the UM phenotype. In the future, studies employing site-directed mutagenesis experiments with the allelic variants combined with the exposure to differently sized and shaped ligands could provide a clearer picture for some of the findings. Experimental

determination of the 3D structure of allelic variants, as well as the full-length protein, could help to validate our results. In conclusion, we discovered several structural differences in the variants that could explain their altered metabolic activity compared to the wild-type on an atomic level with replica simulations displaying satisfying convergence. In the age of personalized medicine, the conclusions from this study could be integrated into virtual screening protocols, which are almost exclusively guided by the molecular docking of compounds into the active site, thereby neglecting the influence of tunnels. Furthermore, atomistic insights into the structure and function of the enzyme tunnels could be of interest in the field of enzyme engineering where CYP2D6 is being used as a biocatalyst and biosensor.

Acknowledgment

M. Smieško and C.Don designed the research with contribution of A. Fischer. Mr. Fischer, as part of his M.Sc dissertation, carried out the study, analysis and had the largest contribution in writing of the manuscript.

REFERENCES

1. Tracy TS, Chaudhry AS, Prasad B, Thummel KE, Schuetz EG, Zhong X -b., et al. Interindividual Variability in Cytochrome P450-Mediated Drug Metabolism. Drug Metab Dispos [Internet]. 2016;44(3):343–51.

2. Ingelman-Sundberg M. Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity. Pharmacogenomics J [Internet]. 2005;5(1):6–13.

3. Kumar S. Engineering Cytochrome P450 Biocatalysts for Biotechnology, Medicine, and Bioremediation. Vol. 6, Expert opinion on drug metabolism & toxicology. 2010. p. 115–31.

4. Caudle KE, Dunnenberger HM, Freimuth RR, Peterson JF, Burlison JD, Whirl-Carrillo M, et al. Standardizing terms for clinical pharmacogenetic test results: Consensus terms from the Clinical Pharmacogenetics Implementation Consortium (CPIC). Genet Med. 2017;19(2):215–23.

5. Ma MK, Woo MH, McLeod HL. Genetic basis of drug metabolism. Am J Heal Pharm [Internet]. 2002;59(21):2061–9.

6. Sakuyama K, Sasaki T, Ujiie S, Obata K, Mizugaki M, Ishikawa M, et al. Functional Characterization of 17 CYP2D6 Allelic Variants (CYP2D6 . 2 , 10 , 14A – B , 18 , 27 , 36 , 39 , 47 – 51 , 53 – 55 , and 57). Pharmacology. 2008;36(12):2460–7.

7. Zhou S-F. Polymorphism of Human Cytochrome P450 2D6 and Its Clinical Significance. Clin Pharmacokinet [Internet]. 2012;48(11):689–723.

8. Han KM, Chang HS, Choi IK, Ham BJ, Lee MS. CYP2D6 P34S polymorphism and outcomes of escitalopram treatment in Koreans with major depression. Psychiatry Investig. 2013;10(3):286–93.

9. Kagimoto M, Heim M, Kagimoto K, Zeugin T, Meyer U. Multiple Mutations of the Human Cytochrome P450IID6 (CYP2D6) in Poor Metabolizers of Debrisoquine. J Biol Chem. 1990;265(26):17209–14.

10. Wright GEB, Niehaus DJH, Drögemöller BI, Koen L, Gaedigk A, Warnich L. Elucidation of CYP2D6 genetic diversity in a unique African population: Implications for the future application of pharmacogenetics in the xhosa population. Ann Hum Genet. 2010;74(4):340–50.

11. Contreras A V, Monge-Cazares T, Alfaro-Ruiz L, Hernandez-Morales S, Miranda-Ortiz H, Carrillo-Sanchez K, et al. Resequencing, haplotype construction and identification of novel variants of CYP2D6 in Mexican Mestizos. Pharmacogenomics. 2011 May;12(5):745–56.

12. De Waal PW, Sunden KF, Furge LL. Molecular dynamics of CYP2D6 polymorphisms in the absence and presence of a mechanism-based inactivator reveals changes in local flexibility and dominant substrate access channels. PLoS One. 2014;9(10).

13. Don CD, Smieško M. Out-compute drug side effects: Focus on cytochrome P450 2D6 modeling. WIREs Comput Mol Sci. 2018;e1366.

14. Rowland P, Blaney FE, Smyth MG, Jones JJ, Leydon VR, Oxbrow AK, et al. Crystal structure of human cytochrome P450 2D6. J Biol Chem. 2006;281(11):7614–22.

15. Butler CR, Ogilvie K, Martinez-Alsina L, Barreiro G, Beck EM, Nolan CE, et al. Aminomethyl-Derived Beta Secretase (BACE1) Inhibitors: Engaging Gly230 without an Anilide Functionality. J Med Chem. 2017;60(1):386–402.

16. Brodney MA, Beck EM, Butler CR, Barreiro G, Johnson EF, Riddell D, et al. Utilizing Structures of CYP2D6 and BACE1 Complexes to Reduce Risk of Drug-Drug Interactions with a Novel Series of Centrally Efficacious BACE1 Inhibitors. J Med Chem. 2015;58(7):3223–52.

17. Wang A, Savas U, Hsu M, Stout CD, Johnson EF. Crystal Structure of Human Cytochrome P450 2D6 with

Prinomastat Bound. J Biol Chem. 2012;287(14):10834–43.
18. Wang A, Stout CD, Zhang Q, Johnson EF. Contributions of ionic interactions and protein dynamics to cytochrome P450 2D6 (CYP2D6) substrate and inhibitor binding. J Biol Chem. 2015;290(8):5092–104.

19. Monk BC, Tomasiak TM, Keniya M V, Huschmann FU, Tyndall JDA, O'Connell JD, et al. Architecture of a single membrane spanning cytochrome P450 suggests constraints that orient the catalytic domain relative to a bilayer. Proc Natl Acad Sci [Internet]. 2014;111(10):3865–70.

20. Berka K, Paloncýová M, Anzenbacher P, Otyepka M. Behavior of human cytochromes P450 on lipid membranes. J Phys Chem B. 2013;117(39):11556–64.

21. Cojocaru V, Winn PJ, Wade RC. The ins and outs of cytochrome P450s. Biochim Biophys Acta - Gen Subj [Internet]. 2007;1770(3):390–401.

22. Jeřábek P, Florián J, Martínek V. Lipid molecules can induce an opening of membrane-facing tunnels in cytochrome P450 1A2. Phys Chem Chem Phys [Internet]. 2016;18(44):30344–56.

23. Baylon JL, Lenov IL, Sligar SG, Tajkhorshid E. Characterizing the membrane-bound state of cytochrome P450 3A4: Structure, depth of insertion, and orientation. J Am Chem Soc. 2013;135(23):8542–51.

24. Berka K, Hendrychová T, Anzenbacher P, Otyepka M. Membrane position of ibuprofen agrees with suggested access path entrance to cytochrome P450 2C9 active site. J Phys Chem A. 2011;115(41):11248–55.

25. Gora A, Brezovsky J, Damborsky J. Gates of enzymes. Chem Rev. 2013;113(8):5871–923.

26. Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, et al. CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. PLoS Comput Biol. 2012;8(10):23–30.

27. Salsbury Jr FR. Molecular Dynamics Simulations of Protein Dynamics and their relevance To Drug Discovery. Curr Opin Pharmacol [Internet]. 2011;10(6):738–44.

28. Werck-Reichhart D, Feyereisen R. Cytochromes P450: a success story. Genome Biol [Internet]. 2000;1(6):REVIEWS3003.

29. Fishelovitch D, Shaik S, Wolfson HJ, Nussinov R. How does the reductase help to regulate the catalytic cycle of cytochrome P450 3A4 using the conserved water channel? J Phys Chem B. 2010;114(17):5964–70.

30. Bridges A, Bruenke L, Chang Y-T, Vasker IA, Loew G, Waskell L. Identification of the Binding Site on Cytochrome P450 2B4 for Cytochrome b5 and Cytochrome P450 Reductase. J Biol Chem. 1998;273(27):17036–49.

31. Ducassou L, Dhers L, Jonasson G, Pietrancosta N, Boucher JL, Mansuy D, et al. Membrane-bound human orphan cytochrome P450 2U1: Sequence singularities, construction of a full 3D model, and substrate docking. Biochimie. 2017;140:166–75.

32. Cojocaru V, Balali-Mood K, Sansom MSP, Wade RC. Structure and dynamics of the membrane-bound cytochrome P450 2C9. PLoS Comput Biol. 2011;7(8).

33. Madhavi Sastry G, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des. 2013;27(3):221–34.

34. Bateman A, Martin MJ, O'Donovan C, Magrane M, Alpi E, Antunes R, et al. UniProt: The universal protein knowledgebase. Nucleic Acids Res. 2017;45(D1):D158–69.

35. Handa K, Nakagome I, Yamaotsu N, Gouda H, Hirono S. In Silico Study on the Inhibitory Interaction of Drugs with Wild-type CYP2D6.1 and the Natural Variant CYP2D6.17. Drug Metab Pharmacokinet [Internet]. 2014;29(1):52–60.

36. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI. OPM: Orientations of proteins in membranes database. Bioinformatics. 2006;22(5):623–5.

37. Bowers K, Chow E, Xu H, Dror R, Eastwood M, Gregersen B, et al. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. ACM/IEEE SC 2006 Conf [Internet]. 2006;(November):43–43.

38. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. J Mol Biol [Internet]. 2001;305(3):567–80.

39. Sapay N, Guermeur Y, Deléage G. Prediction of amphipathic in-plane membrane anchors in monotopic proteins using a SVM classifier. BMC Bioinformatics [Internet]. 2006;7:255.

40. Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, et al. PredictProtein - An open resource for online prediction of protein structural and functional features. Nucleic Acids Res. 2014;42(W1):337–43.

41. Gaedigk A, Ingelman-Sundberg M, Miller NA, Leeder JS, Whirl-Carrillo M, Klein TE. The Pharmacogene Variation (PharmVar) Consortium: Incorporation of the Human Cytochrome P450 (CYP) Allele Nomenclature Database. Clin Pharmacol Ther. 2017;0(0):4–6.

42. 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(14):4629–37.

43. Ohta Y, Kawato S, Tagashira H, Takemori S, Kominami S. Dynamic Structures of Adrenocortical Cytochrome P-450 in Proteoliposomes and Microsomes: Protein Rotation Study. Biochemistry. 1992;31(50):12680–7.

44. Bayburt TH, Sligar SG. Single-molecule height measurements on microsomal cytochrome P450 in nanometer-scale phospholipid bilayer disks. Proc Natl Acad Sci. 2002;99(10):6725–30.

45. De Lemos-Chiarandini C, Frey AB, Sabatini DD, 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(2):209–19.

46. Brown CA, Black SD. Membrane Topology of Mammalian Cytochromes P-450 from Liver Endoplasmic Reticulum. J Biol Chem. 1989;264(15):4442–9.

47. Durrant JD, Votapka L, Sørensen J, Amaro RE. POVME 2.0: An enhanced tool for determining pocket shape and volume characteristics. J Chem Theory Comput. 2014;10(11):5047–56.

48. Okonechnikov K, Golosova O, Fursov M, Varlamov A, Vaskin Y, Efremov I, et al. Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics. 2012;28(8):1166–7.

49. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22(22):4673–80.

50. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189–91.

51. Kozlikova B, Sebestova E, Sustr V, Brezovsky J, Strnad O, Daniel L, et al. CAVER Analyst 1.0: Graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. Bioinformatics. 2014;30(18):2684–5.

52. Scott EE, Wolf CR, Otyepka M, Humphreys SC, Reed JR, Henderson CJ, et al. The role of protein-protein and protein-membrane interactions on P450 function. Drug Metab Dispos. 2016;44(4):576–90.

53. Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. J Phys Chem A. 2001;105(43):9954–60.

54. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982;157(1):105–32.

55. Li W, Shen J, Liu G, Tang Y, Hoshino T. Exploring coumarin egress channels in human cytochrome p450 2a6 by random acceleration and steered molecular dynamics simulations. Proteins Struct Funct Bioinforma. 2011;79(1):271–81.

56. Ahn T, Yun CH, Oh DB. Involvement of nonlamellar-prone lipids in the stability increase of human cytochrome P450 1A2 in reconstituted membranes. Biochemistry. 2005;44(25):9188–96.

57. Oscarson M, Hidestrand M, Johansson I, Ingelman-Sundberg M. A combination of mutations in the CYP2D6*17 (CYP2D6Z) allele causes alterations in enzyme function. Mol Pharmacol. 1997;52:1034–40.

58. Bapiro TE, Hasler JA, Ridderström M, Masimirembwa CM. The molecular and enzyme kinetic basis for the diminished activity of the cytochrome P450 2D6.17 (CYP2D6.17) variant: Potential implications for CYP2D6 phenotyping studies and the clinical use of CYP2D6 substrate drugs in some African populations. Biochem Pharmacol. 2002;64(9):1387–98.

59. Lüdemann SK, Lounnas V, Wade RC. How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways. J Mol Biol. 2000;303(5):813–30.

60. Kim J, Lim YR, Han S, Han JS, Chun YJ, Yun CH, et al. Functional influence of human CYP2D6 allelic variations: P34S, E418K, S486T, and R296C. Arch Pharm Res. 2013;36(12):1500–6.

61. Yamazaki S, Sato K, Suhara K, Sakaguchi M, Mihara K, Omura T. Importance of the proline-rich region following signal-anchor sequence in the formation of correct conformation of microsomal cytochrome P-450s. J Biochem. 1993;114(5):652–7.

62. Shiraishi T, Hosokawa M, Kobayashi K, Tainaka H, Yamaura Y, Taguchi M, et al. Effects of G169R and P34S substitutions produced by mutations of CYP2D6*14 On the functional properties of CYP2D6 expressed in V79 cells. Drug Metab Dispos. 2002;30(11):1201–5.

63. Shen H, He MM, Liu H, Wrighton SA, Wang L, Guo B, et al. Comparative metabolic capabilities and inhibitory profiles of CYP2D6.1, CYP2D6.10, and CYP2D6.17. Drug Metab Dispos. 2007 Aug;35(8):1292–300.

64. He X, He N, Ren L, Ouyang Y, Zhang N, Ma Y, et al. Genetic polymorphisms analysis of CYP2D6 in the Uygur population. BMC Genomics. 2016;17(1):1–9.

65. Masuda K, Hashimoto H, Tamagake K, Okuda Y, Tsuzuki D, Isobe T, et al. Changes in the Enzymatic Properties of CYP2D6 by the Substitution of Phenylalanine at Position 120 by Alanine. 2004;50(5):503–10.

66. Flanagan JU, Maréchal J-D, Ward R, Kemp CA, McLaughlin LA, Sutcliffe MJ, et al. Phe120 contributes to the regiospecificity of cytochrome P450 2D6: mutation leads to the formation of a novel dextromethorphan metabolite. Biochem J [Internet]. 2004;380(Pt 2):353–60.

67. McLaughlin LA, Paine MJI, Kemp CA, Maréchal JD, Flanagan JU, Ward CJ, 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(46):38617–24.

Supporting Information

Table S4.1 **The stages of Desmond MD relaxation and simulation prior to the actual MD simulation.**

Desmond Stage	Procedure
Stage 1	Task (reading files, initializing parameters)
Stage 2	Simulate, Brownian Dynamics NVT, T = 10 K, small time steps, and restraints on solute heavy atoms, 100 ps
Stage 3	Simulate, NVT, T = 10 K, small time steps, and restraints on solute heavy atoms, 12 ps
Stage 4	Simulate, NPT, T = 10 K, and restraints on solute heavy atoms, 12 ps
Stage 5	Solvate pocket
Stage 6	Simulate, NPT and restraints on solute heavy atoms, 12 ps
Stage 7	Simulate, NPT and no restraints, 24 ps

The seven stages were derived from the log file of a Desmond MD simulation.

Table S4.2 **Predictions for membrane-bound residues from web-based protocols.**

Protocol	Prediction N-terminus
TMHMM v2.0	A5-H24
AmphipaSeek	V7-D21
TMSEG	L6-H24

Three web based protocols were employed to get an estimate of the membrane-bound residues in the transmembrane helix of CYP2D6. The predictions were used to place the helix in a POPC membrane.

Table S4.3 **Experimental validation of the CYP2D6 WT model adapted from Cojocaru et al.**

Residues in CYP2D6 (CYP2C9*)	Result	Correlation	Experimental result*	Experimental method*
1-35 (1-30)	M-HG-C	+	Inaccessible	Site-directed antibody [48]
18-32 (17-28)	M-HG-C [b]	0	Accessible	Site-directed antibody [49]
27-41 (23-37)	HG-C-HG-M	+	Accessible	Site-directed antibody [48]
43-51 (39-47)	M-HG-C	+	Inaccessible	Site-directed antibody [48]
64-75 (60-71)	C-HG-M	+	Accessible	Site-directed antibody [48]
96-101 (92-97)	C	-	Inaccessible	Site-directed antibody [48]
111-122 (107-115)	HG-C	+	Accessible	Site-directed antibody [48]
129-138 (121-130)	C	+	Accessible	Site-directed antibody [48]
194-201 (185-192)	C	+	Accessible	Site-directed antibody [48]
218-229 (210-222)	HG-M	+	Inaccessible	Site-directed antibody [48]
231-238 (224-231)	M-HG	-	Accessible	Site-directed antibody [48]
322-330 (314-322)	C	+	Accessible	Site-directed antibody [48]
404-414 (397-407)	C	+	Accessible	Site-directed antibody [48]
40 (36)	M	+	M	Trp fluorescence quenching [46]
73 (69)	C [a]	0	M	Trp fluorescence quenching [46]
387 (380)	C	-	M	Trp fluorescence quenching [46]
84, 128, 198, 246, 355 (80,120, 189, 239, 347)	C	+	C (HG)	Trp fluorescence quenching [46]
232 (225)	M	-	C (HG)	Trp fluorescence quenching [46]

1. Column: Residues in CYP2D6 data based on sequence alignment to CYP2C9. The corresponding residues in CYP2C9 are shown in brackets.
2. Column: Location of the residues in our model related to the membrane. The order of the location is given from N-terminus to C-terminus. The residues can either be positioned in the membrane (M), headgroup region (HG), or cytosol (C).
3. Column: Correlation of our results with the experimental data: (+) indicates a correlation, (0) indicates inconclusive results, and (-) indicates a missing correlation.
4. Column: Experimental results on the location and accessibility of the related protein regions. Data from site-directed antibodies is divided in "accessible" and "inaccessible", while data from tryptophan fluorescence quenching directly indicates the location of the residues.
5. Column: The experimental methods used to produce the mentioned data on the residue location. The referred references are listed in the main article.

*data directly adapted from Cojocaru et al.[18]
[a] not inside membrane, but buried inside the protein.
[b] fewer than 10% of the residues inside the range match the experimental data.

Table S4.4 **Values of the control simulations.**

System	Simulation time	Comment [a]	Heme tilt angle	Burying depth [b]	Location of residues [c]		Overall Appearance
					Q27	R28	
AnchorSim	200 ns	Used for final system	n/a	n/a	HG	HG	angled, not straight
Anchor1	129 ns	Different burying depth	n/a	n/a	HG	S	angled, straight
Anchor2	40 ns	Different burying depth	n/a	n/a	HG	S	angled, not straight
Anchor3	200 ns	Different burying depth	n/a	n/a	HG	S	angled, not straight
Anchor4	200 ns	Different burying depth	n/a	n/a	HG	HG	angled, straight
Anchor 5	200 ns	Different burying depth	n/a	n/a	HG	HG	angled, not straight
MemAs	100 ns	Used for final system	64.4°	41.4 Å	-	-	
Mem1	100 ns	Different burying depth	77.9°	37.1 Å	-	-	
Mem2	100 ns	Different burying depth	56.1°	39.9 Å	-	-	
CYP2DT WT	300 ns	Used for final system	44.4°	36.1 Å	HG	HG	
Fusion1	47 ns	Different linkage [d]	64.3°	36.3 Å	S	S	
Fusion2	65 ns	Different linkage [d]	39.5°	41.0 Å	HG	HG	
Fusion3	200 ns	Different linkage [d]	50.1°	34.7 Å	HG	HG	anchor moved away from protein

[a] Background of the simulations.

[b] Final value after the whole simulation.

[c] The location of two selected residues at the protein-membrane interface was determined in the last frame of the respective simulation. Divided in HG (head groups) and S (solvent).

[d] Structures at different time points of the AnchorSim simulation were covalently linked to the catalytic domain.

Table S4.5 **RMSD differences of the replica simulations to assess structural convergence.**

Compared Simulations [a]	Backbone RMSD (Å) [b]	Heavy atom RMSD (Å) [b]
Replica 1 to Replica 2	0.59	0.72
Replica 1 to CYP2D6 WT	0.41	0.51
Replica 2 to CYP2D6 WT	0.52	0.67
Average CYP2D6 WT	0.51	0.63
Replica 3 to Replica 4	0.70	0.89
Replica 3 to CYP2D6*2	1.26	1.43
Replica 4 to CYP2D6*2	1.20	1.32
Average CYP2D6*2	1.05	1.21
Replica 5 to Replica 6	0.92	1.00
Replica 5 to CYP2D6*4	1.97	2.02
Replica 6 to CYP2D6*4	1.55	1.64
Average CYP2D6*4	1.48	1.55
Replica 7 to Replica 8	0.92	1.03
Replica 7 to CYP2D6*10	0.64	0.76
Replica 8 to CYP2D6*10	0.74	0.89
Average CYP2D6*10	0.76	0.89
Replica 9 to Replica 10	0.69	0.83
Replica 9 to CYP2D6*17	0.88	1.02
Replica 10 to CYP2D6*17	0.80	0.91
Average CYP2D6*17	0.79	0.92
Replica 11 to Replica 12	0.63	0.84
Replica 11 to CYP2D6*53	0.77	0.89
Replica 12 to CYP2D6*53	0.62	0.70
Average CYP2D6*53	0.67	0.81

To assess the structural convergence of the simulations we compared average structures from the 100 ns trajectories and compared their RMSD on two different atomic levels.

[a] The production phase was used in the assessment for the simulations that were no replicas.

[b] Hydrogen atoms were excluded for this calculation.

Table S4.6 **Tunnel opening frequency calculated for production simulations and replicas.**

Compared simulations	Tunnel opening frequency (%)					
	2b	2c	2e	4	S	W
CYP2D6 WT	73.0	20.2	13.8	9.0	5.4	0.0
Replica 1	97.6	48.6	9.6	19.2	9.2	0.8
Replica 2	82.2	73.4	16.4	16.6	4.4	0.0
Average Replicas	89.9	61.0	13.0	17.9	6.8	0.4
CYP2D6*2	94.2	78.0	23.4	35.2	0.0	0.0
Replica 3	65.2	73.6	21.8	46.2	3.8	0.2
Replica 4	92.4	80.8	6.8	n/a	14.6	0.6
Average Replicas	78.8	77.2	14.3	n/a	9.2	0.4
CYP2D6*4	45.6	37.4	29.6	37.2	3.2	0.0
Replica 5	79.0	53.0	41.9	58.4	4.6	0.2
Replica 6	63.0	10.4	21.6	34.2	3.0	0.0
Average Replicas	71.0	31.7	31.8	46.3	3.8	0.1
CYP2D6*10	98.8	0.0	4.0	39.0	0.0	0.0
Replica 7	97.6	0.0	0.6	n/a	0.0	0.0
Replica 8	n/a	n/a	n/a	26.2	0.6	0.0
Average Replicas	n/a	n/a	n/a	n/a	0.3	0.0
CYP2D6*17	100.0	1.6	1.2	54.2	0.4	0.2
Replica 9	100.0	0.8	0.2	41.6	0.2	0.2
Replica 10	100.0	1.6	1.0	41.8	0.2	0.4
Average Replicas	100.0	1.2	0.6	41.7	0.2	0.3
CYP2D6*53	98.6	1.2	15.0	65.2	11.8	0.0
Replica 11	98.8	0.0	27.0	83.3	2.8	0.0
Replica 12	98.2	0.6	20.0	25.0	0.8	0.0
Average Replicas	98.5	0.3	23.5	54.2	1.8	0.0

The tunnel opening frequencies in percent for six highest ranked tunnels. A tunnel was considered to be open if its bottleneck radius was above 1.2 Å.

n/a: The clustering performed by CAVER did not allow the correct calculation of parameters of the tunnel in the respective simulation.

Table S4.7 **Average hydropathic index of tunnel lining residues.**

Tunnel	Average hydropathic index
2a	0.51*
2b	0.95±0.32
2c	1.32±0.29
2d	1.39±0.20
2e	0.55±0.37
3	0.71*
4	1.00±0.28
5	0.71±0.35
6	0.61*
Water	0.88±0.19
Solvent	0.88±0.28

We calculated the average hydropathic index of the tunnel lining residues if a tunnel cluster was present in at least 50 frames of the related simulation. The values were averaged over the variants and the wild-type.

* no standard deviation given since tunnel fitting the calculation criteria only occurred in one variant

Table S4.8 **Average bottleneck radii of the variants for production simulations and replicas.**

Compared simulations	Average bottleneck radii of tunnels (Å)			
	2b	**2c**	**2e**	**4**
CYP2D6 WT	1.4 ± 0.2	1.5 ± 0.2	1.1 ± 0.1	1.1 ± 0.1
Replica 1	1.7 ± 0.2	1.3 ± 0.3	1.1 ± 0.1	1.1 ± 0.2
Replica 2	1.4 ± 0.2	1.4 ± 0.3	1.1 ± 0.1	1.1 ± 0.2
CYP2D6*2	1.8 ± 0.2	1.4 ± 0.2	1.2 ± 0.2	1.4 ± 0.3
Replica 3	1.8 ± 0.3	1.4 ± 0.2	1.2 ± 0.2	1.7 ± 0.4
Replica 4	1.8 ± 0.3	1.4 ± 0.2	1.1 ± 0.2	n/a
CYP2D6*4	1.2 ± 0.2	1.2 ± 0.2	1.1 ± 0.1	1.2 ± 0.2
Replica 5	1.4 ± 0.2	1.3 ± 0.2	1.2 ± 0.2	1.3 ± 0.2
Replica 6	1.3 ± 0.2	1.1 ± 0.1	1.1 ± 0.2	1.1 ± 0.2
CYP2D6*10	1.9 ± 0.2	0.0 ± 0.0	1.1 ± 0.2	1.3 ± 0.3
Replica 7	1.8 ± 0.2	0.0 ± 0.0	1.0 ± 0.1	n/a
Replica 8	n/a	n/a	n/a	1.3 ± 0.2
CYP2D6*17	1.8 ± 0.2	1.0 ± 0.1	1.0 ± 0.1	1.4 ± 0.2
Replica 9	1.9 ± 0.2	1.0 ± 0.1	1.0 ± 0.1	1.4 ± 0.2
Replica 10	1.8 ± 0.2	1.0 ± 0.1	1.0 ± 0.1	1.4 ± 0.3
CYP2D6*53	2.6 ± 0.3	1.1 ± 0.3	1.1 ± 0.2	1.3 ± 0.2
Replica 11	2.7 ± 0.2	1.0 ± 0.1	1.2 ± 0.2	1.4 ± 0.2
Replica 12	2.5 ± 0.2	1.0 ± 0.2	1.1 ± 0.2	1.1 ± 0.2

The average bottleneck radii of enzyme tunnels in the wild-type and the five variants. We chose to compare the four highest ranked tunnels. The average values are given with standard deviation.
n/a: The clustering performed by CAVER did not allow the correct calculation of parameters of the tunnel in the respective simulation.

Table S4.9 **Major bottleneck residues calculated for production phases and replicas.**

Compared runs	Number of tunnels per residue					
	120	305	308	370	374	483
Production phases	5	3	3	4	5	3
Replica 1	5	2	3	2	4	2
Replica 2	4	3	1	4	5	4

Bottleneck residues involved in the regulation of multiple tunnels are shown together with the number of tunnels they are involved in. Note that F120 was involved as a bottleneck residue in tunnel 2b, 2c, 2e, 4, and W.

Table S4.10 **Average number of hydrogen bonds for production phases and replicas.**

Compared simulations	Average number of hydrogen bonds per residue							
	34	91	94	107	120	122	296	486
CYP2D6 WT	0.6 ± 0.5	1.2 ± 0.8	3.4 ± 0.9	2.1 ± 0.5	2.5 ± 0.6	1.3 ± 0.5	2.5 ± 0.7	1.9 ± 0.5
Replica 1	0.7 ± 0.5	1.2 ± 0.8	3.6 ± 0.9	2.9 ± 0.8	2.0 ± 0.1	1.6 ± 0.6	2.8 ± 0.7	2.8 ± 0.7
Replica 2	0.7 ± 0.5	1.7 ± 0.5	4.1 ± 0.9	2.1 ± 0.3	2.0 ± 0.1	1.6 ± 0.6	2.9 ± 0.7	2.1 ± 0.4
CYP2D6*2	-	-	-	-	-	-	0.7 ± 0.9	2.1 ± 0.4
Replica 3	-	-	-	-	-	-	1.6 ± 1.0	2.5 ± 0.5
Replica 4	-	-	-	-	-	-	1.3 ± 1.0	2.4 ± 0.5
CYP2D6*4	0.4 ± 0.5	0.9 ± 0.5	3.2 ± 0.6	-	-	-	-	1.8 ± 0.5
Replica 5	0.0 ± 0.2	1.3 ± 0.5	3.6 ± 0.7	-	-	-	-	2.4 ± 0.5
Replica 6	0.4 ± 0.6	1.3 ± 0.5	3.5 ± 0.7	-	-	-	-	2.3 ± 0.5
CYP2D6*10	0.9 ± 0.4	-	-	-	-	-	-	2.1 ± 0.5
Replica 7	0.9 ± 0.2	-	-	-	-	-	-	2.3 ± 0.6
Replica 8	0.9 ± 0.3	-	-	-	-	-	-	2.0 ± 0.2
CYP2D6*17	-	-	-	1.1 ± 0.7	-	-	1.1 ± 0.5	2.6 ± 0.8
Replica 9	-	-	-	1.5 ± 0.7	-	-	1.4 ± 1.0	3.1 ± 0.3
Replica 10	-	-	-	1.6 ± 0.8	-	-	1.7 ± 1.0	3.0 ± 0.3
CYP2D6*53	-	-	-	-	2.4 ± 1.0	2.3 ± 0.7	-	-
Replica 11	-	-	-	-	2.0 ± 0.6	2.5 ± 0.8	-	-
Replica 12	-	-	-	-	2.0 ± 0.4	2.2 ± 0.8	-	-

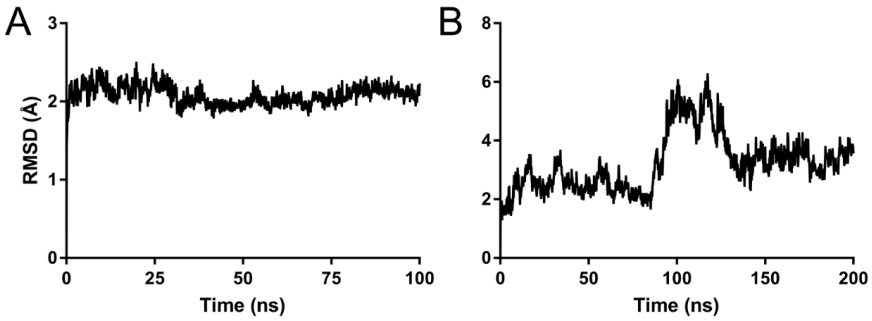
The data was calculated for all mutated amino acids in the wild-type as well as for the mutated amino acids in the variants.
The values are given with standard deviation.

Table S4.11 **Active site volume calculated with POVME for production simulations and replicas.**

Compared simulations	Volume (Å^3)	Average Replicas
CYP2D6 WT	674.9 ± 83.6	
Replica 1	634.0 ± 64.9	589.7 ± 62.6
Replica 2	545.4 ± 52.9	
CYP2D6*2	739.6 ± 76.4	
Replica 3	776.4 ± 64.5	898.0 ± 172.0
Replica 4	1019.6 ± 95.6	
CYP2D6*4	398.6 ± 46.1	
Replica 5	514.0 ± 77.1	455.2 ± 83.2
Replica 6	396.3 ± 40.7	
CYP2D6*10	883.0 ± 80.7	
Replica 7	947.3 ± 80.5	978.0 ± 43.3
Replica 8	1008.6 ± 66.8	
CYP2D6*17	687.9 ± 66.1	
Replica 9	557.8 ± 55.1	628.1 ± 99.4
Replica 10	698.4 ± 60.3	
CYP2D6*53	1092.3 ± 108.3	
Replica 11	1074.9 ± 88.2	1086.3 ± 16.1
Replica 12	1097.7 ± 66.7	

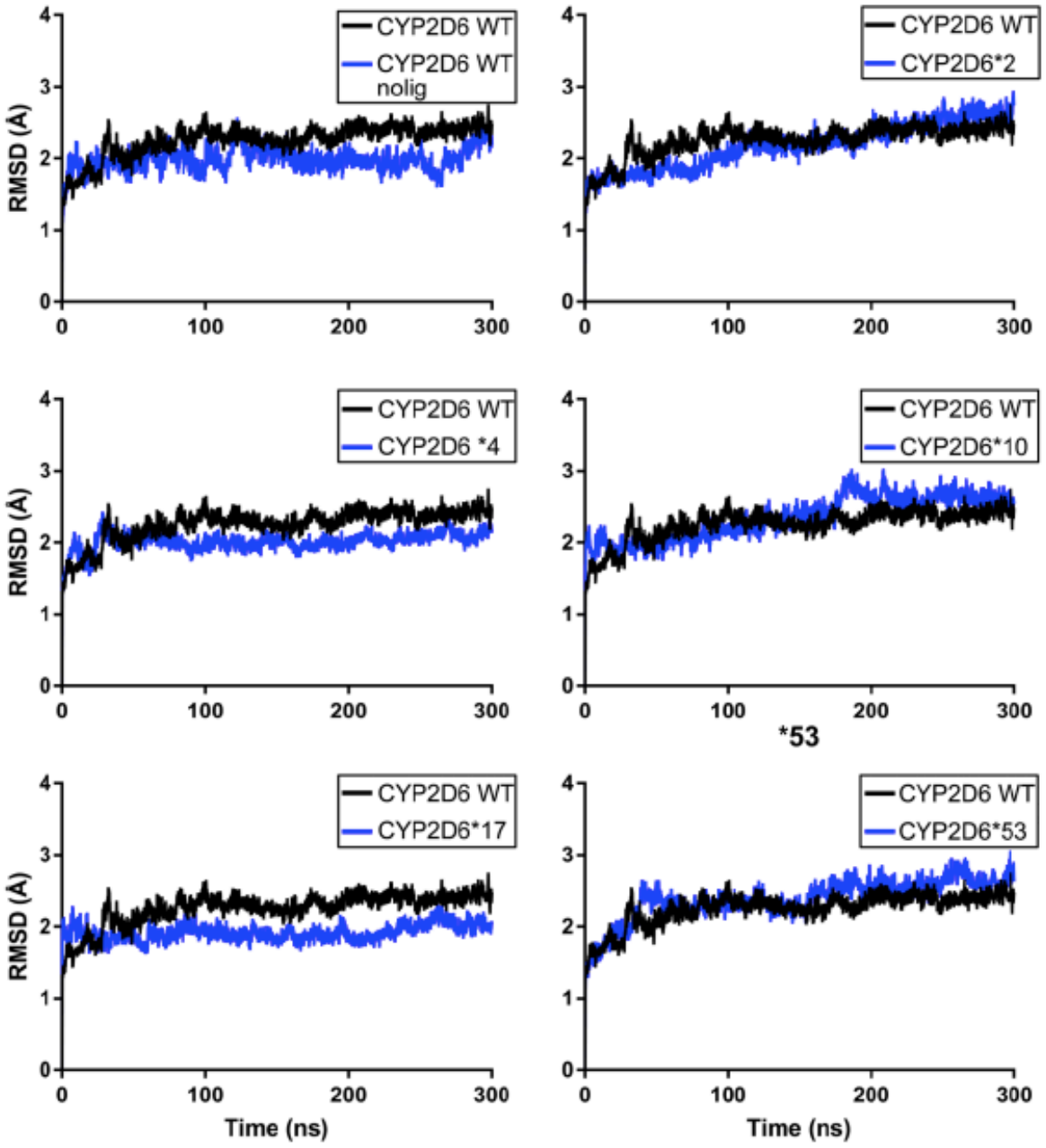
The volumes of the active site cavities in CYP2D6 WT and its related variants. The volumes were estimated with POVME (v2.0) based on the last 100 frames of the production phase. The values are given with standard deviation and average values for the replicas have been calculated.

Figure S4.1 **Backbone RMSD plots of the simulations MemAs and AnchorSim.**



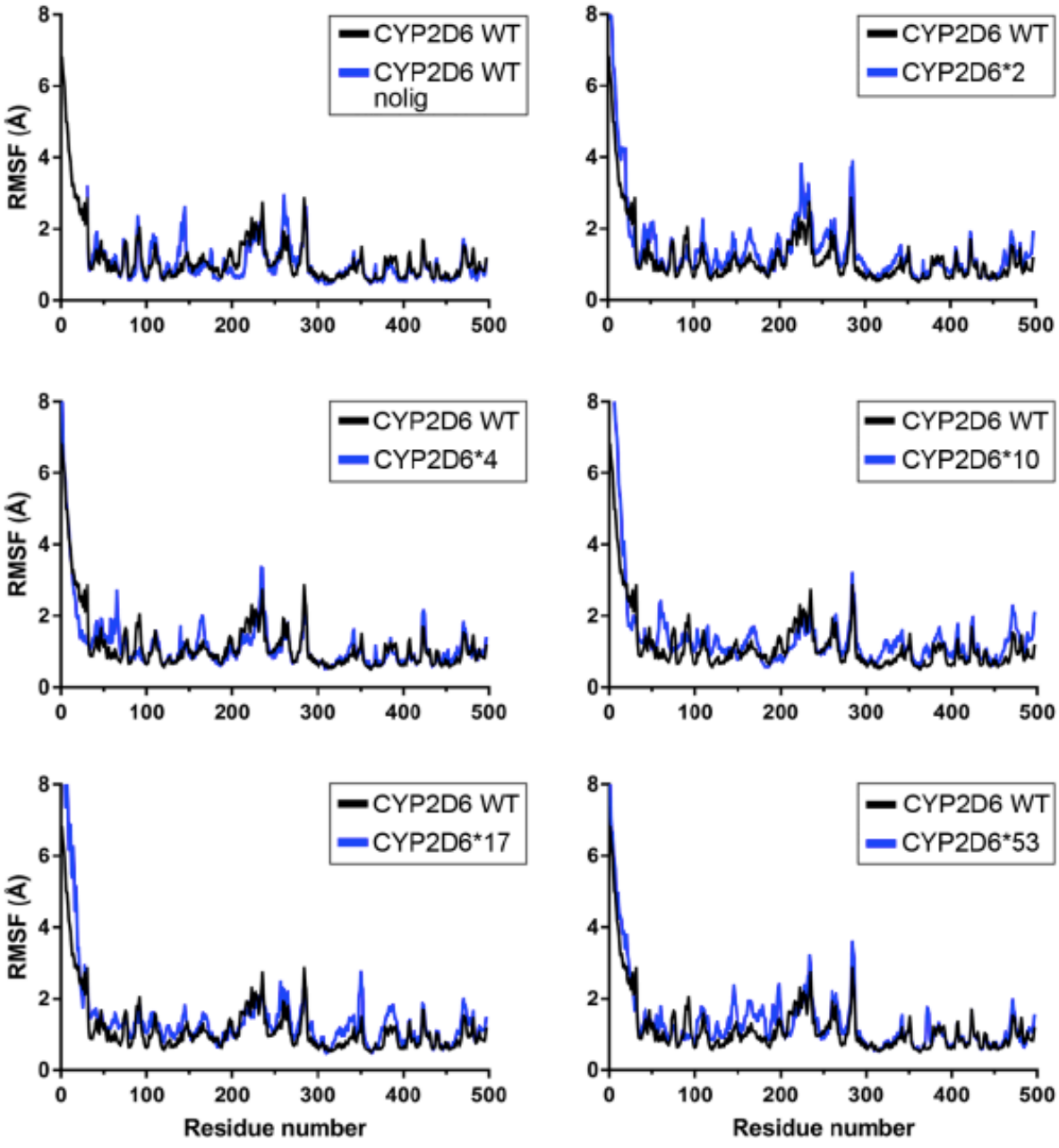
The backbone RMSD values for (A) the MemAs simulation and (B) the AnchorSim simulation. The values are shown for the whole duration of the simulation.

Figure S4.2 **Backbone RMSD profiles of the membrane-bound simulations and the WT nolig simulations of CYP2D6.**



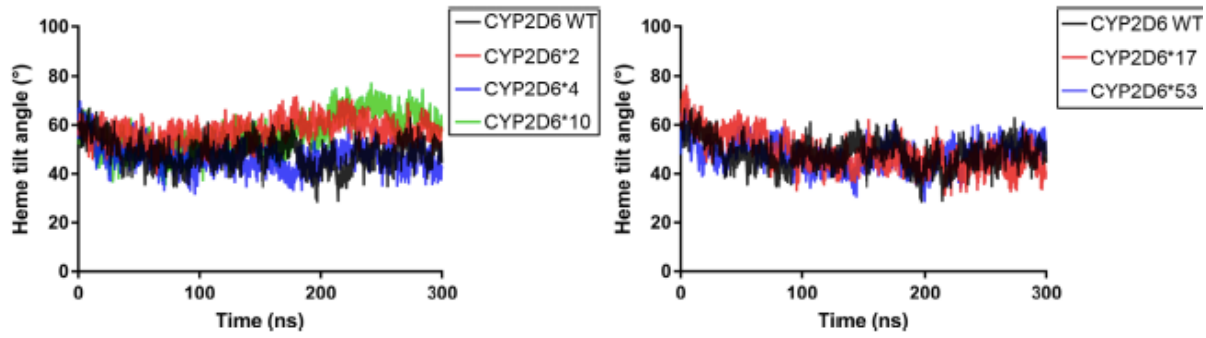
The backbone RMSD values of membrane-bound simulations as well as the WT nolig simulation. The values are shown for the whole 300 ns simulation time.

Figure S4.3 **Backbone RMSF profiles of membrane-bound simulations and the WT nolig simulation.**



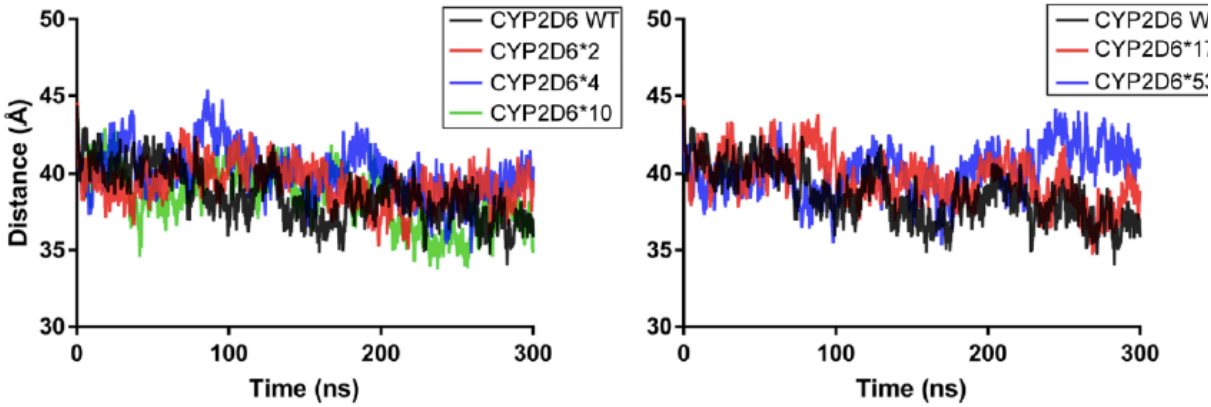
The backbone RMSF values of membrane-bound simulations, as well as the WT nolig simulation. The values are shown for the whole 300 ns simulation time.

Figure S4.4 **Heme tilt angle of membrane-bound simulations of CYP2D6.**



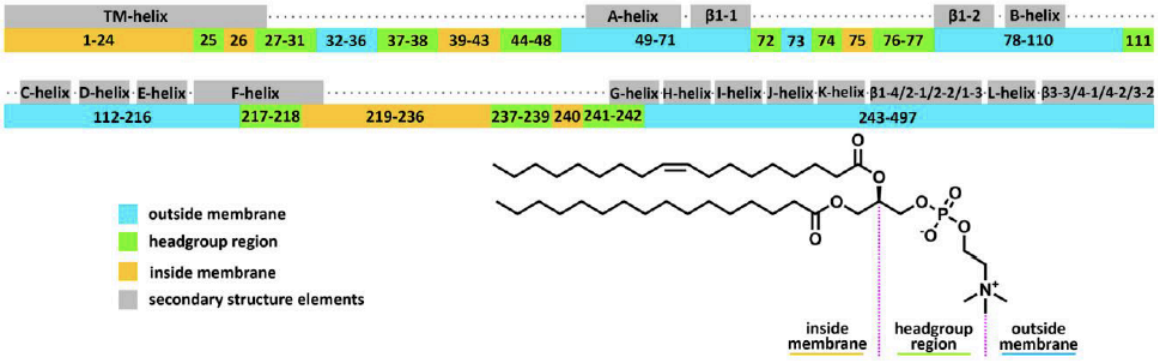
The heme tilt angle of the membrane-bound simulations. It is defined as the angle between the heme plane and the membrane normal corresponding to the z-axis. The values are shown for the whole simulation time of 300 ns.

Figure S5. **Burying depth of membrane-bound systems of CYP2D6.**



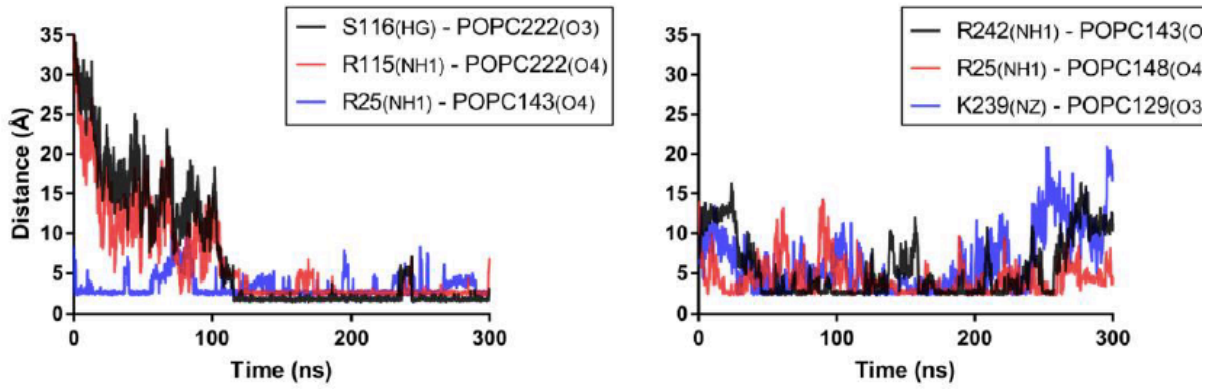
The burying depth of the enzyme in the membrane was calculated according to Ducassou et al. It is defined as the distance between the mass centers of POPC C1 atoms and protein α-carbons. The values are shown for the whole simulation time of 300 ns.

Figure S4.6 **Residues in contact with the membrane, the head group region, or the cytosol in the wild-type model of CYP2D6.**



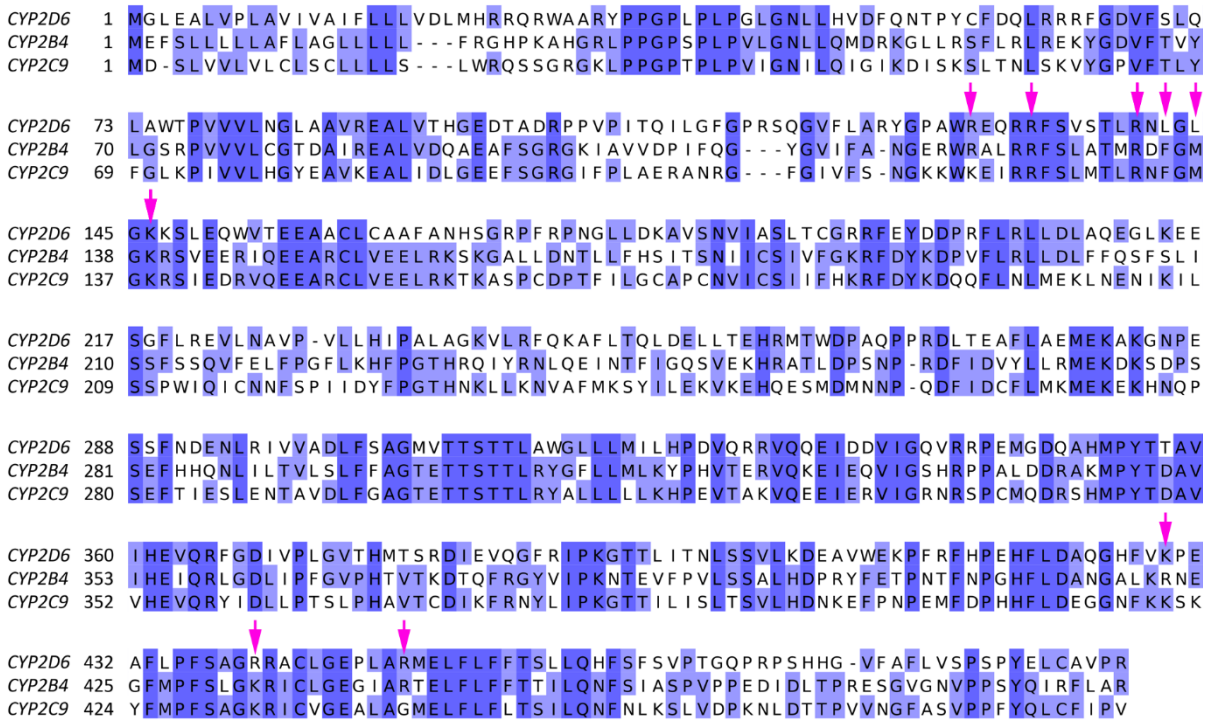
Residues of the CYP2D6 WT model that are in contact with the membrane, the head group region, or the cytosol. We defined he head group region to be between mass centers of the C2-atoms and the N-atoms of the POPC molecules. The localization of the residue was determined considering the position of its α-carbon atom.

Figure S4.7 **Distance between residues and POPC molecules and selected protein residues.**



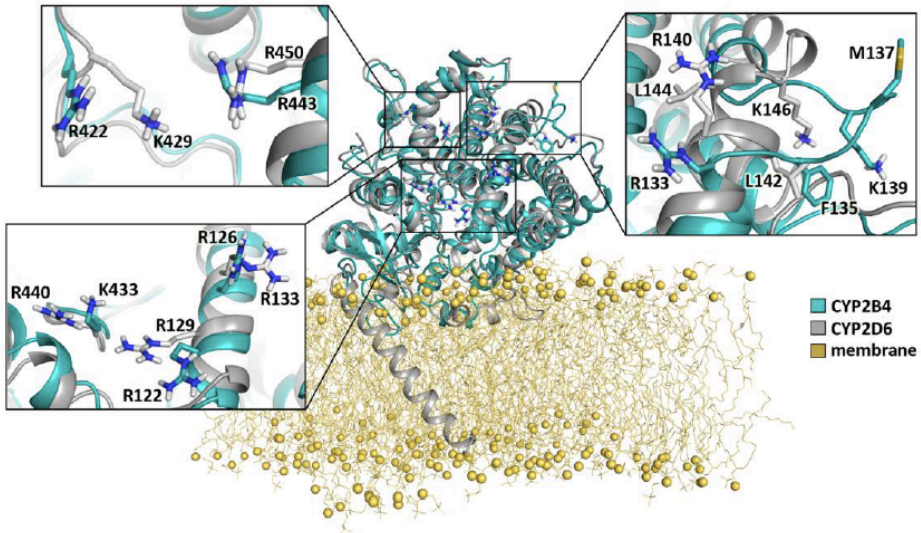
The distance between selected atoms of protein residues and membrane molecules plotted against the simulation time. Distances in the range of 2 Å indicate a possible interaction between the residues.

Figure S4.8 **Sequence alignment of CYP2D6, CYP2B4, and CYP2C9.**



A sequence alignment of CYP2D6, CYP2B4, and CYP2C9. This was done in order to identify the residues of the POR binding site and to translate experimental results for the model validation. The residues corresponding to the POR binding site are highlighted with pink arrows.

Figure S4.9 **Structural alignment of CYP2B4 to CYP2D6 with focus on the POR binding site.**



Protein structure alignment of CYP2B4 to CYP2D6. This was done to identify the binding site of the POR on CYP2D6. The residues in CYP2B4 were established in experiments. The structure of CYP2D6 represents a snapshot after 200 ns simulation time.

Figure S4.10 **Spatial distribution of enzyme tunnels in CYP2D6 WT.**

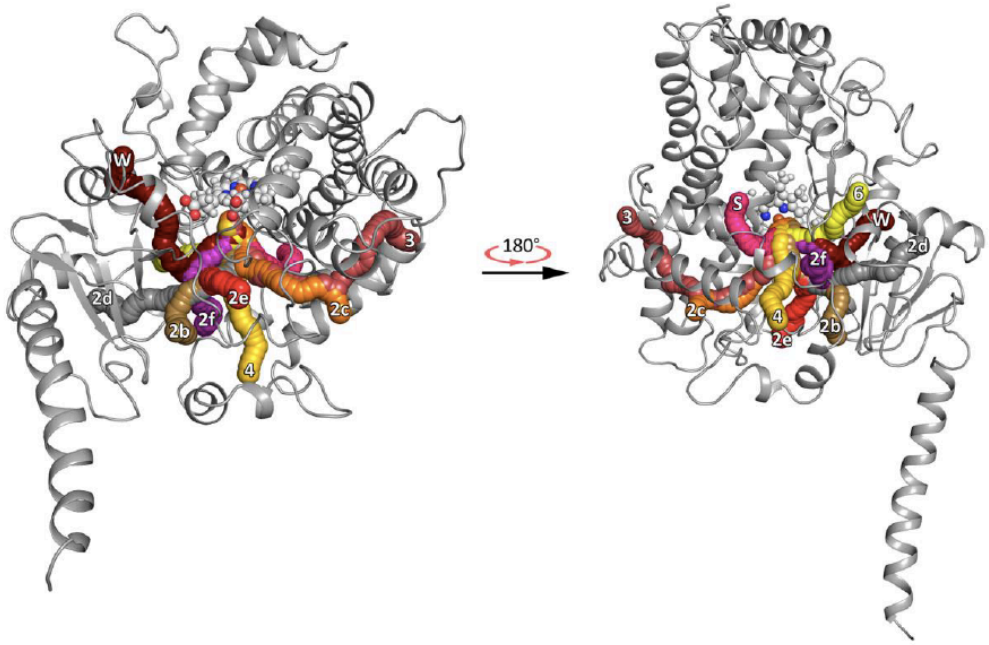


Figure S4.11 **Identified tunnels in all production simulations.**

	Tunnels												
Variants	1	2ac	2b	2c	2d	2e	2f	3	4	5	6	S	W
CYP2D6 WT inhibitor	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CYP2D6 WT nolig	x	x	✓	✓	✓	✓	x	✓	✓	x	✓	✓	✓
CYP2D6 WT	x	x	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓
CYP2D6*2	x	x	✓	✓	x	✓	x	✓	✓	✓	✓	✓	✓
CYP2D6*4	x	x	✓	✓	✓	✓	x	✓	✓	✓	✓	✓	✓
CYP2D6*10	x	x	✓	x	✓	✓	x	x	✓	✓	✓	✓	✓
CYP2D6*17	x	x	✓	✓	✓	✓	x	x	✓	x	✓	✓	✓
CYP2D6*53	x	x	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Identified enzyme tunnels in all production simulations. We considered a tunnel to be present if it occurred in at least five frames during the production phase. We observed 13 of 14 known tunnels in the WT inhibitor simulation.

Chapter 5

Pharmacogenetics *in silico* study of CYP2D6 focused on the pharmacovigilance of herbal antidepressants

Based on the submitted manuscript:

CYP2D6 Pharmacogenetics *in silico* study focused on the pharmacovigilance of herbal antidepressants.

Charleen G. Don & Martin Smieško

Abstract

The annual increase in depression worldwide together with an upwards trend in the use of alternative medicine as treatment asks for developing reliable safety profiles of herbal based medicine. A considerable risk on adverse reactions exists when herbal remedies are combined with prescription medication. Around 25% of the drugs, including many antidepressants, depend on the activity of CYP2D6 for their metabolism and corresponding efficacy. Therefore, probing CYP2D6 inhibition by the active substances in herbal based medicine within the wild-type enzyme and clinically relevant allelic variants is crucial to avoid toxicity issues. In this *in silico* study several compounds with herbal origin suggested to have antidepressant activity were analysed on their CYP2D6 wild-type and *CYP2D6*53* inhibition potential using molecular docking. In addition, several pharmacokinetic properties were evaluated to assess their probability to cross the blood brain barrier and subsequently reach sufficient brain bioavailability for the modulation of central nervous system targets as well as characteristics which may hint towards potential safety issues.

Introduction

Depression as a mood disorder has significantly increased (~18%) during the last decade affecting over 300 million people as estimated by the world health organisation (WHO) (1). Concurrently, an upward trend is seen in the use of antidepressants worldwide [2–4] in parallel with the use of complementary and alternative medicine (CAM) [5–7] . A recent study reported that in Europe (21 surveyed countries) 30% of the affected individuals used CAM to treat depression of which 10% included herbal medicine (natural products) [7]. Antidepressants as neurotherapeutics must be able to cross the blood-brain barrier (BBB) in order to modulate the function of the central nervous system (CNS). The compound exchange between the blood and the nervous tissue is strictly controlled by a continuous layer of endothelial cells which are kept together by tight junctions with a total estimated surface of around 20 m^2 [8]. From a medicinal chemistry perspective, physiochemical properties of CNS modulators (e.g. molecular weight, lipophilicity, hydrogen acceptor and donor counts) therefore

need to fall within a narrower range compared to regular drugs in order to be absorbed, dissolved and permeate across the brain-barrier to reach the CNS (passive transport, non-energy required). More hydrophilic modulators (or other compounds) may still cross the barrier through carrier systems, transporters or endocytosis (active-transport, ATP required) [9,10]. However, good BBB permeation does not guarantee that the efficacious concentration of the compound in the CNS compartment is reached [11]. For the design of a novel neurotherapeutically acting drug or natural-derived compound, predicting its BBB permeability, partitioning, P450 cytochrome association (inhibitor/substrate/none) and off-target binding in the early development stage is essential in order to develop a compound that will have sufficient brain bioavailability ensuring that the desirable therapeutic effect is acquired and maintained at a concentration sufficient to modulate the CNS system without inducing adverse reactions. Safety profiles of antidepressant drugs are studied by many groups and unfortunately are known to include also serious side effects [12–14]. The safety profiles of herbal medicine products (from plant sources) are much less studied despite of being frequently concomitantly administered with another (synthetic) drug and bear a considerable risk for herbal-induced toxicity issues [15–17]. Herbal-drug interactions (HDI) can lead to harmful adverse events as they may alter the pharmacokinetics (PK; adsorption, distribution, metabolism or excretion) and pharmacodynamics (PD; pharmacological effect is changed in a synergetic, additive or antagonistic way) profile of the concomitantly administered drug [18].

The prediction of interactions of the natural compounds with one major class of drug metabolizing enzymes, cytochromes P450, especially the family member CYP2D6, is of particular interest. This isoform is responsible for biotransformation of about 25% of all the marketed drugs, and displays a very high polymorphism rate [19,20]. Inhibition or induction of CYP2D6 metabolism can alter the pharmacokinetic profile of the concomitantly administered drug and potentially can lead to toxicity or affect the drug efficacy [17]. During the catalytic reaction, interference of the substrate binding to the heme, the binding of molecular oxygen, or the biotransformation step in which the substrate is oxidized are prone to induce adverse reactions upon inhibition [21]. Based

on these three interference mechanisms, inhibitors can be divided into three categories; (i) reversible binders (competitive or noncompetitive), (ii) quasi-irreversible binders (also referred to as suicide inhibitors) in which the inhibitors interact directly with the heme-iron and (iii) mechanism-based inhibitors, which irreversibly bind to the protein and accelerate degradation or oxidative fragmentation of the heme [21–23]. Compounds which inhibit the enzyme before the oxidative events occur are usually reversible competitive or noncompetitive inhibitors. Mechanism-based or suicide inhibitors more often act upon the oxygen transfer event or subsequent to this step [22]. The latter category can have severe consequences; the complete inactivation of the enzyme means that for several hours to days the plasma concentrations of drugs which depend on the enzyme activity to become metabolized will increase, which in turn increases the risk on adverse reactions [24]. Previous identified CYP2D6 mechanism-based inhibitor drugs include paroxetine, 3,4-Methylenedioxy methamphetamine (MDMA), (1-[(2-ethyl-4-methyl-1H(-EMTPP-imidazol-5-yl)-methyl]-4-[4-(trifluoromethyl)-2-pyridinyl]piperazine (EMTPP) and 5-Fluoro-2-[4-[(2-phenyl-1H-imidazol-5-yl)methyl]-1-piperazinyl]pyrimidine (SCH66712) [25–27]. Also compounds from several natural sources such as St. John's wort, common sage and goldenseal/berberine have been identified to exercise a moderate to strong inhibition effect on CYP2D6 wild-type (WT) and also on allelic variants (e.g. *CYP2D6*10*) [28–30]. Hence there is an urgent need for cost- and time-efficient methods which can help to determine the safety profile not only of (new) drugs but also of natural compounds with regard to their potential to inhibit CYP2D6 WT and clinically relevant allelic variants. An advantageous combination of methods used for identification of herbal compounds that potentially inhibit CYP2D6 consists of an *in silico* approach for a first fast screening, whereas resulting hits can be further verified by experimental *in vitro/in vivo* studies [31,32]. Recently, Hochleiter and co-workers successfully demonstrated the potential and relevance of a combined *in silico - in vitro* based workflow (including pharmacophore screening and docking) to identify new compounds derived from a natural source that inhibited CYP2D6 WT [31].

In this *in silico* study the safety profile of several compounds from herbs which are known or suggested to have antidepressant activity are assessed by: (i) determining their inhibitory effect on CYP2D6 WT and the clinical relevant *CYP2D6*53* (F120I, A122S) allelic variant [33] and (ii) assessing their physiochemical properties along with several toxicity-related descriptors in order to define their brain bioavailability potential and off-target binding. The *CYP2D6*53* allelic variant has been only identified among the Japanese with an allele frequency of 0.2% [33]. Moreover, it is the only allelic variant associated with increased metabolism activity (4-fold increase in CL_{int} value) for the typical CYP2D6 substrate bufuralol. The increased activity is suggested to be caused by the F120I mutation which is positioned close to the heme. Therefore, investigating and comparing the docking poses of the wild-type and *CYP2D6*53* can provide insight regarding the role of the F120I mutation. The safety profiles compiled for each herbal compound will give a first indication if any toxicity issues might arise when they are used to treat depression, especially in multidrug therapy. It has to be mentioned that the focus of this study is on herbal compounds which are drug-like in terms of size and physiochemical properties. The herbal compounds which were not included in the analysis because they did not comply with the library filtering criteria applied, might still be suitable as antidepressant as they might cross the BBB and modulate the CNS system using any of the active transport mechanisms. However, evaluating their safety profile is much more complex and this falls outside the scope of our study.

Materials and methods

Compound Library selection

The anti-depressant natural compound library was obtained from ChemFaces (accessed July 2019), a high-purity natural products manufacturer (Wuhan, P.R. China). The 3D coordinate files were obtained from ChemSpider. For the 19 out of 51 compounds which passed filtering, a literature search was performed using PubMed and Scopus to verify their suggested antidepressant activity.

CYP2D6 structure selection

For the available co-crystalized structures (Protein Data bank (PDB) IDs: 3QM4, 3TDA, 4WNT, 4WNU, 4WNV, 4WNW, 3TBG) it has been shown that the all-atom root-mean-square deviation (RMSD) of the superimposed binding sites is lower than 1 Å [34]. A major motivation of this study is to help avoiding the CYP2D6 toxicity within a co-therapy context, hence competitive binding. At the moment, there is only one protein-ligand complex structure available with a co-crystalized substrate molecule: thioridazine (a typical antipsychotic drug, PDB ID: 4WNW). Therefore, 4WNW was primarily used for docking and the binding mode analysis of the herbal compounds. However, two additional structures (PDB IDs: 4WNT and 4WNU), co-crystalized with the natural inhibitors ajmalicine and quinidine respectively, were used as well in order to verify consensus of the most favorably scored binding mode.

Library and protein preparation

The structures were processed using the standardized ligand preparation procedure as implemented in the software LigPrep (Schrödinger LLC.) The preparation procedure included bond order assessment, tautomeric state, protonation evaluation (at pH 7.4), and chemical structure consistency checks. For all the compounds, the program MacroModel was used for the minimization of the starting geometries. The three selected CYP2D6 crystal structures (PDB IDs: 4WNW, 4WNU and 4WNT, chain A) were retrieved from the PDB database and processed using the Protein Preparation Wizard of Maestro small-molecule drug discovery suite (v. 2017-2). Missing residues and hydrogen atoms were added (at pH 7.4), bond orders were assigned, and the co-crystalized ligand and co-factors were removed. No crystal waters are resolved in 4WNW and the crystal waters in 4WNU and 4WNT were deleted. Compound state I was modelled (Fe^{3+} bound to O^{2-}, zero-order bond). The distance between the iron and the oxygen atoms was set to 1.97 Å. For the allelic variant *CYP2D6*53* the two residues of the wild type structure were mutated (F120I, A122S) subsequently the mutated structure was fully minimized.

Toxicokinetic parameters and descriptor calculations

For each studied ligand, the QikProp program by Schrödinger was used to calculate the Lipinski's rule of 5 (hydrogen bond donor and acceptor counts, molecular weight, logarithm of the partition coefficient), Veber rules (polar surface area, number of rotatable bonds) and several other absorption, distribution, metabolism and excretion (ADME) and toxicity related descriptors offered by Schrödinger (QP $\log_{HERG}$, QP P_{Caco}, QP $\log_{BB}$, QP P_{MDCK}, Human Oral Absorption and CNS activity score). The models behind the calculations assume absorption through passive permeation. The descriptions can be found in **Table S5.6**. Most psychoactive (CNS) drugs need to cross the BBB for the modulation of a target neurotransmitter system [35]. Improved BBB permeation is usually obtained for such drugs by following narrower Lipinski ranges compared to other drugs [36,37]. Two prominent CNS studies which defined the pharmacokinetic properties (including Lipinski's rules) were used as filtering guideline and an overview of them can be found in **Table 5.1.** In addition, the online tool Molinspiration (www.molinspiration.com/cgi-bin/properties) was used to assess structural similarity and thus potential biological activity towards six important drug classes (G-protein coupled receptor (GPCR) ligands, ion channel modulators, kinase inhibitors, nuclear receptor ligands, protease inhibitors, enzyme inhibitors).

Autodock smina docking

Smina, a fork of Autodock Vina (v.1.1.2) was used for docking. Smina is focused on improving scoring and minimization and includes several convenient functions which can be accessed from the command line such as calculating the box dimensions based on one or several existing ligand(s) [38]. The prepared library and CYP2D6 prepared protein structures (WT and *CYP2D6*53*) were used for the docking. The residues F112, F120 (or I120 in the variant), E211, E215, E216, R221, Q244, R296, I297, D301, S304, and F483 of CYP2D6 were defined as flexible residues (side chains) during the docking. The random seed number was set to 0 and an additional 8 Å buffer space was added to the auto-generated box. All other settings were kept at their default values.

Results and discussion

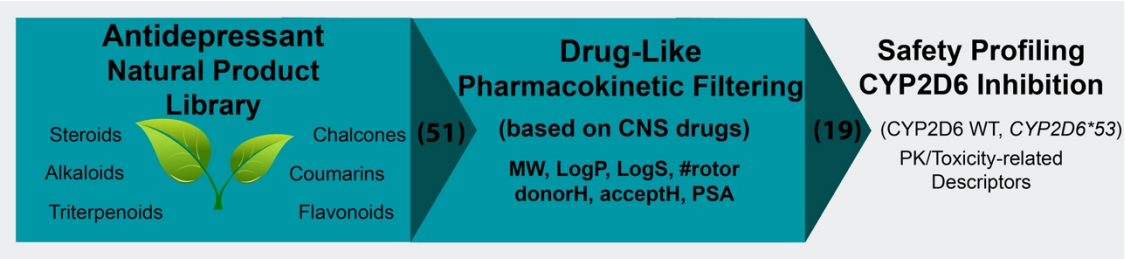


Figure 5.1 **The *in silico* safety profiling approach of the natural product antidepressant library.**

Pharmacokinetics descriptor based filtering

The ChemFaces antidepressant compound library contained 51 molecules for which all pharmacokinetic descriptors were calculated using the program QikProp (Schrödinger LLC.). Filtering criteria were based on pharmacokinetic property guidelines specific for CNS therapeutic agents (**Figure 5.1**). Earlier CNS pharmacokinetics analysis studies demonstrated that CNS drugs fit within a smaller range (**Table 5.1** – Lipinski non-CNS column) compared to the general Lipinski guidelines (**Table 5.1** – Lipinski CNS column) [36].

Table 5.1 **Guidelines for the physicochemical properties associated with improved blood-brain-barrier penetration for CNS drug design.**

	Ghose[a]		Pajouhesh[b]	Wager[c]	Lipinski	
Descriptor	**min**	**max**		**(max)**	**CNS[d]**	**non-CNS**
MW	141	452	< 450	≤ 360 (500)	≤ 400	< 500
donorHB	0	3	< 3	≤ 0.5 (3.5)	≤ 3	< 5
accptHB	1	8	< 7	-	≤ 7	< 10
log P o/w	0.16	6	< 5	≤ 3 (5)	≤ 5	< 5
log S	-0.4	0.5	-	-	-	-
PSA	3.8	109	60 - 70	40 – 90* (120)	-	< 140
#rotor	0	8	< 8	-	-	< 10

a) based on 317 approved CNS drugs, b) based on consensus of several studies discussed in the paper, c) derived from a set of 1500 drugs filtered from United States Adopted Names (USAN) or International Nonproprietary names (INN) for good CNS penetration by Lipinski.

Pajouhesh et al. provided guidelines for CNS drugs based on the consensus of several studies discussed in their paper [36]. Other CNS focused physicochemical analyses

found similar reference values with some variation (**Table 5.1**). Ghose et al. calculated for 317 CNS and 627 non-CNS approved drugs the corresponding physicochemical properties and the preferred ranges including several pharmacokinetic descriptors using QikProp [37]. Wager et al. analysed 119 CNS drugs and 108 Pfizer CNS candidates [39]. An overview of all proposed CNS guidelines can be found in **Table 5.1**. For filtering of the compounds, the consensus values of Pajouhesh were applied as guidelines. In addition, if one of the descriptors was lying outside the guideline range, the other proposed ranges (Ghose or Wagner, **Table 5.1**) were used to validate if the value was still acceptable. Compounds with one violations of the Pajouhesh guidelines are highlighted in **Table 5.2**. It must be mentioned that the compounds which do not comply with the used filter criteria might still be able to cross the BBB and modulate the CNS system through an active transport.

Table 5.2 **Pharmacokinetic properties of the natural antidepressants after filtering for drug-like properties.** The bold highlighted compounds have one or two violation(s) of the Pajouhesh CNS criteria of which at least one is still within the acceptable reference value range proposed by Ghose.

Compound	MW	# rotor	HbD	HbA	QP logPo/w	QP logS	PSA
(-)-Cytisine	190.2	0	1	5	0.7	-0.8	45.1
4-Hydroxyisoleucine	147.2	5	4	5	-2.5	0.0	90.6
5-Isopropyl-2 methylphenol	150.2	2	1	1	3.3	-2.3	21.4
Auraptenol	260.3	5	1	5	2.2	-2.6	60.7
Chelidonic acid	184.1	2	2	7	-0.6	-1.0	139.5
D-(-)-Synephrine	167.2	5	3	4	0.2	-0.3	57.8
Honokiol	266.3	7	2	2	5.0	-4.3	42.3
Isorhynchopylline	384.5	4	1	8	2.6	-4.0	84.7
L-Theanine	174.2	6	4	6	-3.0	0.2	111.0
Magnolol	266.3	7	2	2	5.0	-4.2	42.3
Naringenin	272.3	3	2	4	1.6	-3.4	100.3
Orcinol	124.1	2	2	2	0.8	0.1	45.1
Piperine	285.3	5	0	5	3.3	-3.5	48.0
Protopine	353.4	0	0	7	1.7	-1.1	59.7
Psoralidin	336.3	4	2	5	3.0	-5.0	92.2
Salvigenin	328.3	4	0	5	3.3	-4.1	77.1
Scopoletin	192.2	2	1	4	0.8	-1.7	70.5
Trans-Methylisoeugenol	178.2	3	0	2	2.8	-3.7	15.8
Cannabidiol (CBD)	314.5	7	2	2	5.3	-6	39.1

After filtering, the remaining 19 compounds (**Figure 5.2**) were docked in CYP2D6 WT and the *CYP2D6*53* allelic variant for the three prepared CYP2D6 structures (PDB IDs: 4WNW, 4WNU and 4WNT) and several pharmacokinetic descriptors were calculated.

In addition, the literature was searched if any evidence existed on CYP2D6 inhibition (**Table S5.1**), P450 (major) isoform metabolism (**Table S5.2**), and their ability to cross the BBB (**Table S5.3**).

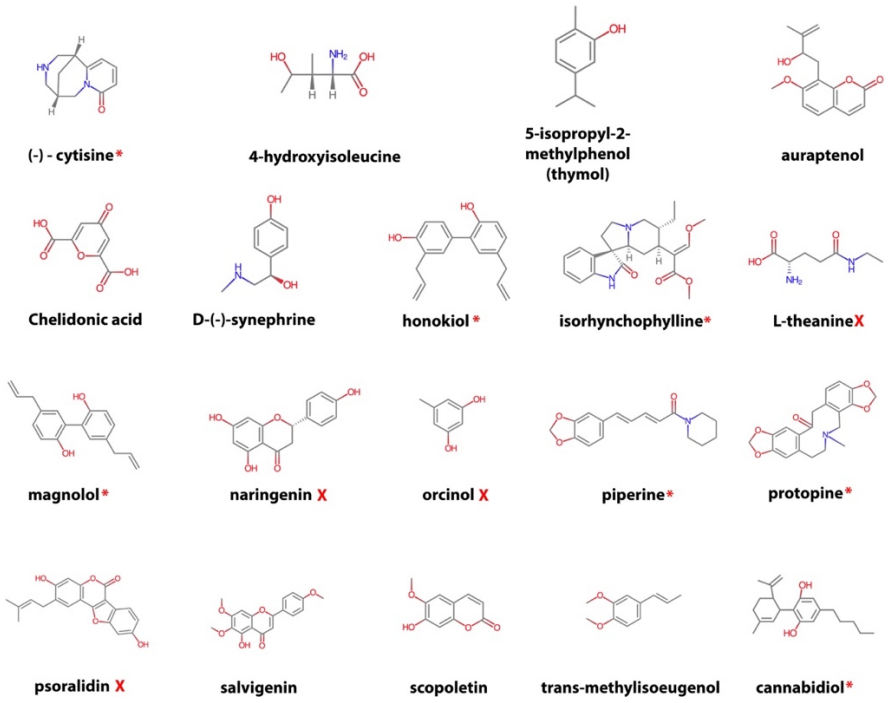


Figure 5.2 **Overview of the selected natural compounds.** The red asterisk indicates that previous research found CYP2D6 inhibition (potent to weak) activity for the compound (see table S1 for the reference), the red cross no CYP2D6 inhibition activity. For the remaining compounds no CYP2D6 inhibition data could be found.

CYP2D6 inhibition

CYP2D6 WT inhibition activity (varying from weak to potent inhibition) data was found for 7 compounds, 4 compounds were identified to have no inhibition activity, and for 8 no data could be found (**Table S5.1**). Though the neolignans honokiol and magnolol present in magnolia bark extract have been considered to be safe in use by various authorities [40] , evidence exists on CYP2D6 inhibition for magnolol (IC_{50} 65.4 μM) [41] and weak inhibition for honokiol (K_i 12 μM) [42]. Cannabidiol (CBD) one of the three major components in the cannabis plant has been identified as potent atypical inhibitor (IC_{50} 6.52 μM) of CYP2D6 [43]. Furthermore, for piperine weak inhibition (IC_{50} 3.2 μM) and protopine potent competitive inhibition (K_i 78 nM) has been determined [44,45]. Moderate CYP2D6 WT inhibition was found for (-)-cytisine (IC_{50} 28.9 μmol/L) and

isorhynchophylline (IC_{50} 44.1 µmol/L) [30]. Literature search confirmed for 13 compounds that they are able to cross the BBB (**Table S5.3**). Whether they reach a high enough concentration to acquire and maintain their bioactivity within the brain remains an open question and needs to be confirmed by additional clinical studies. Experimental evidence of CYP-mediated metabolism was found for nine of the compounds, one has been assigned no CYP-metabolism dependence, and for other nine no data was found (**Table S5.2**). No literature was found that assigned CYP2D6 as major isoform for metabolism for one of the compounds.

Inhibition binding modes from Molecular Docking

The top 10 binding poses were evaluated on potential CYP2D6 inhibition. The distance of the binding pose (distal or proximal) between the heme and the ligand was evaluated as well as the functional group and closest atom to the iron heme. If the distance between the heme-iron and any atom of the ligand was bellow 6 Å it was assigned proximal binding, otherwise it was assigned as distal. Several residues in the binding pocket are essential for ligand binding. Glu216 and/or Asp301 are known to act as first anchoring point by forming salt bridges with the commonly protonated aliphatic nitrogen atom (at physical pH) present in most typical CYP2D6 substrates. Subsequently Phe120 which resides in close proximity of the heme can further steer the orientation of the ligand by interacting with the aromatic part of the substrate through π-π stacking interactions [46–48]. An example of inhibition can be observed in the crystal CYP2D6 structure (PDB ID: 4WNU) where quinidine (inhibitor) binds distal from the heme and its protonated nitrogen forms electrostatic interactions with both E216 and D301 [48]. Based on this knowledge docking binding modes were assigned to potentially inhibit CYP2D6 if (i) a functional group (known not to be normally metabolized) was interacting directly with the iron heme, (ii) the distal binding from the heme prevented other ligands to reach the heme and/or blocked interaction with D301 and/or E216. Furthermore, the binding profile of the pose was evaluated on the presence of potential electrostatic interactions, hydrogen bonding, and hydrophobic interactions. The docking results can be found in **Table 5.3** and a detailed overview of

the different interactions types of the best binding modes can be found in Tables S5.4 and S5.5.

Table 5.3 **Overview CYP2D6 WT and *CYP2D6*53* docking results.** The top 10 docking poses were evaluated on potential CYP2D6 inhibition by evaluating its binding distance towards the heme (proximal or distal) together with its binding pocket interaction profile.

	CYP2D6 WT			CYP2D6*53		
Compounds	**prox**	**dis**	**none**	**prox**	**dis**	**none**
(-)-Cytisine		■				
4-Hydroxyisoleucine						
5-Isopropyl-2-methylphenol	■					
S-Auraptenol		■			■	
Chelidonic acid						
D-(-)-Synephrine	■				■	
Honokiol	■	■		■	■	
Isorhynchopylline						
L-Theanine						
Magnolol	■	■		■		
Naringenin						
Orcinol						
Piperine	■			■		
Protopine	■			■		
Psoralidin						
Salvigenin						
Scopoletin	■			■		
Trans-Methylisoeugenol						
Cannabidiol (CBC)	■			■		

Within the top 10, the most productive pose(s) (based on the mentioned criteria above) proximal and/or distal from the heme were selected. An interaction profile for the herbal compounds that were binding (proximal and/or distal) can be found in the supplementary data (Figures S5.1 and S5.2). From the 7 compounds for which experimental evidence was found on CYP2D6 WT inhibition activity, the binding poses of piperine, protopine, honokiol, magnolol, and cannabidiol were found to potentially inhibit CYP2D6 WT and *CYP2D6*53*; the binding mode of piperine in the wild-type was more strongly stabilized by hydrophobic interactions with F120 than with I120 in *CYP2D6*53* (Figures 5.3 and 5.5B). Protopine formed mainly electrostatic/hydrophobic interactions with the heme group. In CYP2D6 WT π-π stacking was observed with F483 whereas a hydrogen bond with Q244 was formed in *CYP2D6*53* (Figure 5.3).

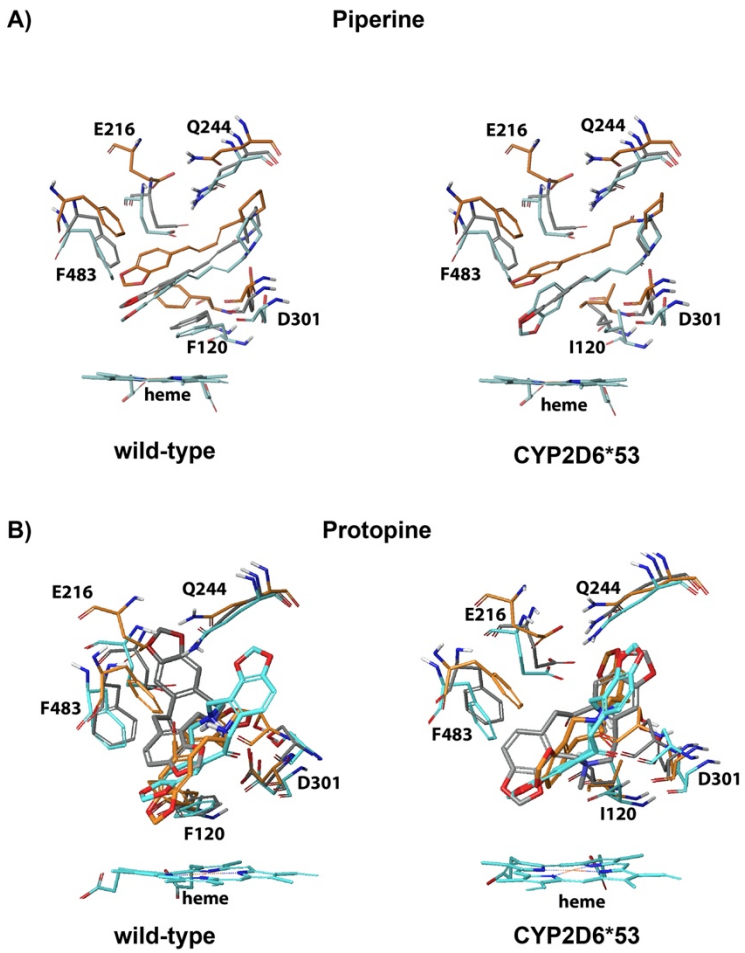


Figure 5.3 **CYP2D6 WT and CYP2D6*53 inhibition binding modes for A) piperine and B) protopine.** The best scored poses are shown docked into 4WNW (cyan), 4WNU (orange) and 4WNT (grey). To keep a clearer overview, interactions are not indicated and only the heme of 4WNW is displayed.

The binding modes observed (proximal and also distal) for honokiol and magnolol in CYP2D6 WT and *CYP2D6*53* were such that one of the two ethylene tails pointed towards the heme and one of the hydroxyl groups formed electrostatic interactions with either D301 or a stabilizing hydrogen bond with S304 (**Figures 5.4 and S5.2**). In addition, for CYP2D6 WT, hydrophobic interactions with F120 and one of the phenyl rings was observed.

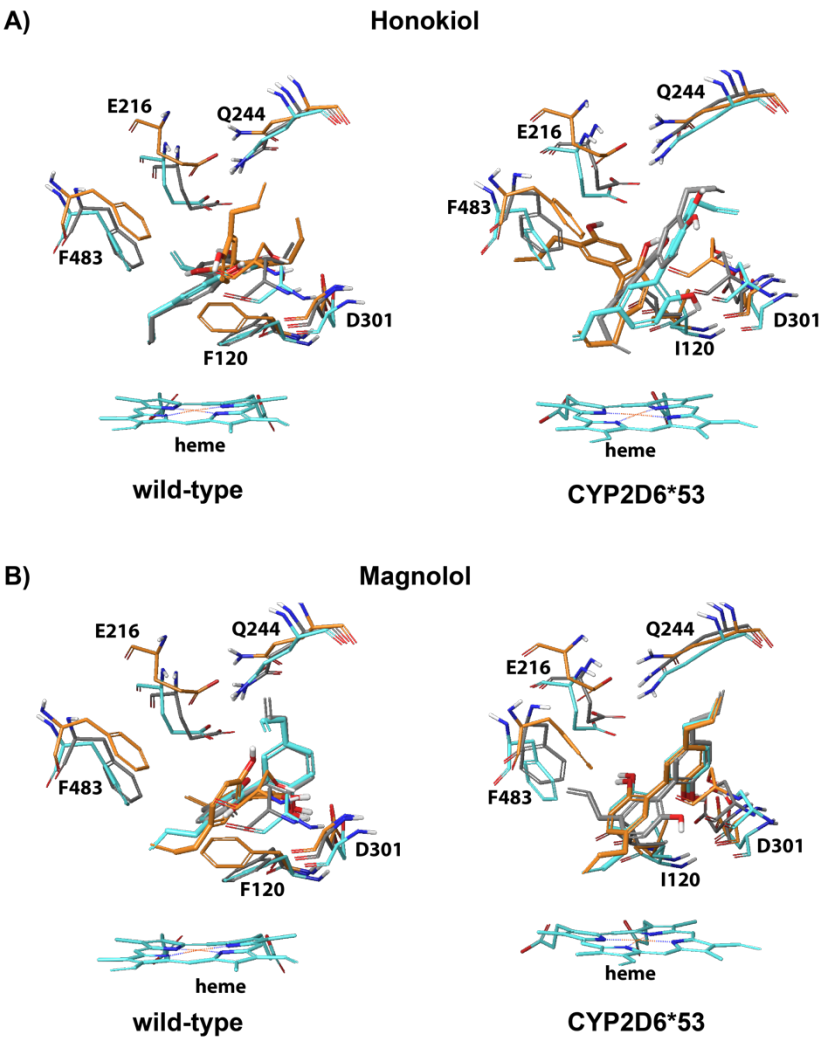


Figure 5.4 **CYP2D6 WT and CYP2D6*53 inhibition binding modes for A) Honokiol and B) Magnolol.** The best scored poses are shown docked into 4WNW (cyan), 4WNU (orange) and 4WNT (grey). Interactions were excluded to keep a clearer overview and only the heme of 4WNW is displayed.

For cannabidiol, the best scored binding modes pointed the aliphatic pentyl tail towards the heme and one of the two hydroxyl groups was stabilized by either a hydrogen bond (S304, Q244) or electrostatic interaction with D301 in CYP2D6 WT and *CYP2D6*53* (**Figure S5.1B**). In CYP2D6 WT stabilizing hydrophobic interactions between F120 and the aromatic ring were found for honokiol and magnolol (**Figure S5.2**). For cytisine, a binding pose pointing towards potential inhibition was only observed in the CYP2D6 WT; the protonated nitrogen of cytisine formed an electrostatic interaction with D301. For isorhynchophylline no reasonable binding poses that could explain its inhibitory effect on CYP2D6 were generated.

The compounds L-theanine, orcinol, naringenin and psoralidin for which experimental studies confirmed no CYP2D6 WT inhibition (**Figure 5.2, Table S5.1**), docking results did neither show any decisive inhibition binding modes in CYP2D6 WT or *CYP2D6*53*. For the remaining 8 compounds for which no experimental CYP2D6 inhibition data was found, 4-hydroxyisoleucine, chelidonic acid, salvigenin, and trans-methylisoeugenol did not show any CYP2D6 inhibition binding modes in both CYP2D6 WT or *CYP2D6*53*. The proximal binding poses of thymol (5-isopropyl-2-methylphenol) formed hydrophobic interactions with F120 and hydrophobic/electrostatic interactions with the heme group in CYP2D6 WT. For the distal binding modes of (S)-auraptenol, the hydroxyl group formed a hydrogen bond with S304 or/and also electrostatic interactions with D301 or E216 in CYP2D6 WT or *CYP2D6*53* (**Figure S5.1A**). In addition, π-π stacking between the hydrophobic ring and F120 was observed for the highest ranked pose in CYP2D6 WT. For D-(-)-synephrine proximal binding was found in CYP2D6 WT with mainly electrostatic stabilizing interactions (E216 and Q244) whereas proximal binding *in CYP2D6*53* with the protonated nitrogen forming electrostatic interactions with D301 (**Figure S5.1B**). Scopoletin showed a similar proximal binding mode in which the oxygen pointed in the direction of the heme group in both CYP2D6 WT and *CYP2D6*53* (**Figure S5.1B**).

Pharmacokinetic and toxicity-related descriptors

Several Several pharmacokinetic descriptors calculated with QikProp can be found in **Tables 5.4 and 5.5**. Chelidonic acid, L-theanine and naringenin have all a PSA above 100 Å, and 4-hydroxyisoleucine and isorhynchophylline exceed the number of acceptable hydrogen bond donors or acceptors for CNS drugs which could decrease their ability to permeate the BBB (**Table 5.2**), although experimentally evidence for all except chelidonic acid exists that they cross the BBB (**Table S5.3**). From the toxicity-related calculated QikProp descriptors (**Table 5.4**), 8 compounds have one or more value(s) lying outside the recommended regions. 4-Hydroxyisoleucine, chelidonic acid and L-theanine have QP P_{caco} and QP P_{MDCK} predicted values indicating potential issues with cell permeability, which is also indicated by their predicted low human oral

absorption and inactive (below 0) CNS score. Naringenin and psoralidin are also predicted to be CNS inactive. Honokiol, magnolol, isorhynchophylline, psoralidin and salvigenin are predicted to interact with the hERG channel. Furthermore, the online prediction platform Molinspiration (www.molinspiration.com) was used for prediction of a bioactivity score towards six of the most important drug classes (GPCRs, ion channel modulators, kinase inhibitors, nuclear receptor ligands, protease inhibitors, enzyme inhibitors) based on the structure similarity with typical class representatives (Table 5.5).

Table 5.4 **QikProp calculated toxicity-related descriptors.**

Compound	QP log HERG	QPP Caco (nm/s)	Qlog BB	QPP MDCK (nm/s)	QPlog Khsa	HOA (%)	CNS
(-)-Cytisine	-3.8	490.6	0.4	253.5	-0.35	79	1
4-Hydroxyisoleucine	-1.5	**15.4**	-0.6	**7.6**	-0.99	34	**-1**
5-Isopropyl-2-methylphenol	-3.6	3696.9	0.1	2032.9	0.06	100	1
Auraptenol	-4.0	1322.6	-0.5	669.3	-0.17	96	0
Chelidonic acid	0.3	**1.3**	-1.8	0.6	-1.18	**26**	**-2**
D-(-)-Synephrine	-4.8	192.5	-0.3	92.2	-0.61	69	0
Honokiol	**-5.8**	1614.8	-0.7	830.4	0.57	100	0
Isorhynchophylline	**-6.1**	263.4	-0.3	129.4	0.20	86	1
L-Theanine	-1.0	**5.6**	-1.1	4.3	-1.28	**23**	**-2**
Magnolol	**-5.7**	1716.8	-0.6	887.3	0.57	100	0
Naringenin	-5.0	130.0	-1.4	54.5	-0.02	74	**-2**
Orcinol	-3.3	911.7	-0.4	447.6	-0.54	85	0
Piperine	-4.8	3980.1	-0.1	2201.8	-0.02	100	0
Protopine	-4.4	1297.8	0.7	725.4	-0.42	93	2
Psoralidin	**-5.7**	311.8	-1.3	140.4	0.37	89	**-2**
Salvigenin	**-5.2**	1504.0	-0.5	769.0	0.15	100	0
Scopoletin	-3.8	626.8	-0.6	298.6	-0.48	82	0
Trans-Methylisoeugenol	-3.9	9906.0	0.1	5899.3	0.10	100	1
Cannabidiol (CBD)	-3.8	490.6	0.4	253.5	-0.35	79	1

A bioactivity prediction score above 0 indicates a high similarity to existing active compound. Considering the fact that GPCRs are paramount for regulation of mood, pain, cognition, and neurotransmitter release through synaptic transmission, predicted bioactivity towards those receptors would be expected for the modeled compounds [49]. Several CNS drugs (e.g. tricyclics such as amitriptyline) modulate ion channel activity to reduce depression in human [50,51] hence a bioactivity score for ion channel modulation above 0 would be considered as a positive result. For the remaining four drug classes (kinase inhibitors, nuclear receptor ligands, protease

inhibitors, enzyme inhibitors) no similarity (activity) is desired. **Table 5.5** shows that cytisine, isorhynchophylline, naringenin, protopine, and CBD are predicted to be similar (active) towards the GPCR ligand and/or the ion channel modulator drug class. Honokiol, naringenin, psoralidin, and CBD are predicted to be similar (active) towards one of the four undesired drug classes.

Table 5.5 **Bioactivity prediction score for the 19 compounds towards six important drug classes.** Favorable predication scores are highlighted yellow, unfavorable red.

Compound | Drug class	GPCR ligand	ion channel modulator	kinase inhibitor	nuclear receptor ligand	protease inhibitor	enzyme inhibitor
(-)-Cytisine	**-0.58**	**0.39**	**-0.75**	**-1.1**	**-0.62**	**-0.25**
4-Hydroxyisoleucine	-0.72	-0.26	-1.31	-0.96	-0.71	-0.31
5-Isopropyl-2-methylphenol	-1.02	-0.51	-1.15	-0.7	-1.25	-0.56
Auraptenol	-0.24	-0.56	-0.73	0.14	-0.56	0.11
Chelidonic acid	-1.05	-0.56	-0.93	-1.02	-0.92	-0.4
D-(-)-Synephrine	-0.39	0.07	-0.79	-0.51	-0.88	-0.04
Honokiol	**0.04**	**0.06**	**-0.08**	**0.32**	**-0.2**	**0.13**
Isorhynchopylline	**0.26**	**0.19**	**-0.3**	**0.03**	**-0.22**	**0.01**
L-Theanine	-0.53	-0.15	-1.15	-1.42	-0.08	-0.4
Magnolol	-0.01	0.05	-0.15	0.2	-0.23	0.07
Naringenin	**0.03**	**-0.2**	**-0.26**	**0.42**	**-0.12**	**0.21**
Orcinol	-2.26	-1.64	-2.35	-2.1	-2.59	-1.77
Piperine	0.15	-0.18	-0.13	-0.13	-0.1	0.04
Protopine	**0.2**	**0.07**	**-0.35**	**-0.24**	**-0.07**	**0.17**
Psoralidin	**-0.2**	**-0.09**	**-0.17**	**0.53**	**-0.15**	**0.21**
Salvigenin	-0.11	-0.27	0.15	0.13	-0.29	0.11
Scopoletin	-1	-0.65	-0.95	-0.81	-1.16	-0.24
Trans-Methylisoeugenol	-0.95	-0.53	-0.98	-0.72	-1.2	-0.53
Cannabidiol (CBD)	**0.35**	**-0.14**	**-0.48**	**0.38**	**-0.19**	**0.33**

Discussion

In this study the safety profile of 19 natural compounds was evaluated on CYP2D6 inhibition using an *in silico* approach. Inhibition was investigated for CYP2D6 WT and the allelic variant *CYP2D6*53* which is associated with increased metabolism activity for bufuralol. The variant contains two amino acid mutations (F120I, A122S) of which F120I is positioned in close vicinity of the heme. Comparison of the docking binding poses of the wild-type and variant can provide insight on the function of the mutation in modulating the accessibility of the compound to the heme. In addition, it may give an indication regarding the impact of allelic variants with amino acid mutations close to the heme on inducing inhibition. Moreover, it is generally accepted that accurate site of metabolism prediction requires to probe the accessibility and the reactivity of

the ligand atoms [52]. However, for the aim of this study in which the focus is on inhibition, using a time-efficient structure-based method such as docking to investigate the compound-CYP interactions and subsequently deduce its inhibition potential has previously shown to be a valuable pre-screening strategy [31,34,53]. Subsequently, *in vitro* and *in vivo* safety studies can be performed for examination of the smaller compound set. The results from the natural compound docking correlated 6 out of 7 experimentally confirmed CYP2D6 WT inhibitors; honokiol, magnolol, cannabidiol, piperine, protopine and cytisine. In the allelic variant *CYP2D6*53*, potent inhibition binding modes were observed as well except for cytisine. The docking poses of (S)-auraptenol and scopoletin in CYP2D6 WT and *CYP2D6*53* pointed towards potential inhibition activity as well. 5-Isopropyl-2-methylphenol formed electrostatic interactions with the prosthetic heme group in CYP2D6 WT only. The mutation F120I enabled for some compounds (especially D-(-)-synephrine, scopoletin, (-)-cytisine) to get closer to the heme-oxygen. These compounds have in common that they are relatively small and less flexible, which restricts their possible binding modes to the heme-oxygen compared to larger more flexible compounds such as for example cannabidiol or magnolol.

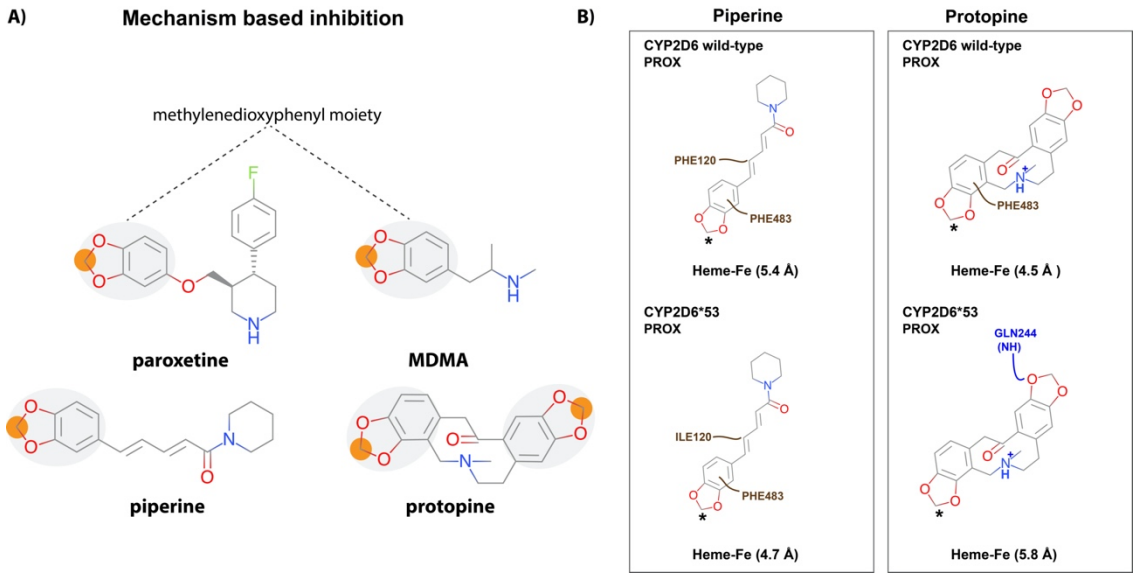


Figure 5.5 **Mechanism based CYP2D6 inhibition. A)** Piperine and protopine both contain a methylenedioxyphenyl moiety which has been associated with mechanism-based inhibition of CYP2D6 (e.g. paroxetine and MDMA are CYP2D6 MBI). **B)** The best scored binding mode (in 2D) for each is shown in CYP2D6 WT and CYP2D6*53. The asterisk indicates the atom closest to the heme-iron. Color code: hydrophobic interactions; brown, electrostatics; purple, and hydrogen bonds; blue.

An interesting observation was made for the alkaloids piperine and protopine both containing a methylenedioxyphenyl moiety (**Figure 5.5A**). The highest scored docking pose of piperine and protopine in both CYP2D6 WT and *CYP2D6*53* positioned the moiety in proximal vicinity of the heme (**Figures 5.5B and 5.3**).

Conclusion

Natural remedies are often associated with a safe use. Buying herbal supplements seems to be harmless as most of the time no special warnings can be found in the accompanying information. However, an increasing trend of using natural compounds as single antidepressant or in polypharmacy needs to be handled with caution. Toxicity issues may arise depending on the dose and therapy and its P450 metabolism dependence if the herbal compound or a complex product is not given safety clearance by the FDA or another established authority. CYP2D6 inhibition or off-target binding are particularly dangerous. The *in silico* results of this study indicate for several of the natural compounds suggested to be used as antidepressant that a potential increased risk may exist on adverse reactions triggered by CYP2D6 inhibition or another off-target analysed (kinase inhibitors, nuclear receptor ligands, protease inhibitors, enzyme inhibitors, hERG binding). From the 19 natural compounds analysed, 9 indicated no CYP2D6 inhibition for both CYP2D6 WT and *CYP2D6*53* (4-hydroxyisoleucine, chelidonic acid, isorhynchopylline, L-theanine, naringenin, orcinol, psoralidin, salvigenin, trans-methylisoeugenol). The other 10 natural compounds ((-)-cytisine, 5-isopropyl-2-methylphenol, s-auraptenol, D-(-)-synephrine, honokiol, magnolol, piperine, protopine, scopoletin and cannabidiol) showed clearly potential for CYP2D6 WT and/or *CYP2D6*53* inhibition. If administered with a concomitant drug which depends on CYP2D6 activity this may lead to adverse reactions. Further experimental investigations are required to confirm the outcome of the *in silico* docking, the off-target predictions and the toxicity descriptors. Especially hERG binding should be evaluated for honokiol, magnolol, isorhynchopylline, psoralidin and salvigenin.

References

1. Bashar MA, Mehra A, Aggarwal AK. Integrating mental health into primary care for addressing depression in a rural population: An experience from North India. Indian J Psychiatry. 2019 Jun;61(3):319–21.

2. Abbing-Karahagopian V, Huerta C, Souverein PC, de Abajo F, Leufkens HGM, Slattery J, et al. Antidepressant prescribing in five European countries: application of common definitions to assess the prevalence, clinical observations, and methodological implications. Eur J Clin Pharmacol. 2014 Jul;70(7):849–57.

3. Forns J, Pottegård A, Reinders T, Poblador-Plou B, Morros R, Brandt L, et al. Antidepressant use in Denmark, Germany, Spain, and Sweden between 2009 and 2014: Incidence and comorbidities of antidepressant initiators. Journal of Affective Disorders. 2019 Apr 15;249:242–52.

4. Iacobucci G. NHS prescribed record number of antidepressants last year. BMJ. 2019 Mar 29;364:l1508.

5. Kemper KJ, Gardiner P, Birdee GS. Use of Complementary and Alternative Medical (CAM) Therapies Among Youth with Mental Health Concerns. Acad Pediatr [Internet]. 2013 [cited 2019 Nov 19];13(6). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3834537/

6. Asher GN, Gerkin J, Gaynes BN. Complementary Therapies for Mental Health Disorders. Med Clin North Am. 2017 Sep;101(5):847–64.

7. Kemppainen LM, Kemppainen TT, Reippainen JA, Salmenniemi ST, Vuolanto PH. Use of complementary and alternative medicine in Europe: Health-related and sociodemographic determinants. Scand J Public Health. 2018 Jun;46(4):448–55.

8. Wong AD, Ye M, Levy AF, Rothstein JD, Bergles DE, Searson PC. The blood-brain barrier: an engineering perspective. Front Neuroeng [Internet]. 2013 Aug 30 [cited 2019 Sep 19];6. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3757302/

9. Sanchez-Covarrubias L, Slosky LM, Thompson BJ, Davis TP, Ronaldson PT. Transporters at CNS Barrier Sites: Obstacles or Opportunities for Drug Delivery? Curr Pharm Des. 2014;20(10):1422–49.

10. Vendel E, Rottschäfer V, de Lange ECM. Improving the Prediction of Local Drug Distribution Profiles in the Brain with a New 2D Mathematical Model. Bull Math Biol. 2019 Sep 1;81(9):3477–507.

11. Dong X. Current Strategies for Brain Drug Delivery. Theranostics. 2018 Feb 5;8(6):1481–93.

12. Santarsieri D, Schwartz TL. Antidepressant efficacy and side-effect burden: a quick guide for clinicians. Drugs Context [Internet]. 2015 Oct 8 [cited 2019 Sep 13];4. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4630974/

13. Carvalho AF, Sharma MS, Brunoni AR, Vieta E, Fava GA. The Safety, Tolerability and Risks Associated with the Use of Newer Generation Antidepressant Drugs: A Critical Review of the Literature. PPS. 2016;85(5):270–88.

14. Wang S-M, Han C, Bahk W-M, Lee S-J, Patkar AA, Masand PS, et al. Addressing the Side Effects of Contemporary Antidepressant Drugs: A Comprehensive Review. Chonnam Med J. 2018 May;54(2):101–12.

15. Ekor M. The growing use of herbal medicines: issues relating to adverse reactions and challenges in monitoring safety. Front Pharmacol. 2014 Jan 10;4:177.

16. Vahabi S, Eatemadi A. Phyto-anesthetics: A mini-review on herb–anesthesia drug interactions. Biomedicine & Pharmacotherapy. 2016 Dec 1;84:1885–90.

17. Singh A, Zhao K. Chapter Nine - Herb–Drug Interactions of Commonly Used Chinese Medicinal Herbs. In: Zeng B-Y, Zhao K, editors. International Review of Neurobiology [Internet]. Academic Press; 2017 [cited 2019 Sep 12]. p. 197–232. (Neurobiology of Chinese Herb Medicine; vol. 135). Available from: http://www.sciencedirect.com/science/article/pii/S0074774217300193

18. Hermann R, Richter O von. Clinical Evidence of Herbal Drugs As Perpetrators of Pharmacokinetic Drug Interactions. Planta Med. 2012 Sep;78(13):1458–77.

19. Zhou S-F. Polymorphism of human cytochrome P450 2D6 and its clinical significance: Part I. Clin Pharmacokinet. 2009;48(11):689–723.

20. Don CG, Smieško M. Out-compute drug side effects: Focus on cytochrome P450 2D6 modeling. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2018;8(5):e1366.

21. Correia MA. Ortiz de Montellano, P. R., (2015) Cytochrome P450: Structure, Mechanism, and Biochemistry, 4rd Edition, Chapter 7, pp 177-259.

22. Kalgutkar AS, Obach RS, Maurer TS. Mechanism-based inactivation of cytochrome P450 enzymes: chemical mechanisms, structure-activity relationships and relationship to clinical drug-drug interactions and idiosyncratic adverse drug reactions. Curr Drug Metab. 2007 Jun;8(5):407–47.

23. Watanabe A, Takakusa H, Kimura T, Inoue S, Kusuhara H, Ando O. Difference in Mechanism-Based Inhibition of Cytochrome P450 3A4 and 3A5 by a Series of Fluoroquinolone Antibacterial Agents. Drug Metab Dispos. 2016 Jan 1;dmd.116.073783.

24. Subehan null, Usia T, Kadota S, Tezuka Y. Mechanism-Based Inhibition of Human Liver Microsomal Cytochrome P450 2D6 (CYP2D6) by Alkamides of Piper nigrum. Planta Med. 2006 Apr;72(6):527–32.

25. Sridar C, Kent UM, Notley LM, Gillam EMJ, Hollenberg PF. Effect of tamoxifen on the enzymatic activity of human cytochrome CYP2B6. J Pharmacol Exp Ther. 2002 Jun;301(3):945–52.

26. Hollenberg PF, Kent UM, Bumpus NN. Mechanism-based inactivation of human cytochromes p450s: experimental characterization, reactive intermediates, and clinical implications. Chem Res Toxicol. 2008 Jan;21(1):189–205.

27. Livezey M, Nagy LD, Diffenderfer LE, Arthur EJ, Hsi DJ, Holton JM, et al. Molecular analysis and modeling of inactivation of human CYP2D6 by four mechanism based inactivators. Drug Metab Lett. 2012 Mar;6(1):7–14.

28. Hellum BH, Nilsen OG. The in vitro Inhibitory Potential of Trade Herbal Products on Human CYP2D6-Mediated Metabolism and the Influence of Ethanol. Basic & Clinical Pharmacology & Toxicology. 2007;101(5):350–8.

29. Gurley BJ, Swain A, Hubbard MA, Williams DK, Barone G, Hartsfield F, et al. Clinical assessment of CYP2D6-mediated herb-drug interactions in humans: Effects of milk thistle, black cohosh, goldenseal, kava kava, St. John's wort, and Echinacea. Mol Nutr Food Res. 2008 Jul;52(7):755–63.

30. Qu Q, Qu J, Han L, Zhan M, Wu L, Zhang Y, et al. Inhibitory effects of phytochemicals on metabolic capabilities of CYP2D6*1 and CYP2D6*10 using cell-based models in vitro. Acta Pharmacol Sin. 2014 May;35(5):685–96.

31. Hochleitner J, Akram M, Ueberall M, Davis RA, Waltenberger B, Stuppner H, et al. A combinatorial approach for the discovery of cytochrome P450 2D6 inhibitors from nature. Sci Rep [Internet]. 2017 Aug 14;7. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5556109/

32. Tm V, K I, K V, N KB, R MK, Gp D. Inhibitory Potency of Selected Therapeutic Bioactive Molecules of Standardized Terminalia arjuna (Roxb.) Extract on CYP3A4 and CYP2D6: Exploring Possible Herb-Drug Interactions. Natural Products Chemistry & Research. 2017;5(4):1–7.

33. Sakuyama K, Sasaki T, Ujiie S, Obata K, Mizugaki M, Ishikawa M, et al. Functional Characterization of 17 CYP2D6 Allelic Variants (CYP2D6.2, 10, 14A–B, 18, 27, 36, 39, 47–51, 53–55, and 57). Drug Metab Dispos. 2008 Dec 1;36(12):2460–7.

34. Martiny VY, Carbonell P, Chevillard F, Moroy G, Nicot AB, Vayer P, et al. Integrated structure- and ligand-based in silico approach to predict inhibition of cytochrome P450 2D6. Bioinformatics. 2015 Dec 15;31(24):3930–7.

35. He Q, Liu J, Liang J, Liu X, Li W, Liu Z, et al. Towards Improvements for Penetrating the Blood–Brain Barrier—Recent Progress from a Material and Pharmaceutical Perspective. Cells [Internet]. 2018 Mar 23 [cited 2019 Nov 19];7(4). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5946101/

36. Pajouhesh H, Lenz GR. Medicinal Chemical Properties of Successful Central Nervous System Drugs. NeuroRx. 2005 Oct;2(4):541–53.

37. Ghose AK, Herbertz T, Hudkins RL, Dorsey BD, Mallamo JP. Knowledge-Based, Central Nervous System (CNS) Lead Selection and Lead Optimization for CNS Drug Discovery. ACS Chem Neurosci. 2011 Nov 2;3(1):50–68.

38. Koes DR, Baumgartner MP, Camacho CJ. Lessons Learned in Empirical Scoring with smina from the CSAR 2011 Benchmarking Exercise. Journal of Chemical Information and Modeling. 2013 Aug 26;53(8):1893–904.

39. Wager TT, Hou X, Verhoest PR, Villalobos A. Moving beyond Rules: The Development of a Central Nervous System Multiparameter Optimization (CNS MPO) Approach To Enable Alignment of Druglike Properties. ACS Chem Neurosci. 2010 Mar 25;1(6):435–49.

40. Sarrica A, Kirika N, Romeo M, Salmona M, Diomede L. Safety and Toxicology of Magnolol and Honokiol. Planta Med. 2018 Nov;84(16):1151–64.

41. Zhang J, Chen Z, Huang X, Shi W, Zhang R, Chen M, et al. Insights on the Multifunctional Activities of Magnolol [Internet]. BioMed Research International. 2019 [cited 2019 Sep 12]. Available from: https://www.hindawi.com/journals/bmri/2019/1847130/

42. Jeong H-U, Kong TY, Kwon SS, Hong S-W, Yeon SH, Choi J-H, et al. Effect of honokiol on cytochrome P450 and UDP-glucuronosyltransferase enzyme activities in human liver microsomes. Molecules. 2013 Sep 3;18(9):10681–93.

43. Yamaori S, Okamoto Y, Yamamoto I, Watanabe K. Cannabidiol, a major phytocannabinoid, as a potent atypical inhibitor for CYP2D6. Drug Metab Dispos. 2011 Nov;39(11):2049–56.

44. Li J, Gödecke T, Chen S-N, Imai A, Lankin D, Farnsworth NR, et al. In vitro metabolic interactions between black cohosh (Cimicifuga racemosa) and tamoxifen via inhibition of cytochromes P450 2D6 and 3A4. Xenobiotica [Internet]. 2011 Aug 9 [cited 2019 Sep 12]; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3565011/

45. Shamsi S, Tran H, Tan RSJ, Tan ZJ, Lim LY. Curcumin, Piperine, and Capsaicin: A Comparative Study of Spice-Mediated Inhibition of Human Cytochrome P450 Isozyme Activities. Drug Metab Dispos. 2017;45(1):49–55.

46. Rowland P, Blaney FE, Smyth MG, Jones JJ, Leydon VR, Oxbrow AK, et al. Crystal structure of human cytochrome P450 2D6. J Biol Chem. 2006 Mar 17;281(11):7614–22.

47. Wang B, Yang L-P, Zhang X-Z, Huang S-Q, Bartlam M, Zhou S-F. New insights into the structural characteristics and functional relevance of the human cytochrome P450 2D6 enzyme. Drug Metab Rev. 2009;41(4):573–643.

48. Wang A, Stout CD, Zhang Q, Johnson EF. Contributions of Ionic Interactions and Protein Dynamics to Cytochrome P450 2D6 (CYP2D6) Substrate and Inhibitor Binding. J Biol Chem. 2015 Feb 20;290(8):5092–104.

49. Huang Y, Thathiah A. Regulation of neuronal communication by G protein-coupled receptors. FEBS Letters. 2015 Jun 22;589(14):1607–19.

50. Kaczorowski GJ, McManus OB, Priest BT, Garcia ML. Ion Channels as Drug Targets: The Next GPCRs. J Gen Physiol. 2008 May;131(5):399–405.

51. Lodge NJ, Li Y-W. Ion channels as potential targets for the treatment of depression. Curr Opin Drug Discov Devel. 2008 Sep;11(5):633–41.

52. Tyzack JD, Kirchmair J. Computational methods and tools to predict cytochrome P450 metabolism for drug discovery. Chem Biol Drug Des. 2019 Apr;93(4):377–86.

53. Dutkiewicz Z, Mikstacka R. Structure-Based Drug Design for Cytochrome P450 Family 1 Inhibitors. Bioinorg Chem Appl. 2018;2018:3924608.

54. Liston HL, DeVane CL, Boulton DW, Risch SC, Markowitz JS, Goldman J. Differential time course of cytochrome P450 2D6 enzyme inhibition by fluoxetine, sertraline, and paroxetine in healthy volunteers. J Clin Psychopharmacol. 2002 Apr;22(2):169–73.

55. Rodgers JT, Davydova NY, Paragas EM, Jones JP, Davydov DR. Kinetic mechanism of time-dependent inhibition of CYP2D6 by 3,4-methylenedioxymethamphetamine (MDMA): Functional heterogeneity of the enzyme and the reversibility of its inactivation. Biochemical Pharmacology. 2018 Oct 1;156:86–98.

56. Ferreira LG, dos Santos RN, Oliva G, Andricopulo AD. Molecular Docking and Structure-Based Drug Design Strategies. Molecules. 2015 Jul 22;20(7):13384–421.

57. Ballante F. Protein-Ligand Docking in Drug Design: Performance Assessment and Binding-Pose Selection. Methods Mol Biol. 2018;1824:67–88.

58. Śledź P, Caflisch A. Protein structure-based drug design: from docking to molecular dynamics. Current Opinion in Structural Biology. 2018 Feb 1;48:93–102.

59. Cournia Z, Allen B, Sherman W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. J Chem Inf Model. 2017 Dec 26;57(12):2911–37.

60. Mobley DL, Gilson MK. Predicting Binding Free Energies: Frontiers and Benchmarks. Annu Rev Biophys. 2017 22;46:531–58.

61. Koukos PI, Xue LC, Bonvin AMJJ. Protein–ligand pose and affinity prediction: Lessons from D3R Grand Challenge 3. J Comput Aided Mol Des. 2019;33(1):83–91.

62. Friesner RA, 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.

63. Olah J, Mulholland AJ, Harvey JN. Understanding the determinants of selectivity in drug metabolism through modeling of dextromethorphan oxidation by cytochrome P450. Proceedings of the National Academy of Sciences. 2011 Apr 12;108(15):6050–5.

64. Reinhard FGC, Visser SP de. Biodegradation of cosmetics products: A computational study of Cytochrome P450 metabolism of phthalates. In 2017.

65. Shityakov S, Neuhaus W, Dandekar T, Förster C. Analysing molecular polar surface descriptors to predict blood-brain barrier permeation. Int J Comput Biol Drug Des. 2013;6(1–2):146–56.

66. Smith QR. Transport of Glutamate and Other Amino Acids at the Blood-Brain Barrier. J Nutr. 2000 Apr 1;130(4):1016S-1022S.

67. Ioakimidis L, Thoukydidis L, Mirza A, Naeem S, Reynisson J. Benchmarking the Reliability of QikProp. Correlation between Experimental and Predicted Values. QSAR & Combinatorial Science. 2008;27(4):445–56.

Supporting Information

Table S5.1 **Overview of literature sources for CYP2D6 inhibition.**

Table S5.2 **Overview of literature sources for P450 metabolism.**

Table S5.3 **Overview of literature sources for blood brain barrier (BBB) crossing.**

Figure S5.1 **The best scored binding modes (in 2D) for A) (-)-Cytisine, (S)-auraptenol and 5-isopropyl-2-methylphenol and B) Scopoletin, cannabidiol (CBD) and D-(-)-synephrine are shown in CYP2D6 WT and CYP2D6*53.**

Figure S5.2 **The best scored binding modes (in 2D) for honokiol and magnolol are shown in CYP2D6 WT and CYP2D6*53.**

Table S5.4 **Overview of the various types of interactions found in the best binding modes for CYP2D6 WT.**

Table S5.5 **Overview of the various types of interactions found in the best binding modes for CYP2D6*53.**

Table S5.6 **Overview QikProp descriptors with Schrödinger explanation.**

Table S5.1. **Overview of literature sources for CYP2D6 inhibition.** If literature for a compound on the topic could not be found, this is indicated as "unknown".

Natural Product	CYP2D6 inhibition	Reference
(-)-Cytisine	moderate	[29]
4-Hydroxyisoleucine	unknown	-
5-Isopropyl-2-methylphenol	unknown	-
Auraptenol	unknown	-
Chelidonic acid	unknown	-
D-(-)-Synephrine	unknown	-
Honokiol	moderate	[48]
Isorhynchophylline	moderate	[29]
L-Theanine	none	[57]
Magnolol	moderate	[47]
Naringenin	none	[65]
Orcinol	none	[49]
Piperine	weak	[44]
Protopine	potent	[43]
Psoralidin	none	[70]
Salvigenin	unknown	-
Scopoletin	unknown	-
Trans-Methylisoeugenol	unknown	-
cannabidiol (CBC)	potent	[49]

Table S5.2. **Overview of literature sources for P450 metabolism.** If literature for a compound on the topic could not be found, this is indicated as "unknown".

Natural Product	P450 metabolism (major isoforms)	Reference
(-)-Cytisine	unknown	-
4-Hydroxyisoleucine	unknown	-
5-Isopropyl-2-methylphenol	CYP2A6	[63]
Auraptenol	unknown	-
Chelidonic acid	unknown	-
D-(-)-Synephrine	unknown	-
Honokiol	CYP1A2, CYP2C8, CYP2C9, CYP2C19	[58]
Isorhynchophylline	CYP2D, CYP1A1, CYP1A2, CYP2C	[40]
L-Theanine	none	[57]
Magnolol	CYP2E1, CYP3A4, CYP1A2	[74]
Naringenin	P450-dependent	[66]
Orcinol	unknown	-
Piperine	unknown	-
Protopine	CYP2D1, CYP2C11	[55]
Psoralidin	CYP2C19	[70]
Salvigenin	unknown	-
Scopoletin	unknown	-
Trans-Methylisoeugenol	P450-dependent	[62]
cannabidiol (CBC)	CYP2C19, CYP3A4	[60]

Table S5.3. **Overview of literature sources for crossing of blood brain barrier (BBB).** If literature for a compound on the topic could not be found, this is indicated as "unknown".

Natural Product	BBB	Reference
(-)-Cytisine	poor	[59]
4-Hydroxyisoleucine	unknown	-
5-Isopropyl-2-methylphenol	unknown	-
Auraptenol	unknown	-
Chelidonic acid	unknown	-
D-(-)-Synephrine	poor	[61]
Honokiol	√	[64]
Isorhynchophylline	√	[72]
L-Theanine	√	[71]
Magnolol	√	[53]
Naringenin	√	[56]
Orcinol	√	[40]
Piperine	√	[68]
Protopine	√	[69]
Psoralidin	unknown	-
Salvigenin	√	[54]
Scopoletin	√	[67]
Trans-Methylisoeugenol	unknown	-
cannabidiol (CBC)	√	[73]

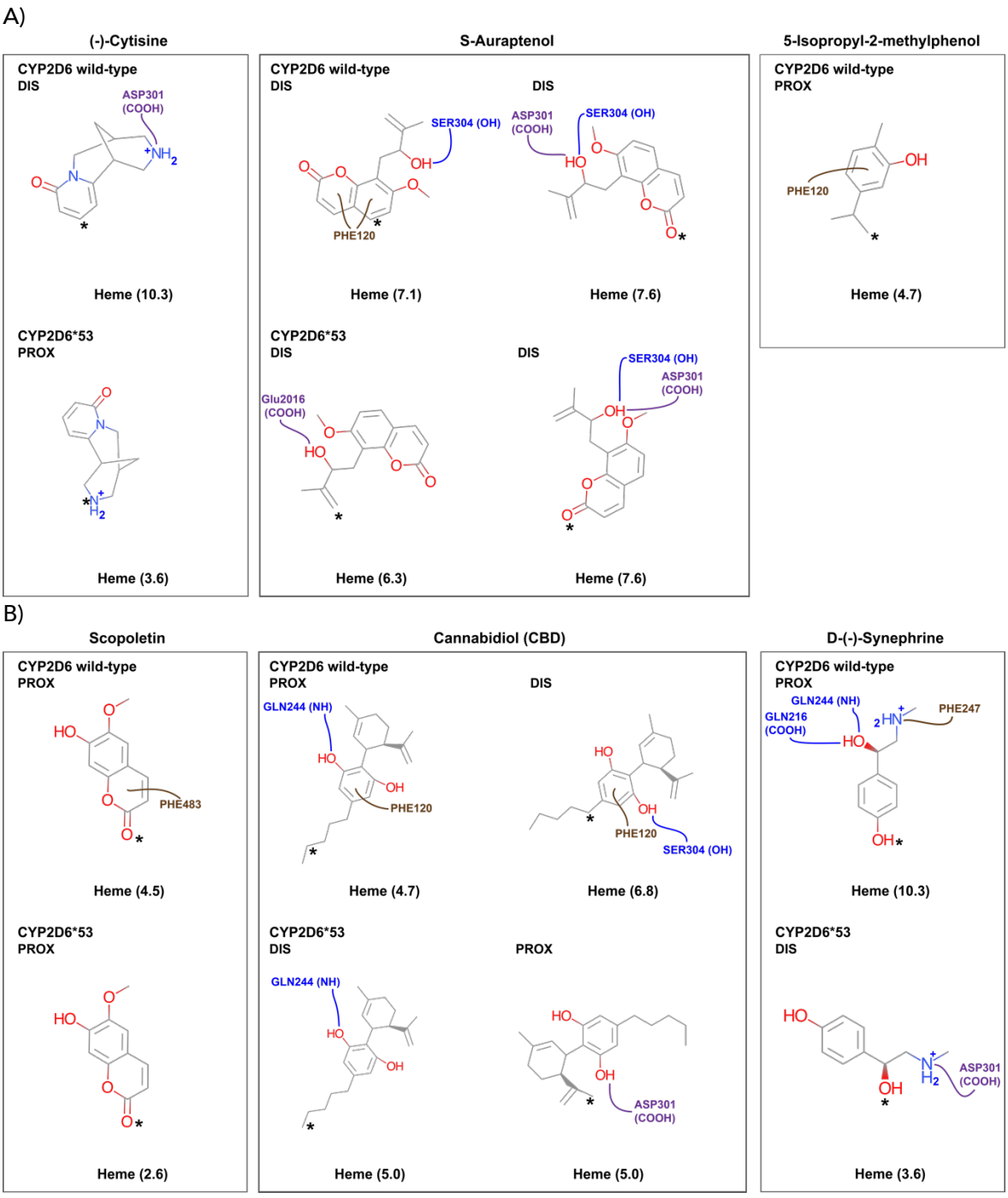


Figure S5.1 **The best scored binding modes (in 2D) for A) (-)-cytisine, (S)-auraptanol and 5-isopropyl-2-methylphenol and B) Scopoletin, cannabidiol (CBD) and D-(-)-synephrine are shown in CYP2D6 WT and CYP2D6*53.** The asterisk indicates the atom closest to the heme-iron. Color code: hydrophobic interactions; brown, electrostatics; purple, and hydrogen bonds; blue. Proximal: distance between heme-Fe and closest atom < 6 Å and distal; > 6 Å.

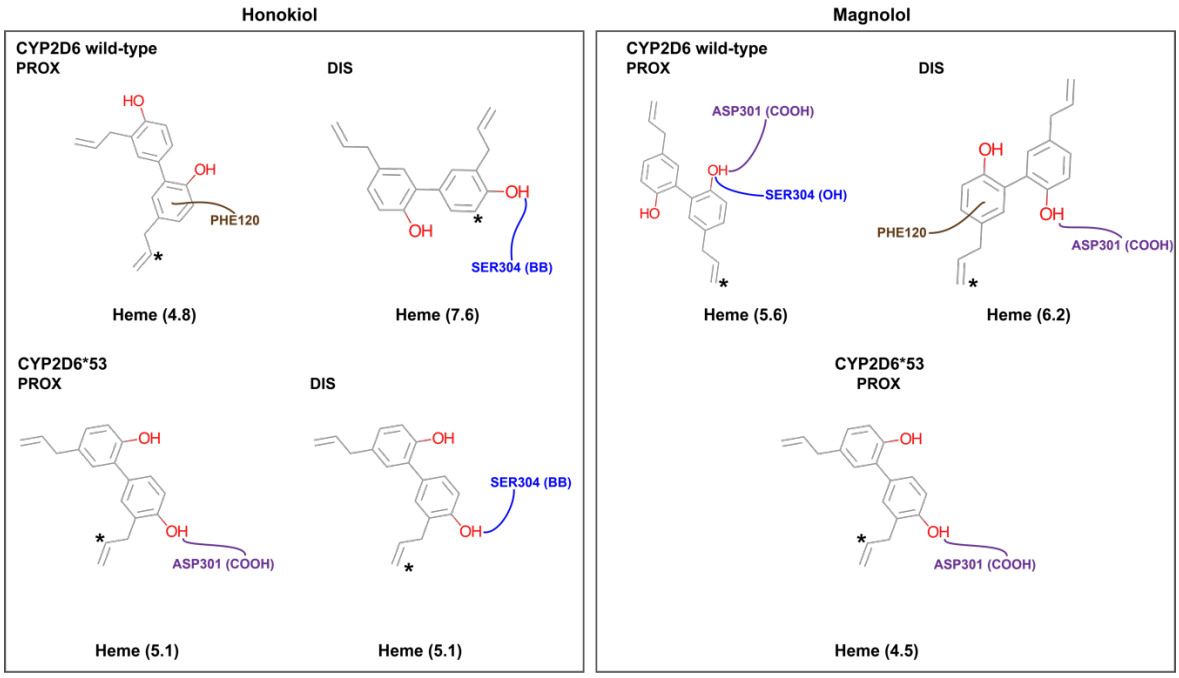


Figure S5.2 **The best scored binding modes (in 2D) for honokiol and magnolol are shown in CYP2D6 WT and CYP2D6*53.** The asterisk indicates the atom closest to the heme-iron. Color code: hydrophobic interactions; brown, electrostatics; purple, and hydrogen bonds; blue. Proximal: distance between heme-Fe and closest atom < 6 Å and distal; > 6 Å.

Table S5.4 **Overview of the various types of interactions found in the best binding modes for CYP2D6 WT.**

Natural Product		4WNW WT		Distance	Heme
		AA	Type	(Å)	(Å)
Thymol		F120	π stacking	~3.8	4.7
s-auraptenol		D301	ionic	2.3	7.6
		S304	H-bond	2.4	
D-(-)-synephrine	**01**	D301	ionic	2.6	6.1
		F120	π-stacking	~ 4.7	
	02	E216	ionic	2.3	5.1
		Q244	H-bond	2.3	
		F247	π stacking	~ 4.3	
Honokiol	**01**	F120	π stacking	~ 4.3	4.7
	02	F120	π stacking	~ 3.9	7.6
		S304	H-bond	2.5	
Magnolol	**01**	D301	ionic	1.9	6.0
		F120	π stacking	~ 4.0	
	02	D301	ionic	2.6	5.4
		S304	H-bond	2.3	
Piperine		F120	π stacking	~3.8	5.4
		F483	π stacking	~ 4.0	
Protopine		Q244	H-bond	2.4	5.9
		F120	π stacking	3.1	
		F120	π stacking	6.5	
		F483	π stacking	3.5	
Scopoletin		S304	H bond	2.4	5.3
		F120	π stacking	~ 3.8	
Cannabidiol	**01**	Q244	H-bond	2.3	4.7
		F120	π stacking	3.6	
	02	S304	H-bond	2.4	7.8
		F120	π stacking	~ 4.0	

Table S5.5 **Overview of the various types of interactions found in the best binding modes for CYP2D6*53.**

Natural Product		4WNW V53		Distance	Heme
		AA	Type	(Å)	(Å)
Thymol		F120	π stacking	~3.8	4.7
(S)-auraptenol		D301	ionic	2.3	7.6
		S304	H-bond	2.4	
D-(-)-synephrine	**01**	D301	ionic	2.6	6.1
		F120	π-stacking	~ 4.7	
	02	E216	ionic	2.3	5.1
		Q244	H-bond	2.3	
		F247	π stacking	~ 4.3	
Honokiol	**01**	F120	π stacking	~ 4.3	4.7
	02	F120	π stacking	~ 3.9	7.6
		S304	H-bond	2.5	
Magnolol	**01**	D301	ionic	1.9	6.0
		F120	π stacking	~ 4.0	
	02	D301	ionic	2.6	5.4
		S304	H-bond	2.3	
Piperine		F120	π stacking	~3.8	5.4
		F483	π stacking	~ 4.0	
Protopine		Q244	H-bond	2.4	5.9
		F120	π stacking	3.1	
		F120	π stacking	6.5	
		F483	π stacking	3.5	
Scopoletin		S304	H bond	2.4	5.3
		F120	π stacking	~ 3.8	
Cannabidiol	**01**	Q244	H-bond	2.3	4.7
		F120	π stacking	3.6	
	02	S304	H-bond	2.4	7.8
		F120	π stacking	~ 4.0	

Table S5.6 **Overview QikProp descriptors with Schrödinger explanation.**

Descriptor	Explanation (Schrödinger – QikProp)
QPlogHERG	Predicted IC_{50} value for blockage of HERG K+ channels
QPPCaco	Predicted apparent Caco-2 cell permeability in nm/sec. Caco2 cells are a model for the gut-blood barrier. QikProp predictions are for non-active transport.
QPlogBB	Predicted brain/blood partition coefficient. Note: QikProp predictions are for orally delivered drugs so, for example, dopamine and serotonin are CNS negative because they are too polar to cross the blood-brain barrier
QPPMDCK	Predicted apparent MDCK cell permeability in nm/sec. MDCK cells are considered to be a good mimic for the blood brain barrier. QikProp predictions are for non-active transport.
QPlogKhsa	Prediction of binding to human serum albumin
HOA (%)	Predicted human oral absorption on 0 to 100% scale. The prediction is based on a quantitative multiple linear regression model.
CNS	Predicted central nervous system activity on a –2 (inactive) to +2 (active) scale.

Chapter 6

Conclusions and future needs in CYP2D6 research

This PhD work explored several aspects of CYP2D6 pharmacogenetics at the atomic level using molecular modeling. Various computational techniques were applied in order to gain a clearer picture of how amino acid mutations within the CYP2D6 binding pocket can alter enzyme activity and its selectivity. Deciphering the key factors and their contribution in the overall catalytic cycle is essential in order to develop accurate metabolism and toxicity prediction tools. Such tools can be used by the research community to extend our current CYP2D6 knowledge and also by medicinal chemists working in the CADD department to evaluate - from early stage on - the chances that a lead compound might interact with CYP2D6 WT or clinical relevant allelic variants. This insight can be valuable for the rational design of new drugs and lower or prevent treatment failure due to an unfavorable metabolic profile.

In this work we have shown that a reliable site of metabolism prediction requires the assessment of the energetic and steric accessibility of the potential metabolic hot spots in the molecule within the active site and the intrinsic reactivity of the metabolic hot spot towards the heme-oxygen. The interplay between the residues within the binding pocket and the ligand can modulate its access and orientation towards the heme as observed in the increased activity variant CYP2D6*53 (F120I, A122S) during 1 microsecond MD simulations. By studying the hydroxylation of bufuralol using QM/MM, we found that the amino acid mutation F120I in close vicinity of the heme led to a lower activation barrier for the major site of metabolism. It has been found in previous studies that in smaller binding pockets in which the ligand can adopt only a limited number of conformations, the steric factors become more relevant. On the other hand, in a larger binding pocket, the ligand adopts multiple conformations which are in thermodynamic equilibrium, hence the kinetic factor – the activation barrier - will be more important in directing the selectivity. Also, the large flexibility of CYP2D6 requires for each ligand that the binding pocket is properly adjusted by for example induced fit, a short MD simulation based minimization or ensemble docking. The MD simulations of membrane-anchored CYP2D6 showed that the formation of tunnels, which provide the access and egress pathways for the ligands to and from the heme, can be seen as a compound pre-filtering step. The residues lining the tunnel

will favor particular ligands over others (e.g. ligands with a more hydrophobic nature, small to medium sized). This is also demonstrated by our finding that several natural herbal-derived products were, just like drugs, able to bind within the active site and inhibit CYP2D6 wild-type and and the allelic variant CYP2D6*53. This shows that as long as the compounds share the molecular characteristics that a particular CYP isoform favors, there is a higher probability that they will enter the CYP, reach the active site and the heme. Hence, there is a need to perform for any compound, rational designed or derived from a natural source, a profound safety assessment including P450 inhibition screening before approving it as a supplement or natural/herb product based therapy.

As of today, considerable progress has been made in the development of *in silico* CYP2D6 predictive tools especially by application of QSAR, pharmacophore modeling, molecular docking, MD simulations, and their integrative use. Nevertheless, the complete framework in which these elegant enzymes operate is large and several challenges still need to be addressed in future. Within the context of structure-based metabolism and toxicity predictions, improvements can be expected for the handling of the enzyme flexibility and the incorporation of polymorphism data by using methods such as MD simulations or also elastic networks. Furthermore, considering the level of expertise often required to be able to use such tools, tailoring the approaches into automated implemented protocols, accessible at a user-friendly webserver provides benefit also to less-experienced users. A relevant question which adds to the complexity of the toxicity predictions is the observation that more than one molecule can bind simultaneously within the active site of several isoforms including CYP2D6 and CYP3A4. Little information is available on how this will impact the binding and enzyme activity. Further studies tackling this phenomenon need to be performed in order to know how to deal with this in predictive tools. Furthermore, the outcome of computational toxicity predictions need to be interpreted relative and not as absolute answer. Any computational outcome is based on a model and every model has its inherent limitation. For instance, dose or exposure rates are difficult parameters to include into the model. Hence, these techniques will not be able to

completely replace the *in vivo* and *in vitro* approaches for predicting toxicity. The emerging position of the *in silico* Drug, Metabolism and Pharmacokinetics (DMPK) expert which works with medicinal chemists, computational chemists and DMPK scientists demonstrates the value of synergistic collaboration in order to intertwine both approaches more closely and improve the safety of the drug.

Summary

Our body is daily challenged by a plethora of chemical substances coming from diverse sources; cosmetics that are applied to the skin (e.g. phthalates, parabens, and volatile fragrances), contaminated food (e.g. with pesticides; organophosphate, organochlorine, or carbamate), drugs causing unforeseen adverse reactions, air pollution (e.g. ozone, carbon monoxide, and lead), pollutants that are difficult to remove from the wastewater (e.g. drugs and heavy metals), and many others found sometimes at unexpected sources such as children toys or cashier bills. The elimination of such toxins occurs mainly in the liver. Here, a large enzyme family, the P450 cytochromes (CYPs), is responsible for the breakdown (metabolism) of xenobiotics (foreign molecules: drugs, environmental pollutants, carcinogens, and other chemicals) and endogenous metabolites (molecules produced by the body: e.g. fatty acids, steroids, prostaglandins). Although the compounds are structurally diverse, they all share a rather lipophilic character. For elimination from the body, the compounds need to be transformed into more hydrophilic (water soluble) molecules. This is the task of CYPs; the chemical reactions catalyzed by the CYPs make compounds more hydrophilic by adding and/or transforming functional groups, e.g. addition of a hydroxyl group to a carbon atom through oxidation. Subsequently, clearance from our body occurs mainly through excretion by the kidneys. Unravelling parts of the complex framework in which CYPs operate and selectively catalyze compounds has started about 60 years ago. Although several discoveries have been paramount for building our P450s knowledge, many aspects related to the interplay between the enzyme and its environment and the major driving forces of the catalysis reaction within each isoform remain unclear.

One family member, CYP2D6, has been the main focus of many research studies primarily due to its high polymorphism rate (gene modification) which can lead to serious toxicity issues especially within the context of drug therapies. The research described in this work confirms how beautifully complex the enzymes are and nevertheless provides some essential clues regarding the CYP2D6 framework and its function. This new knowledge may contribute to the development of more accurate

in silico CYP P450 mediated metabolism and toxicity prediction and modeling software. Last but not least it can potentially aid in the design of more efficient CYP biocatalysts.

In **Chapter 1** a basic introduction is given which should provide the reader with the fundamental knowledge needed to understand the following chapters. As start, an overview of the earliest P450 experiments that led to the discovery of the versatile P450 enzyme family is given. Subsequently, P450 family member CYP2D6 is introduced, the consequences of CYP2D6 pharmacogenetics on clinical treatment and the significance of exploring CYP2D6-drug interactions as early as possible within the the drug design process are highlighted. Following section focuses on the P450 catalytic reaction cycle together with a detailed explanation of the P450 hydroxylation reaction pathways. After that, the state-of-the-art modeling techniques used to study CYP2D6 and its environment follow. This section is divided into ligand-based and structure-based methods. The last part of this chapter contains a comprehensive description of the fundamental theory underlying molecular dynamics (MD) simulations and quantum mechanics/ molecular mechanics (QM/MM) calculations as they were the major computational methods applied within this thesis research.

In **Chapter 2**, CYP2D6 pharmacogenetics is investigated by determining the impact of amino acid mutations, located in close vicinity of the reactive heme center, on the enzyme activity. Two CYP2D6 variants (increased and no activity) are used in addition to the wild-type, and bufuralol, a typical CYP2D6 substrate, to study the aliphatic and aromatic hydroxylation mechanism within different active site compositions. Experimental research identified one major and two minor sites of metabolism (SoM). One additional site is included as a negative control. In the first part of the study, molecular docking is used for the generation of the poses and well-tempered metadynamics (WT-MTD) simulation is employed to investigate the free energy surface (FES) profile for each SoM within each CYP2D6 variant. The two collective variables (CVs) chosen to drive the WT-MTD simulation are correlated to the transition state geometry and thus can provide insight on the following questions:

(i) *is the lowest energy pose of the substrate in the active site also is the pose in which the monitored SoM is positioned favourably to the expected transition state (TS) geometry*, and (ii*) is it possible to distinguish major and minor SoMs based on the minima in the free energy profiles of each SoM*.

Assessment of the FES profiles indicates that neither the first nor the second question can be answered by solely comparing the FES profiles of each SoM within each allelic variant. Keeping in mind the accuracy limitations of each individual force field describing the free energy landscape, this knowledge may point towards reliability issues of SoM toxicity predictions solely based on docking methods. In the second part of the study, the contribution of the intrinsic reactivity of the substrate within different active site environments on the overall catalytic reaction is explored. In order to properly account for the reactivity, a quantum mechanics method is required and considering the large size of the system in question, a QM/MM approach is most frequently used. The starting structure for each SoM and the control site within the wild-type and increased activity allelic variant CYP2D6*53 could be obtained from the WT-MTD simulation performed before. The QM region comprised of around 150 atoms and the MM part the remaining solvated protein. Using a starting structure with a geometry closest to the transition state structure, a M06-2X/LACV3P* QM/MM minimization was run and from the optimized structure the intermediate structure was generated and also QM/MM minimized. Subsequently a transition search was performed to localize the transition state geometry. Single point energy calculations (M06-2X/LACV3P**) on each minimized structure enabled to define the activation barrier for each SoM. The QM/MM results demonstrate first of all that it is crucial to define the stereoelectronic complementarity of a ligand for an accurate CYP2D6 metabolism prediction. Secondly, valuable insight was gained regarding the F120I amino acid mutation within the active site that can selectively steer the efficiency of the C-H bond activation; the conserved F120I mutation led to a lower activation barrier for the major S1 and minor S2 SoM. Nevertheless, the enantioselectivity and stereoselectivity remained the same. Hence, we anticipate that a mutation with a smaller amino acids, is needed to influence the enantioselectivity as experimentally

observed for the F120A mutation. In **Chapter 3**, apo and holo MD simulations of one microsecond were performed in order to evaluate the impact of the amino acid location (active site or substrate recognition site) and the type (conservation) on the dynamic behaviour of the wild-type and five allelic variants with normal (one), increased (one), decreased (two) and inactive (one) activity. The motivation for this study was to improve our current knowledge regarding the structure-function activity correlation of the wild-type and the five selected allelic variants. Such information can be of great value for the rational design of drugs and can aid in the further development of CYP2D6 toxicity modeling tools. The selection of the variants was based on their clinical relevance and also if they shared at least one common mutation with another variant (all four do so except CYP2D6*53). Also the majority or all of the amino acid mutations within one variant were located in or within close vicinity of the binding pocket (with CYP2D6*4 as exception). Furthermore, to assess the impact of substrate or inhibitor on the dynamics of the enzyme fold three substrates (bufuralol, tamoxifen and veliparib) and two inhibitors (prinomastat and quinidine) were used. Analysis of the trajectories indicated that the presence of a ligand (substrate or inhibitor) generally had a stabilizing effect on the enzyme structure. Furthermore, investigation of the hydrogen bonding networks in the binding pocket revealed that especially non-conserved amino acid mutations located very close to the heme (e.g. T107I or A122S positioned at the BC-loop), can modulate the flexibility of two important loops (BC- and FG loop) which facilitate the access from three of the major access tunnels (2c/2e/2b) to the reactive heme center and vice versa. In comparison to the wild-type, missing or additional hydrogen bonds led to increased flexibility of one of the loops (depending on the mutation location) and subsequently influenced the ligand pathway positively (more accessible) or negatively (more steric hindrance). We proposed that such a mechanism may contribute to increased (CYP2D6*53) or decreased (CYP2D6*17) enzyme activity. During the course of the project, it was recognized that the presence of the membrane to which CYP2D6 is normally anchored in its native environment will likely have a stabilizing impact on the structure and therefore also influences the dynamics of the pathways which steer the ligand to and

from the active site. At this point, Mr. André Fischer joined our research group to perform his master thesis project. He was very motivated and agreed on performing similar MD simulations as described in this chapter but then with CYP2D6 anchored to the membrane. Hence in **Chapter 4**, the results of the work he did during his master project under my supervision can be found. In order to define how the membrane influences the dynamics of CYP2D6 wild-type and the five allelic variants, membrane-anchored CYP2D6 models were constructed, evaluated and used for subsequent MD simulations. Next, an in-depth analysis of the tunnels and the overall fold was performed to identify differences which may mechanistically explain the changes in enzyme activity. The most striking findings were the following; for the allelic variants associated with decreased activity, lower opening frequencies of the most prominent tunnels (2b/2c) providing access to the heme could be observed. For the allelic variants associated with increased activity, it could be shown that for one of the major tunnels (2b) an enlargement of the bottleneck part (minimal diameter) occurred (related to the F120I mutation), thereby improving the access to the heme, thus promoting a higher ligand turn-over. Furthermore, smaller active site volumes could be correlated with decreased or no enzyme activity variants whereas larger ones with increased enzyme activity. Compared to the study in Chapter 3, the membrane had a stabilizing impact on the enzyme, especially its overall fold and decreased the average fluctuation of the FG - and BC loops. Both studies confirmed the crucial function of the FG- and BC loops to steer the closing and opening mechanism of the tunnels in close vicinity of the heme. **Chapter 5** focuses as an application study on analysing the *in silico* safety profile of several herbal-based compounds suggested to bear antidepressant activity, in particular within the context of CYP2D6 inhibition. Synthetic antidepressants are one of the major drug categories which depend on CYP2D6 activity before reaching any therapeutic effect. Since several years there is an upward trend in the depression occurrence within the society and also in the number of people that use herbal/natural products in combination therapy as a treatment. However, accurate safety profiling studies, as required for any approved drug, are frequently incompletely performed if not completely lacking for natural products. This may pose

a considerable risk to the patient, as for example inhibition of CYP2D6 may lead to accumulation of the concomitantly administered drug and cause severe side effects. In addition, CYP2D6 pharmacogenetics is an additional risk factor which can have a significant effect on the response and the clearance of the substrate drug. Hence, this computational study was performed to indicate which of the natural product compounds should be prioritized for further experimental safety studies as they likely increase the chance on adverse reactions, in particular through CYP2D6 inhibition. An initial dataset of 51 herbal-based products suggested to have antidepressant activity was compiled. Subsequently, the products which have little chance, based on their physicochemical parameters, to reach the brain through passive transport were filtered out. Neurotherapeutic drugs required to cross the blood brain barrier (BBB) in order to modulate the central nervous system (CNS). For the design of CNS drugs, an adjusted version of the famous Lipinski rules was compiled before that increase their chance to be absorbed and reach the brain through passive transport. These guidelines were used to filter the initial dataset and 19 herbal-based anti-depressant products passed and were further analysed. First, a literature study was performed to find any evidence regarding BBB permeability, CYP2D6 inhibition and P450 metabolism. Subsequently, all compounds were docked into CYP2D6 wild-type and the increased activity allelic variant CYP2D6*53. The top 10 best scored poses were visually inspected to determine if they could act (bind) as an inhibitor. In addition, several toxicity-related descriptors were calculated that could provide hints on off-target binding. For several of the natural compounds studied the outcome points towards an increased risk on adverse reactions due to CYP2D6 inhibition or another off-target analysed. Further experimental studies are required to confirm these indications. The final chapter, **Chapter 6**, presents the most important findings from this thesis and the future perspectives on the directions of the CYP2D6 pharmacogenetic modeling field are given. The major challenges are described as well as potential approaches that can aid in accelerating new CYP2D6 discoveries and translating them into valuable metabolism and toxicity modeling tools.

Abbreviations

CYP2D6	Cytochrome P450 2D6
MD	Molecular Dynamics
WT-MTD	Well Tempered Metadynamics
QM/MM	Quantum Mechanics/Molecular Mechanics
DFT	Density Function Theory
CV	Collective Variable
FES	Free Energy Surface
SoM	Site of Metabolism
SRS	Site Recognition Residues
POR	Cytochrome P450 Oxidoreductase
PM	Poor Metabolizer
IM	Intermediate Metabolizer
NM	Normal Metabolizer
UM	Ultrarapid Metabolizer
3D	Three-Dimensional
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation

List of Publications

[1] **Don, C.G**. und Smieško, M. (2019) «In silico pharmacogenetics CYP2D6 study focused on the pharmacovigilance of herbal antidepressants», *Front. Pharmacol.*, 11;683. doi: 10.3389/fphar.2020.00683.

[2] Roel-Touris, J., **Don, C.G**., Vargas Honorato, R, Rodrigues, P.G.L.M., und Bonvin, A.M.J.J., (2019) «Less is more: Coarse-grained integrative modeling of large biomolecular assemblies with HADDOCK», J. Chem. Theory Comput., 9b00310. doi: 10.1021/acs.jctc.9b00310.

[3] **Don, C.G**. und Smieško, M. (2018) «Out-compute drug side effects: Focus on cytochrome P450 2D6 modeling», *Wiley Interdisciplinary Reviews: Computational Molecular Science*. Wiley, S. e1366. doi: 10.1002/wcms.1366.

[4] Fischer, A., **Don, C.G**., und Smieško, M. (2018) «Molecular Dynamics Simulations Reveal Structural Differences among Allelic Variants of Membrane-Anchored Cytochrome P450 2D6», *Journal of chemical information and modeling*. American Chemical Society, 58(9), S. 1962-1975. doi: 10.1021/acs.jcim.8b00080.

[5] **Don, C. G.** und Smieško, M. (2018) «Microsecond MD simulations of human CYP2D6 wild-type and five allelic variants reveal mechanistic insights on the function», *PloS one*, 13(8), S. e0202534. doi: 10.1371/journal.pone.0202534.

[6] Smieško, M., **Don, C. G.**, Meuwly, R., Kucsera, S. und Brüschweiler, B. J. (2018) «Large-scale in silico screening of compounds contained in printing inks for food packaging materials», *Toxicology Letters*, 295, S. 96.

[7] **Don, C.G.** und Riniker, S. (2014) «Scents and sense: in silico perspectives on olfactory receptors», *Journal of computational chemistry*, 35(32), S. 2279-2287. doi: 10.1002/jcc.23757.

Submitted manuscript

[8] **Don, C.G.** und Smieško, M. (2020) « Deciphering reaction determinants of altered-activity CYP2D6 variants by well-tempered metadynamics simulation and QM/MM calculations»

Curriculum Vitae

Charleen Don was born on the 10th of June 1988 in Hilversum, the Netherlands.

In 2006 she obtained her high school diploma from Luzac College (accelerated program which offers the last two years of high school into 1 year). That same year she entered Utrecht University to study Arts and German. The next year, she went to study law at the University of Lucerne. Than she enrolled again at Utrecht University to start with undergraduate studies in Chemistry and a minor in Entrepreneurship. In the last year of her bachelor studies (2011) she joined the Computational Structural Biology Group of Prof. Dr. Alexandre Bonvin which mainly focuses on the characterization of biomolecular interactions.

In 2014 she received her Master in Science degree (Biochemistry, GPA 4.0). During her Masters two research projects were carried out, one with Prof. Dr. Alexandre Bonvin at Utrecht University, and one with Prof. Dr. Philippe Hünenberger at the laboratory of Physical Chemistry, ETH Zürich.

After that, she performed two more research projects of which the first project was with Dr. Guido Capitani at the Laboratory for Biomolecular Research, Paul Scherer Institut, and the second one with Prof. Dr. Sereina Riniker at the the laboratory of Physical Chemistry, ETH Zürich.

In February 2017 she joined the Computational Pharmacy Group of (at that time) PD Dr. Martin Smieško. The results of this research have been published in several scientific journals and are described in this Ph.D. thesis. During her PhD studies, Charleen has been involved as teaching assistant in several teaching courses and practica including Modern Drug Design and Modeling of Adverse Effects and supervised several bachelor and master students.