Heinrich Heine
Universität
Düsseldorf

# Molecular determinants underlying substrate promiscuity and polymorphism in drug-metabolizing CYP enzymes

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von
**Daniel Fritz Walter Becker**
aus Berlin

Düsseldorf, Januar 2024

Aus dem Institut für Pharmazeutische und Medizinische Chemie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstattende:
1. Prof. Dr. Holger Gohlke
2. Prof. Dr. Vlada Urlacher

Tag der mündlichen Prüfung: 29.01.2024

# EIDESSTATTLICHE ERKLÄRUNG

Ich, Daniel Fritz Walter Becker, versichere an Eides statt, dass die Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Diese Dissertation wurde in der vorgelegten oder einer ähnlichen Form noch bei keiner anderen Institution eingereicht, und es wurden bisher keine erfolglosen Promotionsversuche von mir unternommen.

Düsseldorf, im Januar 2024

**To my family**

*“There is no royal road to science, and only those who do not dread the fatiguing climb of its steep paths have a chance of gaining its luminous summits.”*

– Karl Marx

# TABLE OF CONTENTS

# LIST OF PUBLICATIONS

This thesis is based on the following research papers and manuscripts:

**Becker, D.;** Bharatam, P. V.; Gohlke, H.

*F/G Region Rigidity is Inversely Correlated to Substrate Promiscuity of Human CYP Isoforms Involved in Metabolism*

J. Chem. Inf. Model. **2021,** 61, 4023-4030**.**

Impact Factor reported in 2021: 6.162

---

**Becker, D.;** Bharatam, P. V.; Gohlke, H.

*Molecular mechanisms underlying single nucleotide polymorphism-induced reactivity decrease in CYP2D6*

Manuscript in preparation

---

Additionally authored publications during this thesis:

Twizerimana, A. P.; Scheck, R.; **Becker, D.;** Zhang, Z.; Wammers, M.; Avelar, L.; Pflieger, M.; Häussinger, D.; Kurz, T.; Gohlke, H.; Münk, C.

*Cell Type-Dependent Escape of Capsid Inhibitors by Simian Immunodeficiency Virus SIVcpz*

J. Virol. **2020**, 94, e01338-20.

Impact Factor reported in 2020: 5.103

---

Jaguva Vasudevan, A. A.; **Becker, D.;** Luedde, T.; Gohlke, H.; Münk, C.

*Foamy Viruses, Bet, and APOBEC3 Restriction*

Viruses **2021**, 13, 504.

Impact Factor reported in 2021: 5.818

---

Twizerimana, A. P.; **Becker, D.;** Zhu, S.; Luedde, T.; Gohlke, H.; Münk, C.

*The cyclophilin A binding loop of the capsid regulates the human TRIM5α sensitivity of nonpandemic HIV-1*

PNAS **2023,** 120, e2306374120.

Impact Factor reported in 2023: 12.779

---

# PROCEEDINGS AND CONFERENCE CONTRIBUTIONS

**Becker, D.;** Bharatam, P. V.; Gohlke, H.

*F/G-region rigidity is inversely correlated to substrate promiscuity of human CYP isoforms involved in metabolism.*

**DPhG Annual Meeting**, online, September 28 - October 1, **2021.**

---

**Becker, D.;** Bharatam, P. V.; Gohlke, H.

*Molecular determinants underlying substrate promiscuity and polymorphism in drug-metabolizing CYP enzymes*.

**NIC Symposium 2022**, Forschungszentrum Jülich, September 29 - 30, **2022**

---

Schott Verdugo, S.; Mirzaeigarakani, T.; Graf von Westarp, W.; Rothkranz, B.; **Becker, D.;** Schwaneberg, U.; Jupke, A.; Rother, D.; Gohlke, H.

*Continuous, stable processes for the sustainable enzymatic production of chiral amino alcohols integrating downstream processing*

**BioSC Forum**, Forschungszentrum Jülich, May 17, **2023**

---

**Becker, D.;** Schott Verdugo, S.; Gohlke, H.

*Prediction of thermal weak spots in* Ap*PDC by Constraint Network Analysis to enable more efficient enzyme optimization*

**EMBO Workshop Computational structural biology**, EMBL Heidelberg, December 6 - 9, **2023**

---

# ABBREVIATIONS

| | |
|---|---|
| ABC | ATP-binding cassette |
| ADME | absorption, distribution, metabolism, and excretion |
| AKR | aldol-keto reductase |
| BSEP | bile-salt export pump |
| CNA | Constraint Network Analysis |
| COMT | catechol O-methyl transferase |
| CYP | Cytochrome P450 |
| CG | comparison/control group |
| CPR | Cytochrome P450 reductase |
| ER | endoplasmic reticulum |
| FAD | Flavine dinucleotide |
| FIRST | Floppy Inclusion and Rigidity Substructure Topology |
| FMN | Flavine mononucleotide |
| FMO | flavin-containing monooxygenase |
| GSH | glutathione |
| GST | glutathione *S*-transferase |
| HEM | heme |
| MAO | monoamine oxidase |
| MATE1 | toxin extrusion protein 1 |
| MBD | membrane binding domain |
| MD | Molecular dynamics |
| MRP | multidrug resistance-related protein |
| NADH | Nicotinamide adenine dinucleotide |

| | |
|---|---|
| NADPH | Nicotinamide adenine dinucleotide phosphate |
| NAPQI | *N*-acetyl-*p*-benzo-quinone imine |
| NAT | *N*-acetyltransferase |
| NTCP | $Na^+$-dependent taurocholate cotransporting polypeptide |
| nonCG | not in the control group |
| OAT | organic anion transporter |
| OATP | organic anion transporting polypeptide |
| P-gp | P-glycoprotein |
| RMSD | root-mean-square deviation |
| RMSF | root mean-square fluctuation |
| SLC | solute carrier |
| SNP | single nucleotide polymorphism |
| SULT | sulfotransferase |
| TPMT | thiopurine *S*-methyl transferase |
| UGT | UDP-glucuronosyltransferase |

# ZUSAMMENFASSUNG

Der Abbau von Arzneistoffen im menschlichen Körper ist ein entscheidender Faktor in der Arzneimitteltherapie. Dieser Abbau wird meist als Arzneimittelmetabolismus bezeichnet und umfasst die Phasen 0-III. In den Phasen 0 und III spielen Transportproteine eine entscheidende Rolle, während in Phase I Oxidoreduktasen und in Phase II Transferasen maßgeblich am Arzneimittelmetabolismus beteiligt sind. Innerhalb der Phase I sind Enzyme der Cytochrom P450 Familie (CYPs) von besonderer Bedeutung, da sie 95% aller Reaktionen dieser Phase katalysieren und deshalb häufig Ursache für Arzneistoffinteraktionen sind. Darüber hinaus führt eine untypische CYP-Aktivität oft zur Unter- oder Überdosierung von Arzneistoffen.

Daher habe ich CYPs in der vorliegenden Arbeit mittels computergestützter Methoden untersucht. Ich habe Modelle entwickelt, die dazu beitragen, das Verhalten von CYPs im Arzneimittelmetabolismus besser zu verstehen und veränderte CYP-Aktivität vorherzusagen.

In der ersten Studie habe ich untersucht, warum zwölf CYPs in der Lage sind, den Großteil aller Arzneistoffe zu oxidieren. Ein besonderer Fokus wurde auf CYP3A4 gelegt, da dieses Enzym 27 % aller Arzneimittelabbaureaktionen katalysiert. Das Substratspektrum von CYP3A4 gilt daher als besonders breit. Mit meiner Arbeit konnte ich die Bedeutung der F/G-Region bestätigen, die maßgeblich die Promiskuität von humanen CYPs bestimmt, und ein quantitatives Modell entwickeln, welches den Zusammenhang zwischen der CYP-Substratpromiskuität und der Rigidität der F/G-Region beschreibt.

In der zweiten Studie habe ich das Enzym CYP2D6 eingehend untersucht. CYP2D6 katalysiert zwar nur 6% aller Arzneimittelabbaureaktionen, jedoch sind mittlerweile über 160 verschiedene genetische Varianten bekannt, deren Enzymaktivität signifikant variiert. Für die 20 am häufigsten vorkommenden Mutationen habe ich den Einfluss der Strukturänderung auf die Aktivität der Varianten ermittelt. Mittels Molekulardynamiksimulationen konnte ich vier molekulare Mechanismen identifizieren, die zu einer geringen CYP2D6 Aktivität führen. Basierend auf meinen Simulationen habe ich einen Entscheidungsbaum erstellt, der eine Einordnung der Enzymaktivität für bislang nicht experimentell untersuchte Varianten ermöglicht.

Zusammengenommen habe ich zwei Modelle von CYP-Enzymen erstellt, die sowohl klinische Beobachtungen auf atomistischer Ebene erklären als auch die Vorhersage von CYP-Enzymaktivitäten ermöglichen.

# ABSTRACT

The degradation of drugs in the human body is a decisive factor in drug therapy. This degradation is usually referred to as drug metabolism and comprises phases 0-III. In phases 0 and III, transport proteins play a decisive role, while in phase I oxidoreductases and in phase II transferases are significantly involved in drug metabolism. Within phase I, enzymes of the cytochrome P450 family (CYPs) are of particular importance, as they catalyze 95% of all reactions in this phase and are, therefore, often the cause of drug interactions. In addition, atypical CYP activity often leads to under- or overdosing of drugs.

Therefore, I have investigated CYPs in this thesis using computational methods. I have developed models that help to better understand the behavior of CYPs in drug metabolism and predict altered CYP activity.

In the first study, I investigated why twelve CYPs are able to oxidize the majority of all drugs. A particular focus was placed on CYP3A4, as this enzyme catalyzes 27 % of all drug degradation reactions. Therefore, the substrate spectrum of CYP3A4 is considered particularly broad. With my work, I confirmed the importance of the F/G region, which determines the promiscuity of human CYPs, and developed a quantitative model that describes the relationship between CYP substrate promiscuity and the rigidity of the F/G region.

In the second study, I examined the enzyme CYP2D6 in detail. Although CYP2D6 catalyzes only 6% of all drug degradation reactions, more than 160 different genetic variants are now known. Within those variants, enzyme activity varies significantly. For the 20 most frequently occurring mutations, I have determined the influence of the structural change on the activity of the variants. Using molecular dynamics simulations, I was able to identify four molecular mechanisms that lead to low CYP2D6 activity. Based on my simulations, I created a decision tree that allows the classification of enzyme activity for variants that have not yet been studied experimentally.

To conclude, I have created two models of CYP enzymes that both explain clinical observations at the atomistic level and enable the prediction of CYP enzyme activities.

# 1 INTRODUCTION

Throughout evolutionary history, organisms were forced to process molecules they do not produce themselves and which are potentially harmful.[1] Due to the potential harm of such xenobiotics (ξένος xénos - "foreign" und βίος bíos - "life"), mechanisms have evolved to excrete them from the organism.[2] Xenobiotics are processed differently to substances produced or taken up by the body as needed. Since Xenobiotics are highly diverse, their excretion requires a system capable of handling a wide range of chemical functionalities.[3]

Hence, the increasing complexity of organisms during evolution was paralleled by an increasing complexity in the excretory machinery for xenobiotics.[4] In the case of humans, evolution led to the development of the liver, which plays a central role in the xenobiotic excretion process.[5] The liver is placed close to the gastrointestinal tract. All substances absorbed during digestion pass through the liver, as the portal vein connects the gastrointestinal tract to the liver (Figure 1). The liver actively transports xenobiotics from the blood in the portal vein to the hepatocytes.[6, 7] In the hepatocytes, xenobiotics are oxidized and conjugated before being excreted into the bile or the tissue.[8]

The oxidation process in the liver is mainly catalyzed by cytochrome P450 (CYP) enzymes.[9, 10] CYP enzymes primarily act as monooxygenases, and the most essential catalyzed reaction is the oxidation of non-activated C-H bonds.[9, 11] This reaction enhances the hydrophilicity of drugs and other xenobiotics, which accelerates excretion. However, it can either inactivate drugs or, conversely, lead to biologically active metabolites.[12] 57 CYP enzymes are encoded in the human genome.[13] Only twelve play a relevant role in human drug metabolism; all these isoforms belong to three CYP families.[14] Among the twelve relevant CYP enzymes, CYP3A4 plays a unique role because of its involvement in ~27% of all drug metabolizing reactions catalyzed by CYP enzymes.[15] Due to its prominent position and high promiscuity, CYP3A4 is a major cause of drug-drug interactions.[16, 17] The molecular reason for the high promiscuity of CYP3A4 is not fully understood.

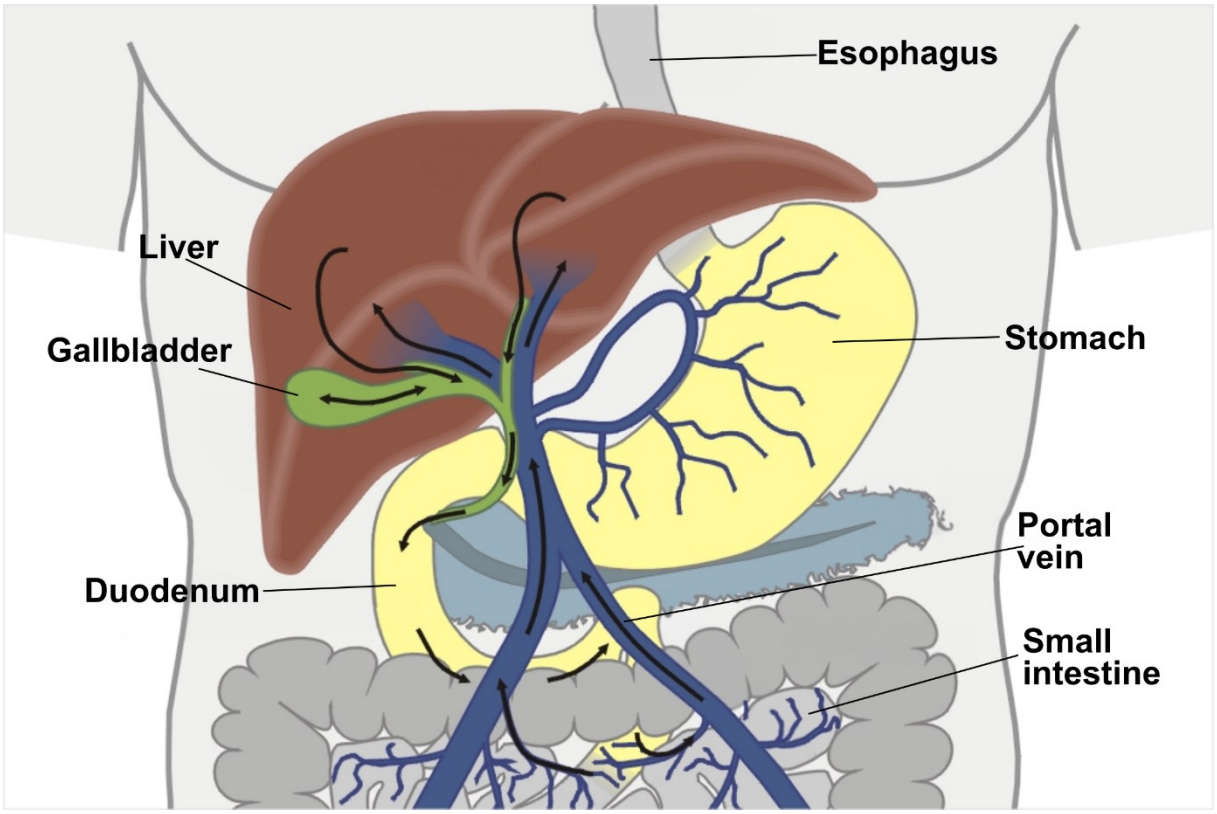



*Figure 1: The portal vein transports absorbed molecules from the gastrointestinal tract to the liver. Bile gets excreted from liver cells and is concentrated in the gallbladder. The bile from the gallbladder is transported to the duodenum, where it neutralizes the acidic milieu of the stomach and aids in fat digestion.[18] This picture is adapted from https://commons.wikimedia.org/wiki/File:Grafik_enterohepatischerkreislauf.jpg from the user Sansculotte. The figure is under the Creative Commons Attribution-Share Alike 2.5 Generic license.*

The CYP enzymes are a superfamily of heme-thiolate enzymes identified in all kingdoms of life and in viruses.[19, 20] Within the superfamily, the enzymes do not only share a cofactor (heme in complex with a cysteine (Figure 2)) but also a common fold.[21] This is remarkable since the sequence identity between CYP enzymes, even within one species, is often below 40%.[22]

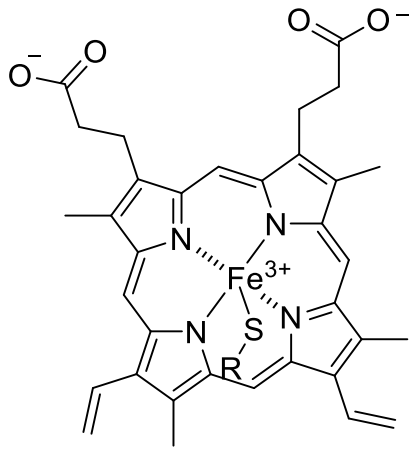



*Figure 2: Structure of heme B, the cofactor in CYP enzymes. The iron in the center forms a complex with the porphyrin moiety and the sulfur of a cysteine.*

In the last 3500 years of human development, humans started systematically treating patients depending on their illness. The first known systematic approach was the Ebers Papyrus, composed around 1550 BCE, which described remedies to be given to a patient depending on the disease.[23, 24] Many other books followed that gave physicians advice on how to treat patients. Prominent examples include the books of Galen in Antiquity and the works of Hildegard von Bingen in the Middle Ages.[25, 26] Those books were still influenced by religious ideas and based on dogmatic theories.[27] In the 17th century, medicine reinvented the concept of empiricism and opened the door to modern medicine.[28] At the end of the last century, it became more and more evident that differences in drug metabolism were also a reason for a lack of response to therapy.[29] Also, the correlation between drug response and ethnical group, sex, and age was elucidated at this time.[30] The combination of interindividual differences in drug response and differences between ethnical groups led to a modern understanding of drug metabolism, including different phenotypes caused by gene duplicates or different genotypes.[31]

Taking these differences into account is particularly important for CYP2D6,which is involved in human drug metabolism and takes part in 6% of all drug metabolizing reactions catalyzed by CYP enzymes.[15] The isoform is considered to be highly polymorphic: At least 165 genetic variations are known.[14] This genetic variability leads to a variability in enzyme activity. Differences in the activity of CYP2D6 can lead to severe side effects, such as apnea under therapy with codeine[10] or cardiac arrest under imipramine therapy.[32]

The differences in activity can be caused by overexpression (rapid metabolizer), decreased protein stability, decreased binding affinity, or reduced electron transfer to the active site of CYP2D6 (poor metabolizer). Nowadays, databases of activity-decreasing mutations are available, and genotypes are taken into account in clinical therapies.[33, 34] However, how mutations influence metabolism on an atomistic level is still not fully understood.

In this thesis, two projects aim to elucidate different aspects of the human drug metabolism catalyzed by the CYP enzyme. The first question to be addressed is which molecular mechanism leads to the broad substrate promiscuity of drug-metabolizing CYP enzymes and how these mechanisms differ and lead to different promiscuity of the CYP enzymes (Publication I). The second question to be addressed is how mutations can influence the enzyme activity of CYP2D6 (Publication II). In other words, what are the molecular mechanisms that lead to decreased activity, and which criteria can be used to predict poor metabolizers.

# 2 BACKGROUND

The role of CYP enzymes in the human drug metabolism is crucial. The enzymes are members of one enzyme superfamily and take part in the human drug metabolism. However, due to a key role in drug degradation they still cause severe problems in pharmacotherapy. In the following chapter, we will have a deeper look into the role of CYPs in human drug metabolism and their general structure. I will first give insights in the human drug metabolism and its interaction with other parts of pharmacokinetics, followed by a deeper insight into the structure of CYPs and the reactions catalyzed by CYPs.

## 2.1 Human drug metabolism

### 2.1.1 ADME concept and the role of the liver in human drug metabolism

The interaction of the human body with a drug is commonly divided into four stages, i.e., **a**bsorption, **d**istribution, **m**etabolism, and **e**xcretion (ADME).[35] For example, the active pharmaceutical ingredient of an orally taken drug is absorbed (**a**bsorption) from the gastrointestinal tract into the portal vein, which passes through the liver (**d**istribution), where the first round of metabolism takes place (**m**etabolism) (first pass effect) (Figure 1). Metabolites from the liver are then excreted via the bile (**e**xcretion). Since in most of the cases only a fraction of the drug is metabolized when passing by the liver, drug molecules reach the blood cycle and act at the desired tissue. After primary metabolization in the liver, the drug and/or its metabolites are excreted via fecal or renal routes.[36] The liver filters ~1.6 L $min^{-1}$ of blood for xenobiotics for a healthy adult, while the kidney filters only ~130 mL $min^{-1}$.[37, 38] This underlines the importance of the liver as blood filter. Hepatocytes that contain drug metabolizing CYP enzymes comprise up to 80 % of the total cell population of the human liver.[39] In addition to drug metabolism, they carry out most liver functions: protein synthesis, carbohydrate transformation, and bile formation.[40]

ADME is a part of pharmacokinetics and is of general interest because many clinical trials fail due to pharmacokinetic shortcomings of the investigated compounds.[41] The ADME concept is crucial for modeling the pharmacokinetics of a drug molecule.[42] Each part of ADME can be modeled independently, which allows predicting a drug's overall dose- and time-dependent behavior.[43] The absorption process is traditionally regarded as first-order kinetic process that depends on the oral absorption rate. In the case of

poor solubility of the drug or pronounced first pass metabolism, more complex models are used.[44] The distribution of a drug is modeled by the central volume of distribution, which represents the apparent volume in which a drug is distributed after absorption. It is calculated based on the amount of drug absorbed and the measured plasma concentration. The volume of distribution increases with the lipophilicity of a drug, because more lipophilic drugs tend to distribute to compartments other than the central compartment of the human body.[45] Drug clearance includes metabolism and excretion.[46] It depends on the drug concentration in the blood, i.e., the central volume of distribution, and the blood flow passing through the clearance organs, e.g., kidney and liver. Clearance is calculated based on the extraction coefficient, which is tissue- and drug-dependent and corresponds to the drug fraction eliminated.[47]

Consideration of metabolism is important in drug development not only because of pharmacokinetic models, but also because CYP enzymes can be involved in metabolic intoxication, e.g., in the metabolism of benzo[*a*]pyrene (Figure 3).[48] Benzo[*a*]pyrene itself is not toxic, but its metabolites are highly carcinogenic. The resulting (+)benzo[*a*]pyrene-7,8-dihydrodiol-9,10-epoxide can form covalent adducts with the DNA bases guanine and adenine.[49] The adduct disrupts the double helical structure and causes mutations that lead to cancer.

Not all drugs need to undergo the full ADME. Some drugs with poor oral bioavailability, such as vancomycin, are administered almost exclusively by injection, so absorption becomes pharmacokinetically neglectable for these drugs.[50] Gabapentin, on the other hand, is not metabolized at all and excreted renally in unaltered form.[51] However, for most marketed drugs, all aspects of pharmacokinetics are essential and a strong determinant of therapeutical outcome.[15, 52]

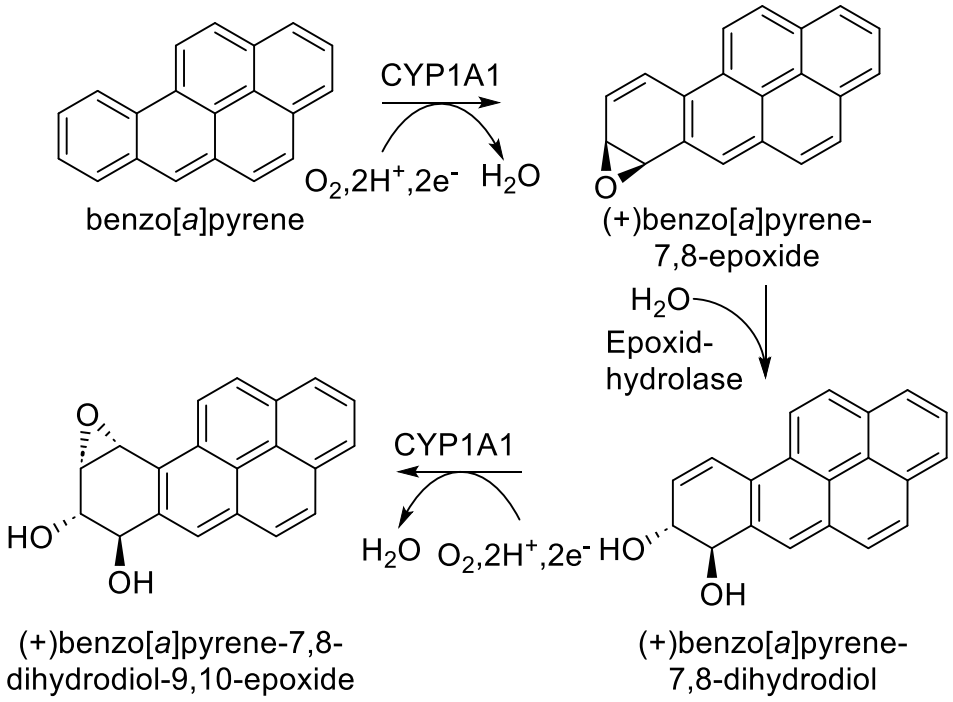



*Figure 3: Metabolic intoxication of benzo[a]pyrene by CYP1A1. The resulting (+)benzo[a]pyrene-7,8-dihydrodiol-9,10-epoxide is highly toxic and believed to be the primary cause of lung cancer in smokers. The process consists of three enzymatic reactions. In the first step, CYP1A1 oxidizes benzo[a]pyrene to (+)benzo[a]pyrene-7.8-epoxide. In the second step, the epoxide is metabolized by epoxide-hydrolase to (-)benzo[a]pyrene-7.8-dihydrodiol. The acting carcinogen is formed in the third step: CYP1A1 catalyzes a second oxidation to (+)benzo[a]pyrene-7,8-dihydrodiol-9,10-epoxide.*

### 2.1.2 Phases of drug metabolism

The human drug metabolism consists of 4 phases: Phase 0 describes the uptake of a drug from the bloodstream to the hepatocyte, Phase I describes redox and hydrolysis reactions that are performed by enzymes, Phase II describes the conjugation of the drug with a polyatomic group, and Phase III describes the process of excretion of the drug from the liver cell (Figure 4).[53] The numbering scheme – starting with phase 0, so index 0 – is uncommon for natural sciences and caused by the fact that phase I and II were introduced first as part of the ADME concept, and later on, the product elimination was explored and added to the scheme as phase III.[54] After product egress, the substrate/drug uptake was uncovered and consequently added to the scheme as phase 0.[7]

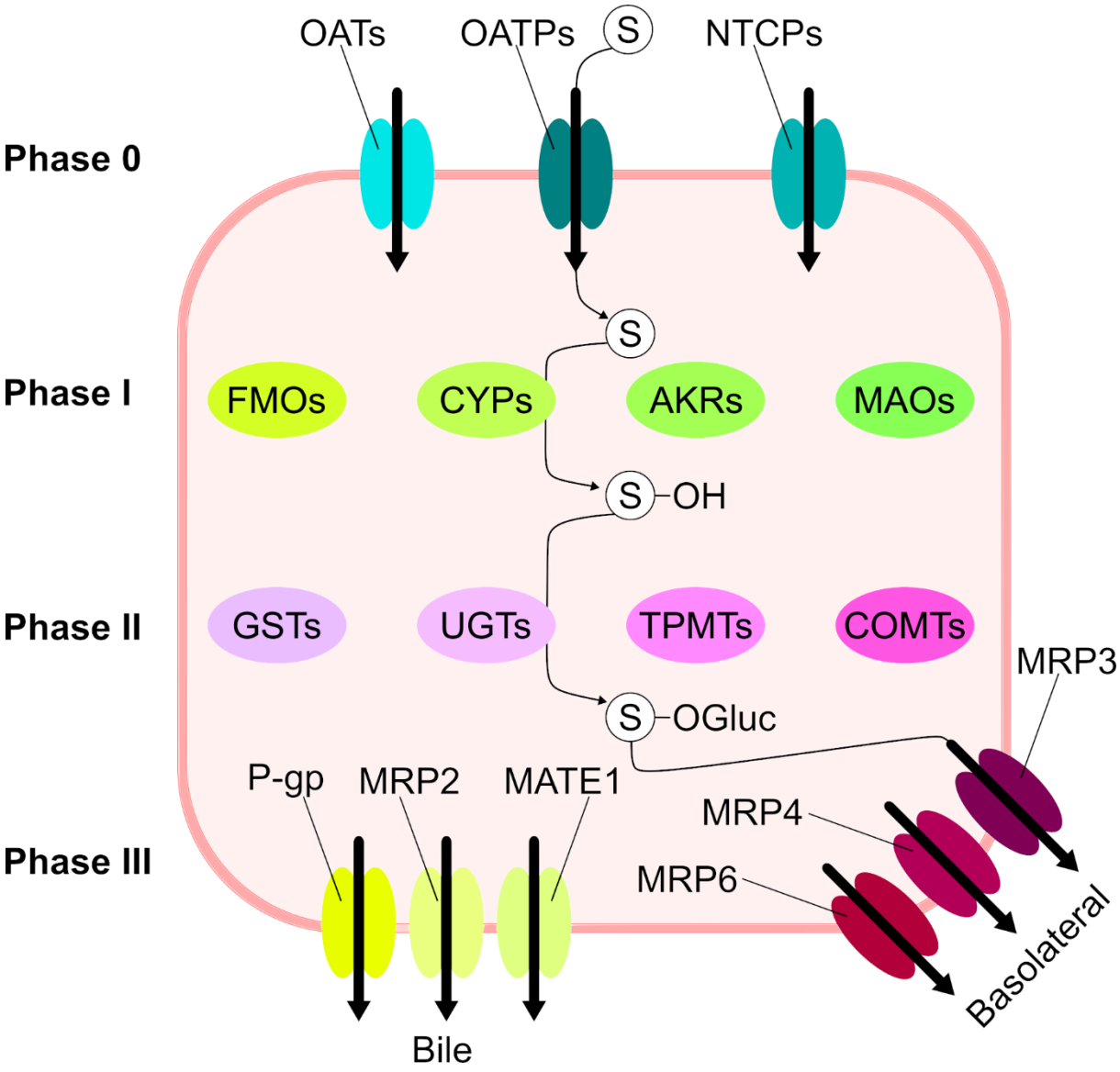



*Figure 4: Schematic representation of human drug metabolism. Positions of transporters are defined by the phase of metabolism and don't reflect position in real hepatocytes. In phase 0, the substrate S is actively transported into the hepatocyte. In phase I, reduction and oxidation reactions take place. In phase II conjugation reactions, use the substrate's de novo generated functional groups. The conjugated substrate is then excreted in phase III. For abbreviations, see text.*

Each phase is associated with a specific set of proteins that either perform transport or catalyze the reactions in that phase. The localization of the transporters is critical because hepatocytes are polarized cells, with one side facing the blood vessel (sinusoids) and the other facing the bile canaliculi.

In general, drug transporters in the liver can be divided into two families – the solute carrier (SLC) family and the ATP-binding cassette (ABC) transporter family, with the

former mainly playing a role in phase 0, while the latter act in phase III. In phase 0, the uptake of drugs is performed by the liver-specific $Na^+$-dependent taurocholate cotransporting polypeptide (NTCP), organic anion transporters (OAT2, OAT5, and OAT7), and organic anion transporting polypeptides (OATP1B1, OATP1B3, and OATP2B1).[55] These transporters actively (co-)transport drugs into the hepatocyte and make them available in the cytosol.[6] All transporters of phase 0 are SLCs. NTCP is the carrier for statin uptake in liver cells.[56] OATs transport weak organic acids; most of the ten known OATs are expressed in the kidney. OAT2, OAT5, and OAT7 are exceptions and show predominant expression in the hepatocytes.[57, 58] Common substrates for OATPs are weak organic acids with molecular weights >300, zwitterionic drugs, and neutral linear and cyclic peptides.[59]

Phase I involves the reduction, oxidation, or hydrolysis of a drug molecule. These reactions are mainly performed by CYP enzymes (~95%) that are located at the endoplasmic reticulum of the hepatocyte, but also flavin-containing monooxygenases (FMO; ~2%), aldo-keto reductases (AKR; ~2%), and monoamine oxidases (MAO; ~1%) perform reactions in this phase.[15] Like CYP enzymes, FMOs are located at the endoplasmic reticulum, while MAOs are located at the mitochondrial membrane. [60-62] The modifications done in phase I are often necessary to deactivate a drug and create a phase II reaction site. However, phase I reactions can also be exploited in drug development for targeted activation of promoieties in prodrugs. A prodrug is a molecule, which does not act as a drug (or at much lower activity) but has a pharmacologically active metabolite. An example of this concept is clopidogrel, which is oxidized by CYP enzymes in the liver to 2-oxo clopidogrel, and the resulting thiolactone is hydrolyzed. The ring opening leads to an accessible -SH group, which then forms a disulfide bond with the target of clopidogrel: $P2Y_{12}$.[63]

In phase II, the enzymes involved are mainly transferases because this phase aims to perform a conjugation reaction. The major enzymes are UDP-glucuronosyltransferases (UGTs), sulfotransferases (SULTs), *N*-acetyltransferases (NATs), glutathione *S*-transferases (GSTs), and methyltransferases, e.g., thiopurine *S*-methyl transferase (TPMT) and catechol *O*-methyl transferase (COMT).[64] Within this phase, UGTs are the most important enzyme class because approximately 40 to 75% of all xenobiotics undergo a glucuronidation reaction catalyzed by UGTs.[65] Generally, the transfer of polyaromatic groups detoxifies the drug and makes it more soluble. In

the case of paracetamol, an example for drug interactions caused by CYP enzymes, limitations in phase II narrow the therapeutic range (the dosage difference between therapeutical dosage and toxic dosage) (Figure 5).[66] Paracetamol is partially oxidized by CYP2E1 to its toxic metabolite *N*-acetyl-*p*-benzo-quinone imine (NAPQI).[67] GSTs are deactivating NAPQI by conjugation. For the conjugation, reduced glutathione (GSH) is necessary.[68] If the amount of NAPQI exceeds the available GSH, the highly reactive NAPQI binds covalently to different molecules in the liver cell and can cause acute hepatic necrosis.[69, 70]

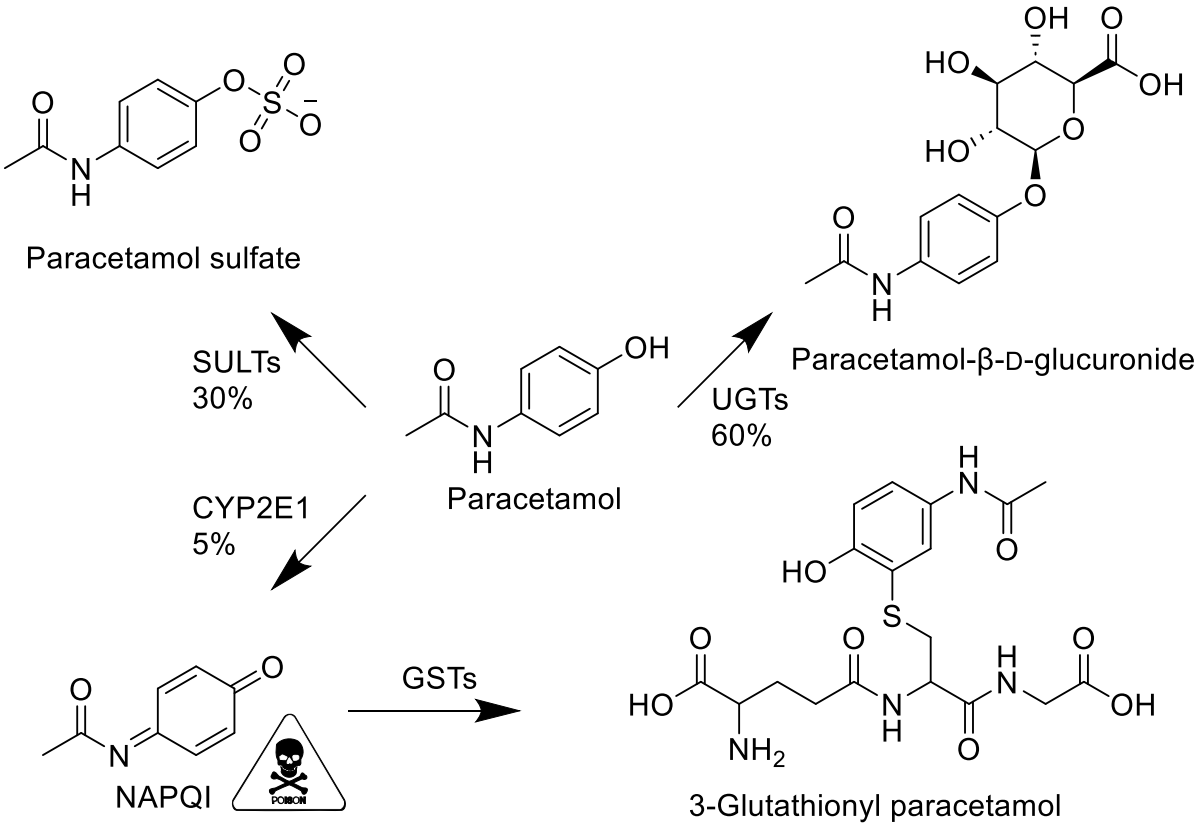



*Figure 5: Paracetamol is mainly metabolized in phase II metabolism (90%). Only 5% is metabolized by CYP2E1, but the metabolite produced by CYP2E1, NAPQI, is toxic. If a sufficient amount of GSH is available, it is conjugated by GSTs to 3-Glutathionyl paracetamol. If the pool of GSH is empty, NAPQI will react with different molecules in the hepatocyte and cause acute hepatic necrosis.*

Phase III transporter proteins are often ABC transporters in contrast to phase 0 transporters.[71] While multidrug resistance-related protein (MRP) 3, MRP4, and MRP6 face the basolateral membrane of the hepatocyte, P-glycoprotein (P-gp), bile salt export pump (BSEP), MRP2, and multidrug and toxin extrusion protein 1 (MATE1), an SLC, face the apical site. [55] P-gp is noteworthy in this phase because it acts like a

vacuum cleaner and leads to an efflux of drugs from various tissues – often pumping molecules out of the cell membrane before they even enter the cell completely. In hepatocytes, P-gp pumps drug metabolites into the bile, which is then concentrated in the gallbladder and discharged into the duodenum after food uptake.[8]

## 2.2 Cytochrome P450 enzymes

### 2.2.1 Cytochrome P450 as monooxygenases in Phase I drug metabolism

As mentioned above, CYP enzymes are a superfamily of enzymes named for the characteristic absorption band at 450 nm under CO atmosphere.[72] All CYP enzymes share the heme group in the active site but show a vast sequential diversity (Figure 2). This diversity also accounts for the unique naming scheme of this superfamily: CYPs are named by a family in Arabic numbers, e.g., 1, which shares at least 40% sequence identity, and a subfamily in Latin letter, e.g. A, which shares at least 55% sequence identity, and an individual number of the CYP enzyme in the subfamily of the organism in Arabic numbers, e.g., 1.[22] So, "human CYP1A1" describes one CYP enzyme sufficiently. In the human drug metabolism, only twelve members of the first three families oxidize drug molecules in the liver, the CYP1, CYP2, and CYP3 families (Figure 6).[14]

As noted above, CYP enzymes play a major role in phase I, catalyzing about 95% of all reactions in this phase.[15] CYPs are monooxygenases, so the catalyzed reaction is the transfer of an oxygen to the substrate (a drug in the case of drug metabolism). In some cases, the transfer of an oxygen leads to a destabilization of the substrate, which results in multiple reaction mechanisms catalyzed by CYPs (Table 1).[62, 73]

*Table 1: Drug metabolism reactions catalyzed by CYP enzymes.*

| Reaction catalyzed by CYP | Example substrate | Reference |
|---|---|---|
| **Hydrocarbon oxidation** | Ibuprofen | Carlile *et al.*[74] |
| **Oxidation adjacent to a heteroatom** | Chloramphenicol | Pohl *et al.*[75] |
| **Heteroatom oxidation** | Nicotine | Majumdar[76] |
| **Olefine and acetylene oxidation** | Butene | Vaz *et al*.[77] |
| **Aromatic ring oxidation** | Atorvastatin | Tekes *et al.*[78] |
| **Carbon-carbon bond cleavage** | Nabumetone | Varfaj *et al.*[79] |

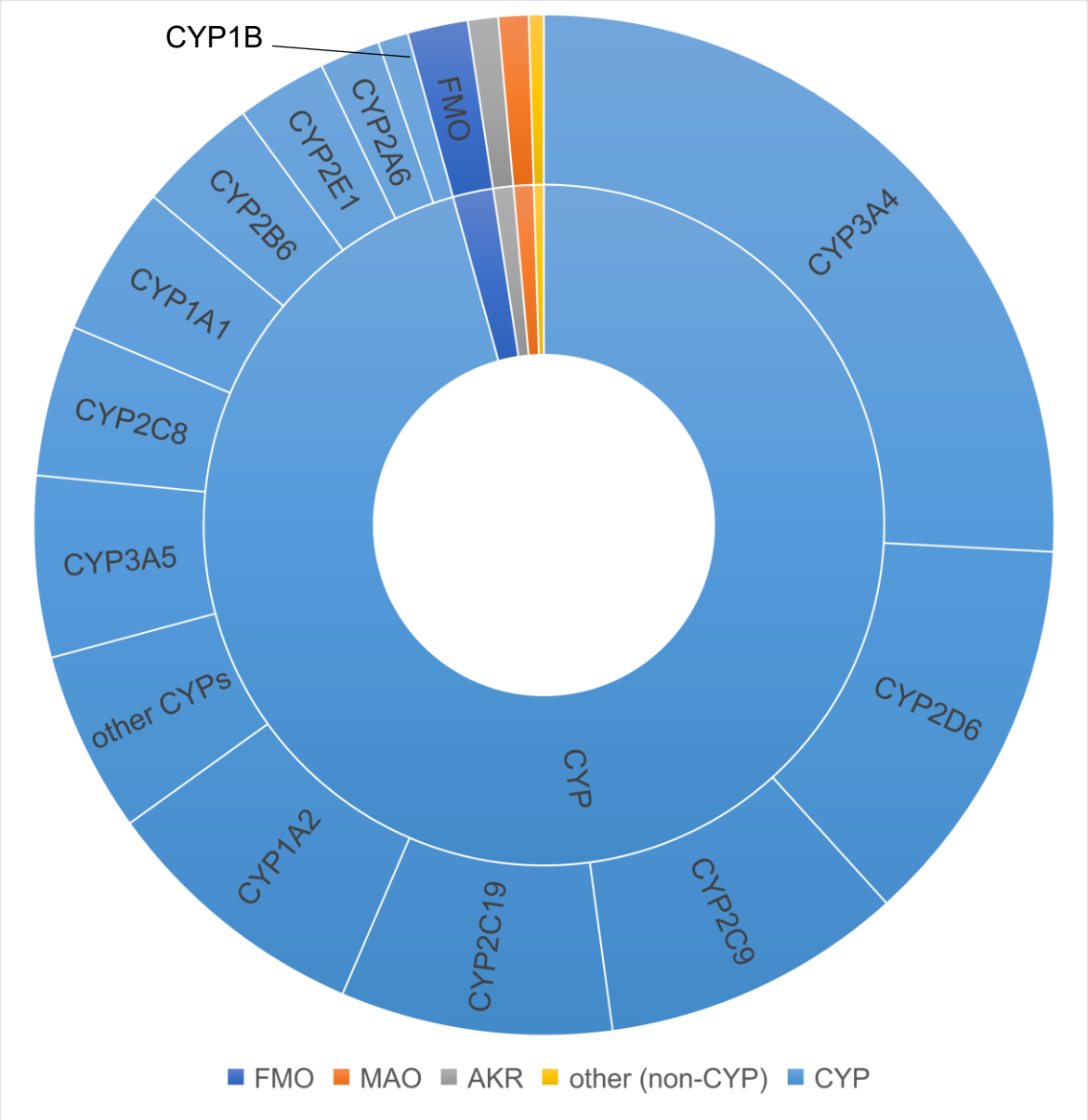



*Figure 6: Human oxidoreductases that are part of drug metabolism. The share of each protein is based on the number of reactions the protein catalyzes in human drug metabolism. CYP3A4 participates most in the human drug metabolism. The Sunburst plot is based on Table 1 of ref[15].*

Interestingly, even within these twelve drug-metabolizing CYP enzymes, there is a wide variation in their contribution to metabolic reactions: CYP1B1 participates in ~1% of drug-metabolizing reactions, while CYP3A4 is involved in 27%. CYP3A4 is thus more promiscuous than CYP1B1, but the latter still oxidizes multiple drugs, making it more promiscuous than other proteins.

### 2.2.2 Clinical relevance of CYP enzymes

Due to the pivotal role that CYP enzymes play in the metabolism of drugs in humans, variations in individual CYP activity can influence the efficacy of drug therapies.[10] A higher level of activity leads to underdosing if the applied molecule acts directly as drug or to overdosing if the drug is a prodrug that needs to be activated by CYP. Variation in CYP activity is a particular problem with CYP2D6. Interindividual differences in this isoform are common, and patients are divided into poor metabolizers, intermediate metabolizers, extensive metabolizers, and ultrarapid metabolizers.[80] If a patient has been assigned to such a class, it should be considered in all future drug therapy regimens to prevent side effects.[34] Even a single nucleotide polymorphism, where the change of one DNA base results in the change of a single amino acid, is sufficient to cause poor metabolism.[81]

A prime example of the clinical relevance of CYP2D6 polymorphism is the metabolism of codeine.[82] Codeine is used as an antitussive agent and analgesic[83], but has little activity itself until it is activated to morphine by CYP2D6-catalyzed demethylation in the liver. If a poor metabolizer receives a standard dose of codeine, therapy will have little effect, whereas the same dose in an ultrarapid metabolizer may result in an overdose that can lead to apnea. Therefore, antitussive or analgetic therapy may be fatal for ultrarapid metabolizers. Another example for the relevance of CYP2D6 is imipramine, a tricyclic antidepressant. An overdose of tricyclic antidepressants can lead to cardiac arrest. Imipramine is activated by CYP2C19 and its active metabolite is only metabolized by CYP2D6.[84] Patients on imipramine therapy receive 25% - 180% of the standard dose depending on the CYP2D6 activity.[32]

### 2.2.3 Oxidation reaction catalyzed by CYP enzymes and its reduction by Cytochrome P450 reductase

As mentioned, CYP enzymes are monooxygenases that catalyze the transfer of an oxygen. In this process, the CYP enzyme undergoes a complex reaction cycle, which was elucidated in 2000 by Schlichting *et al.* (Figure 7) for cytochrome P450cam from *Pseudomonas putida*.[85, 86] This cycle is now widely accepted also for eukaryotic CYP enzymes.[73]

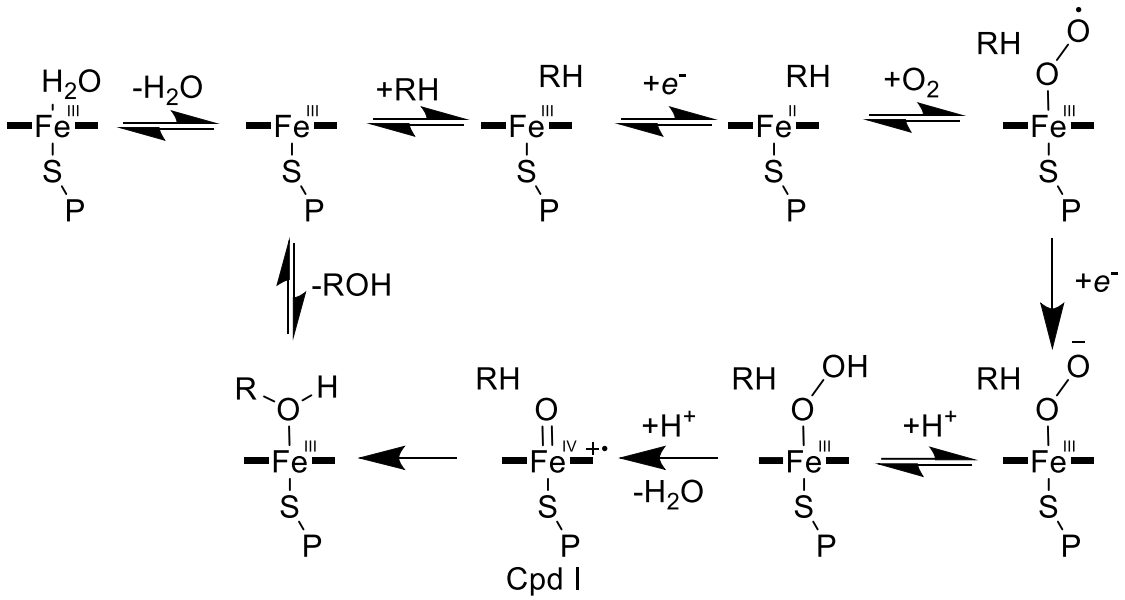



*Figure 7: Schematic representation of the CYP reaction cycle. The -Fe- represents the iron in the center of the heme, while RH represents the substrate. The different reaction partners necessary for the reaction cycles are indicated above the arrows. The required electrons ($e^-$) are provided by donor proteins, e.g., Cytochrome P450 reductase. Other needed molecules ($H_2O$, $O_2$, $H^+$) will diffuse from the cytosol into the active site. Figure is based on Ortiz* et al.[73]

Unlike other enzyme classes, the substrate has no significant mechanistic involvement in the reaction cycle. The catalytic machinery is centered around the iron in the heme group. Substrate binding replaces the water and stabilizes a five-coordinated state of the iron – a cysteine binds to the iron at the coordination site opposite the substrate coordination site. Consequently, the first electron is transferred to the iron, changing the oxidation state to II. Iron II can bind molecular oxygen to form a six-coordinated radical complex. Transfer of a second electron and a proton leads to protonated peroxyl species. In some cases, the substrate is already oxidized by this species. Removal of water follows the addition of a second proton. The oxyferryl/porphyrin radical cation is called Compound I (Cpd I) and determines the outcome in most CYP-catalyzed oxidation. From Cpd I, the oxygen is transferred to the substrate. The product and heme group may form an intermediate six-coordinated complex until a water molecule replaces the product.

This reaction cycle requires a reaction partner that can transfer single electrons. In the case of the most human CYPs, this is the NADPH-dependent Cytochrome P450 reductase (CPR).[87] CPR is a multidomain protein containing a membrane binding domain (MBD), a flavine mononucleotide domain (FMD), and a flavine adenine

dinucleotide domain (FAD). As the name MBD implies, CPR is bound to a membrane, in the case of humans, the ER membrane. The electron transfer from NAPHD to the heme iron involves multiple conformational changes (Figure 8). At the beginning of the cycle, CPR is in the open conformation.[88] When NADPH binds to the FAD, CPR changes to the closed conformation. The hydride is transferred to the flavine adenine dinucleotide, which is in the closed conformation ~4 Å away from the flavine mononucleotide, so the electrons are transferred directly to the FMN with a rate of ~55 $s^{-1}$ in humans.[89, 90] After electron transfer, CPR dissociates and is again available in the open conformation.

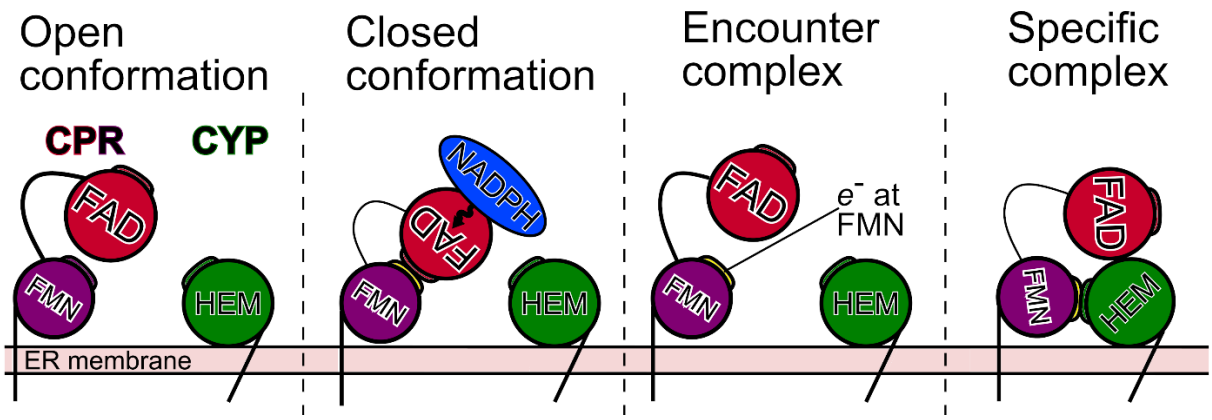



*Figure 8: Schematic representation of CPR-CYP-complex formation in the ER membrane. The cofactor abbreviations are written in the according domains, which are colored differently. Flavine mononucleotide (FMN) domain: purple, flavine dinucleotide (FAD) domain: red, heme (HEM) containing CYP: green. Open conformation is present in the apo state of the CPR. Closed conformation is preferred upon Nicotinamide adenine dinucleotide phosphate (NADPH) binding. A hydride can be transferred to FAD in this conformation, which induces the interflavin electron transfer. Encounter complex explores the ER membrane and associates the FMN domain with a CYP. A specific complex is built after further conformational changes. Now, electrons can be transferred from FMN to HEM. Figure adopted and modified from ref [91].*

In microbial cells, the redox pathway is the same, but the electrons can come from different reductases and Nicotinamide adenine dinucleotide (NADH) can act as cofactor. It is often not known which redox partner is interacting with which CYP.[87]

### 2.2.4 CYP enzymes in biotechnology

Due to the variety of catalyzed reactions described in section 2.2.1, CYPs are also of high interest in biotechnology.[92] In traditional synthesis approaches, the functionalization of non-activated C-H bonds is challenging due to the missing

selectivity.[93] With CYPs, it is possible to derive in a one step synthesis highly valuable products from relatively cheap precursor molecules, because the oxidation takes place regio- and stereoselectivly.[94]

CYPs accept many different substrates; as already shown for human CYPs, it is common for one CYP to catalyze reactions of multiple substrates. Besides this, CYPs have several advantages in biotechnology: [95]

- CYPs operate in mild – aqueous – conditions, as many other proteins.
- CYPs are a well-studied enzyme superfamily not only in biotechnology but also in many other fields due to their key role in cellular systems.
- CYPs can be expressed in established host organisms, e.g., *Escherichia coli, Pseudomonas putida, Saccharomyces cerevisiae,* and *Pichea postoris.*
- Large number of available CYPs offers possibilities for screening.

On the downside, the high potential of CYPs is limited due to different circumstances: [95]

- Monooxygenases show in comparison with other enzyme classes low turnover rates.
- The two electrons that sustain the catalytic cycle are usually derived from NAD(P)H, which is expensive.
- CYPs require a suitable redox partner, which is often unknown in the case of bacteria.
- Uncoupling NAD(P)H oxidation and product formation can lead to reactive oxygen species, e.g., if electrons are transferred when water is bound. This leads to a consumption of NAD(P)H without product formation and may lead to an oxidation of the heme or the protein.
- Industrial applications are so far limited to whole-cell systems, leading to limitations regarding substrate and product concentrations.

The most common CYP in biotechnology is P450BM3 from *Bacillus megaterium* (CYP102A1).[96] P450BM3 is a natural fusion of a heme domain and a diflavin reductase domain.[97] The combination of both domains leads to high catalytic activity. Moreover, P450BM3 exhibits high solubility and high expression levels in *Escherichia coli*, making it an ideal candidate for industrial bioengineering applications. Wildtype P450BM3 is not able to metabolize drug-like compounds, though, as its natural substrates are fatty

acids. However, optimization of the enzyme by rational design and directed evolution lead to variants that oxidize drugs and small molecules of commercial interest, e.g., testosterone, dextromethorphan, and propranolol.[98, 99]

Rational design and directed evolution as enzyme optimization techniques, are exclusively based on the target enzyme, and thus limited in their ability to produce major changes in an enzyme's activity. To overcome this limitation, an approach called SCHEMA developed by Frances Arnold and coworkers was introduced.[100] SCHEMA is based on structure-guided DNA recombination to obtain chimeras with completely novel functions. It was applied to P450BM3 and its homologs CYP102A2 and CYP102A3. The novel function also included the ability to oxidize verapamil and astemizole.[101] This highlights the high potential of P450BM3 and its variants in biotechnology.

Another important CYP in biotechnology is CYP101A1 from *Pseudomonas putida,* also known as P450cam. The native substrate of P450cam is (+)-camphor, which gets oxidized to 5-*exo*-hydroxycamphor. By enzyme engineering, it is possible to extend P450cam substrate spectrum to, e.g., other terpenes, such as (+)-α-pinene, alkanes, styrene, polychlorinated benzene, and other aromatic compounds.[102] In these studies, four "hot spots" were identified, determining the substrate specificity of P450cam (F87, Y96, L244, and V247). Modifying these residues, it is also possible to oxidize propane to propan-2-ol. The substrate propane is of special interest, because until then only oxidation of *n*-alkanes towards pentane and longer alkanes was possible. This oxidation is catalyzed by the so-called EB-variant of P450cam, which is mutated as follows: F87W, Y96F, T101L, and V247L.[103] The ability to oxidize even gases demonstrates the nearly unlimited potential substrate spectrum of CYPs.

As demonstrated in the examples above, CYPs have vast potential for applications in biotechnology. Different strategies were developed to overcome their limitations. In particular, in the last decades, multiple techniques have been introduced to provide tailor-suited CYP for various reactions. CYPs undoubtedly have a big future in biotechnological applications.

### 2.2.5 The general structure of CYP enzymes in mammals

As mentioned above, the heme group in complex with a cysteine defines the active site of CYP enzymes. The overall fold of CYP enzymes is highly conserved within the

superfamily, even between different kingdoms.[21] The first CYP crystal structure of CYP101A1 was solved in 1985 and refined in 1987 (PDB-ID: 2CPP).[104, 105] Already in this structure, many important structural elements are visible, such as the long I-helix close to the heme and the F- and G-helix bounding the active site (Figure 9A). These elements are present in all CYP enzymes, but the relative orientation of these helices to each other differs between isoforms, even if both are of prokaryotic origin (Figure 9A, B). The structural element not present in prokaryotes is the transmembrane helix (TM-helix). In contrast to prokaryotic CYPs, eukaryotic CYPs are generally membrane-bound. The first complete crystal structure of a eukaryotic CYP was that of CYP51A1 from *Saccharomyces cerevisiae*, resolved in 2014 (PDB-ID: 4LXJ) (Figure 9C,D).[106]

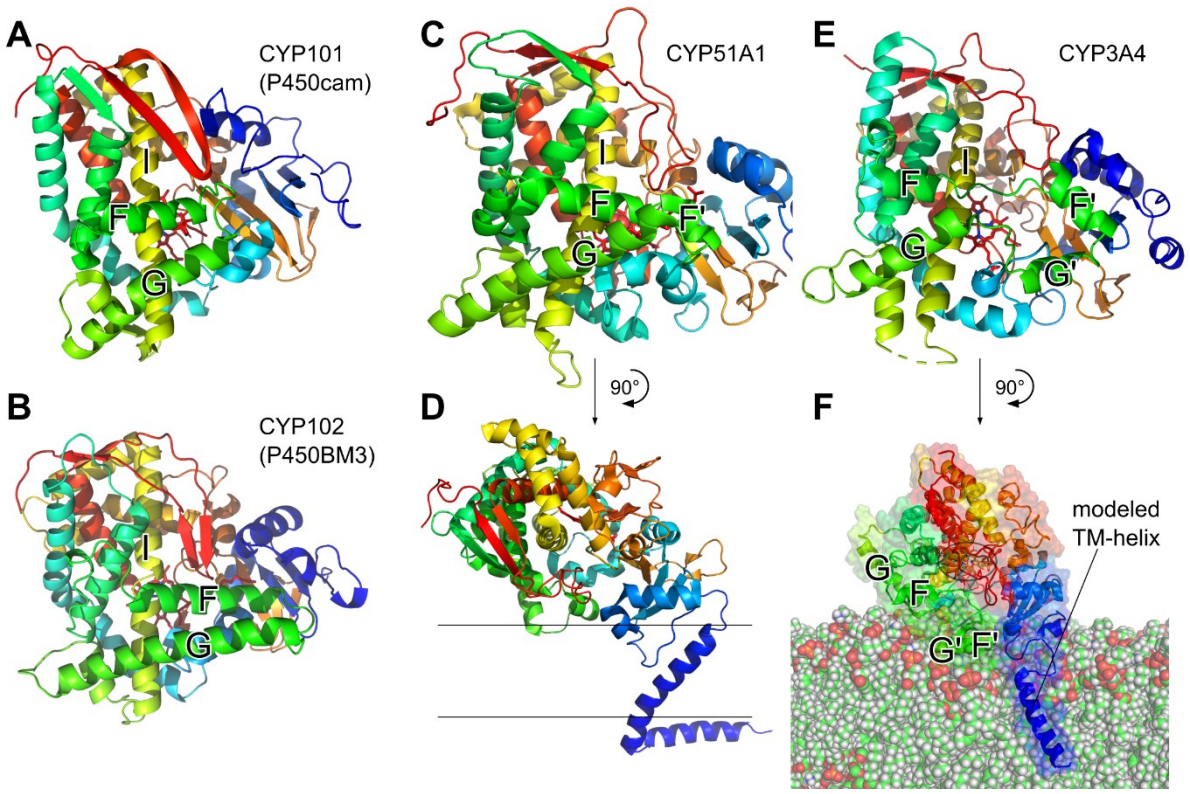



*Figure 9: Multiple CYP enzyme structures.* ***A:*** *CYP101 from the first resolved crystal structure of a CYP enzyme (PDB-ID: 2CPP).* ***B:*** *Structure of P450BM3 heme domain from* Priestia megaterium *(PDB-ID: 1FAG)[107]. P450BM3 is commonly used in biotechnology as platform for biocatalysis.[108, 109]* ***C:*** *CYP51A1 from* Saccharomyces cerevisiae *is a membrane-bound CYP enzyme. The shown crystal structure is the first solved crystal structure of a complete eukaryotic CYP enzyme (PDB-ID: 4LXJ)[106].* ***D:*** *Complete structure of C from a side perspective. Two lines schematically represent the membrane position as suggested by the crystallographers.* ***E:*** *Structure of human CYP3A4, bottom perspective. F/G-region is split into four helices (PDB-ID: 1TQN)[110].* ***F:*** *Model of CYP3A4 embedded into a membrane*

Structures of human drug-metabolizing CYP enzymes' globular domain were already crystallized (Figure 9E).[110-118] These structures already provide insight into the general shape of human drug-metabolizing CYP enzymes. Although the overall scaffold is similar to that of prokaryotic CYP enzymes, the orientation of the F/G-region shows large differences between the presented CYPs of prokaryotic origin compared to human CYP3A4.

The structure of CYP51A1 of *Saccharomyces cerevisiae* allowed a general model for eukaryotic CYP enzymes to be derived. In eukaryotic CYP enzymes, the TM-helix at the N-terminus holds the protein close to the enzyme (Figure 9F). The F- and G-helices are partially lipophilic and interact with the membrane surface. Around the F- and G-helices, multiple channels were detected.[119] This structural model suggests that the lipophilic drug molecule enters the CYP enzyme from the membrane, is oxidized at the active site, and the more hydrophilic metabolite exits through a channel to the membrane surface near the cytosol.[120, 121]

### 2.2.6 Current explanations for differences in CYP promiscuity.

As described in section 2.2.1, there are big differences between the twelve drug metabolizing CYPs when considering their part in metabolism. Explaining these differences is a major goal of CYP research. Early molecular docking-based analyses indicated that the extent of promiscuity might be influenced, at least in part, by the amino acid residues located in the dome region of CYPs (e.g., CYP3A4 – hydrophobicity, CYP2D6 – acidity, CYP2C9 – basicity-dominated residues) (Figure 10A,B).[122, 123] Further investigations based on crystal structures and computational simulations underscore the importance of the F/G-region, comprising the F-, F'-, G'-, and G-helices (for further details on the structural features of CYP enzymes see section 0), for substrate binding.[124-126] The F/G-region is at the beginning of a tunnel that is considered to be the substrate entrance channel (Figure 10C).[120] Such tunnels can act as filters and have been found to influence both substrate specificity and catalytic mechanism.[127-130]

All the studies mentioned above defined promiscuity as taking part in metabolism. A quantitative measurement was not used to relate promiscuity and structural features of the CYP enzymes.

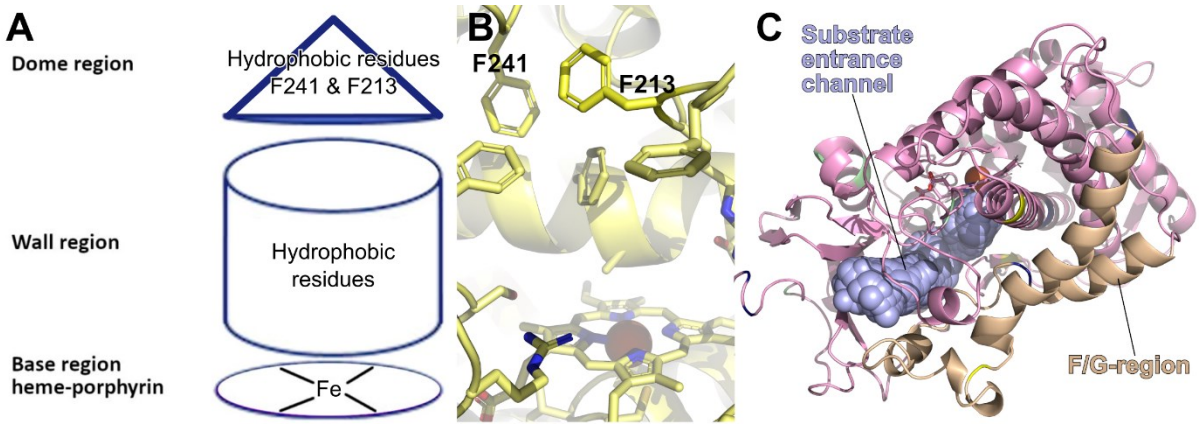



*Figure 10:* ***A:*** *Schematic representation of the pocket in CYP3A4 taken from Ramesh* et al.[131]*. The dome of CYP3A4 is formed by F241 and F213. The wall region is formed by different hydrophobic residues and the base is formed by the heme-molecule.* ***B:*** *Dome region of CYP3A4 from a structural model. F241 and F213 are labeled and at the upper part of the pocket. Multiple phenylalanines form the wall region.* ***C:*** *The F/G-region of CYP2D6 (marked in beige) is part of the substrate entrance channel. The substrate entrance channel was analyzed with CAVER[132] and is depicted as purple spheres.*

# 3 SCOPE OF THE THESIS

In the previous section, the ADME concept, the function of the liver, and the metabolism performed by CYP enzymes were presented. The relevance and current state of knowledge about CYP enzymes involved in human drug metabolism were depicted throughout the previous chapters. CYP enzymes play a crucial role in the metabolism of many drugs, as they are involved in ~95 % of all phase I reactions. I also showed that CYP3A4 participates in more drug metabolism reactions than all other human CYP enzymes. The reasons for the unique role of CYP3A4 were qualitatively described in crystallographic and modeling studies (section 2.2.1). Section 2.2.2 highlighted the clinical relevance of CYP enzymes in general and with regard to their polymorphism. Particular emphasis was placed on the challenges in pharmacotherapy associated with CYP2D6 polymorphism and the occurrence of poor metabolizers in the population. The general structure of human CYP enzymes and the importance of CPR were elucidated: The reaction cycle of the CYP enzyme is based on the oxidation state of the iron in the heme-thiolate complex as it depends on the two electrons spent by the CPR after substrate binding (section 2.2.3).

However, a quantitative model of the relationship between the structural differences within the group of human drug-metabolizing enzymes and their promiscuity remained elusive. To close this gap, extensive molecular dynamic simulations were performed in Publication I to generate structural ensembles of multiple drug-metabolizing CYP enzymes. The rigidity of each isoform was investigated based on the structural ensembles derived from MD. The results from the Constraint Network Analysis (CNA) were combined with the promiscuity index $I_{cat}$ (as described in section 4.3.3) to obtain a quantitative model. The derived model describes the relationship between the F/G-region rigidity and the substrate promiscuity of a CYP isoform.

Furthermore, an atomistic model suitable to explain the effects of CYP2D6 polymorphism was missing. This topic was addressed by me in Publication II. The twenty most common mutations were identified, the respective variants were modeled, and molecular dynamics simulations were performed. Based on the obtained trajectories, molecular mechanisms leading to poor CYP2D6 metabolism were identified and criteria for classifying a variant as poor were derived. Overall four key interactions were identified: a) reduced enzyme stability caused by lower rigidity of the central helix, b) reduced substrate binding due to lower or no availability of two central

phenylalanines, c) decreased binding to CPR due to less basic residues on the surface, and d) decreased electron transfer to the iron ion due to higher distances of the residue R132 to the central atom of the heme group. To quantify these key interactions, criteria were defined and subsequently analyzed, e.g., the availability of the interaction partner with CPR was evaluated by the stability of an inner molecular salt bridge. A variant is classified as poor if it does not fulfill all four of these criteria. I generated a decision tree based on the developed criteria, which allows to estimate the effect of a newly identified mutation on CYP2D6.

# 4 METHODS

In this work, various computational methods were applied to investigate human CYP enzymes. In the vast field of computational biology, varying approaches with different limitations are used to model biological systems and processes. In the following sections, a brief overview of these methods is provided.

## 4.1 Molecular dynamic simulations

Molecular dynamic simulations, a technique developed in the late 70s for biomolecules,[133] is a simulation approach to sample the movement of atoms in a system over time. During such simulations, the interactions between atoms are approximated by simple potentials derived from Newtonian physics such as harmonic potentials, Coulomb potentials, and Lennart-Jones-potentials. An example of such a potential is the harmonic potential describing bonds between two atoms, which has the physical interpretation of two masses connected by a spring. This simplification offers the option to investigate routinely large biological systems like proteins, potentially consisting of more than a million of atoms, which could not be investigated by more accurate approaches like quantum mechanics.[134]

In an MD simulation, a trajectory is depicted as a consecutive series of system conformations. This series is obtained by iteratively integrating over the acting forces in the system for an integral time step ($\Delta t$) and thus updating atomic positions and momenta accordingly (Figure 11).

During the time interval $\Delta t$, the determined forces acting on the atoms remain constant. As a result, if $\Delta t$ is chosen too long, artifacts and instabilities will occur in the trajectory. If $\Delta t$ is chosen to low, the computational costs (time, energy consumption) will increase. As a rule of thumb, the maximal possible value of $\Delta t$ must be one order of magnitude lower than the highest vibrational frequency.[135] In biological systems, the most rapid vibration is the translational vibration of bonds between heavy atoms and a hydrogen atom. These vibrations usually occur with a frequency of $10^{14}$ $s^{-1}$ and thus limiting $\Delta t$ to a maximum of ~1 fs.[136] However, these vibrations do not influence the conformational changes of the overall system. Therefore, different algorithms, such as SHAKE[137] or LINCS[138], were developed to artificially slow down these vibrations, allowing for $\Delta t$ > 1 fs. Combining the SHAKE algorithm with hydrogen mass

repartitioning, time steps of up to 4 fs can be achieved increasing the achievable computational efficiency.[139, 140]

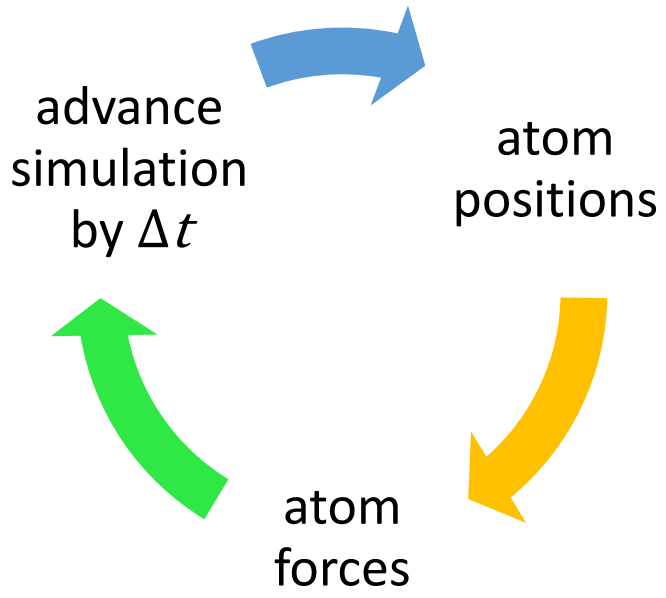



*Figure 11: MD simulations consist of a repetitive performance of three steps, which depend on each other. The forces and potentials acting on each atom are calculated numerically based on the atoms' positions. Then, the atom forces are propagated as acceleration, which result in changes of velocity vectors, and the new velocities are applied to the atoms according to the defined simulation time step. Afterwards, the cycle is repeated until the investigated process was sampled sufficiently.*

Different software suites are available for MD simulations of biological systems: currently AMBER,[141] CHARMM,[142] GROMACS,[143] and NAMD[144] are the most popular for simulating biological systems.[136] In this work, the GPU implementation of the Particle Mesh Ewald (PME) method within AMBER was used due to its superior performance.[145]

In MD simulation suites, force fields describe the mechanical and electrical potentials and their parameters used in the simulation. These force fields are derived based on quantum chemical calculations. E.g., a bond stretching movement's (free) energy profile is derived from quantum mechanical calculations. Then, the energy profile is transferred by fitting it into the force field terms. These parameterizations are performed for a set of similar molecules, e.g., amino acids, and provided as one

collection of parameters for the respective class of molecules. More general force fields are available for molecules that do not appear regularly in simulations. Within this work, the following force fields were used:

- ff14SB for proteins[146]
- LIPID17 for phospholipids in membranes[147]
- OPC as water model[148]
- GAFF2 as general force field[149]

Force fields in AMBER follow the scheme shown in Equation 1. Stretching of bonds and bending of angles are approximated by harmonic potentials. Torsion barriers are approximated by a cosine term, which defines barriers to the rotation around a rotatable bond. Moreover, out-of-plane bending potential is described by this term. For non-bonded interactions, a 12-6 Lennard-Jones potential is used. It describes the van-der-Waals interactions. The Coulomb potential describes electrostatic interactions.[150, 151]

$$E_{system} = \sum_{\text{bonds}} K_r\left(r - r_{\text{eq}}\right)^2 + \sum_{\text{angles}} K_\theta\left(\theta - \theta_{\text{eq}}\right)^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}\left[1 + cos(n\phi - \gamma)\right] + \sum_{i<j} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}} \qquad \textit{Equation 1}$$

Trajectories obtained from MD simulations provide atomistic insights into the conformational space explored by a protein. They can be further analyzed using a wide range of methods, e.g., calculation of the root-mean-square deviation (RMSD) of the atomic positions allows measuring structural diversity within the conformational ensemble. The root-mean-square fluctuation (RMSF) of $C_\alpha$-atoms allows to estimate the residue-wise differences in movement. Finally, energetic differences between states can be estimated by calculating the potential of mean force using the weighted ensemble method.[152] These analyses can be performed with AmberTools.[153]

## 4.2 Constraint Network Analysis

Generally speaking, proteins are marginally stable.[154] The structures are organized in rigid and flexible regions.[155] Analyzing the hierarchy between these regions provides an understanding of the relationship between (thermo-) stability and function. To get

this insight into the stability of protein structure, Constraint Network Analysis (CNA) was introduced by Radestock and Gohlke.[156, 157] The approach was then further developed into the CNA software suite by Pfleger et al.[158] The CNA software suite is a front- and backend to the software FIRST (Floppy Inclusion and Rigidity Substructure Topology).[159] In CNA, a biomolecule is considered as a constraint network: atoms are the nodes of the network, and covalent and noncovalent bonds are constraints between the nodes (Figure 12A).[159] Noncovalent interactions, e.g., hydrogen bonds, salt bridges, hydrophobic tethers, and stacking interactions, are the main causes of biomolecular stability. By gradually removing noncovalent constraints from a starting network representation of a biomolecule derived from FIRST according to a cutoff energy $E_{cut}$, a sequence of network states [σ] is generated that forms a 'constraint dilution trajectory'.[156, 158] Performing rigidity analysis[160] on such a trajectory reveals a hierarchy of structural stability that reflects the modular structure of biomolecules in terms of secondary, tertiary, and quaternary structure.[130]

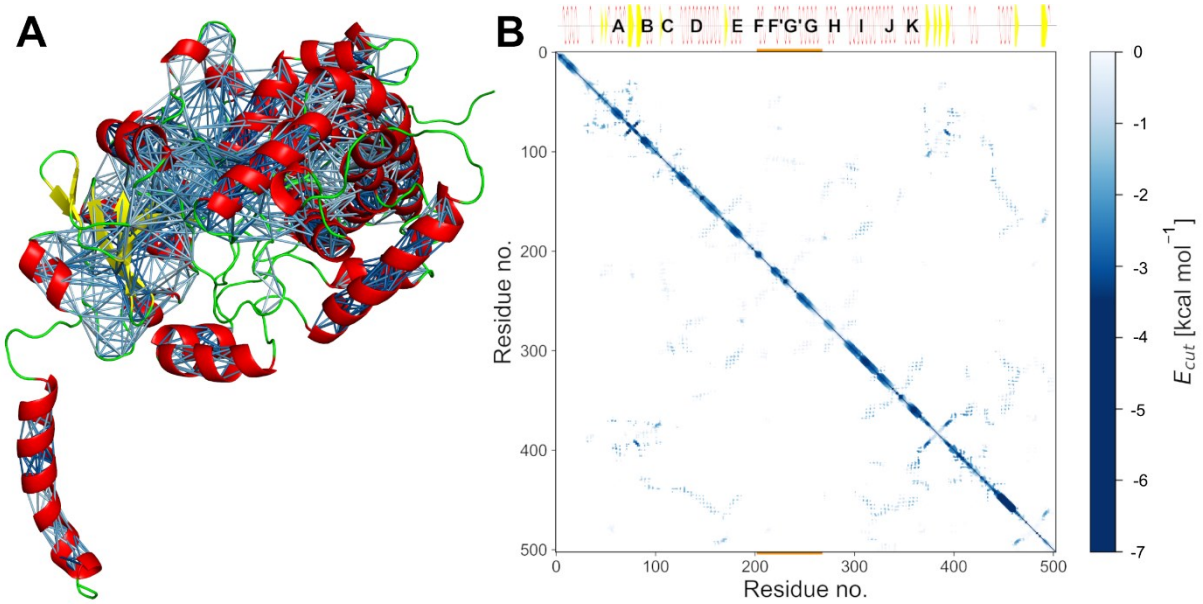



*Figure 12: Exemplary Constraint Network Analysis of CYP3A4. A: The rigid constraints projected on the structure of CYP3A4. Bars between $C_\alpha$ atoms represent the strength of the rigid contact (darker blue: stronger contact). Secondary structure elements are colored: red - helics, yellow - beta-sheets, and green – loops. B: Neighbor stability map derived by CNA. The dot color indicates the strength of the rigid contacts between all residues. The color bar is shown on the left. Secondary structure elements are indicated on the top.*

Using a per-residue decomposition scheme to identify the extent to which single residues contribute to the structural stability, a neighbor stability map

($rc_{ij,\text{neighbor}}(E_{\text{cut}}(\sigma))$) is derived that comprises information collected over all network states [σ] along the constraint dilution trajectory in that they monitor the persistence of rigid contacts for pairs of residues during the dilution process (Figure 12B). [130, 161, 162]

From the stability map, the sum of energies associated with rigid contacts between residue *i* to all other residues $E_{i,\text{CNA}}$ can be obtained (Equation 2).

$$E_{\text{i,CNA}} = \sum_{j=1}^{n} rc_{ij,neighbor} \qquad \textit{Equation 2}$$

From $E_{i,\text{CNA}}$, the average of the chemical potential of rigid contacts between the *n* residues within a region ($\bar{E}_{\text{region,CNA}}$) can be calculated, where "region" is either the entire protein or a region of interest, e.g., the F/G-region (Equation 3).

$$\bar{E}_{\text{region,CNA}} = \frac{1}{n} \sum_{i=region_{start}}^{region_{end}} E_{\text{i,CNA}} \qquad \textit{Equation 3}$$

An ensemble generated by MD simulations can be used as input to improve the robustness of CNA.[162] In ensemble analysis, each structure undergoes the thermal dilution process, and the results are then interpreted statistically.

## 4.3 Substrate promiscuity of an enzyme

### 4.3.1 Enzymatic/Catalytic definition of promiscuity

Enzyme's substrate promiscuity describes the ability of an enzyme to work with unexpected substrates, e.g., drugs. A substrate promiscuity typically refers to the broadness of the substrate spectrum and acceptance of larger molecules as substrates. [163]

### 4.3.2 Diversity of a substrate dataset

Any investigation of the enzyme's promiscuity depends on the set of substrates as the basis of the calculation. If the set is too homogenous, the measure's validity will decrease. To tackle this, Nutschel *et al.* developed an approach to estimate the diversity of a given substrate set.[164]

To assess the similarity of the substrate set, the maximum pairwise Tanimoto-Combo distance score $\delta_{TC,ij}$ for compound *i* versus *j* is calculated. $\delta_{TC,ij}$ can have values

between 0 for identical and 2 for totally dissimilar molecules. The Tanimoto-Combo distance score accounts for shape and chemical complementarity between 3D structures as determined by the Rapid Overlay of Chemical Structures approach.[165] From this matrix, the mean pairwise Tanimoto-Combo distance score $\langle\delta_{TC,i}\rangle$ of a substrate *i* to all other substrates in the data set is computed. $\langle\delta_{TC,i}\rangle$ is then averaged for all substrates $\langle\delta_{TC}\rangle$.[130]

As a point of reference, we performed this analysis on 500 samples of 55 randomly selected entries pulled from the ZINC drug-like database. We clustered the compounds according to the abovementioned procedure and calculated $\langle\delta_{TC}\rangle$ for all 500 samples, which is in average at 1.12 (Figure 13). Thus, if a set of compounds shows $\langle\delta_{TC}\rangle > 1.12$, it is more diverse than the ZINC drug-like database. Complete linkage clustering is also performed on the pairwise distance matrix calculated for the substrate, such that all compounds in a cluster have a maximum Tanimoto Combo distance of 1.0, equivalent to $\delta_{ij} = 1.0$. That way, the ratio between the number of clusters and the number of substrate compounds correlates with the chemical diversity of the compounds. This approach was applied to the dataset of substrate compounds of Nath *et al.* used in Publication I.[130, 166]

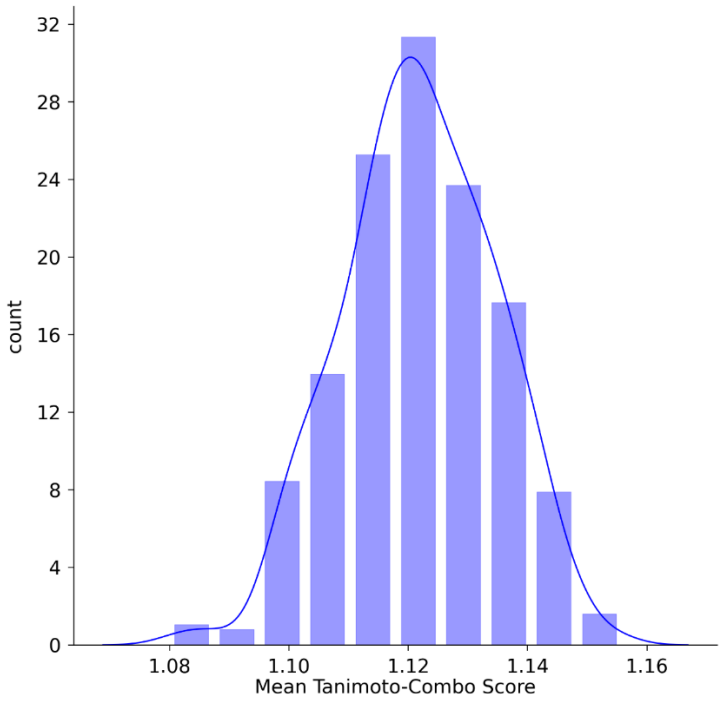



*Figure 13: Distribution of the means of mean Tanimoto-Combo distance scores (<$\delta_{TC}$>) from 500 randomly drawn samples of the ZINC drug-like database.[167] <$\delta_{TC}$> for each randomly drawn sample are differing between 1.08 and 1.16. To obtain these values, we pulled 500 times randomly 55 out of 981,259,785 substances from the ZINC database and calculated <$\delta_{TC}$> for each sample, afterwards, we calculated the mean of means per pull to provide information about the overall diversity in a random pull. Figure taken from ref[130].*

### 4.3.3 Promiscuity index

In the scientific background we discussed the special role of CYP3A4 due to its large part in all drug metabolizing reactions. The connection between metabolizing multiple different drugs and defining a protein as promiscuous is used in the literature.[14] However, this approach lacks a quantitative relationship. An enzyme A that oxidizes multiple chemically highly similar substrates is intuitively less promiscuous than an enzyme B, which oxidizes the same number of substrates but has much more chemical diversity in the substrate set.

To overcome this problem, Nath and Atkins introduced the quantitative index of substrate promiscuity ($I_{cat}$).[168] The index is based on information theory by using entropy as a measurement for diversity. This approach is also used to measure the

diversity of ecosystems.[169, 170] Entropy for a set of *N* outcomes with the probability of $p_i$ for each of the outcomes is defined by Equation 4:

$$H = -\sum_{i=1}^{N} p_i \log p_i \qquad \textit{Equation 4}$$

To obtain $p_i$, Nath et al. used the enzyme's catalytic efficiency (*e*). *e* can be obtained from the turnover number ($k_{cat}$) and the Michaelis constant ($K_M$) (Equation 5).

$$e = \frac{k_{\mathrm{cat}}}{K_{\mathrm{M}}} \qquad \textit{Equation 5}$$

From there, a $p_i$ for chemical processes can be conceptualized. In biochemical systems, $p_i$ describes the probability of substrate *i* being the first metabolized substrate when an enzyme is exposed to equal concentrations of all *N* substrates (Equation 6).

$$p_i = \frac{e_i}{\sum_{j=1}^{N} e_j} \qquad \textit{Equation 6}$$

If this $p_i$ is put into the entropy equation and the entropy is normalized, the promiscuity index ($I_{cat}$) is derived (Equation 7). This index gives the information if an enzyme is highly promiscuous ($I_{cat}$ = 1) or perfectly specific ($I_{cat}$ = 0).

$$I_{cat} = -\frac{1}{\log_{10} N} \sum_{i=1}^{N} \frac{e_i}{\sum_{j=1}^{N} e_j} \log_{10} \frac{e_i}{\sum_{j=1}^{N} e_j} \qquad \textit{Equation 7}$$

$I_{cat}$ provides an elegant measure of substrate promiscuity if a set of enzymes is tested against a diverse set of substrates. For CYP enzymes, such a diverse dataset is available.[166]

# 5 F/G-REGION RIGIDITY

Daniel Becker[a], Prasad V. Bharatam[b], Holger Gohlke[a,c,1]

[a]Mathematisch-Naturwissenschaftliche Fakultät, Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany

[b]Department of Medicinal Chemistry, National Institute of Pharmaceutical Education and Research (NIPER), Sector 67, S. A. S. Nagar, Mohali, 160 062 Punjab, India

[c]John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre (JSC), Institute of Biological Information Processing (IBI-7: Structural Biochemistry), and Institute of Bio- and Geosciences (IBG-4: Bioinformatics), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

The content of this chapter is taken from “F/G-Region Rigidity is Inversely Correlated to Substrate Promiscuity of Human CYP Isoforms Involved in Metabolism.”[130] Word-by-word citations are not marked explicitly.

## 5.1 Author Contributions

HG designed the study; DB performed computations; DB and HG analyzed results; DB and HG wrote the manuscript; PVB revised the manuscript; HG and PVB secured funding.

## 5.2 Background

As pointed out above (Section 2.2.1), only 12 isoforms out of the 57 CYP enzymes encoded in the human genome metabolize about 90% of all xenobiotics.[15] These are CYP1A1, CYP1A2, CYP1B1, CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP3A4, and CYP3A5. Although all are substrate-promiscuous enzymes, they show recognizable differences. CYP3A4 contributes the most to xenobiotics metabolism by being involved in 20% of all known CYP transformations.[15] Conversely, CYP1B1 contributes only to 3% of all known CYP transformations.[15] However, the number of xenobiotic metabolism pathways a CYP can contribute to is not the best indicator for its substrate scope, which also depends on the enzyme’s catalytic efficiency towards specific substrates. Although CYP3A5 generally catalyzes the same reactions as CYP3A4, it does so often at lower rates, such that CYP3A5 only

contributes to 4% of all known CYP transformations.[15] In a previous study, a molecular docking-based analysis pointed out that the degree of promiscuity may be partly determined by the type of amino acid residues present in the dome region of the CYPs (CYP3A4 – hydrophobicity-dominated, CYP2D6 – acidic-dominated, CYP2C9 – basicity-dominated residues).[122, 123]

Previous work based on crystal structures and computational studies stressed the importance of the F/G-region, consisting of the F-, F'-, G'-, and G-helices, involved in substrate binding[124-126] and in the formation of a substrate entrance channel.[120] Such a tunnel acts as a selectivity filter and has been found to influence both substrate specificity and catalytic mechanism.[127-129] Accordingly, one study conducted on the three isoforms CYP3A4, CYP2C9, and CYP2A6 hypothesized that higher mobility, deduced from movements of the $C_\alpha$ atoms in the F/G-region correlates with higher substrate promiscuity.[171] Enzyme structural dynamics, besides its role in catalysis[172, 173], and allosteric regulation[174-177], has also been recognized for other systems as an essential mechanism regulating substrate promiscuity.[178] To our knowledge, no further studies comparatively addressed the relationship between enzyme dynamics and promiscuity for the other CYP isoforms reported in the literature. Furthermore, the approaches described in these studies did not investigate any quantitative connection between substrate scope and structural rigidity of CYP isoforms involved in metabolism. Here, we applied MD simulations and CNA to an extensive repertoire of CYP isoforms and show that the structural rigidity of the F/G-region is inversely correlated to the enzymes' substrate scope.

## 5.3 Results

For example, the constraint dilution trajectories of CYP3A4, CYP3A5, CYP2C8, and CYP1A2 are shown in Figure 14. Considering that the segregation of substructural parts at lower $E_{cut}$ values indicates that these parts are more weakly coupled to the remainder of the CYP structure than regions with higher $E_{cut}$ values, a qualitative ranking of the CYP isoforms based on local structural stability was performed (Figure 13). CYP3A4 shows that the F/G-region is the first part of the globular domain to segregate at $E_{cut}$ = -0.25 to -0.45 kcal mol$^{-1}$ from the largest rigid cluster. These observations are in line with findings reported for different enzymes where helices at the substrate entrance tunnel are also less stable, e.g., in aldolases[179], phosphate synthases[180], and peptidases[181], as well as in previous work on CYP enzymes[171]. In

CYP3A5, segregation of the F/G-region from the largest rigid cluster appears at lower $E_{cut}$ values between -0.40 and -0.60 kcal mol$^{-1}$. For CYP2C8 and CYP1A2, this segregation occurs in the range of -0.95 to -1.15 kcal mol$^{-1}$ and at -1.05 to -1.75 kcal mol$^{-1,}$ respectively. Hence, the F/G region is structurally the least stable in CYP3A4 and the most stable in CYP1A2. Finally, rigidity analysis reveals that the anchor helix segregates at the smallest $E_{cut}$ in all four cases. This indicates that it is only weakly coupled to the globular part, in line with observations from our MD simulations and ref. [121] which showed that the globular part can move relative to the anchor helix.

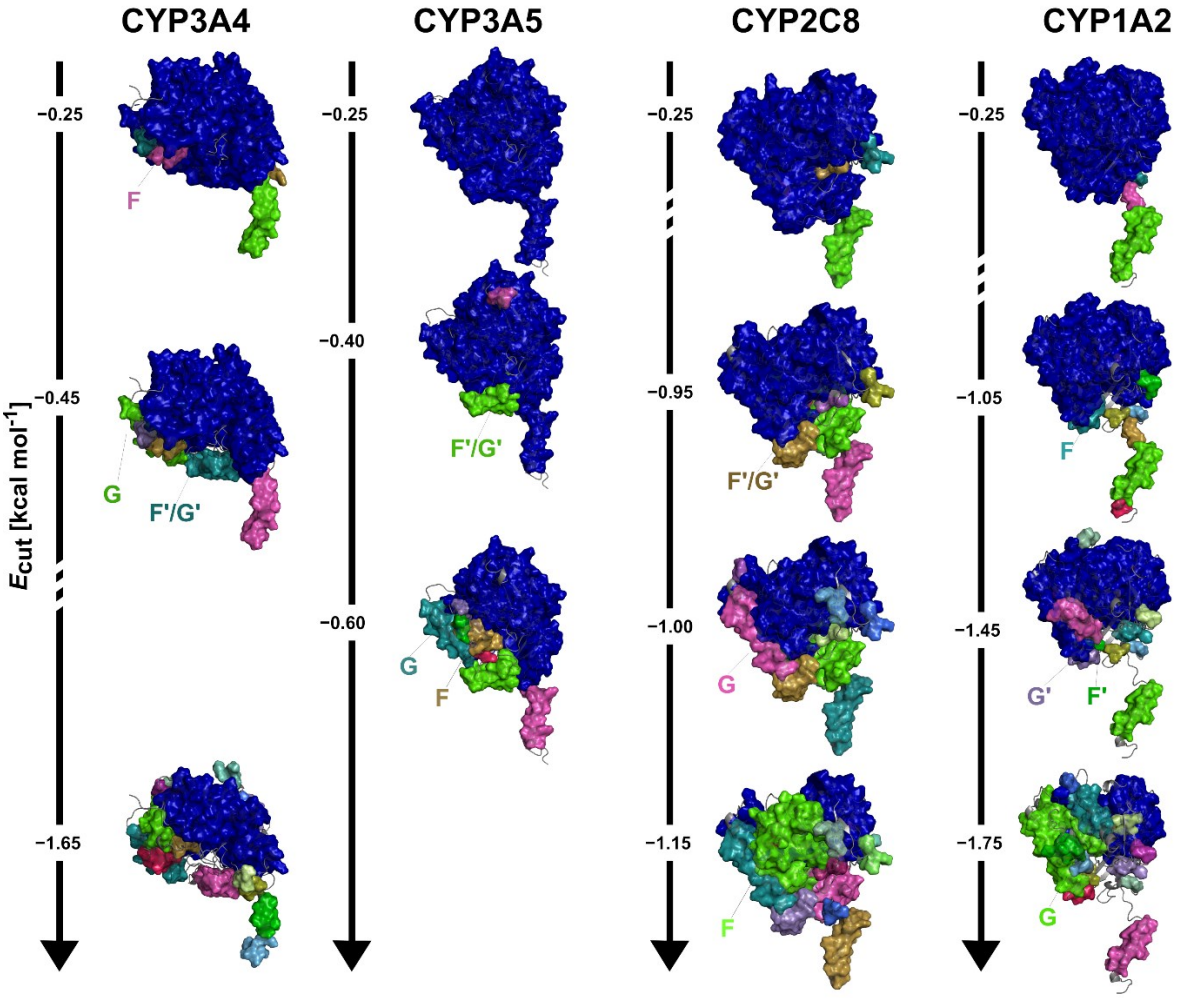



*Figure 14: Constraint network analysis of CYP3A4, CYP3A5, CYP2C8, and CYP1A2 and rigid cluster decompositions along constraint dilution trajectories. Rigid clusters of the CYP isoforms at different $E_{cut}$ values are colored blue, green, pink, and cyan in descending order respective to their size. Left: In CYP3A4, the F-helix forms its separated rigid cluster at $E_{cut}$ = −0.25 kcal mol$^{-1}$, G-helix, and F'/G'-helix form their separated rigid clusters at $E_{cut}$ = −0.45 kcal mol$^{-1}$. At $E_{cut}$ = 1.65 kcal mol$^{-1}$, the largest rigid cluster of CYP3A4 is still covering most of the globular part. Middle-left: In CYP3A5, the F'-helix and G'-helix form their*

*separated rigid clusters at $E_{cut}$ = −0.40 kcal mol$^{-1}$, the F-helix and the G-helix form their separated cluster at $E_{cut}$ = −0.60 kcal mol$^{-1}$ and the largest rigid cluster of CYP3A5 is still covering most of the globular part. Middle-right: In CYP2C8, the F'-helix and G'-helix form their separated rigid cluster at $E_{cut}$ = −0.95 kcal mol$^{-1}$. The G-helix forms its separated cluster at $E_{cut}$ = −1.00 kcal mol$^{-1}$. At $E_{cut}$ = 1.15 kcal mol$^{-1}$, the F-helix forms its separated cluster, and the largest rigid cluster of CYP2C8 covers only the active site in the globular part. Complete segregation of the F/G-region is not recognizable for CYP2C8. Right: In CYP1A2, the F-helix forms its separated rigid cluster at $E_{cut}$ = −1.05 kcal mol$^{-1}$. The F'-helix and G'-helix form their separated cluster at $E_{cut}$ = −1.45 kcal mol$^{-1}$. At $E_{cut}$ = 1.75 kcal mol$^{-1}$, the G-helix forms its separated cluster, and the largest rigid cluster of CYP1A2 covers only the active site in the globular part. These examples qualitatively depict that the F/G-region of more promiscuous CYP isoforms, such as CYP3A4 and CYP3A5, are less structurally stable than that of more specific isoforms. Furthermore, differences in structural stability are revealed for sequentially close isoforms. The rigid clusters are calculated based on the neighbour stability map of one exemplary MD trajectory, respectively, as described in ref[182].*

Differences in the structural stability of the F/G-region occur not only between sequentially different CYP isoforms (Figure 14) but also for close sequence homologs, such as CYP3A4 and CYP3A5 (sequence identity/similarity 84%/90%).

In all CYP isoforms investigated in this study, the F/G region is most weakly coupled to the remainder of the globular domain. Between both sequentially different isoforms and homologous ones, differences in the structural stability of this region are revealed by rigidity analysis. These differences qualitatively relate to the substrate scope of the isoforms: more promiscuous isoforms such as CYP3A4 show less structurally stable F/G-regions and *vice versa*.

To quantify the relation between substrate scope and structural rigidity, we computed $\bar{E}_{FG,CNA}$ (Equation 3) as a measure for how well residues in the F/G-region form rigid contacts and correlated it to the promiscuity index $I_{cat}$ (Equation 7) introduced by Nath and Atkins.[168]

A good and significant inverse correlation ($R^2$ = 0.85 and $p$ < 0.01, Wald test) between isoforms promiscuity and respective structural stability of the F/G region is obtained for eight studied CYP isoforms for which $k_{cat}$ and $K_M$ values are available in ref.[166] (Figure 15).

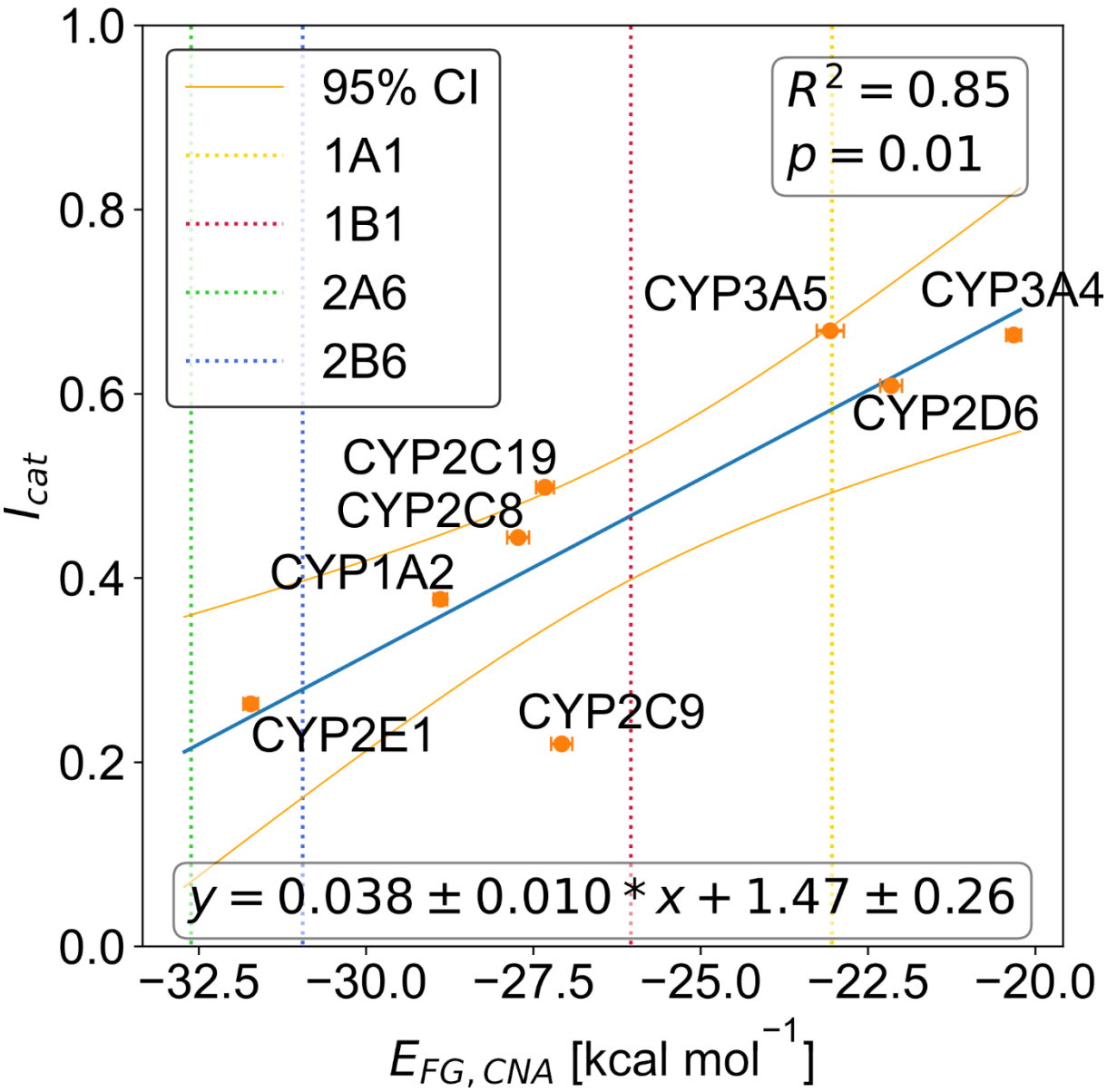



*Figure 15: The promiscuity of CYP isoforms is inversely correlated to the structural rigidity of their F/G regions. CYP2C9 can be considered an outlier since this isoform's globular part moved away from the membrane during the MD simulations. Vertical dotted lines indicate calculated $\bar{E}_{FG,CNA}$ values for CYP isoforms for which no experimental data was available for computing $I_{cat}$. 95% confidence interval limits are shown in orange. Error bars denote the SEM obtained from averaging over five replicas, ten for CYP2E1 and CYP3A4, respectively, per-trajectory results. The formula of the predictive fit is given at the bottom of the diagram.*

In contrast, no significant correlation can be found in $\bar{E}_{region,CNA}$ (Equation 3) of the whole enzymes or other regions (the I helix, which is the most extended helix in the center of CYP enzymes, or the B/C-region, which is also part of the main entrance channel).

Using the fit obtained in Figure 14, we used the $\bar{E}_{FG,CNA}$ values computed for CYP1A1, CYP1B1, CYP2A6, and CYP2B6 to predict their respective $I_{cat}$ values. The predicted $I_{cat}$ values of 0.58, 0.47, 0.22, and 0.28, respectively, allow us to classify these CYP isoforms as being lowly to moderately promiscuous. This finding is in agreement with experimental data showing that CYP3A5 generally catalyzes the same reactions as CYP3A4 but almost always at lower rates, which leads to similar $I_{cat}$ values that are associated with a slightly lower $\bar{E}_{FG,CNA}$ of CYP3A5 than CYP3A4 (Figure 15).

## 5.4 Conclusion and significance

We derived a novel quantitative relation for predicting the promiscuity of CYP isoforms, as demonstrated for CYP1A1, CYP1B1, CYP2A6, and CYP2B6. The difference in $\bar{E}_{FG,CNA}$ between CYP1A1 and CYP1A2 is larger than that observed for other isoforms of the same subfamily, such as CYP3A4 and CYP3A5, but still in the range found for CYP2C8 and CYP2C19. Isoforms of the same subfamily can differ markedly in their substrate scopes, and our approach can detect that. However, the model cannot be used to predict the ability of a specific CYP to metabolize a given substrate.

In summary, our results emphasize that the structural rigidity of the F/G-region can be used to classify CYP isoforms involved in metabolism according to their promiscuity. Our model may be generalized to novel CYP enzymes, e.g., found in meta-genome approaches for potential use in bioorganic chemistry or biotechnology.

# 6 MOLECULAR MECHANISMS IN CYP2D6

Daniel Becker[a], Prasad V. Bharatam[b], Holger Gohlke[a,c,1]

[a]Mathematisch-Naturwissenschaftliche Fakultät, Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany

[b]Department of Medicinal Chemistry, National Institute of Pharmaceutical Education and Research (NIPER), Sector 67, S. A. S. Nagar, Mohali, 160 062 Punjab, India

[c]John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre (JSC), Institute of Biological Information Processing (IBI-7: Structural Biochemistry), and Institute of Bio- and Geosciences (IBG-4: Bioinformatics), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

The content of this chapter is taken from “Molecular Mechanisms Underlying Single Nucleotide Polymorphism-induced Reactivity Decrease in CYP2D6”. Word-by-word citations are not marked explicitly.

## 6.1 Author Contributions

HG designed the study; DB performed computations; DB and HG analyzed results; DB and HG wrote the manuscript; PVB revised the manuscript; HG and PVB secured funding.

## 6.2 Background

CYP 2D6 is crucial in human drug metabolism. It catalyzes the oxidation of specific lipophilic drugs, occurring as phase I biotransformation before excretion (see sections 2.2.1 and 2.2.2). Already in 1977, Mahgoub *et al.* described that differences in the hydroxylation rate of debrisoquine, an antihypertensive drug, depend on a single autosomal gene,[183] called later CYP2D6 according to its position in the CYP gene superfamily.[184] Since then, CYP2D6 polymorphism has raised increasing awareness during the development of small-molecule drugs because CYP polymorphism can lead to underdosing – leading to a failure of therapy or intoxication by drug metabolites – or overdosing – leading to intoxication by the drug itself.[10]

Differential mechanisms drive the differences in CYP2D6-related metabolism. The "very extensive" or "ultra-metabolizer" phenotype is caused mainly by gene

duplication.[185] This duplication results in the overexpression of the CYP2D6 enzyme. Around 7% of Caucasians show an "ultra-metabolizer" phenotype.[186] Although an explanation at the genetic level for the "ultra-metabolizer" phenotype is often appropriate, mutations in the CYP2D6 gene can also lead to higher enzyme activity.[187] The "poor-metabolizer" and "non-metabolizer" phenotypes are established either at the genetic level, probably related to regulatory factors[188], or at the protein level, in which a (combination of) single nucleotide polymorphisms (SNPs) changes the enzymatic activity[189]. Deciphering the molecular mechanisms underlying the SNP-induced reactivity decrease of CYP2D6 is of utmost interest from a fundamental and drug development point of view.

Effects on CYP2D6 reactivity related to different alleles, including combinations of mutations, were investigated previously in wet lab experiments and computational studies.[189-192] Due to the many naturally occurring alleles, only investigations of a subgroup of alleles have been possible. In these studies, different aspects of how mutations affect the catalysis steps have been pointed out. I) Decreased activity can be correlated with an instability of the expressed enzyme variant.[193] II) Mutations of F120 and F483 affect the substrate binding.[194-196] III) The arginine and lysine are crucial in the interaction between CYP and CPR. Allele CYP2D6*31 (combining SNPs R296C, R440H, and S486T; listed in the Pharmacogene Variation Consortium database) has a mutation in one of these positions, likely underlying that this CYP2D6 variant has no enzymatic activity *in vivo* and *in vitro*.[33] IV) The bridging arginine between FMN of CPR that supports the electron transfer [197] is another crucial residue, as a substitution probably prevents electron transfer to the heme, leading to inactivity. However, to our knowledge, no mutagenesis of this residue has been documented.

In this work, we aim to characterize the effect of mutations found in naturally occurring alleles on CYP2D6 activity as a function of its structural dynamics at the atomistic level. Due to the high number of natural mutants, our study focuses on the twenty most common SNPs reported in the global population (Figure 16). We performed extensive all-atom MD simulations with a cumulated simulation time of 200 µs. On this basis, we pursue a mechanism- and hypothesis-driven evaluation to detect if and how a given SNP could lead to a decreased reactivity due to a change in structural flexibility.

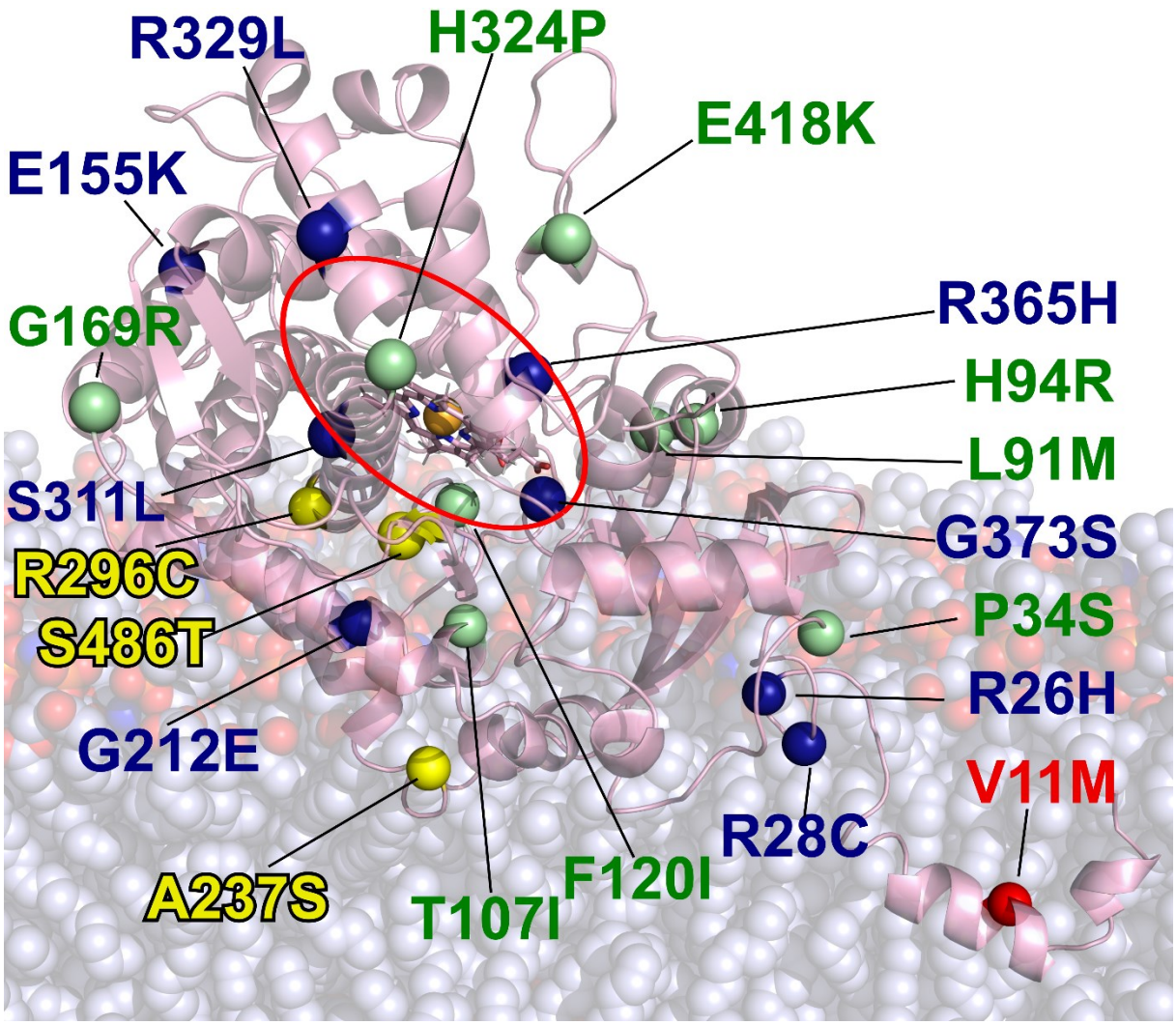



*Figure 16: Overview of the 20 most frequent SNPs mapped onto CYP2D6 (pink). Mutations that lead to poor metabolism are colored green, variants with no effect are colored yellow, V11M leading to an ultra-rapid metabolization is colored red, and variants with unknown effects are colored blue. $C_\alpha$ atoms of the positions are shown as spheres. The orange sphere in the center indicates the heme iron in the center of the porphyrin ring, which is shown as sticks.*

For this study, we assumed that mutations documented as extensive and ultrarapid can be used as a comparison group (CG) to identify the atomistic mechanisms leading to decreased activity in the other variants. The variants were investigated as to different parameters relevant to the catalyzed reaction, i.e., enzyme stability, substrate binding, binding of CPR, electron transfer, and substrate access. We then defined criteria for determining whether a given parameter significantly differs from the CG. The derived criteria could also be applied to characterize other variants not included in this dataset.

## 6.3 Results

*Identification of poor metabolizing CYP2D6 variants*

The most common SNPs were identified based on the Uniprot entry (Uniprot-ID: P10635),[198] which reports a total of 50 SNPs. We restricted ourselves to the 20 most frequent mutations with a relative frequency of at least $10^{-4}$ in the global population. This frequency was taken from the dbSNP database.[199] Data from ClinVar,[200] PharmVar,[33], and the Human Gene Mutation Databank[201] were used to assess the clinical relevance of the mutations.

*Enzyme stability*

The importance of the I-helix, as the central helix in the enzyme 3D structure, entails that if the enzyme is destabilized in this region, it loses its function.[202] Thus, we use the structural stability of the I-helix as the most relevant metric to determine if a reduced CYP2D6 activity could be correlated with a decreased stability of this region. Compared to the CG, the chemical potential (Equation 3) predicted for this region and averaged over all other variants ($\bar{E}_{\mathrm{I,CNA}}^{lim.}$ = -41.57 kcal mol$^{-1}$) is significantly different even when considering the maximum standard deviation of the variants that are not in the comparison group (nonCG) ($p < 0.01$, two-sided t-test). Thus, variants with $E_{\mathrm{I,CNA}} > E_{\mathrm{I,CNA}}^{lim.}$ are considered to lead to poor drug metabolism.

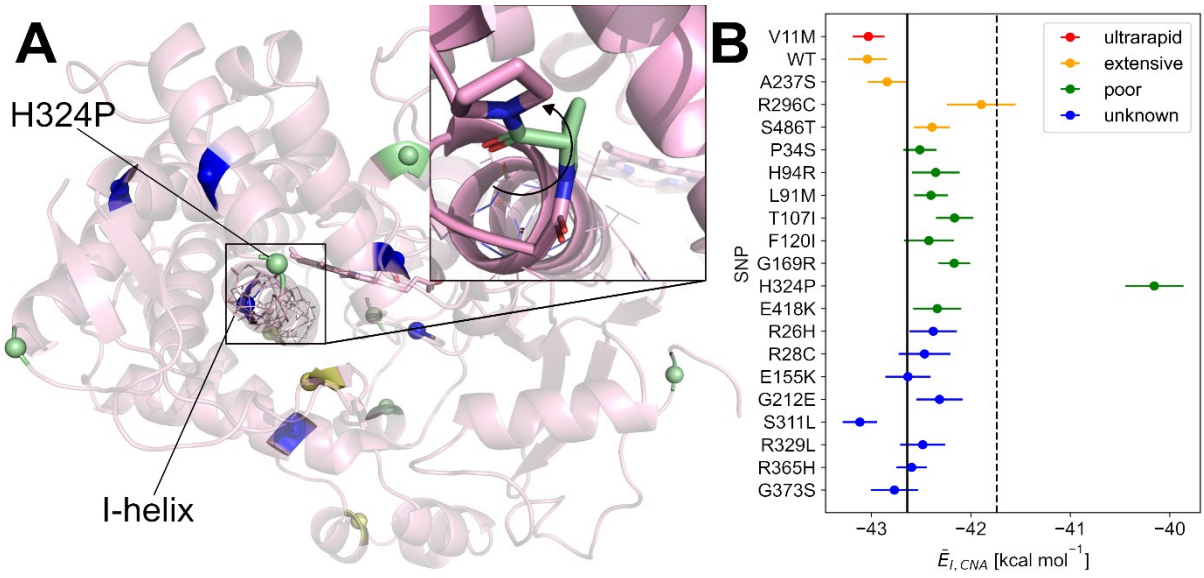



*Figure 17:* ***Structural instability of the I-helix leads to poor metabolism of the CYP2D6 H324P. A:*** *The structure of the globular part of CYP2D6 (pink) is shown with the I-helix marked with backbone lines and positions of SNPs, as shown in Figure 16. The heme group is shown as sticks. The H324P variant is shown with the (angle indicated) in the blowup.* ***B:*** *Average chemical potential of rigid contacts of I-*

*helix residues ($\bar{E}_{I,CNA}$) shown for each investigated SNP. For all variants except H324P, the value is approximately between -43 and -42 kcal mol$^{-1}$. Only H324P differs significantly ($p < 0.01$, two-sided t-test) from CG, leading to a decreased structural stability of the I-helix. The continuous vertical line denotes the mean of $\bar{E}_{I,CNA}$ of the comparison group (Equation 3); the error bars represent the SEM determined by error propagation along ten independent simulations. The dashed vertical line denotes the limit for a significantly different value ($p < 0.01$, two-sided t-test).*

*Substrate binding*

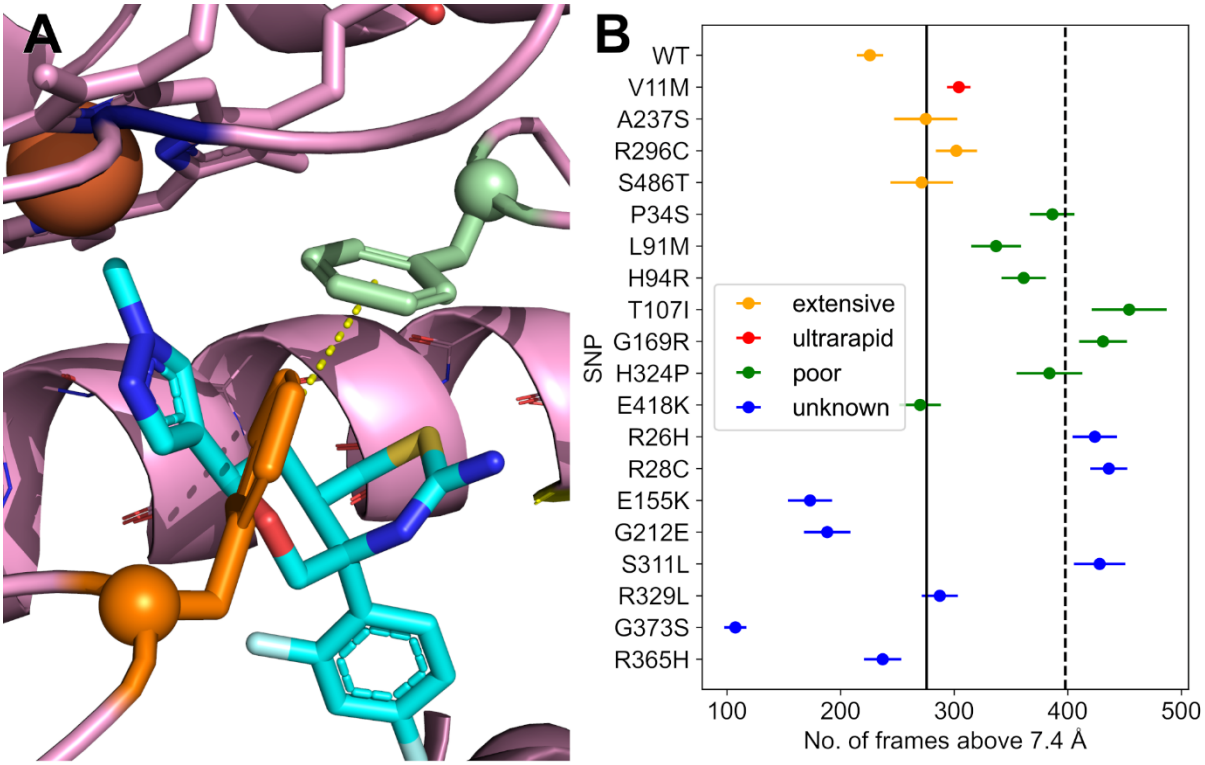



*Figure 18: **Importance of F120 and F483 in substrate binding. A:** Zoom into the crystal structure of the active site of CYP2D6 (pink) binding inhibitor BACE-1 (cyan) (PDB-ID: 4XRY). The heme and ligand are shown as sticks; F120 is marked in green, and F483 is marked in orange. The iron within the heme is shown as a sphere. **B:** The number of frames with a distance between F120 and F483 above 7.4 Å is depicted for all investigated variants except F120I. The distance was measured between the phenyl rings. The continuous vertical line denotes the mean distance between F120 and F483 of the comparison group; the error bars represent the SEM determined by error propagation along ten independent simulations. The dashed vertical line denotes the limit for a significantly different value ($p < 0.01$, two-sided t-test).*

As substrate binding is a crucial step in the reaction process, we hypothesized that the substrate binding free energy should be the second most relevant parameter influencing CYP2D6 activity. In the crystal structure of CYP2D6 bound to BACE1 (PDB

ID: 4XRY), both F120 and F483 stabilize the aromatic substrate (Figure 18A). In this structure, the distance between the ring moieties of these residues is 7.4 Å. Assuming that this distance is relevant for substrate binding, we investigated how often this distance is found above 7.4 Å in our MD trajectories. Compared to the CG, an occurrence of 4% was predicted for the other variants and is found to be significantly different ($p < 0.01$, two-sided t-test) (Figure 18B). Thus, variants for which a [F120-F143] ring distance above 7.4 Å is seen in more than 4% of the total simulation time are considered to lead to poor metabolism.

*Binding to CPR*

The transfer of an electron from CPR to the CYP heme is mandatory for the reaction cycle of CYP enzymes (see Section 2.2.3). Binding between both proteins is a prerequisite, which we consider the third most relevant parameter influencing CYP2D6 activity. The unavailability of R450 of CPY2D6 as an interaction partner in the interface caused by an intramolecular salt bridge formation with E150 was analyzed with respect to the average chemical potential of this interaction, $\bar{E}_{\text{R450-E150,CNA}}$. Compared to the CG, the average chemical potential of the other variants, i.e., $\bar{E}_{\text{R450-E150,CNA}}^{lim.}$ = -1.0 kcal mol$^{-1}$, is found to be significantly different ($p < 0.01$, two-sided t-test). Thus, variants with $E_{\text{R450-E150,CNA}} < E_{\text{R450-E150,CNA}}^{lim.}$ lead to poor metabolism.

*Electron transfer via R132*

We considered the transfer of an electron mediated by R132 as a fourth relevant parameter influencing CYP2D6 activity (Figure 19). As a geometrical criterion, this parameter was quantified by the distance between iron and R132 monitored during the MD simulations. Compared to the CG, an average distance of 7.91 Å was calculated for the other variants and found to be significantly different ($p < 0.01$, two-sided t-test). Thus, variants with a distance > 7.91 Å are considered to lead to poor metabolism.

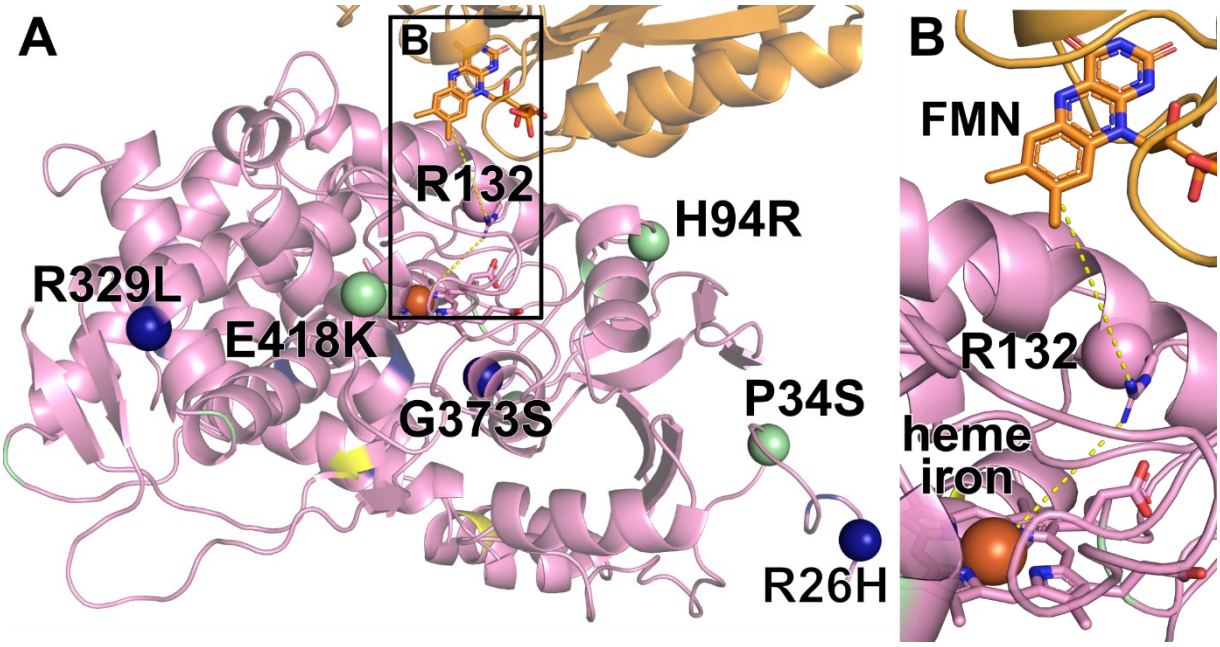



*Figure 19:* ***R132 plays a crucial role in electron transfer from CPR to CYP2D6. A:*** *Interface of CPR (yellow) and CYP2D6 (pink) derived by docking using the HADDOCK web server [203]. Heme and flavin mononucleotide are shown as sticks, as is R132. The iron within the heme is shown as a sphere. All Cα-atoms of substitutions in variants that show a significantly higher distance between iron and R132 are shown as a sphere.* ***B:*** *Close-up view of the interface between CYP2D6 and CPR. The route of the electron transfer, according to ref.[197] is indicated with dashed yellow lines. Thus, the distance between R132 and the heme iron is essential for the likelihood of an electron transfer.*

*Identification of new poor metabolizing isoforms*

As previously described, we analyzed four different parameters which influence CYP2D6-catalyzed reactions. These metrics can be ranked by order of importance and associated to classification rules depending on their value compared to the CG. The resulting decision tree (Figure 20A) aims to predict whether or not a given CYP2D6 substitution leads to poor metabolism.

Following this decision tree (Figure 20A), variants of the CG and those that lead to known poor metabolism are correctly classified (Figure 20B). Applying this decision tree to variants with unknown effects leads to the predictions that variants with unknown effect R26H, R28C, S311L, R329L, and G373S result in poor drug metabolism.

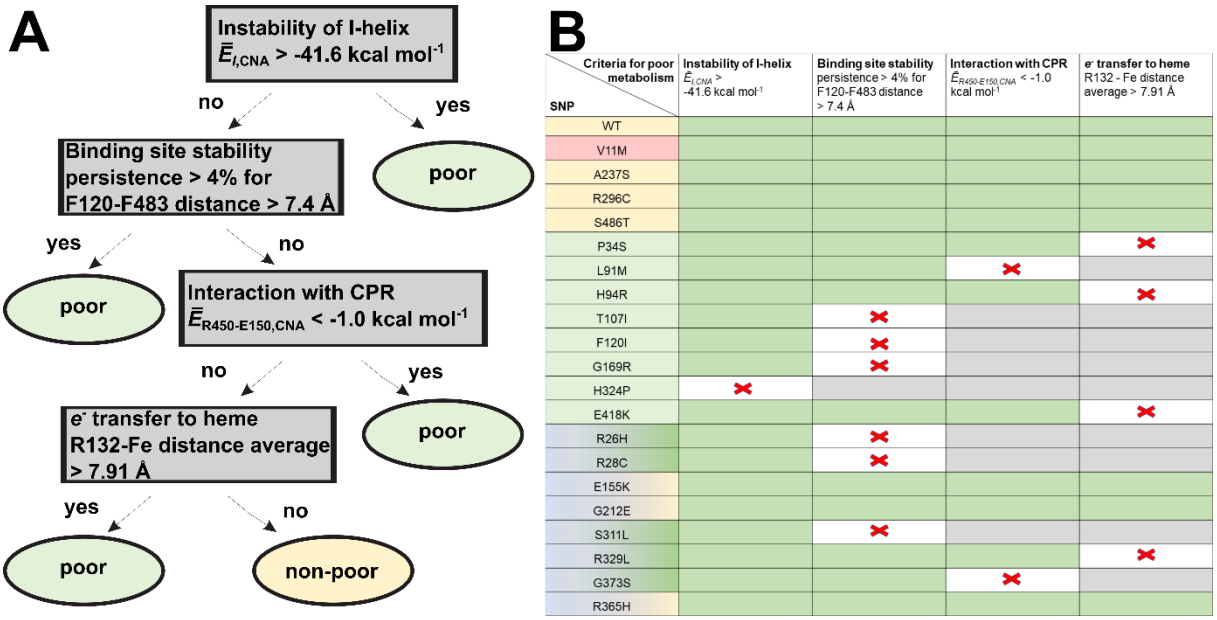



*Figure 20:* ***Decision tree for the classification of CYP variants with unknown effect. A:*** *Decision tree used to distinguish poorly metabolizing variants from the CG.* ***B:*** *Classification results for the 20 variants considered in this study according to the four chosen metrics. The variants showing values for these metrics that differ significantly from the control group are marked with a red cross. A green box is used in the opposite scenario. A grey box indicates that the condition was not investigated in the decision tree. As a result, R26H, R28C, and S311L are predicted to lead to poor metabolism due to diminished binding site stability, whereas G373S is predicted to lead to poor metabolism due to reduced interaction with CPR. The first column is colored according to the (predicted) effect on the metabolism (poor metabolism: green; variants with no effect: yellow; V11M leading to an ultra-rapid metabolization: red; variants with unknown effects are colored with a color gradient from blue to the predicted effect on metabolism (green: poor; yellow: non-poor).*

## 6.4 Conclusion and significance

According to our analyses, the F/G-region stability and the substrate promiscuity are unchanged regardless of the investigated SNPs. This observation fits the clinical practice given that a patient identified as a poor metabolizer should have a poor metabolism irrespective of the CYP2D6 substrate considered.[34] In such cases, the drug regime must be modified once for all substrates and not for each substrate separately.

The derived mechanisms and the associated criteria can be used to identify the effects of unknown SNPs and scrutinize all known variants as to the molecular origins that lead to the clinical effect. This approach can especially be helpful for clinical cases with low incidence.

In summary, we identified four distinct mechanisms that lead to poor drug metabolism in the case of 20 known CYP2D6 variants with the help of extensive MD simulations and Constraint Network Analysis. Our model shall allow us to predict the impact of novel SNPs of CYP2D6 on drug metabolism. We anticipate that this tool may be used to guide recommendations to modify drug regimes.

# 7 SUMMARY AND PERSPECTIVE

Throughout this thesis, phase I of the human drug metabolism was investigated in two separate studies. The first study aims to identify the molecular factors governing substrate promiscuity of CYP enzymes in general while focusing on the special role of CYP3A4 within the group of CYP enzymes involved in human drug metabolism. The second study focused on CYP2D6, which is of particular interest to pharmaceutical research and drug development due to its high polymorphism.

Human drug metabolism is mainly driven by twelve CYP enzymes belonging to families 1, 2, and 3. CYP3A4 plays a unique role because it takes part in ~27% of all drug metabolizing reactions catalyzed by CYP enzymes.[15] To quantify the promiscuity of enzymes, Nath *et al*. invented the promiscuity index ($I_{cat}$), used in this work to detect correlations between CYPs structural features and their promiscuity.[168] Studies based on crystal structures and computational simulations emphasize the role of the F/G-region, consisting of the F-, F'-, G'-, and G-helices for substrate binding[124-126]. The F/G-region is at the beginning of the substrate entrance channel.[120] We developed a quantitative approach between CYP substrate promiscuity and F/G-region rigidity in **Publication I**. For the first time, this model quantitively describes such a relationship, which has so far been suggested qualitatively based on three isoforms in another study.[126] Thus, we provide a general understanding of the differences in promiscuity in human drug-metabolizing CYP enzymes.

Further investigations would be needed to validate the presented model for highly selective human CYPs or CYPs from other organisms. Such models will help to derive a more general understanding of CYP selectivity. In highly selective CYP enzymes, other selectivity mechanisms probably play a role. From P450cam, we already know that selectivity depends partly on four important residues (Section 2.2.4). Such a rule probably also exists for human CYPs, and it would be possible to predict not only the promiscuity but also the substrate spectra.

CYP2D6 is highly interesting due to the massive variety of genetic variants, which result in different clinical outcomes. The isoform takes part in 6% of all drug metabolizing reactions catalyzed by CYP enzymes, and the differences in CYP2D6 metabolism can cause severe effects in therapy by over- and underdosing.[15, 34, 83] In **Publication II,** we identified four molecular mechanisms that lead to poor metabolism

of CYP2D6 variants. Our model shall allow us to predict if newly identified SNPs of CYP2D6 will lead to poor metabolism, which may be used for recommendations to modify drug regimes. In this study, we used extensive MD simulations combined with Constraint Network Analysis. This approach is resource-intensive, so further investigations are needed to derive simpler models for rapidly identifying CYP2D6 poor metabolizers.

Furthermore, the approach of systematically investigating molecular mechanisms that lead to changes in enzyme activity could be applied to other CYPs and other enzyme classes. Our work investigated mechanisms that potentially appear in all other CYP isoforms. An investigation of CYPs used in biotechnology, i.e., P450BM3, would provide a tool to predict *in silico* if a planned mutation to change oxidation sides leads to poor enzyme activity.

# 8 ACKNOWLEDGEMENTS

First, I would like to thank my supervisor, **Prof. Dr. Holger Gohlke,** for the opportunity to conduct my research in his group. I am very grateful for the opportunities he has given me, both academically through an exciting main project and interesting side projects, as well as in my personal development through teaching responsibilities and my stay in India.

Special thanks also go to **Prof. Bharatam**, who was a great collaboration partner and my host in India. He and his working group gave me a deep insight into the Indian culture, and I am always happy when I think back to these three months.

I would also like to thank my second supervisor, **Prof. Dr. Vlada Urlacher,** for reviewing this work and for her excellent reviews in the field of CYP enzymes, which gave me a deep insight into the topic.

A big thank you goes to all my colleagues and friends from the CPClab and CBClab. Especially, I thank **Dr. Benoit David,** who introduced me to the world of molecular dynamic simulations (and the French way of storing cheese). Deep thanks are due to **Dr. Christoph G. W. Gertzen**, **Dr. Christopher Pfleger**, **Dr. David Bickel**, **Dr. Michele Bonus**, and **Dr. Stephan Schott Verdugo**. Each has accompanied and shaped my scientific development through discussion and debate.

Thanks also to my (former) office mates: **Dr. Benoit David**, **Birte Schmitz**, **Dr. Christina Gohlke**, **Jesko Kaiser**, **Katharina Klostermann,** and **Marco Lapsien**. This journey would not have been possible with this office as spot to work focused.

Teaching with the wrong team does not work out, so I am thankful that my teaching duties were always shared with outstanding and motivated colleagues: **Dr. Barbara Gioffreda**, **Dr. Daniel Mulnaes**, **Dr. David Bickel**, **Jesko Kaiser**, **Korana Mudrovic**, **Lisa Kersten**, **Nicola Porta**, **Pablo Cea**, **Pegah Golchin**, **Rocco Gentile**, **Dr. Stefanie Brands** and **Yu Lin Ho**.

Besides research and teaching, the administration of the IT infrastructure was my duty, and here I have to thank the cpcsudo group for the steep learning curve and the all-time interesting jour fixe: **Jesko Kaiser**, **Korana Mudrovic**, **Peter Sippel**, and **Dr. Stephan Schott Verdugo**.

In addition, I want to thank **Annika Behrendt**, **Bastian Schepers**, **Filip König**, **Dr. Jonas Dittrich**, **Julian Greb**, **Lucas Raei**, **Dr. Sabahudin Ahmad**, and **Susanne Hermanns** for enriching my social life and several scientific discussions.

Thanks to all **my collaborators,** Prof. Dr. Bharatam, Prof. Dr. Carsten Münk, Dr. Augustin Penda Twizerimana, and Samima Zaman.

I sincerely thank my WPP and master/diploma students, **Alena Endres**, **Enis Abdyli**, **Kai Haas**, and **Luca Przygoda** for their excellent work on CYP2D6, the exciting comparison of TopModel and AlphaFold, and the interessting investigations in structural rigidity of proteins. It was more than a pleasure to work with you.

I owe my deepest gratitude and appreciation to my **fiancée**, who supported my work and tolerated my phases of evening workaholism.

Last but not least, I would like to thank **my whole family**, all my **friends,** and the **Musketeers** for their support along my journey through life.

# 9 REPRINT PERMISSIONS

Publication I

# 11 REFERENCES


1. Williams, R. J., A system's view of the evolution of life. *J R Soc Interface* 2007, 4, 1049-70.
2. Oesch, F.; Arand, M. Role of Individual Enzymes in the Control of Genotoxic Metabolites. In *Molecular and Applied Aspects of Oxidative Drug Metabolizing Enzymes*, Arinç, E.; Schenkman, J. B.; Hodgson, E., Eds.; Springer US: Boston, MA, 1999, pp 211-220.
3. Qadir, A.; Hashmi, M. Z.; Mahmood, A. Xenobiotics, Types, and Mode of Action. In *Xenobiotics in the Soil Environment: Monitoring, Toxicity and Management*, Hashmi, M. Z.; Kumar, V.; Varma, A., Eds.; Springer International Publishing: Cham, 2017, pp 1-7.
4. Kovalev, I. E.; Shipulina, N. V., Covalent binding of xenobiotics with proteins in the organism as the mechanism of adaptation. *Pharmaceutical Chemistry Journal* 1996, 30, 665-670.
5. Klaassen, C. D.; Plaa, G. L., Biliary Excretion of Xenobiotics. *CRC Critical Reviews in Toxicology* 1975, 4, 1-29.
6. Döring, B.; Petzinger, E., Phase 0 and phase III transport in various organs: Combined concept of phases in xenobiotic transport and metabolism. *Drug Metab. Rev.* 2014, 46, 261-282.
7. Petzinger, E.; Geyer, J., Drug transporters in pharmacokinetics. *Naunyn-Schmiedeberg's Arch. Pharmacol.* 2006, 372, 465-475.
8. Potter, B. J., Biliary and Fecal Drug Excretion. 2014.
9. Guengerich, F. P., Common and Uncommon Cytochrome P450 Reactions Related to Metabolism and Chemical Toxicity. *Chem. Res. Toxicol.* 2001, 14, 611-650.
10. Nebert, D. W.; Russell, D. W., Clinical importance of the cytochromes P450. *Lancet* 2002, 360, 1155-1162.
11. Ortiz de Montellano, P. R., Hydrocarbon Hydroxylation by Cytochrome P450 Enzymes. *Chem. Rev.* 2010, 110, 932-948.
12. Anzenbacher, P.; Anzenbacherová, E., Cytochromes P450 and metabolism of xenobiotics. *Cell. Mol. Life Sci.* 2001, 58, 737-747.
13. Nelson, D. R.; Nebert, D. W. Cytochrome P450 ( CYP ) Gene Superfamily. In *Encyclopedia of Life Sciences*; 2018, pp 1-19.
14. Guengerich, F. P. Human Cytochrome P450 Enzymes. In *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Ortiz de Montellano, P. R., Ed.; Springer International Publishing: Cham, 2015, pp 523-785.
15. Rendic, S.; Guengerich, F. P., Survey of Human Oxidoreductases and Cytochrome P450 Enzymes Involved in the Metabolism of Xenobiotic and Natural Chemicals. *Chem. Res. Toxicol.* 2015, 28, 38-42.
16. Gallo, P.; De Vincentis, A.; Pedone, C.; Nobili, A.; Tettamanti, M.; Gentilucci, U. V.; Picardi, A.; Mannucci, P. M.; Incalzi, R. A., Drug–drug interactions involving CYP3A4 and p-glycoprotein in hospitalized elderly patients. *European Journal of Internal Medicine* 2019, 65, 51-57.
17. Pelkonen, O.; Hakkola, J.; Hukkanen, J.; Turpeinen, M., CYP-associated drug–drug interactions: A mission accomplished? *Arch. Toxicol.* 2020, 94, 3931-3934.
18. Silbernagl, S., *Taschenatlas Physiologie*. 8., überarb. und erw. Aufl. ed.; Thieme: Stuttgart u.a., 2012.
19. Lamb, D. C.; Lei, L.; Warrilow Andrew, G. S.; Lepesheva Galina, I.; Mullins Jonathan, G. L.; Waterman Michael, R.; Kelly Steven, L., The First Virally Encoded Cytochrome P450. *J. Virol.* 2009, 83, 8266-8269.
20. Sharma, V.; Shing, B.; Hernandez-Alvarez, L.; Debnath, A.; Podust, L. M., Domain-Swap Dimerization of Acanthamoeba castellanii CYP51 and a Unique Mechanism of Inactivation by Isavuconazole. *Mol. Pharmacol.* 2020, 98, 770-780.
21. Poulos, T. L.; Johnson, E. F. Structures of Cytochrome P450 Enzymes. In *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Ortiz de Montellano, P. R., Ed.; Springer International Publishing: Cham, 2015, pp 3-32.
22. Nebert, D. W.; Adesnik, M.; Coon, M. J.; Estabrook, R. W.; Gonzalez, F. J.; Guengerich, F. P.; Gunsalus, I. C.; Johnson, E. F.; Kemper, B.; Levin, W.; Phillips, I. R.; Sato, R. Y. O.; Waterman, M. R., The P450 Gene Superfamily: Recommended Nomenclature. *DNA* 1987, 6, 1-11.

23. Ebers, G., Papyrus Ebers. *Zeitschrift für Ägyptische Sprache und Altertumskunde* 1873, 11, 41-46.
24. Carpenter, S.; Rigaud, M.; Barile, M.; Priest, T. J.; Perez, L.; Ferguson, J. B. J. B. C., The Ebers Papyrus. 2006, 3.
25. Bernardi, G. Hildegard von Bingen (1098–1179). In *The Unforgotten Sisters: Female Astronomers and Scientists before Caroline Herschel*, Bernardi, G., Ed.; Springer International Publishing: Cham, 2016, pp 49-54.
26. Nutton, V., Galen and Roman Medicine: or can a Greek become a Latin? *European Review* 2012, 20, 534-542.
27. Ranhel, A. S.; Mesquita, E. T., The middle ages contributions to cardiovascular medicine. *Brazilian journal of cardiovascular surgery* 2016, 31, 163-170.
28. Cook, H. J. Victories for Empiricism, failures for theory: Medicine and science in the seventeenth century. In *The Body as Object and Instrument of Knowledge: Embodied Empiricism in Early Modern Science*; Springer: 2010, pp 9-32.
29. Eichelbaum, M., Progress in clinical pharmacology over the last 20 years. *Br. J. Clin. Pharmacol.* 2003, 55, 458-9.
30. Lin, K.-M.; Poland, R. E.; Lesser, I. M., Ethnicity and psychopharmacology. *Culture, Medicine and Psychiatry* 1986, 10, 151-165.
31. Zanger, U. M.; Schwab, M., Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol. Ther.* 2013, 138, 103-141.
32. Hicks, J. K.; Swen, J. J.; Thorn, C. F.; Sangkuhl, K.; Kharasch, E. D.; Ellingrod, V. L.; Skaar, T. C.; Müller, D. J.; Gaedigk, A.; Stingl, J. C., Clinical Pharmacogenetics Implementation Consortium Guideline for CYP2D6 and CYP2C19 Genotypes and Dosing of Tricyclic Antidepressants. *Clinical Pharmacology & Therapeutics* 2013, 93, 402-408.
33. Gaedigk, A.; Ingelman-Sundberg, M.; Miller, N. A.; Leeder, J. S.; Whirl-Carrillo, M.; Klein, T. E.; the PharmVar Steering, C., The Pharmacogene Variation (PharmVar) Consortium: Incorporation of the Human Cytochrome P450 (CYP) Allele Nomenclature Database. *Clinical Pharmacology & Therapeutics* 2018, 103, 399-401.
34. Crews, K. R.; Monte, A. A.; Huddart, R.; Caudle, K. E.; Kharasch, E. D.; Gaedigk, A.; Dunnenberger, H. M.; Leeder, J. S.; Callaghan, J. T.; Samer, C. F.; Klein, T. E.; Haidar, C. E.; Van Driest, S. L.; Ruano, G.; Sangkuhl, K.; Cavallari, L. H.; Müller, D. J.; Prows, C. A.; Nagy, M.; Somogyi, A. A.; Skaar, T. C., Clinical Pharmacogenetics Implementation Consortium Guideline for CYP2D6, OPRM1, and COMT Genotypes and Select Opioid Therapy. *Clinical Pharmacology & Therapeutics* 2021, 110, 888-896.
35. Li, Y.; Meng, Q.; Yang, M.; Liu, D.; Hou, X.; Tang, L.; Wang, X.; Lyu, Y.; Chen, X.; Liu, K.; Yu, A. M.; Zuo, Z.; Bi, H., Current trends in drug metabolism and pharmacokinetics. *Acta pharmaceutica Sinica. B* 2019, 9, 1113-1144.
36. Patrick, G. L.; Spencer, J., *An introduction to medicinal chemistry*. 4th ed ed.; Oxford University Press Oxford: Oxford, 2009.
37. Leblanc, M.; Roy, L. F.; Villeneuve, J.-P.; Malo, B.; Pomier-Layrargues, G.; Legault, L., Liver Blood Flow in Chronic Hemodialysis Patients. *Nephron* 2008, 73, 396-402.
38. Nahler, G. glomerular filtration rate (GFR). In *Dictionary of Pharmaceutical Medicine*; Springer Vienna: Vienna, 2009, pp 81-81.
39. Schulze, R. J.; Schott, M. B.; Casey, C. A.; Tuma, P. L.; McNiven, M. A., The cell biology of the hepatocyte: A membrane trafficking machine. *J. Cell Biol.* 2019, 218, 2096-2112.
40. Kaur, I.; Vasudevan, A.; Rawal, P.; Tripathi, D. M.; Ramakrishna, S.; Kaur, S.; Sarin, S. K. Primary Hepatocyte Isolation and Cultures: Technical Aspects, Challenges and Advancements. Bioengineering, 102023.
41. Singh, S.; Singh, D. B.; Gautam, B.; Singh, A.; Yadav, N. Chapter 19 - Pharmacokinetics and pharmacodynamics analysis of drug candidates. In *Bioinformatics*, Singh, D. B.; Pathak, R. K., Eds.; Academic Press: 2022, pp 305-316.
42. Ferreira, L. L. G.; Andricopulo, A. D., ADMET modeling approaches in drug discovery. *Drug Discov. Today* 2019, 24, 1157-1165.

43. Göller, A. H.; Kuhnke, L.; Montanari, F.; Bonin, A.; Schneckener, S.; ter Laak, A.; Wichard, J.; Lobell, M.; Hillisch, A., Bayer's in silico ADMET platform: a journey of machine learning over the past two decades. *Drug Discov. Today* 2020, 25, 1702-1709.
44. Yim, D. S.; Choi, S.; Bae, S. H., Predicting human pharmacokinetics from preclinical data: absorption. *Translational and clinical pharmacology* 2020, 28, 126-135.
45. Smith, D. A.; Beaumont, K.; Maurer, T. S.; Di, L., Volume of Distribution in Drug Design. *J. Med. Chem.* 2015, 58, 5691-5698.
46. Sirianni, G. L.; Pang, K. S., Organ Clearance Concepts: New Perspectives on Old Principles. *J. Pharmacokinet. Biopharm.* 1997, 25, 449-470.
47. Rowland, M.; Pang, K. S., Commentary on "The Universally Unrecognized Assumption in Predicting Drug Clearance and Organ Extraction Ratio". *Clinical Pharmacology & Therapeutics* 2018, 103, 386-388.
48. Fox, C. H.; Selkirk, J. K.; Price, F. M.; Croy, R. G.; Sanford, K. K.; Cottler-Fox, M., Metabolism of Benzo(a)pyrene by Human Epithelial Cells in Vitro. *Cancer Res.* 1975, 35, 3551-3557.
49. Volk, D. E.; Thiviyanathan, V.; Rice, J. S.; Luxon, B. A.; Shah, J. H.; Yagi, H.; Sayer, J. M.; Yeh, H. J. C.; Jerina, D. M.; Gorenstein, D. G., Solution Structure of a Cis-Opened (10R)-N6-Deoxyadenosine Adduct of (9S,10R)-9,10-Epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene in a DNA Duplex. *Biochemistry* 2003, 42, 1410-1420.
50. Moellering, R. C., Jr, Pharmacokinetics of vancomycin. *J. Antimicrob. Chemother.* 1984, 14, 43-52.
51. McLean, M. J., Clinical pharmacokinetics of gabapentin. *Neurology* 1994, 44, S17-22; discussion S31-2.
52. Alqahtani, M. S.; Kazi, M.; Alsenaidy, M. A.; Ahmad, M. Z., Advances in Oral Drug Delivery. 2021, 12.
53. Marin, J. J. G.; Serrano, M. A.; Monte, M. J.; Sanchez-Martin, A.; Temprano, A. G.; Briz, O.; Romero, M. R., Role of Genetic Variations in the Hepatic Handling of Drugs. *Int. J. Mol. Sci.* 2020, 21.
54. Ishikawa, T., The ATP-dependent glutathione S-conjugate export pump. *Trends Biochem. Sci.* 1992, 17, 463-468.
55. Giacomini, K. M.; Huang, S.-m.; Tweedie, D. J.; Benet, L. Z.; Brouwer, K. L.; Chu, X.; Dahlin, A.; Evers, R.; Fischer, V.; Hillgren, K. M.; Hoffmaster, K. A.; Ishikawa, T.; Keppler, D.; Kim, R. B.; Lee, C. A.; Niemi, M.; Polli, J. W.; Sugiyama, Y.; Swaan, P. W.; Ware, J. A.; Wright, S. H.; Wah Yee, S.; Zamek-gliszczynski, M. J.; Zhang, L., Membrane transporters in drug development. *Nature Reviews. Drug Discovery* 2010, 9, 215-36.
56. Greupink, R.; Dillen, L.; Monshouwer, M.; Huisman, M. T.; Russel, F. G. M., Interaction of fluvastatin with the liver-specific Na+-dependent taurocholate cotransporting polypeptide (NTCP). *European Journal of Pharmaceutical Sciences* 2011, 44, 487-496.
57. Burckhardt, G.; Burckhardt, B. C. In Vitro and In Vivo Evidence of the Importance of Organic Anion Transporters (OATs) in Drug Therapy. In *Drug Transporters*, Fromm, M. F.; Kim, R. B., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2011, pp 29-104.
58. Burckhardt, G., Drug transport by Organic Anion Transporters (OATs). *Pharmacol. Ther.* 2012, 136, 106-130.
59. Obaidat, A.; Roth, M.; Hagenbuch, B., The Expression and Function of Organic Anion Transporting Polypeptides in Normal Tissues and in Cancer. *Annual Review of Pharmacology and Toxicology* 2012, 52, 135-151.
60. Philpot, R. M.; Atta-Asafo-Adjei, E.; Nikbakht, K.; Burnett, V.; Lawton, M. P. Emergence of the Flavin-Containing Monooxygenase Gene Family: Molecular Biology, Structure, and Function. In Molecular Aspects of Oxidative Drug Metabolizing Enzymes, Berlin, Heidelberg, 1995//, 1995; Arinç, E.; Schenkman, J. B.; Hodgson, E., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 1995; pp 209-224.
61. Datta, C.; Bhattacharjee, A., Role of monoamine oxidase A (MAO-A) in cardiac aging. *Journal of Cardiology and Cardiovascular Sciences* 2020, 4.
62. Meunier, B.; de Visser, S. P.; Shaik, S., Mechanism of Oxidation Reactions Catalyzed by Cytochrome P450 Enzymes. *Chem. Rev.* 2004, 104, 3947-3980.

63. Sangkuhl, K.; Klein, T. E.; Altman, R. B., Clopidogrel pathway. *Pharmacogenetics and Genomics* 2010, 20.
64. Jancova, P.; Anzenbacher, P.; Anzenbacherova, E., Phase II drug metabolizing enzymes. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* 2010, 154, 103-116.
65. Peter, G. W.; Peter, I. M.; Jayanta Roy, C.; Chantal, G.; Philip, A. G.; Yuji, I.; Antony, J. H.; Fay, K. K.; Perry, M. K.; Namita Roy, C.; Joseph, K. R., GLUCURONIDATION AND THE UDP-GLUCURONOSYLTRANSFERASES IN HEALTH AND DISEASE. *Drug Metab. Dispos.* 2004, 32, 281.
66. Sharma, C. V.; Mehta, V., Paracetamol: mechanisms and updates. *Continuing Education in Anaesthesia Critical Care & Pain* 2013, 14, 153-158.
67. Manyike, P. T.; Kharasch, E. D.; Kalhorn, T. F.; Slattery, J. T., Contribution of CYP2E1 and CYP3A to acetaminophen reactive metabolite formation. *Clinical Pharmacology & Therapeutics* 2000, 67, 275-282.
68. Coles, B.; Wilson, I.; Wardman, P.; Hinson, J. A.; Nelson, S. D.; Ketterer, B., The spontaneous and enzymatic reaction of N-acetyl-p-benzoquinonimine with glutathione: A stopped-flow kinetic study. *Arch. Biochem. Biophys.* 1988, 264, 253-260.
69. Mitchell, J. R.; Thorgeirsson, S. S.; Potter, W. Z.; Jollow, D. J.; Keiser, H., Acetaminophen-induced hepatic injury: Protective role of glutathione in man and rationale for therapy. 1974, 16, 676-684.
70. Zhao, L.; Pickering, G., Paracetamol metabolism and related genetic differences. *Drug Metab. Rev.* 2011, 43, 41-52.
71. Nwabufo, C. K., Relevance of ABC transporters in drug development. *Curr. Drug Metab.* 2022, 23, 434-446.
72. Omura, T.; Sato, R., The Carbon Monoxide-binding Pigment of Liver Microsomes: II. SOLUBILIZATION, PURIFICATION, AND PROPERTIES. *J. Biol. Chem.* 1964, 239, 2379-2385.
73. Ortiz de Montellano, P. R. Substrate Oxidation by Cytochrome P450 Enzymes. In *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Ortiz de Montellano, P. R., Ed.; Springer International Publishing: Cham, 2015, pp 111-176.
74. Carlile, D. J.; Hakooz, N.; Bayliss, M. K.; Houston, J. B., Microsomal prediction of in vivo clearance of CYP2C9 substrates in humans. *Br. J. Clin. Pharmacol.* 1999, 47, 625-635.
75. Pohl, L. R.; Nelson, S. D.; Krishna, G., Investigation of the mechanism of the metabolic activation of chloramphenicol by rat liver microsomes: Identification of a new metabolite. *Biochem. Pharmacol.* 1978, 27, 491-496.
76. Majumdar, T. K. Regulated Bioassay of N-Oxide Metabolites Using LC-MS: Dealing with Potential Instability Issues. In *Handbook of LC-MS Bioanalysis*; 2013, pp 461-470.
77. Vaz, A. D. N.; McGinnity, D. F.; Coon, M. J., Epoxidation of olefins by cytochrome P450: Evidence from site-specific mutagenesis for hydroperoxo-iron as an electrophilic oxidant. *Proceedings of the National Academy of Sciences* 1998, 95, 3555-3560.
78. Tekes, K.; Kalasz, H.; Y Hasan, M.; Adeghate, E.; Darvas, F.; Ram, N.; Adem, A., Aliphatic and aromatic oxidations, epoxidation and S-oxidation of prodrugs that yield active drug metabolites. *Current medicinal chemistry* 2011, 18, 4885-4900.
79. Fatbardha, V.; Siti, N. A. Z.; Hyoung-Goo, P.; Victoria, L. C.; James, J. D. V.; Paul, R. O. d. M., Carbon-Carbon Bond Cleavage in Activation of the Prodrug Nabumetone. *Drug Metab. Dispos.* 2014, 42, 828.
80. Khalaj, Z.; Baratieh, Z.; Nikpour, P.; Khanahmad, H.; Mokarian, F.; Salehi, R.; Salehi, M., Distribution of CYP2D6 polymorphism in the Middle Eastern region. *Journal of research in medical sciences : the official journal of Isfahan University of Medical Sciences* 2019, 24, 61.
81. Sakuyama, K.; Sasaki, T.; Ujiie, S.; Obata, K.; Mizugaki, M.; Ishikawa, M.; Hiratsuka, M., 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, 2460-7.
82. Gasche, Y.; Daali, Y.; Fathi, M.; Chiappe, A.; Cottini, S.; Dayer, P.; Desmeules, J., Codeine Intoxication Associated with Ultrarapid CYP2D6 Metabolism. *New England Journal of Medicine* 2004, 351, 2827-2831.
83. Tobias, J. D.; Green, T. P.; Coté, C. J.; Section On, A.; Pain, M.; Committee On, D.; Agarwal, R.; Anderson, C. T. M.; Hardy, C. A.; Honkanen, A.; Rehman, M. A.; Neville, K.; Houck, C. S.; Jones, B.; Paul, I. M.; Sullivan, J. E.; Van Den Anker, J. N., Codeine: Time to Say "No". *Pediatrics* 2016, 138, e20162396.

84. Schenk, P. W.; van Fessem, M. A. C.; Verploegh-Van Rij, S.; Mathot, R. A. A.; van Gelder, T.; Vulto, A. G.; van Vliet, M.; Lindemans, J.; Bruijn, J. A.; van Schaik, R. H. N., Association of graded allele-specific changes in CYP2D6 function with imipramine dose requirement in a large group of depressed patients. *Molecular Psychiatry* 2008, 13, 597-605.
85. Schlichting, I.; Berendzen, J.; Chu, K.; Stock Ann, M.; Maves Shelley, A.; Benson David, E.; Sweet Robert, M.; Ringe, D.; Petsko Gregory, A.; Sligar Stephen, G., The Catalytic Pathway of Cytochrome P450cam at Atomic Resolution. *Science* 2000, 287, 1615-1622.
86. Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I., Structure and Chemistry of Cytochrome P450. *Chem. Rev.* 2005, 105, 2253-2278.
87. Hannemann, F.; Bichet, A.; Ewen, K. M.; Bernhardt, R., Cytochrome P450 systems—biological variations of electron transport chains. *Biochimica et Biophysica Acta (BBA) - General Subjects* 2007, 1770, 330-344.
88. Iyanagi, T.; Xia, C.; Kim, J.-J. P., NADPH–cytochrome P450 oxidoreductase: Prototypic member of the diflavin reductase family. *Arch. Biochem. Biophys.* 2012, 528, 72-89.
89. Bhattacharyya, A. K.; Lipka, J. J.; Waskell, L.; Tollin, G., Laser flash photolysis studies of the reduction kinetics of NADPH: cytochrome P-450 reductase. *Biochemistry* 1991, 30, 759-765.
90. Gutierrez, A.; Paine, M.; Wolf, C. R.; Scrutton, N. S.; Roberts, G. C. K., Relaxation Kinetics of Cytochrome P450 Reductase: Internal Electron Transfer Is Limited by Conformational Change and Regulated by Coenzyme Binding. *Biochemistry* 2002, 41, 4626-4637.
91. Waskell, L.; Kim, J.-J. P. Electron Transfer Partners of Cytochrome P450. In *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Ortiz de Montellano, P. R., Ed.; Springer International Publishing: Cham, 2015, pp 33-68.
92. Urlacher, V. B.; Girhard, M., Cytochrome P450 Monooxygenases in Biotechnology and Synthetic Biology. *Trends Biotechnol* 2019, 37, 882-897.
93. Newhouse, T.; Baran, P. S., If C□H Bonds Could Talk: Selective C□H Bond Oxidation. *Angew. Chem. Int. Ed.* 2011, 50, 3362-3374.
94. Lewis, J. C.; Coelho, P. S.; Arnold, F. H., Enzymatic functionalization of carbon–hydrogen bonds. *Chem. Soc. Rev.* 2011, 40, 2003-2021.
95. Girhard, M.; Bakkes, P. J.; Mahmoud, O.; Urlacher, V. B. P450 Biotechnology. In *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Ortiz de Montellano, P. R., Ed.; Springer International Publishing: Cham, 2015, pp 451-520.
96. Urlacher, V. B.; Girhard, M., Cytochrome P450 monooxygenases: an update on perspectives for synthetic application. *Trends Biotechnol* 2012, 30, 26-36.
97. Narhi, L. O.; Fulco, A. J., Characterization of a catalytically self-sufficient 119,000-dalton cytochrome P-450 monooxygenase induced by barbiturates in Bacillus megaterium. *J. Biol. Chem.* 1986, 261, 7160-7169.
98. van Vugt-Lussenburg, B. M. A.; Damsten, M. C.; Maasdijk, D. M.; Vermeulen, N. P. E.; Commandeur, J. N. M., Heterotropic and homotropic cooperativity by a drug-metabolising mutant of cytochrome P450 BM3. *Biochem. Biophys. Res. Commun.* 2006, 346, 810-818.
99. Otey, C. R.; Bandara, G.; Lalonde, J.; Takahashi, K.; Arnold, F. H., Preparation of human metabolites of propranolol using laboratory-evolved bacterial cytochromes P450. *Biotechnol. Bioeng.* 2006, 93, 494-499.
100. Otey, C. R.; Landwehr, M.; Endelman, J. B.; Hiraga, K.; Bloom, J. D.; Arnold, F. H., Structure-Guided Recombination Creates an Artificial Family of Cytochromes P450. *PLoS Biology* 2006, 4, e112.
101. Li, Y.; Drummond, D. A.; Sawayama, A. M.; Snow, C. D.; Bloom, J. D.; Arnold, F. H., A diverse family of thermostable cytochrome P450s created by recombination of stabilizing fragments. *Nat. Biotechnol.* 2007, 25, 1051-1056.
102. Fasan, R., Tuning P450 Enzymes as Oxidation Catalysts. *ACS Catal.* 2012, 2, 647-666.
103. Bell, S. G.; Stevenson, J.-A.; Boyd, H. D.; Campbell, S.; Riddle, A. D.; Orton, E. L.; Wong, L.-L., Butane and propane oxidation by engineered cytochrome P450cam. *Chem. Commun.* 2002, 490-491.
104. Poulos, T. L.; Finzel, B. C.; Gunsalus, I. C.; Wagner, G. C.; Kraut, J., The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450. *J. Biol. Chem.* 1985, 260, 16122-16130.

105. Poulos, T. L.; Finzel, B. C.; Howard, A. J., High-resolution crystal structure of cytochrome P450cam. *J. Mol. Biol.* **1987**, 195, 687-700.
106. Monk, B. C.; Tomasiak, T. M.; Keniya, M. V.; Huschmann, F. U.; Tyndall, J. D. A.; O'Connell, J. D.; Cannon, R. D.; McDonald, J. G.; Rodriguez, A.; Finer-Moore, J. S.; Stroud, R. M., Architecture of a single membrane spanning cytochrome P450 suggests constraints that orient the catalytic domain relative to a bilayer. *Proceedings of the National Academy of Sciences* **2014**, 111, 3865-3870.
107. Li, H.; Poulos, T. L., The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. *Nature Structural Biology* **1997**, 4, 140-146.
108. Fasan, R.; Chen, M. M.; Crook, N. C.; Arnold, F. H., Engineered Alkane-Hydroxylating Cytochrome P450BM3 Exhibiting Nativelike Catalytic Properties. *Angew. Chem. Int. Ed.* **2007**, 46, 8414-8418.
109. Kokorin, A.; Urlacher, V. B., Artificial Fusions between P450 BM3 and an Alcohol Dehydrogenase for Efficient (+)-Nootkatone Production. *ChemBioChem* **2022**, 23, e202200065.
110. Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F., The Structure of Human Microsomal Cytochrome P450 3A4 Determined by X-ray Crystallography to 2.05-Å Resolution*. *J. Biol. Chem.* **2004**, 279, 38091-38094.
111. Walsh, A. A.; Szklarz, G. D.; Scott, E. E., Human Cytochrome P450 1A1 Structure and Utility in Understanding Drug and Xenobiotic Metabolism*. *J. Biol. Chem.* **2013**, 288, 12932-12943.
112. Sansen, S.; Yano, J. K.; Reynald, R. L.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F., Adaptations for the Oxidation of Polycyclic Aromatic Hydrocarbons Exhibited by the Structure of Human P450 1A2*♦. *J. Biol. Chem.* **2007**, 282, 14348-14355.
113. Wang, A.; Savas, U.; Stout, C. D.; Johnson, E. F., Structural Characterization of the Complex between α-Naphthoflavone and Human Cytochrome P450 1B1*. *J. Biol. Chem.* **2011**, 286, 5736-5743.
114. Yano, J. K.; Denton, T. T.; Cerny, M. A.; Zhang, X.; Johnson, E. F.; Cashman, J. R., Synthetic Inhibitors of Cytochrome P-450 2A6: Inhibitory Activity, Difference Spectra, Mechanism of Inhibition, and Protein Cocrystallization. *J. Med. Chem.* **2006**, 49, 6987-7001.
115. Schoch, G. A.; Yano, J. K.; Wester, M. R.; Griffin, K. J.; Stout, C. D.; Johnson, E. F., Structure of Human Microsomal Cytochrome P450 2C8: EVIDENCE FOR A PERIPHERAL FATTY ACID BINDING SITE*. *J. Biol. Chem.* **2004**, 279, 9497-9503.
116. Wester, M. R.; Yano, J. K.; Schoch, G. A.; Yang, C.; Griffin, K. J.; Stout, C. D.; Johnson, E. F., The Structure of Human Cytochrome P450 2C9 Complexed with Flurbiprofen at 2.0-Å Resolution*. *J. Biol. Chem.* **2004**, 279, 35630-35637.
117. Reynald, R. L.; Sansen, S.; Stout, C. D.; Johnson, E. F., Structural Characterization of Human Cytochrome P450 2C19: ACTIVE SITE DIFFERENCES BETWEEN P450s 2C8, 2C9, AND 2C19*♦. *J. Biol. Chem.* **2012**, 287, 44581-44591.
118. Porubsky, P. R.; Meneely, K. M.; Scott, E. E., Structures of Human Cytochrome P-450 2E1: INSIGHTS INTO THE BINDING OF INHIBITORS AND BOTH SMALL MOLECULAR WEIGHT AND FATTY ACID SUBSTRATES*. *J. Biol. Chem.* **2008**, 283, 33698-33707.
119. Šrejber, M.; Navrátilová, V.; Paloncýová, M.; Bazgier, V.; Berka, K.; Anzenbacher, P.; Otyepka, M., Membrane-attached mammalian cytochromes P450: An overview of the membrane's effects on structure, drug binding, and interactions with redox partners. *J. Inorg. Biochem.* **2018**, 183, 117-136.
120. Cojocaru, V.; Winn, P. J.; Wade, R. C., The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta, Gen. Subj.* **2007**, 1770, 390-401.
121. 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, 11248-11255.
122. Hritz, J.; de Ruiter, A.; Oostenbrink, C., Impact of Plasticity and Flexibility on Docking Results for Cytochrome P450 2D6: A Combined Approach of Molecular Dynamics and Ligand Docking. *J. Med. Chem.* **2008**, 51, 7469-7477.
123. Arfeen, M.; Patel, D. S.; Abbat, S.; Taxak, N.; Bharatam, P. V., Importance of cytochromes in cyclization reactions: quantum chemical study on a model reaction of proguanil to cycloguanil. *J. Comput. Chem.* **2014**, 35, 2047-55.

124. Williams, P. A.; Cosme, J.; Vinković, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H., Crystal Structures of Human Cytochrome P450 3A4 Bound to Metyrapone and Progesterone. *Science* 2004, 305, 683.
125. Hendrychová, T.; Anzenbacherová, E.; Hudeček, J.; Skopalík, J.; Lange, R.; Hildebrandt, P.; Otyepka, M.; Anzenbacher, P., Flexibility of human cytochrome P450 enzymes: molecular dynamics and spectroscopy reveal important function-related variations. *Biochim. Biophys. Acta.* 2011, 1814, 58-68.
126. Otyepka, M.; Berka, K.; Anzenbacher, P., Is There a Relationship Between the Substrate Preferences and Structural Flexibility of Cytochromes P450? *Curr. Drug Metab.* 2012, 13, 130-142.
127. Gora, A.; Brezovsky, J.; Damborsky, J., Gates of Enzymes. *Chem. Rev.* 2013, 113, 5871-5923.
128. Kokkonen, P.; Bednar, D.; Pinto, G.; Prokop, Z.; Damborsky, J., Engineering enzyme access tunnels. *Biotechnol. Adv.* 2019, 37, 107386.
129. Kingsley, L. J.; Lill, M. A., Substrate tunnels in enzymes: Structure–function relationships and computational methodology. *Proteins* 2015, 83, 599-611.
130. Becker, D.; Bharatam, P. V.; Gohlke, H., F/G Region Rigidity is Inversely Correlated to Substrate Promiscuity of Human CYP Isoforms Involved in Metabolism. *J. Chem. Inf. Model.* 2021, 61, 4023-4030.
131. Ramesh, M.; Bharatam, P. V., Importance of hydrophobic parameters in identifying appropriate pose of CYP substrates in cytochromes. *Eur. J. Med. Chem.* 2014, 71, 15-23.
132. Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J., CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLoS Comput. Biol.* 2012, 8, e1002708.
133. McCammon, J. A.; Gelin, B. R.; Karplus, M., Dynamics of folded proteins. *Nature* 1977, 267, 585-590.
134. Hug, S. Classical Molecular Dynamics in a Nutshell. In *Biomolecular Simulations: Methods and Protocols*, Monticelli, L.; Salonen, E., Eds.; Humana Press: Totowa, NJ, 2013, pp 127-152.
135. Barth, E.; Mandziuk, M.; Schlick, T. A separating framework for increasing the timestep in molecular dynamics. In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, van Gunsteren, W. F.; Weiner, P. K.; Wilkinson, A. J., Eds.; Springer Netherlands: Dordrecht, 1997, pp 97-121.
136. Hospital, A.; Goñi, J. R.; Orozco, M.; Gelpí, J. L., Molecular dynamics simulations: advances and applications. *Advances and Applications in Bioinformatics and Chemistry* 2015, 8, 37-47.
137. Miyamoto, S.; Kollman, P. A., Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 1992, 13, 952-962.
138. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 1997, 18, 1463-1472.
139. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 1977, 23, 327-341.
140. Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E., Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* 2015, 11, 1864-1874.
141. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* 1995, 91, 1-41.
142. Brooks, B. R.; Brooks Iii, C. L.; Mackerell Jr, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The biomolecular simulation program. *J. Comput. Chem.* 2009, 30, 1545-1614.
143. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C., GROMACS: Fast, flexible, and free. *J. Comput. Chem.* 2005, 26, 1701-1718.

144. Nelson, M. T.; Humphrey, W.; Gursoy, A.; Dalke, A.; Kalé, L. V.; Skeel, R. D.; Schulten, K., NAMD: a Parallel, Object-Oriented Molecular Dynamics Program. *The International Journal of Supercomputer Applications and High Performance Computing* 1996, 10, 251-268.
145. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* 2013, 9, 3878-3888.
146. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* 2015, 11, 3696-3713.
147. Dickson, C. J.; Madej, B. D.; Skjevik, A. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C., Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* 2014, 10, 865-879.
148. Izadi, S.; Anandakrishnan, R.; Onufriev, A. V., Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* 2014, 5, 3863-3871.
149. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling* 2006, 25, 247-260.
150. Case, D. A.; Cheatham Iii, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *J. Comput. Chem.* 2005, 26, 1668-1688.
151. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* 2004, 25, 1157-1174.
152. Grossfield, A. *WHAM: the weighted histogram analysis method*, self published: Rochester, NY, 2017.
153. Case, D. A.; Aktulga, H. M.; Belfon, K.; Cerutti, D. S.; Cisneros, G. A.; Cruzeiro, V. W. D.; Forouzesh, N.; Giese, T. J.; Götz, A. W.; Gohlke, H.; Izadi, S.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kurtzman, T.; Lee, T.-S.; Li, P.; Liu, J.; Luchko, T.; Luo, R.; Manathunga, M.; Machado, M. R.; Nguyen, H. M.; O'Hearn, K. A.; Onufriev, A. V.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Risheh, A.; Schott-Verdugo, S.; Shajan, A.; Swails, J.; Wang, J.; Wei, H.; Wu, X.; Wu, Y.; Zhang, S.; Zhao, S.; Zhu, Q.; Cheatham, T. E., III; Roe, D. R.; Roitberg, A.; Simmerling, C.; York, D. M.; Nagan, M. C.; Merz, K. M., Jr., AmberTools. *J. Chem. Inf. Model.* 2023.
154. Taverna, D. M.; Goldstein, R. A., Why are proteins marginally stable? *Proteins* 2002, 46, 105-109.
155. Hermans, S. M. A.; Pfleger, C.; Nutschel, C.; Hanke, C. A.; Gohlke, H., Rigidity theory for biomolecules: concepts, software, and applications. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* 2017, 7, e1311.
156. Radestock, S.; Gohlke, H., Exploiting the Link between Protein Rigidity and Thermostability for Data-Driven Protein Engineering. *Eng. Life Sci.* 2008, 8, 507-522.
157. Radestock, S.; Gohlke, H., Protein rigidity and thermophilic adaptation. *Proteins-Structure Function and Bioinformatics* 2011, 79, 1089-1108.
158. Pfleger, C.; Rathi, P. C.; Klein, D. L.; Radestock, S.; Gohlke, H., Constraint Network Analysis (CNA): A Python Software Package for Efficiently Linking Biomacromolecular Structure, Flexibility, (Thermo-)Stability, and Function. *J. Chem. Inf. Model.* 2013, 53, 1007-1015.
159. Jacobs, D. J.; Rader, A. J.; Kuhn, L. A.; Thorpe, M. F., Protein flexibility predictions using graph theory. *Proteins* 2001, 44, 150-165.
160. Jacobs, D. J.; Thorpe, M. F., Generic Rigidity Percolation: The Pebble Game. *Phys. Rev. Lett.* 1995, 75, 4051-4054.
161. Pfleger, C.; Radestock, S.; Schmidt, E.; Gohlke, H., Global and local indices for characterizing biomolecular flexibility and rigidity. *J. Comput. Chem.* 2013, 34, 220-233.
162. Pfleger, C.; Minges, A.; Boehm, M.; McClendon, C. L.; Torella, R.; Gohlke, H., Ensemble- and Rigidity Theory-Based Perturbation Approach To Analyze Dynamic Allostery. *J. Chem. Theory Comput.* 2017, 13, 6343-6357.
163. Bornscheuer, U. T.; Kazlauskas, R. J. Enzymatic Catalytic Promiscuity and the Design of New Enzyme Catalyzed Reactions. In *Enzyme Catalysis in Organic Synthesis*; 2012, pp 1693-1733.

164. Nutschel, C.; Coscolín, C.; David, B.; Mulnaes, D.; Ferrer, M.; Jaeger, K.-E.; Gohlke, H., Promiscuous Esterases Counterintuitively Are Less Flexible than Specific Ones. *J. Chem. Inf. Model.* 2021, 61, 2383-2395.
165. Hawkins, P. C.; Skillman, A. G.; Nicholls, A., Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* 2007, 50, 74-82.
166. Foti, R. S.; Honaker, M.; Nath, A.; Pearson, J. T.; Buttrick, B.; Isoherranen, N.; Atkins, W. M., Catalytic versus Inhibitory Promiscuity in Cytochrome P450s: Implications for Evolution of New Function. *Biochemistry* 2011, 50, 2387-2393.
167. Sterling, T.; Irwin, J. J., ZINC 15 – Ligand Discovery for Everyone. *J. Chem. Inf. Model.* 2015, 55, 2324-2337.
168. Nath, A.; Atkins, W. M., A Quantitative Index of Substrate Promiscuity. *Biochemistry* 2008, 47, 157-166.
169. Mills, A. L.; Wassel, R. A., Aspects of Diversity Measurement for Microbial Communities. *Appl. Environ. Microbiol.* 1980, 40, 578-586.
170. Buckland, S. T.; Magurran, A. E.; Green, R. E.; Fewster, R. M., Monitoring change in biodiversity through composite indices. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2005, 360, 243-254.
171. Skopalík, J.; Anzenbacher, P.; Otyepka, M., Flexibility of Human Cytochromes P450: Molecular Dynamics Reveals Differences between CYPs 3A4, 2C9, and 2A6, which Correlate with Their Substrate Preferences. *J. Phys. Chem. B* 2008, 112, 8165-8173.
172. Eisenmesser, E. Z.; Bosco, D. A.; Akke, M.; Kern, D., Enzyme Dynamics During Catalysis. *Science* 2002, 295, 1520.
173. Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D., A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* 2007, 450, 913-916.
174. Holliday, M. J.; Camilloni, C.; Armstrong, G. S.; Vendruscolo, M.; Eisenmesser, E. Z., Networks of Dynamic Allostery Regulate Enzyme Function. *Structure* 2017, 25, 276-286.
175. Goodey, N. M.; Benkovic, S. J., Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.* 2008, 4, 474-482.
176. Saavedra, H. G.; Wrabl, J. O.; Anderson, J. A.; Li, J.; Hilser, V. J., Dynamic allostery can drive cold adaptation in enzymes. *Nature* 2018, 558, 324-328.
177. Tzeng, S.-R.; Kalodimos, C. G., Protein dynamics and allostery: an NMR view. *Curr. Opin. Struct. Biol.* 2011, 21, 62-67.
178. Nobeli, I.; Favia, A. D.; Thornton, J. M., Protein promiscuity and its implications for biotechnology. *Nat. Biotechnol.* 2009, 27, 157-167.
179. Dick, M.; Weiergraber, O. H.; Classen, T.; Bisterfeld, C.; Bramski, J.; Gohlke, H.; Pietruszka, J., Trading off stability against activity in extremophilic aldolases. *Sci Rep* 2016, 6, 17908.
180. Merz, A.; Yee, M.-c.; Szadkowski, H.; Pappenberger, G.; Crameri, A.; Stemmer, W. P. C.; Yanofsky, C.; Kirschner, K., Improving the Catalytic Activity of a Thermophilic Enzyme at Low Temperatures. *Biochemistry* 2000, 39, 880-889.
181. Fuxreiter, M.; Magyar, C.; Juhász, T.; Szeltner, Z.; Polgár, L.; Simon, I., Flexibility of prolyl oligopeptidase: Molecular dynamics and molecular framework analysis of the potential substrate pathways. *Proteins* 2005, 60, 504-512.
182. Rathi, P. C.; Fulton, A.; Jaeger, K.-E.; Gohlke, H., Application of Rigidity Theory to the Thermostabilization of Lipase A from Bacillus subtilis. *PLoS Comput. Biol.* 2016, 12, e1004754.
183. Mahgoub, A.; Dring, L. G.; Idle, J. R.; Lancaster, R.; Smith, R. L., POLYMORPHIC HYDROXYLATION OF DEBRISOQUINE IN MAN. *The Lancet* 1977, 310, 584-586.
184. Kimura, S.; Umeno, M.; Skoda, R. C.; Meyer, U. A.; Gonzalez, F. J., The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and identification of the polymorphic CYP2D6 gene, a related gene, and a pseudogene. *Am. J. Hum. Genet.* 1989, 45, 889-904.
185. Dahl, M. L.; Johansson, I.; Bertilsson, L.; Ingelman-Sundberg, M.; Sjöqvist, F., Ultrarapid hydroxylation of debrisoquine in a Swedish population. Analysis of the molecular genetic basis. *J. Pharmacol. Exp. Ther.* 1995, 274, 516-20.

186. Løvlie, R.; Daly, A. K.; Molven, A.; Idle, J. R.; Steen, V. M., Ultrarapid metabolizers of debrisoquine: Characterization and PCR-based detection of alleles with duplication of the CYP2D6 gene. *FEBS Lett.* 1996, 392, 30-34.
187. Johansson, I.; Lundqvist, E.; Bertilsson, L.; Dahl, M. L.; Sjöqvist, F.; Ingelman-Sundberg, M., Inherited amplification of an active gene in the cytochrome P450 CYP2D locus as a cause of ultrarapid metabolism of debrisoquine. *Proceedings of the National Academy of Sciences* 1993, 90, 11825-11829.
188. Chiba, K.; Kato, M.; Ito, T.; Suwa, T.; Sugiyama, Y., Inter-individual variability of in vivo CYP2D6 activity in different genotypes. *Drug Metabolism and Pharmacokinetics* 2012, advpub, 1201160353-1201160353.
189. Yu, A.; Kneller, B. M.; Rettie, A. E.; Haining, R. L., 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.
190. Dai, D.-P.; Geng, P.-W.; Wang, S.-H.; Cai, J.; Hu, L.-M.; Nie, J.-J.; Hu, J.-H.; Hu, G.-X.; Cai, J.-P., In vitro functional assessment of 22 newly identified CYP2D6 allelic variants in the Chinese population. *Basic Clin Pharmacol Toxicol* 2015, 117, 39-43.
191. de Groot, M. J.; Wakenhut, F.; Whitlock, G.; Hyland, R., Understanding CYP2D6 interactions. *Drug Discov. Today* 2009, 14, 964-972.
192. Huff, H. C.; Vasan, A.; Roy, P.; Kaul, A.; Tajkhorshid, E.; Das, A., Differential Interactions of Selected Phytocannabinoids with Human CYP2D6 Polymorphisms. *Biochemistry* 2021, 60, 2749-2760.
193. Zanger, U. M.; Fischer, J.; Raimundo, S.; Stüven, T.; Evert, B. O.; Schwab, M.; Eichelbaum, M., Comprehensive analysis of the genetic factors determining expression and function of hepatic CYP2D6. *Pharmacogenetics and Genomics* 2001, 11.
194. Flanagan, J. U.; MarÉChal, J.-D.; Ward, R.; Kemp, C. A.; McLaughlin, L. A.; Sutcliffe, M. J.; Roberts, G. C. K.; Paine, M. J. I.; Wolf, C. R., 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.
195. Keizers, P. H. J.; Lussenburg, B. M. A.; de Graaf, C.; Mentink, L. M.; Vermeulen, N. P. E.; Commandeur, J. N. M., Influence of phenylalanine 120 on cytochrome P450 2D6 catalytic selectivity and regiospecificity: crucial role in 7-methoxy-4-(aminomethyl)-coumarin metabolism. *Biochem. Pharmacol.* 2004, 68, 2263-2271.
196. Lussenburg, B. M. A.; Keizers, P. H. J.; de Graaf, C.; Hidestrand, M.; Ingelman-Sundberg, M.; Vermeulen, N. P. E.; Commandeur, J. N. M., The role of phenylalanine 483 in cytochrome P450 2D6 is strongly substrate dependent. *Biochem. Pharmacol.* 2005, 70, 1253-1261.
197. Prade, E.; Mahajan, M.; Im, S.-C.; Zhang, M.; Gentry, K. A.; Anantharamaiah, G. M.; Waskell, L.; Ramamoorthy, A., A Minimal Functional Complex of Cytochrome P450 and FBD of Cytochrome P450 Reductase in Nanodiscs. *Angew. Chem. Int. Ed.* 2018, 57, 8458-8462.
198. UniProt, C., UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 2019, 47, D506-D515.
199. Sherry, S. T.; Ward, M. H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E. M.; Sirotkin, K., dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001, 29, 308-11.
200. Landrum, M. J.; Chitipiralla, S.; Brown, G. R.; Chen, C.; Gu, B.; Hart, J.; Hoffman, D.; Jang, W.; Kaur, K.; Liu, C.; Lyoshin, V.; Maddipatla, Z.; Maiti, R.; Mitchell, J.; O'Leary, N.; Riley, G. R.; Shi, W.; Zhou, G.; Schneider, V.; Maglott, D.; Holmes, J. B.; Kattman, B. L., ClinVar: improvements to accessing data. *Nucleic Acids Res.* 2020, 48, D835-D844.
201. HGMD (human gene mutation database). In *Encyclopedia of Genetics, Genomics, Proteomics and Informatics*, Rédei, G. P., Ed.; Springer Netherlands: Dordrecht, 2008, pp 873-873.
202. Evert, B.; Griese, E.-U.; Eichelbaum, M., A missense mutation in exon 6 of the CYP2D6 gene leading to a histidine 324 to proline exchange is associated with the poor metabolizer phenotype of sparteine. *Naunyn-Schmiedeberg's Arch. Pharmacol.* 1994, 350, 434-439.
203. van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A. M. J. J., The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* 2016, 428, 720-725.

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

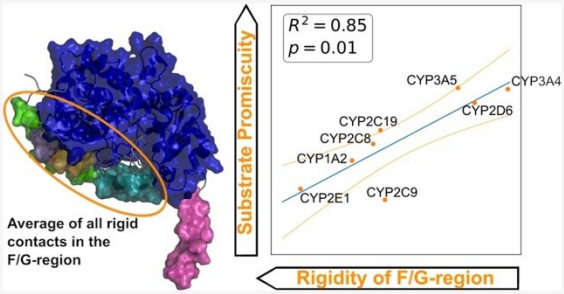
# F/G Region Rigidity is Inversely Correlated to Substrate Promiscuity of Human CYP Isoforms Involved in Metabolism

Daniel Becker, Prasad V. Bharatam, and Holger Gohlke*

Cite This: J. Chem. Inf. Model. 2021, 61, 4023−4030 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** Of 57 human cytochrome P450 (CYP) enzymes, 12 metabolize 90% of xenobiotics. To our knowledge, no study has addressed the relation between enzyme dynamics and substrate promiscuity for more than three CYPs. Here, we show by constraint dilution simulations with the Constraint Network Analysis for the 12 isoforms that structural rigidity of the F/G region is significantly inversely correlated to the enzymes' substrate promiscuity. This highlights the functional importance of structural dynamics of the substrate tunnel.




## INTRODUCTION

Cytochrome P450 (CYP) enzymes play a central role in human drug metabolism.[1,2] They primarily act as monooxygenases, and the most essential catalyzed reaction is the hydroxylation of nonactivated C−H bonds.[3,4] This makes drugs and other xenobiotics more hydrophilic, which facilitates excretion, but also can inactivate drugs or, vice versa, lead to biologically active metabolites.[5] All human CYP enzymes share three common properties. First, CYPs are heme proteins with a similar fold. Second, the active site containing the heme is buried and only accessible via channels. Finally, human CYPs are generally associated with the membrane, usually the membrane of the endoplasmic reticulum, by an N-terminal anchor helix.[6,7] In total, 57 human CYP enzymes are known, but not all are involved in metabolism.[8,9] For some other CYP enzymes, the function is yet unknown.[10,11]

Interestingly, only 12 CYP isoforms metabolize about 90% of all xenobiotics.[1] These are CYP1A1, CYP1A2, CYP1B1, CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP3A4, and CYP3A5. Although all are substrate-promiscuous enzymes, they show recognizable differences, e.g., CYP3A4 contributes the most to xenobiotics metabolism by being involved in 20% of all known CYP transformations.[1] On the other hand, CYP1B1 contributes to only 3% of all known CYP transformations.[1] However, the sheer number of xenobiotics metabolism pathways a CYP contributes to is not the best indicator for its substrate scope. Substrate scope as a functional enzyme parameter should also take the enzyme's catalytic efficiency into account, e.g., although CYP3A5 generally catalyzes the same reactions as CYP3A4, it does so often at lower rates, such that CYP3A5 only contributes to 4% of all known CYP transformations.[1] Earlier, a molecular docking-based analysis pointed out that the degree of promiscuity may be partly determined by the amino acid residues in the dome region of the CYPs (CYP3A4—hydrophobicity-dominated, CYP2D6—acidic-dominated, CYP2C9—basicity-dominated residues).[12,13]

Previous work based on crystal structures and computational studies stressed the importance of the F/G region, consisting of the F, F′, G′, and G helices, for substrate binding[14−16] because the region is at the beginning of a tunnel that is considered the substrate entrance channel.[7] Such tunnels can act as filters and have been found to influence both substrate specificity and catalytic mechanism.[17−19] Accordingly, one study hypothesized for the three isoforms CYP3A4, CYP2C9, and CYP2A6 that higher mobility, deduced from movements of the $C_\alpha$ atoms, in the F/G region correlates with higher substrate promiscuity.[20] Enzyme structural dynamics, in addition to its role in catalysis[21,22] and allosteric regulation,[23−26] has also been recognized for other systems as an important mechanism by which promiscuity can be achieved.[27] Surprisingly, to our knowledge, no further studies addressed the relation between enzyme dynamics and promiscuity of more than three CYP isoforms comparatively. Furthermore, the approaches did not provide a quantitative relation between substrate scope and structural rigidity of CYP isoforms involved in metabolism. Here, we show for the, to our

Received: May 18, 2021
Published: August 9, 2021





https://doi.org/10.1021/acs.jcim.1c00558
J. Chem. Inf. Model. 2021, 61, 4023−4030

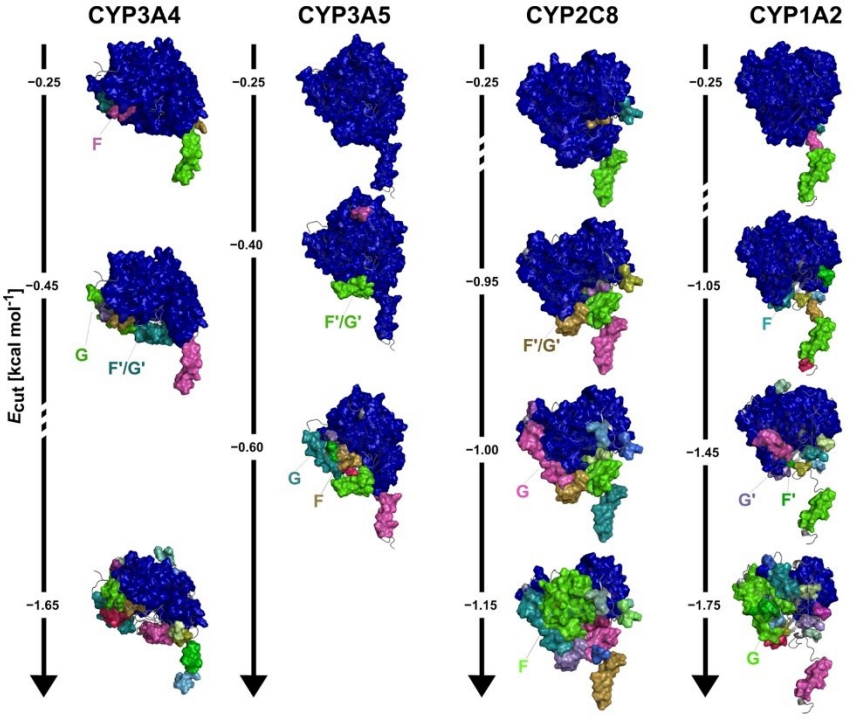



**Figure 1.** Constraint network analysis of CYP3A4, CYP3A5, CYP2C8, and CYP1A2 and rigid cluster decompositions along constraint dilution trajectories. Rigid clusters of the CYP isoforms at different $\overline{E}_{cut}$ values are colored blue, green, pink, and cyan in descending order of their size. Left: In CYP3A4, the F helix forms its own rigid cluster at $\overline{E}_{cut}$ = −0.25 kcal mol$^{-1}$ and G helix and F′/G′ helix form their own rigid clusters at $\overline{E}_{cut}$ = −0.45 kcal mol$^{-1}$. At $\overline{E}_{cut}$ = 1.65 kcal mol$^{-1}$, the largest rigid cluster of CYP3A4 is still covering most of the globular part. Middle-left: In CYP3A5, the F′ helix and G′ helix form their own rigid clusters at $\overline{E}_{cut}$ = −0.40 kcal mol$^{-1}$, the F helix and the G helix form their own cluster at $\overline{E}_{cut}$ = −0.60 kcal mol$^{-1}$, and the largest rigid cluster of CYP3A5 is still covering most of the globular part. Middle-right: In CYP2C8, the F′ helix and G′ helix form their own rigid cluster at $\overline{E}_{cut}$ = −0.95 kcal mol$^{-1}$. The G helix forms its own cluster at $\overline{E}_{cut}$ = −1.00 kcal mol$^{-1}$. At $\overline{E}_{cut}$ = 1.15 kcal mol$^{-1}$, the F helix forms its own cluster, and the largest rigid cluster of CYP2C8 is covering only the active site in the globular part. Complete segregation of the F/G region is not recognizable for CYP2C8. Right: In CYP1A2, the F helix forms its own rigid cluster at $\overline{E}_{cut}$ = −1.05 kcal mol$^{-1}$. The F′ helix and G′ helix form their own cluster at $\overline{E}_{cut}$ = −1.45 kcal mol$^{-1}$. At $\overline{E}_{cut}$ = 1.75 kcal mol$^{-1}$, the G helix forms its own cluster, and the largest rigid cluster of CYP1A2 is covering only the active site in the globular part. These examples qualitatively depict that the F/G region of more promiscuous CYP isoforms, such as CYP3A4 and CYP3A5, are less structurally stable than that of more specific isoforms. Furthermore, differences in structural stability are revealed for sequentially close isoforms. The rigid clusters are calculated based on the neighbor stability map of one exemplary MD trajectory, respectively, as described in ref 55.

knowledge, largest data set of CYP isoforms investigated in this context that the structural rigidity of the F/G region is inversely correlated to the enzymes' substrate scope.

## ■ MATERIALS AND METHODS

**Comparative Modeling of 12 CYP Isoforms.** Comparative models of the investigated CYP isoforms were generated with TopModel using the Protect Templates mode.[28] Specified templates that were not removed during threading are listed in Table S1. The globular part and the transmembrane helix were modeled separately, and their positions on or in the membrane were predicted with CCTop web server.[29,30] Afterward, they were docked together using the predicted membrane positions as a spatial restraint. Target sequences were collected from UniProt (Table S1).[31] The quality of the models was assessed with TopScoreSingle (for details, see Text S1).[32] Protonation states of protein residues were adjusted according to pH 7.4 using the Epik routine[33,34] in Maestro.[35] The heme group was transferred from crystal structures that were specified for the TopModel run by aligning the model in Pymol and copying the heme into the model (Table S1).

**Conformational Sampling.** To improve the robustness of the analyses and quantify the statistical uncertainty of the results, we carried out CNA on ensembles of network topologies generated from five MD trajectories of 1 $\mu$s length for each of the enzymes. For this, the generated structural models were embedded by PACKMOL-Memgen[36] into a membrane with a composition of CHL:DOPC:DSPC:DAPC:-DOPE 10:22:13:19:21, the main lipid components of the human endoplasmic reticulum.[37] The GPU particle mesh Ewald implementation[38] of the AMBER18[39] molecular simulations suite was used with ff14SB parameters[40] for the protein, Lipid17 parameters[41] for the membrane, and OPC as a water model.[42] Parameters for the heme and cysteine residues forming the S−Fe bridge between heme and protein were taken from Shahrokh et al.[43] Because covalent bonds to hydrogens were constrained with the SHAKE algorithm,[44] a time step of 2 fs was used. The cutoff for nonbonded interactions was set to 10 Å. For further details, see Text S2.[45] The structures overall (without the transmembrane helix, as it



https://doi.org/10.1021/acs.jcim.1c00558
J. Chem. Inf. Model. 2021, 61, 4023−4030

moves relative to the globular part) and the F/G regions specifically remain stable with backbone RMSD < 4 Å compared to the starting structures (Figures S1−S4).

**Constraint Network Analysis.** We analyzed static properties, i.e., structural rigidity and its opposite flexibility,[46] of the 12 CYP isoforms predominantly involved in metabolism. The enzymes were represented as constraint networks, where atoms are the nodes and covalent and noncovalent bonds constitute constraints in between.[47] Noncovalent interactions such as hydrogen bonds, salt bridges, hydrophobic tethers, and stacking interactions contribute most to biomolecular stability. The strength of hydrogen bonds and salt bridges was quantified with an empirical energy function.[48] By gradually removing these polar noncovalent constraints from an initial network representation of a biomolecule according to a cutoff energy $E_{cut}$, a succession of network states $\sigma$ is generated that forms a "constraint dilution trajectory".[49,50] For this, hydrogen bonds and salt bridges are removed in the order of increasing strength such that for a network state $\sigma$, only those hydrogen bonds are kept that have an energy $E_{HB} \leq E_{cut}(\sigma)$. Performing rigidity analysis[51] on such a trajectory reveals a hierarchy of structural stability that reflects the modular structure of biomolecules in terms of secondary, tertiary, and supertertiary structure.

Using a per-residue decomposition scheme to identify the extent to which single residues contribute to the structural stability, we derived neighbor stability maps $(rc_{ij,neighbor}(E_{cut}(\sigma)))$ that contain information accumulated over all network states $\sigma$ along the trajectory[52,53] in that they monitor the persistence of rigid contacts for pairs of residues during a constraint dilution process. In the neighbor stability map, for all residue pairs, the $E_{cut}$ value is given at which the rigid contact between two residues is lost, i.e., when these two residues stop belonging to the same rigid cluster.

The dilution process was applied to each protein conformation of a trajectory separately, and the results were averaged to obtain one neighbor stability map per trajectory. In this work, we consider one trajectory as one independent experiment. A chemical potential energy $E_{i,CNA}$ of residue $i$ is then obtained by summation over all rigid contacts (eq S1). Furthermore, we calculated $\overline{E}_{region,CNA}$ as an average over the $\overline{E}_{i,CNA}$ of all residues of a specific region in the protein or even the entire protein. (eq S2). $\overline{E}_{region,CNA}$ was calculated for each MD trajectory separately and then averaged over all replicas. The computations were done with the Constraint Network Analysis (CNA) program (version 4.0) developed by us,[50] which has been applied in the context of protein thermostability,[54−56] allosteric signaling,[53,57] and substitution influences on the function[58,59] before.

**Statistical Analysis and Fitting.** Curve fitting was performed with the SciPy module stats.[60] The function linregress calculates linear least-squares regression of two sets of measurement and applies a Wald test whose null hypothesis is that the slope is zero.

## ■ RESULTS

Exemplarily, constraint dilution trajectories of CYP3A4, CYP3A5, CYP2C8, and CYP1A2 are shown in Figure 1. Considering that segregation of substructural parts at lower $\overline{E}_{cut}$ values indicates that these parts are more weakly coupled to the remainder of the CYP structure, a qualitative ranking of the CYP isoforms is revealed: In CYP3A4, the F/G region is the first part of the globular domain to segregate at $\overline{E}_{cut}$ = −0.25 to −0.45 kcal mol$^{-1}$ from the largest rigid cluster. In contrast to all other helices of the globular domain, which show higher $\overline{E}_{i,CNA}$ values compared to the TM helix, the F/G region shows $\overline{E}_{i,CNA}$ values comparable to the TM helix, which does not interact with the globular domain (Figure 2).

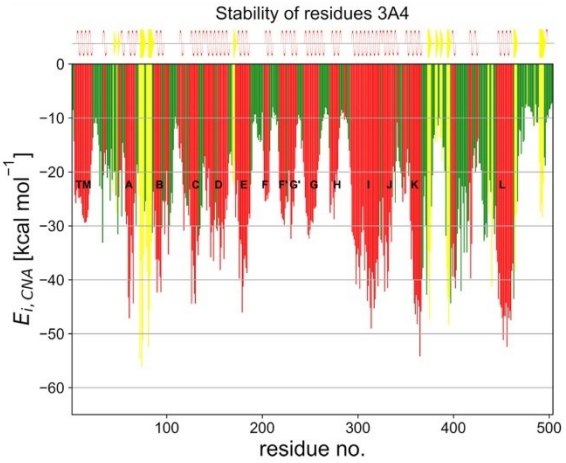



**Figure 2.** Residue-wise potential chemical energy $\overline{E}_{i,CNA}$ of CYP3A4. $\overline{E}_{i,CNA}$ (eq S1) due to all rigid contacts in which a residue is involved. Secondary structure elements identified with DSSP[80] are colored in red for helices, yellow for $\beta$-sheets, and green for loops and given for helices and $\beta$-sheets above the plot. Helices are named according to Williams et al.[14]

This is in accordance with findings for different enzymes where helices at the substrate entrances are also less stable, e.g., in aldolases,[61] phosphate synthases,[62] and peptidases,[63] as well as with previous work on CYP enzymes.[20] In CYP3A5, the segregation appears at slightly lower $\overline{E}_{cut}$ values: the F/G region is the first part of the globular domain to segregate at $\overline{E}_{cut}$ = −0.40 to −0.60 kcal mol$^{-1}$ from the largest rigid cluster. For CYP2C8, segregation of the F/G region occurs in the range of −0.95 to −1.15 kcal mol$^{-1}$ and for CYP1A2 at −1.05 to −1.75 kcal mol$^{-1}$. Hence, the F/G region is structurally the least stable in CYP3A4 and the most in CYP1A2. It is unknown, however, why within the F/G region sometimes the F, G helices and sometimes the F′, G′ helices segregate first. Finally, the rigidity analysis reveals that the anchor helix segregates at the smallest $\overline{E}_{cut}$ in all four cases. This indicates that it is only weakly coupled to the globular part, in line with observations from our MD simulations and ref 64 that the globular part can move relative to the anchor helix (Table S2).

Differences in the structural stability of the F/G region do not only occur between sequentially different CYP isoforms (Figure 1) but also for sequentially close ones as exemplarily depicted for CYP3A4 and CYP3A5 (sequence identity (similarity) 84% (90%) (Figures 1 and 3A)). There, sequence differences particularly involving nonconservative substitutions also occur in the F/G region and lead to higher $\overline{E}_{i,CNA}$ values for the F helix and the G helix of CYP3A5 (Figure 3A,B). As a result, the entire F/G region in CYP3A5 is significantly ($p$ = 0.05, standard independent two-sample $t$-test) more strongly connected to the remainder of the globular domain than in CYP3A4 (Table S3 and Figure 3C). For the F′/G′ region, no such difference is found, indicating that the small sequence differences there (Figure 3A), including exchanges of



https://doi.org/10.1021/acs.jcim.1c00558
J. Chem. Inf. Model. 2021, 61, 4023−4030

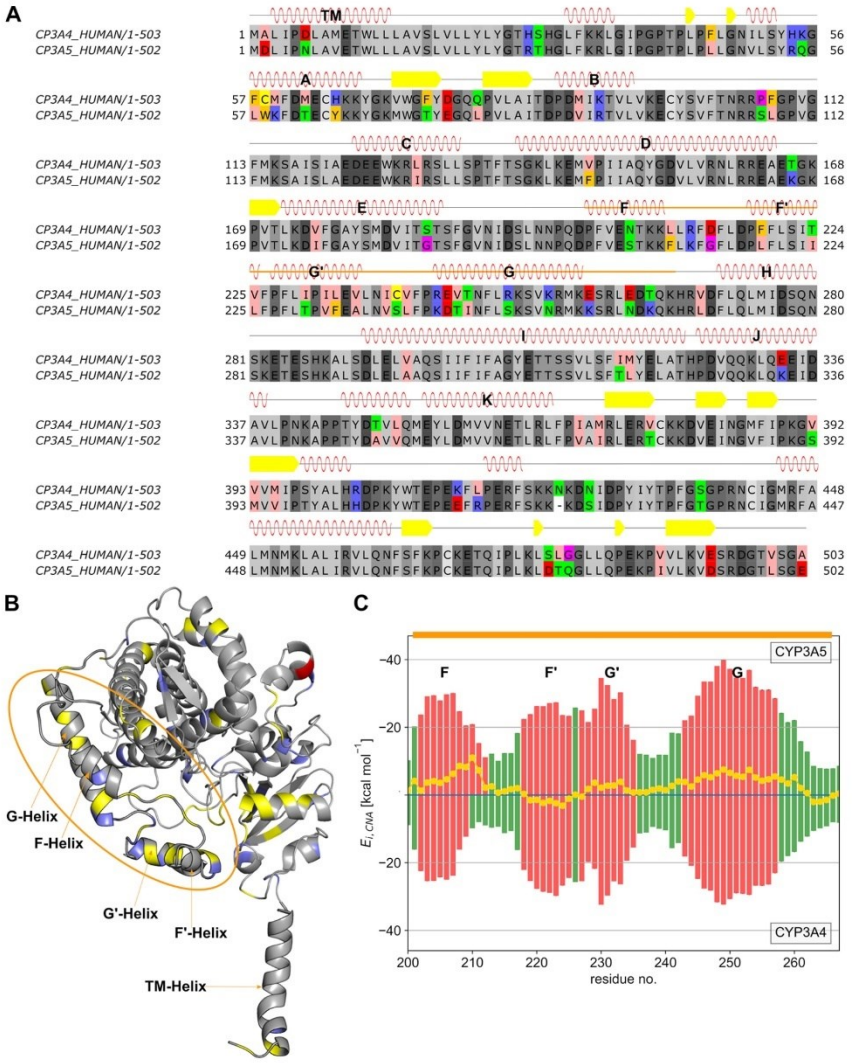



**Figure 3.** Comparison of CYP3A4 and CYP3A5 with respect to sequence differences. In all panels, the F/G region is marked in orange. (A) Sequence alignment of CYP3A4 and CYP3A5; conserved residues are in gray and nonconserved residues are colored according to the zappo color scheme as to the chemical properties of the side chains. (B) Projection of sequence differences from (A) onto the structure of CYP3A4. Gray: same residue type, blue: same chemical properties, yellow: different chemical properties, and red: gap in CYP3A5. Helices are named according to Williams et al.[14] (C) $\overline{E}_{i,CNA}$ (eq S1) of both isoforms in the F/G region (residue 201−267). More negative values indicate a higher rigidity. The yellow line depicts the difference between both isoforms. Differences in the sequences of the F and G helices cause an overall stronger connection to the remainder of the globular domain of CYP3A5, as indicated by the higher $\overline{E}_{i,CNA}$ values. Differences in the sequences of the F′ and G′ helices do not lead to differences in $\overline{E}_{i,CNA}$. The SEM is <0.05 kcal mol$^{-1}$ in all cases and not depicted. In (A) and (C), secondary structure elements identified with DSSP[80] are colored in red for helices, yellow for $\beta$-sheets, and green for loops.

amino acids with those of similar chemical properties, do not lead to differences in the rigidity of this region.

To conclude, in all CYP isoforms exemplarily shown, the F/G region is most weakly coupled to the remainder of the globular domain. Between both sequentially different isoforms and sequentially close ones, differences in the structural stability of this region are revealed by rigidity analysis. These differences qualitatively relate to the substrate scope of the isoforms: more promiscuous isoforms such as CYP3A4 (Table S3) show less structurally stable F/G regions and vice versa.

To quantify the relation between substrate scope and structural rigidity, we computed $\overline{E}_{FG,CNA}$ (eq S2) as a measure for how well residues in the F/G region form rigid contacts (Table S3) and correlated it to the promiscuity index $I_{cat}$ (eq S3) introduced by Nath and Atkins (Table S4).[65] $I_{cat}$ is based on experimentally determined $k_{cat}$ and $K_M$ values for 55 substrates and considers the catalytic efficiency $\overline{e}_i = k_{cat,I}/K_{M,i}$ of an isoform with respect to a substrate $i$. In this approach, information entropy is used to describe how an enzyme's catalytic efficiency evolved toward different substrates, as measured by the probability $p_i$ that a substrate $i$ will be



https://doi.org/10.1021/acs.jcim.1c00558
J. Chem. Inf. Model. 2021, 61, 4023−4030

metabolized by the enzyme ($p_i = e_i / \sum_{i=1}^{N} e_i$).[65] The index is normalized and ranges from 0 (specific) to 1 (promiscuous).[66]

A good and significant inverse correlation ($R^2 = 0.85$ and $p < 0.01$, Wald test) between the isoform promiscuity and the structural stability of the F/G region is obtained for eight CYP isoforms for which $k_{cat}$ and $K_M$ values are available in ref 66 (Figure 4). Notably, no significant correlations are found if

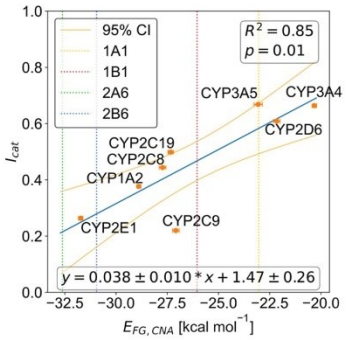



**Figure 4.** Promiscuity of CYP isoforms is inversely correlated to the structural rigidity of their F/G regions. Correlation between promiscuity index ($I_{cat}$) and $\overline{E}_{FG,CNA}$ (eq S2) for CYP1A2, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP3A4, and CYP3A5. As to CYP2C9, considered an outlier, the globular part moved away from the membrane during the MD simulations. Vertical lines indicate calculated $\overline{E}_{FG,CNA}$ values for CYP isoforms for which no experimental data for computing $I_{cat}$ are available. The 95% confidence interval limits are shown in orange. Error bars denote the SEM obtained from five, respectively, ten for CYP2E1 and CYP3A4, per-trajectory results. The equation of the predictive line is given at the bottom of the diagram.

$\overline{E}_{region,CNA}$ (eq S2) of the whole enzymes (Figure S7 and Table S5) or of other regions (the I helix, which is the longest helix in the center of CYP enzymes (Figure S8 and Table S6), or the B/C region, which is also part of the main entrance channel (Figure S9 and Table S7)) are used.

Using the correlation for the F/G region, $\overline{E}_{FG,CNA}$ values computed for CYP1A1, CYP1B1, CYP2A6, and CYP2B6 relate to predicted $I_{cat}$ values of 0.58, 0.47, 0.22, and 0.28, respectively, which classifies these CYP isoforms as being lowly to moderately promiscuous. CYP3A5 generally catalyzes the same reactions as CYP3A4 but almost always at lower rates (Figure S12), which leads to similar $I_{cat}$ values that are associated with a slightly lower $\overline{E}_{FG,CNA}$ of CYP3A5 than CYP3A4 (Figure 4). CYP2C9 deviates the most from the correlation line (Figure 4), which may be caused by the globular part that moved away from the membrane during the MD simulations. This is probably due to strain in the loop in the starting structure that connects the globular and transmembrane domains.

## ■ DISCUSSION

Our results demonstrate that the structural rigidity of the F/G region quantitatively and inversely correlates to the promiscuity of CYP isoforms involved in metabolism. Previously, enzyme structural dynamics has been recognized as an important mechanism by which promiscuity can be achieved,[27] although the size and architecture of the active site may be further determinants of substrate promiscuity.[67] In addition to CYP enzymes, the possibility of dynamically restructuring active sites has also been recognized for other systems underlying their promiscuity.[68−71] Still, examples for the opposite, i.e., conformational changes selected in evolution such that they enhance specificity in molecular recognition, have also been described in ref 72.

We exploited comprehensive experimental information on the substrate promiscuity of CYP isoforms[66] together with computationally efficient rigidity analyses[46,50,73] of comparative models of the isoforms to understand the molecular origin of the observed promiscuity range. Although $I_{cat}$ used here is a functional parameter defined for a specified set of substrates, promiscuity indices for different enzymes are quantitatively comparable if they have been calculated using the same substrate set.[65] Furthermore, substrates that are chemically similar to each other are expected to be metabolized similarly by a CYP isoform; such correlations in the substrate set would reduce the effective CYP promiscuity. Therefore, we computed the mean maximum pairwise Tanimoto−Combo distance score $\delta_i$ of a substrate $i$ to all other substrates in the data set, which ranges from 0 for identical substrates to 2 for dissimilar ones (for details see Text S6); the Tanimoto−Combo distance score accounts for shape and chemical complementary between 3D structures as determined by the Rapid Overlay of Chemical Structures approach. The negatively skewed histogram of $\delta_i$ peaks at 1.20, with an average value $\langle\delta\rangle$ of 1.26 (Figure S14), indicating that a substrate is generally more dissimilar than similar to all others in the data set. For comparison, for randomly drawn compounds from the ZINC database, $\langle\delta\rangle$ peaks at 1.12, indicating that such compounds are more similar to each other than those in our data set (Figure S15).

For CYP enzymes, crystallographic studies and other molecular simulations also demonstrated that more promiscuous CYPs show larger structural plasticity and mobility.[74] However, in these studies, only a few isoforms were compared,[15,20,75] no quantitative relations were derived,[15] or short and likely nonconverged MD simulations were applied.[76] Indirectly, the role of structural plasticity and mobility was also investigated in studies focussing on substrate channels.[76] These studies showed qualitative differences in channel properties of CYP isoforms[77−79] but did not take all 12 isoforms involved in metabolism into account and did not derive a quantitative model.

Our quantitative relation allows predicting the promiscuity for CYP1A1, CYP1B1, CYP2A6, and CYP2B6. The difference in $\overline{E}_{FG,CNA}$ between CYP1A1 and CYP1A2 is larger than that observed for other isoforms of the same subfamily such as CYP3A4 and CYP3A5, but still in the range found for CYP2C8 and CYP2C19. Apparently, isoforms of the same subfamily can differ markedly in their substrate scopes, and our approach is able to detect that. However, the model does not allow predicting if a substrate is metabolized by a specific CYP.

In summary, our results signify that characterizing the structural rigidity of the F/G region can be used to classify CYP isoforms involved in metabolism with respect to their substrate scope. Our model may allow predicting the substrate promiscuity of novel CYP enzymes, e.g., found in metagenome approaches for potential use in bioorganic chemistry or biotechnology.



https://doi.org/10.1021/acs.jcim.1c00558
J. Chem. Inf. Model. 2021, 61, 4023−4030

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.1c00558.

Supporting materials and methods, supporting tables, supporting figures, and supporting references (PDF) A repository with used scripts, MD simulation data, and CNA results are available from researchdata.hhu.de via DOI: 10.25838/d5p-21 (TAR.GZ).

## ■ AUTHOR INFORMATION


### Corresponding Author

**Holger Gohlke** − *Mathematisch-Naturwissenschaftliche Fakultät, Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany; John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre (JSC), Institute of Biological Information Processing (IBI-7: Structural Biochemistry), and Institute of Bio- and Geosciences (IBG-4: Bioinformatics), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany;* orcid.org/0000-0001-8613-1447; Phone: (+49) 211 81 13662; Email: gohlke@uni-duesseldorf.de; Fax: (+49) 211 81 13847

### Authors

**Daniel Becker** − *Mathematisch-Naturwissenschaftliche Fakultät, Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany*

**Prasad V. Bharatam** − *Department of Medicinal Chemistry, National Institute of Pharmaceutical Education and Research (NIPER), Mohali 160062 Punjab, India*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.1c00558


### Author Contributions

H.G. designed the study; D.B. performed computations; D.B. and H.G. analyzed results; D.B. and H.G. wrote the manuscript; P.V.B. revised the manuscript; and H.G. and P.V.B. secured funding.

### Notes

The authors declare no competing financial interest.

All MD input structures, MD logfiles, CNA results, and scripts used to analyze CNA results are provided in the supporting repository available at researchdata.hhu.de/DOI: 10.25838/d5p-21. Further data sets generated and analyzed during the current study are available from the corresponding author on reasonable request. For molecular simulations, the AMBER18 package of molecular simulation codes was used. AMBER18 is available from http://ambermd.org/. For rigidity analysis, the in-house CNA software package v4.0 was used. CNA is available from the corresponding author free of charge for non-for-profit organizations. Furthermore, a web server is available at https://cpclab.uni-duesseldorf.de/cna/.

## ■ ACKNOWLEDGMENTS

The authors are grateful for computational support and infrastructure provided by the "Zentrum für Informations- und Medientechnologie" (ZIM) at the Heinrich Heine University Düsseldorf and the computing time provided by the John von Neumann Institute for Computing (NIC) on the supercomputer JUWELS at Jülich Supercomputing Centre (JSC) (user ID: HKF7, CYP450). The study was supported by Bundesministerium für Bildung und Forschung, Germany (BMBF) and the Department of Biotechnology, India (DBT) through project ReMetaDrug (Funding Number 01DQ19002).

## ■ REFERENCES


(1) Rendic, S.; Guengerich, F. P. Survey of Human Oxidoreductases and Cytochrome P450 Enzymes Involved in the Metabolism of Xenobiotic and Natural Chemicals. *Chem. Res. Toxicol.* **2015**, *28*, 38−42.

(2) Nebert, D. W.; Russell, D. W. Clinical Importance of the Cytochromes P450. *Lancet* **2002**, *360*, 1155−1162.

(3) Ortiz de Montellano, P. R. Hydrocarbon Hydroxylation by Cytochrome P450 Enzymes. *Chem. Rev.* **2010**, *110*, 932−948.

(4) Guengerich, F. P. Common and Uncommon Cytochrome P450 Reactions Related to Metabolism and Chemical Toxicity. *Chem. Res. Toxicol.* **2001**, *14*, 611−650.

(5) Anzenbacher, P.; Anzenbacherová, E. Cytochromes P450 and Metabolism of Xenobiotics. *Cell. Mol. Life Sci.* **2001**, *58*, 737−747.

(6) Black, S. D. Membrane Topology of the Mammalian P450 Cytochromes. *FASEB J.* **1992**, *6*, 680−685.

(7) Cojocaru, V.; Winn, P. J.; Wade, R. C. The Ins and Outs of Cytochrome P450s. *Biochim. Biophys. Acta* **2007**, *1770*, 390−401.

(8) Nelson, D. R. Introductory Remarks on Human Cyps. *Drug Metab. Rev.* **2002**, *34*, 1−5.

(9) Ghosh, D.; Griswold, J.; Erman, M.; Pangborn, W. Structural Basis for Androgen Specificity and Oestrogen Synthesis in Human Aromatase. *Nature* **2009**, *457*, 219−223.

(10) Rieger, M. A.; Ebner, R.; Bell, D. R.; Kiessling, A.; Rohayem, J.; Schmitz, M.; Temme, A.; Rieber, E. P.; Weigle, B. Identification of a Novel Mammary-Restricted Cytochrome P450, Cyp4z1, with Over-expression in Breast Carcinoma. *Cancer Res.* **2004**, *64*, 2357.

(11) Liu, J.; Machalz, D.; Wolber, G.; Sorensen, E. J.; Bureik, M. New Proluciferin Substrates for Human Cyp4 Family Enzymes. *Appl. Biochem. Biotechnol.* **2021**, *193*, 218−237.

(12) Hritz, J.; de Ruiter, A.; Oostenbrink, C. Impact of Plasticity and Flexibility on Docking Results for Cytochrome P450 2d6: A Combined Approach of Molecular Dynamics and Ligand Docking. *J. Med. Chem.* **2008**, *51*, 7469−7477.

(13) Arfeen, M.; Patel, D. S.; Abbat, S.; Taxak, N.; Bharatam, P. V. Importance of Cytochromes in Cyclization Reactions: Quantum Chemical Study on a Model Reaction of Proguanil to Cycloguanil. *J. Comput. Chem.* **2014**, *35*, 2047−2055.

(14) Williams, P. A.; Cosme, J.; Vinković, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H. Crystal Structures of Human Cytochrome P450 3a4 Bound to Metyrapone and Progesterone. *Science* **2004**, *305*, 683.

(15) Hendrychová, T.; Anzenbacherová, E.; Hudeček, J.; Skopalík, J.; Lange, R.; Hildebrandt, P.; Otyepka, M.; Anzenbacher, P. Flexibility of Human Cytochrome P450 Enzymes: Molecular Dynamics and Spectroscopy Reveal Important Function-Related Variations. *Biochim. Biophys. Acta* **2011**, *1814*, 58−68.

(16) Otyepka, M.; Berka, K.; Anzenbacher, P. Is There a Relationship between the Substrate Preferences and Structural Flexibility of Cytochromes P450? *Curr. Drug Metab.* **2012**, *13*, 130−142.

(17) Gora, A.; Brezovsky, J.; Damborsky, J. Gates of Enzymes. *Chem. Rev.* **2013**, *113*, 5871−5923.

(18) Kokkonen, P.; Bednar, D.; Pinto, G.; Prokop, Z.; Damborsky, J. Engineering Enzyme Access Tunnels. *Biotechnol. Adv.* **2019**, *37*, No. 107386.

(19) Kingsley, L. J.; Lill, M. A. Substrate Tunnels in Enzymes: Structure−Function Relationships and Computational Methodology. *Proteins* **2015**, *83*, 599−611.

(20) Skopalík, J.; Anzenbacher, P.; Otyepka, M. Flexibility of Human Cytochromes P450: Molecular Dynamics Reveals Differences

between Cyps 3a4, 2c9, and 2a6, Which Correlate with Their Substrate Preferences. *J. Phys. Chem. B* **2008**, *112*, 8165−8173.
(21) Eisenmesser, E. Z.; Bosco, D. A.; Akke, M.; Kern, D. Enzyme Dynamics During Catalysis. *Science* **2002**, *295*, 1520.
(22) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. A Hierarchy of Timescales in Protein Dynamics Is Linked to Enzyme Catalysis. *Nature* **2007**, *450*, 913−916.
(23) Holliday, M. J.; Camilloni, C.; Armstrong, G. S.; Vendruscolo, M.; Eisenmesser, E. Z. Networks of Dynamic Allostery Regulate Enzyme Function. *Structure* **2017**, *25*, 276−286.
(24) Goodey, N. M.; Benkovic, S. J. Allosteric Regulation and Catalysis Emerge Via a Common Route. *Nat. Chem. Biol.* **2008**, *4*, 474−482.
(25) Saavedra, H. G.; Wrabl, J. O.; Anderson, J. A.; Li, J.; Hilser, V. J. Dynamic Allostery Can Drive Cold Adaptation in Enzymes. *Nature* **2018**, *558*, 324−328.
(26) Tzeng, S.-R.; Kalodimos, C. G. Protein Dynamics and Allostery: An Nmr View. *Curr. Opin. Struct. Biol.* **2011**, *21*, 62−67.
(27) Nobeli, I.; Favia, A. D.; Thornton, J. M. Protein Promiscuity and Its Implications for Biotechnology. *Nat. Biotechnol.* **2009**, *27*, 157−167.
(28) Mulnaes, D.; Porta, N.; Clemens, R.; Apanasenko, I.; Reiners, J.; Gremer, L.; Neudecker, P.; Smits, S. H. J.; Gohlke, H. Topmodel: Template-Based Protein Structure Prediction at Low Sequence Identity Using Top-Down Consensus and Deep Neural Networks. *J. Chem. Theory Comput.* **2020**, *16*, 1953−1967.
(29) Dobson, L.; Reményi, I.; Tusnády, G. E. Cctop: A Consensus Constrained Topology Prediction Web Server. *Nucleic Acids Res.* **2015**, *43*, W408−W412.
(30) Dobson, L.; Reményi, I.; Tusnády, G. E. The Human Transmembrane Proteome. *Biol. Direct* **2015**, *10*, No. 31.
(31) UniProt, C. Uniprot: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2019**, *47*, D506−D515.
(32) Mulnaes, D.; Gohlke, H. Topscore: Using Deep Neural Networks and Large Diverse Data Sets for Accurate Protein Model Quality Assessment. *J. Chem. Theory Comput.* **2018**, *14*, 6117−6126.
(33) Greenwood, J. R.; Calkins, D.; Sullivan, A. P.; Shelley, J. C. Towards the Comprehensive, Rapid, and Accurate Prediction of the Favorable Tautomeric States of Drug-Like Molecules in Aqueous Solution. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 591−604.
(34) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: A Software Program for Pkaprediction and Protonation State Generation for Drug-Like Molecules. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 681−691.
(35) *Schrödinger Release 2019-3: Maestro*; Schrödinger, LLC: New York, NY, 2019.
(36) Schott-Verdugo, S.; Gohlke, H. Packmol-Memgen: A Simple-to-Use, Generalized Workflow for Membrane-Protein-Lipid-Bilayer System Building. *J. Chem. Inf. Model.* **2019**, *59*, 2522−2528.
(37) Yeagle, P. *The Membranes of Cells*, 3rd ed.; Elsevier: Amsterdam, 2016; Vol. *xii*, 439 pp.
(38) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with Amber on Gpus. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878−3888.
(39) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, I. T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Giambasu, G.; Giese, T.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Huang, Y.; Izadi, S.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wilson, L.; Wolf, R. M.; Wu, X.; Xiao, L.; Xiong, Y.; York, D. M.; Kollman, P. A. *Amber 2019*; University of California: San Francisco, 2019.
(40) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14sb: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99sb. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(41) Dickson, C. J.; Madej, B. D.; Skjevik, A. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C. Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014**, *10*, 865−879.
(42) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 3863−3871.
(43) Shahrokh, K.; Orendt, A.; Yost, G. S.; Cheatham, T. E., 3rd Quantum Mechanically Derived Amber-Compatible Heme Parameters for Various States of the Cytochrome P450 Catalytic Cycle. *J. Comput. Chem.* **2012**, *33*, 119−133.
(44) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(45) Rathi, P. C.; Radestock, S.; Gohlke, H. Thermostabilizing Mutations Preferentially Occur at Structural Weak Spots with a High Mutation Ratio. *J. Biotechnol.* **2012**, *159*, 135−144.
(46) Hermans, S. M. A.; Pfleger, C.; Nutschel, C.; Hanke, C. A.; Gohlke, H. Rigidity Theory for Biomolecules: Concepts, Software, and Applications. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2017**, *7*, No. e1311.
(47) Jacobs, D. J.; Rader, A. J.; Kuhn, L. A.; Thorpe, M. F. Protein Flexibility Predictions Using Graph Theory. *Proteins* **2001**, *44*, 150−165.
(48) Dahiyat, B. I.; Benjamin Gordon, D.; Mayo, S. L. Automated Design of the Surface Positions of Protein Helices. *Protein Sci.* **1997**, *6*, 1333−1337.
(49) Radestock, S.; Gohlke, H. Exploiting the Link between Protein Rigidity and Thermostability for Data-Driven Protein Engineering. *Eng. Life Sci.* **2008**, *8*, 507−522.
(50) Pfleger, C.; Rathi, P. C.; Klein, D. L.; Radestock, S.; Gohlke, H. Constraint Network Analysis (Cna): A Python Software Package for Efficiently Linking Biomacromolecular Structure, Flexibility, (Thermo-)Stability, and Function. *J. Chem. Inf. Model.* **2013**, *53*, 1007−1015.
(51) Jacobs, D. J.; Thorpe, M. F. Generic Rigidity Percolation: The Pebble Game. *Phys. Rev. Lett.* **1995**, *75*, 4051−4054.
(52) Pfleger, C.; Radestock, S.; Schmidt, E.; Gohlke, H. Global and Local Indices for Characterizing Biomolecular Flexibility and Rigidity. *J. Comput. Chem.* **2013**, *34*, 220−233.
(53) Pfleger, C.; Minges, A.; Boehm, M.; McClendon, C. L.; Torella, R.; Gohlke, H. Ensemble- and Rigidity Theory-Based Perturbation Approach to Analyze Dynamic Allostery. *J. Chem. Theory Comput.* **2017**, *13*, 6343−6357.
(54) Contreras, F.; Nutschel, C.; Beust, L.; Davari, M. D.; Gohlke, H.; Schwaneberg, U. Can Constraint Network Analysis Guide the Identification Phase of Knowvolution? A Case Study on Improved Thermostability of an Endo-B-Glucanase. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 743−751.
(55) Rathi, P. C.; Fulton, A.; Jaeger, K.-E.; Gohlke, H. Application of Rigidity Theory to the Thermostabilization of Lipase a from *Bacillus subtilis*. *PLoS Comput. Biol.* **2016**, *12*, No. e1004754.
(56) Nutschel, C.; Fulton, A.; Zimmermann, O.; Schwaneberg, U.; Jaeger, K.-E.; Gohlke, H. Systematically Scrutinizing the Impact of Substitution Sites on Thermostability and Detergent Tolerance for Bacillus Subtilis Lipase A. *J. Chem. Inf. Model.* **2020**, *60*, 1568−1584.
(57) Pfleger, C.; Kusch, J.; Kondapuram, M.; Schwabe, T.; Sattler, C.; Benndorf, K.; Gohlke, H. Allosteric Signaling in C-Linker and Cyclic Nucleotide-Binding Domain of Hcn2 Channels. *Biophys. J.* **2021**, *120*, 950−963.
(58) Preising, M. N.; Görg, B.; Friedburg, C.; Qvartskhava, N.; Budde, B. S.; Bonus, M.; Toliat, M. R.; Pfleger, C.; Altmüller, J.; Herebian, D.; Beyer, M.; Zöllner, H. J.; Wittsack, H.-J.; Schaper, J.; Klee, D.; Zechner, U.; Nürnberg, P.; Schipper, J.; Schnitzler, A.; Gohlke, H.; Lorenz, B.; Häussinger, D.; Bolz, H. J. Biallelic Mutation of Human Slc6a6 Encoding the Taurine Transporter Taut Is Linked to Early Retinal Degeneration. *FASEB J.* **2019**, *33*, 11507−11527.



https://doi.org/10.1021/acs.jcim.1c00558
J. Chem. Inf. Model. 2021, 61, 4023−4030

(59) Wifling, D.; Pfleger, C.; Kaindl, J.; Ibrahim, P.; Kling, R. C.; Buschauer, A.; Gohlke, H.; Clark, T. Basal Histamine H4 Receptor Activation: Agonist Mimicry by the Diphenylalanine Motif. *Chem. – Eur. J.* **2019**, *25*, 14613–14624.
(60) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Vijaykumar, A.; Bardelli, A. P.; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G.-L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; de Miranda Cardoso, J. V.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmerer, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y.; SciPy, C.; et al. Scipy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272.
(61) Dick, M.; Weiergraber, O. H.; Classen, T.; Bisterfeld, C.; Bramski, J.; Gohlke, H.; Pietruszka, J. Trading Off Stability against Activity in Extremophilic Aldolases. *Sci. Rep.* **2016**, *6*, No. 17908.
(62) Merz, A.; Yee, M.-c.; Szadkowski, H.; Pappenberger, G.; Crameri, A.; Stemmer, W. P. C.; Yanofsky, C.; Kirschner, K. Improving the Catalytic Activity of a Thermophilic Enzyme at Low Temperatures. *Biochemistry* **2000**, *39*, 880–889.
(63) Fuxreiter, M.; Magyar, C.; Juhász, T.; Szeltner, Z.; Polgár, L.; Simon, I. Flexibility of Prolyl Oligopeptidase: Molecular Dynamics and Molecular Framework Analysis of the Potential Substrate Pathways. *Proteins* **2005**, *60*, 504–512.
(64) 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*, 11248–11255.
(65) Nath, A.; Atkins, W. M. A Quantitative Index of Substrate Promiscuity. *Biochemistry* **2008**, *47*, 157–166.
(66) Foti, R. S.; Honaker, M.; Nath, A.; Pearson, J. T.; Buttrick, B.; Isoherranen, N.; Atkins, W. M. Catalytic Versus Inhibitory Promiscuity in Cytochrome P450s: Implications for Evolution of New Function. *Biochemistry* **2011**, *50*, 2387–2393.
(67) Müller, H.; Becker, A. K.; Palm, G. J.; Berndt, L.; Badenhorst, C. P. S.; Godehard, S. P.; Reisky, L.; Lammers, M.; Bornscheuer, U. T. Sequence-Based Prediction of Promiscuous Acyltransferase Activity in Hydrolases. *Angew. Chem., Int. Ed.* **2020**, *59*, 11607–11612.
(68) Seibert, C. M.; Raushel, F. M. Structural and Catalytic Diversity within the Amidohydrolase Superfamily. *Biochemistry* **2005**, *44*, 6383–6391.
(69) Oppermann, U.; Filling, C.; Hult, M.; Shafqat, N.; Wu, X.; Lindh, M.; Shafqat, J.; Nordling, E.; Kallberg, Y.; Persson, B.; Jörnvall, H. Short-Chain Dehydrogenases/Reductases (SDR): The 2002 Update. *Chem.–Biol. Interact.* **2003**, *143–144*, 247–253.
(70) Fushinobu, S.; Nishimasu, H.; Hattori, D.; Song, H.-J.; Wakagi, T. Structural Basis for the Bifunctionality of Fructose-1,6-Bisphosphate Aldolase/Phosphatase. *Nature* **2011**, *478*, 538–541.
(71) Du, J.; Say, R. F.; Lü, W.; Fuchs, G.; Einsle, O. Active-Site Remodelling in the Bifunctional Fructose-1,6-Bisphosphate Aldolase/Phosphatase. *Nature* **2011**, *478*, 534–537.
(72) Nutschel, C.; Coscolín, C.; David, B.; Mulnaes, D.; Ferrer, M.; Jaeger, K.-E.; Gohlke, H. Promiscuous Esterases Counterintuitively Are Less Flexible Than Specific Ones. *J. Chem. Inf. Model.* **2021**, *61*, 2383–2395.
(73) Krüger, D. M.; Rathi, P. C.; Pfleger, C.; Gohlke, H. CNA Web Server: Rigidity Theory-Based Thermal Unfolding Simulations of Proteins for Linking Structure, (Thermo-)Stability, and Function. *Nucleic Acids Res.* **2013**, *41*, W340–W348.
(74) Hawkins, P. C.; Skillman, A. G.; Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2007**, *50*, 74–82.
(75) Ekroos, M.; Sjögren, T. Structural Basis for Ligand Promiscuity in Cytochrome P450 3a4. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13682–13687.
(76) Schleinkofer, K.; Sudarko; Winn, P. J.; Lüdemann, S. K.; Wade, R. C. Do Mammalian Cytochrome P450s Show Multiple Ligand Access Pathways and Ligand Channelling? *EMBO Rep.* **2005**, *6*, 584–589.
(77) Kingsley, L. J.; Lill, M. A. Ensemble Generation and the Influence of Protein Flexibility on Geometric Tunnel Prediction in Cytochrome P450 Enzymes. *PLoS One* **2014**, *9*, No. e99408.
(78) Berka, K.; Paloncýová, M.; Anzenbacher, P.; Otyepka, M. Behavior of Human Cytochromes P450 on Lipid Membranes. *J. Phys. Chem. B* **2013**, *117*, 11556–11564.
(79) Urban, P.; Lautier, T.; Pompon, D.; Truan, G. Ligand Access Channels in Cytochrome P450 Enzymes: A Review. *Int. J. Mol. Sci.* **2018**, *19*, No. 1617.
(80) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577–2637.



https://doi.org/10.1021/acs.jcim.1c00558
J. Chem. Inf. Model. 2021, 61, 4023–4030

# Supporting Information

# F/G-Region Rigidity is Inversely Correlated to Substrate Promiscuity of Human CYP Isoforms Involved in Metabolism

Daniel Becker[a], Prasad V. Bharatam[b], Holger Gohlke[a,c,1]

[a]Mathematisch-Naturwissenschaftliche Fakultät, Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany

[b]Department of Medicinal Chemistry, National Institute of Pharmaceutical Education and Research (NIPER), Sector 67, S. A. S. Nagar, Mohali, 160 062 Punjab, India

[c]John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre (JSC), Institute of Biological Information Processing (IBI-7: Structural Biochemistry), and Institute of Bio- and Geosciences (IBG-4: Bioinformatics), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

[1] To whom correspondence should be addressed.

Running title: CYP F/G-region rigidity

Keywords: Cytochrome P450, metabolism, rigidity theory, promiscuity

[*]Address: Universitätsstr. 1, 40225 Düsseldorf, Germany.

Phone: (+49) 211 81 13662; Fax: (+49) 211 81 13847

E-mail: gohlke@uni-duesseldorf.de

# Content

# Supporting Methods

## 1. Quality Assessment of CYP Structural Models

TopScoreSingle predicts the structural similarity between a protein model and the native structure using a two-stage deep neural network. TopScoreSingle is used for quality assessment of models on a per-residue level, but also for the whole model; it is bounded between 0 and 1, with lower values indicating models of higher quality. The TopScoreSingle was below 0.3 for all models except for those of CYP1B1 (0.32) and CYP2D6 (0.41). Thus, all models were of good quality. The high value for CYP2D6 is caused by the N-terminal loop. After minimization in the presence of the membrane (see below), the TopScoreSingle value for CYP2D6 dropped to 0.32.

## 2. Molecular Dynamics Simulations

To generate an ensemble of protein conformations for Constraint Network Analysis (CNA), molecular dynamic simulations were performed for all isoforms.

The generated structural models were embedded by PACKMOL-Memgen[1] into a membrane with a composition of CHL:DOPC:DSPC:DAPC:DOPE 10:22:13:19:21, the main lipid components of the human endoplasmic reticulum[2]. The minimum distance of the protein to the box boundaries was set to 15 Å, and the minimum thickness of the water slab above and below the membrane was set to 17.5 Å. Potassium and chloride ions were added in the solvent box with a concentration of ~0.15 M, thereby ensuring the electroneutrality of the simulation systems.

The GPU particle mesh Ewald implementation[3] of the AMBER18[4] molecular simulations suite was used with ff14SB parameters[5] for the protein, Lipid17 parameters[6] for the membrane, and OPC as water model[7]. Parameters for the heme and cystine residues forming the S-Fe bridge between heme and protein were taken from Shahrokh *et al.*[8]. Because covalent bonds to hydrogens were constrained with the SHAKE algorithm[9], a time step of 2 fs was used. The cutoff for nonbonded interactions was set to 10 Å.

For CYP3A4 and CYP2E1, ten independent MD simulations of 1 μs length were performed. For all other investigated isoforms, five independent simulations of 1 μs were performed.

Production runs were prepared as described by Schott-Verdugo et al.[10], but the minimization without restraints was performed for 40,000 steps. Langevin dynamics[11] with a friction coefficient of 1 $ps^{-1}$ were used to keep the temperature at 310 K. Initially, the minimized systems were heated by gradually increasing the temperature from 0 K to 310 K for 280 ps under NVT conditions. The system density was adjusted using NPT conditions at 1 bar for 2 ns. Productions runs were then performed using the same NPT conditions. The pressure was controlled with the Berendsen barostat[12] with semiisotropic pressure scaling, coupling the membrane (x,y) plane.

The tilt angle between the heme and the membrane was calculated as the angle between the heme nitrogens and the z-axis, according to Berka et al.[13]. All analyses were performed with pytraj[14], a Python package binding to the cpptraj program[15].The interaction of the structures with the ER membrane was in accordance with previous work. All models have an average tilt angle between 36.4° and 62.8° (Table S1), in accordance with expected angles between 35° and 80° for systems in a membrane composed of only one component, e.g., DOPC.[16] All isoforms stayed in the membrane with the transmembrane helix and parts of the F/G-region as expected,[17] except CYP2C9. Here, the F/G region moved out of the membrane. The models did not undergo large conformational changes in most simulations (Figure S1, Figure S2). Also, the membrane tilt remained stable during the simulations (Figure S3). The membrane structure remained stable, as monitored by the electron density of the components (Figure S4). These tests were performed to demonstrate the structural invariance of the systems over time. During the MD simulations, the CYP moves relative to the membrane, which is also expected under physiological conditions.

## 3. Constraint Network Analysis

To detect changes in structural rigidity between CYP isoforms, we analyzed ensembles of constraint network topologies based on conformational ensembles saved every 200 ps from unbiased MD simulations. For CYP3A4 and CYP2E1, we generated 50,000 conformations each and for all other isoforms 25,000 conformations. Overall, we investigated 350,000 conformations in this study.

Structural rigidity was analyzed with the CNA software package[18, 19], which is a front and back end to the FIRST software[20]. It was used to construct networks of nodes (atoms) and covalent and noncovalent (hydrogen bonds, salt bridges, and hydrophobic tethers) constraints. The hydrogen bond energy (including salt bridges) is determined from an empirical function[21], while hydrophobic tethers between carbon and sulfur atoms were considered if the distance between these atoms was less than the sum of their van der Waals radii plus a cutoff of 0.25 Å.

Biomolecules generally display a hierarchy of rigidity that reflects the modular structure of biomolecules in terms of secondary, tertiary, and supertertiary structure.[19] This hierarchy can be identified by gradually removing noncovalent constraints from an initial network representation of a biomolecule, which generates a succession of network states σ, resulting in a "constraint dilution trajectory". For that, hydrogen bonds and salt bridges are removed in the order of increasing strength such that for network state $\sigma$, only those hydrogen bonds are kept that have an energy $E_{HB} \leq E_{cut}(\sigma)$.

For all MD-generated snapshots, neighbor stability maps ($rc_{ij,neighbor}$) were derived from "constraint dilution trajectories"[18]. Neighbor stability maps of one isoform were averaged by calculating the mean for each map position (Figure S5). A stability map shows the persistence of rigid contacts for pairs of residues during the bond dilution progress. A rigid contact between two residues $R_{i,j}$ exists if they are in the same rigid cluster. During the constraint dilution trajectory, $E_{cut}$ is identified at which a rigid contact

between two residues is lost; this value is entered into $rc_{ij,neighbor}$.[22] From the stability maps, we calculated $E_{i,\mathrm{CNA}}$ (eq. S1), the sum of energies associated with rigid contacts between residue $i$ to all other residues.

$$E_{\mathrm{i,CNA}} = \sum_{j=1}^{n} rc_{ij,neighbor} \qquad \textbf{(eq. S1)}$$

We also calculated $\bar{E}_{\mathrm{region,CNA}}$ (eq. S2), where "region" is either the entire protein, the F/G region, the B/C region, or the I-helix, as the average of the chemical potential of rigid contacts between the $n$ residues within a region.

$$\bar{E}_{\mathrm{region,CNA}} = \frac{1}{n} \sum_{i=region_{start}}^{region_{end}} E_{\mathrm{i,CNA}} \qquad \textbf{(eq. S2)}$$

## 4. Statistical Analysis and Data Analysis of Computational Results

The standard error of the mean (SEM) for the calculated $\bar{E}_{FG,\mathrm{CNA}}$ values was calculated by considering each trajectory as an independent sample.

Data of CNA analyses were visualized by visualCNA.[23]

All calculations were performed with NumPy.[24] Curve fitting was performed with the SciPy module stats.[25] The sequence alignment was generated with MAFFT[26] in JalView[27].

## 5. Promiscuity Index

Substrate promiscuity was calculated based on the work of Nath *et al.*[28] as implemented by us in python (for details, see supporting repository) (eq. S3).

$$I_{cat} = -\frac{1}{log_{10} N} \sum_{i=1}^{N} \frac{e_i}{\sum_{j=1}^{N} e_j} log_{10} \frac{e_i}{\sum_{j=1}^{N} e_j} \qquad \textbf{(eq. S3)}$$

## 6. Diversity of the Substrate Dataset

Substrates that are chemically similar to each other are expected to be metabolized similarly by a CYP; such correlations in the substrate set would reduce the effective CYP promiscuity. Therefore, the similarity of the 55 substrates was assessed by the maximum pairwise Tanimoto-Combo distance score $\delta_{ij}$ for compound $i$ versus $j$, which is bounded between 0 for identical compounds and 2 for dissimilar ones, and the mean maximum pairwise Tanimoto-Combo distance score $\delta_i$ of a substrate $i$ to all other substrates in the data set; the Tanimoto-Combo distance score accounts for shape and chemical complementary between 3D structures as determined by the Rapid Overlay of Chemical Structures approach.[29]

Complete linkage clustering on the pairwise distance matrix calculated for 55 compounds from $\delta_{ij}$ yielded 35 clusters at a distance of 1.0 (Figure S13), which is equivalent to $\delta_{ij} = 1.0$, indicating that on average less than two substrates share a similarity that is halfway between dissimilar and identical.

For comparison, we performed the same analysis also with 500 randomly selected molecules. To do so, we took samples of 55 randomly selected entries from the ZINC drug like database and performed the same procedure as described above. Additionally, we calculated the average of $\delta_i$ ($<\delta>$) for all samples, which peaks at 1.12. Hence, randomly pulled drug-like molecules are more similar to each other than those in our dataset (Figure S15).

The python code is provided in the supporting repository.

## 7. Bootstrapping

With the results of CYP3A4 and CYP2E1 obtained from ten independent trajectories, we performed bootstrapping to estimate the reliability of results for $\bar{E}_{\mathrm{FG,CNA}}$. Bootstrapping was performed with the Python module itertools by calculating the maximal SEM for all possible combinations (sampling without replacement) for different numbers of simulations. Beyond five independent MD simulations, the maximal SEM was < 0.8 kcal mol$^{-1}$ (Figure S11). This is close to thermal energy at 300 K. Hence, for all other systems, only five independent MD simulations were performed.

## 8. Supplemental Tables

**Table S1:** Uniprot accession codes for investigated isoforms and PDB IDs used as templates for transferring the heme group to generated models.

| Isoform | Uniprot | PDB |
|---|---|---|
| CYP1A1 | P04798 | 4I8V |
| CYP1A2 | P05177 | 2HI4 |
| CYP1B1 | Q16678 | 3PM0 |
| CYP2A6 | P11509 | 2FDV |
| CYP2B6 | P20813 | 5UFG |
| CYP2C8 | P10632 | 1PQ2 |
| CYP2C9 | P11712 | 1R9O |
| CYP2C19 | P33261 | 4GQS |
| CYP2D6 | P10635 | 3TBG |
| CYP2E1 | P05181 | 3E6I |
| CYP3A4 | P08684 | 1TQN |
| CYP3A5 | P20815 | 5VEU |

**Table S2:** Heme tilt angle in CYP isoforms measured in MD simulations.

| **System** | **Tilt angle**[a] |
|---|---|
| **CYP1A1** | 53.7 (1.68) |
| **CYP1A2** | 58.9 (2.09) |
| **CYP1B1** | 46.3 (1.27) |
| **CYP2A6** | 52.9 (2.34) |
| **CYP2B6** | 62.8 (3.00) |
| **CYP2C8** | 53.0 (3.00) |
| **CYP2C9** | 36.4 (5.00) |
| **CYP2C19** | 43.4 (1.87) |
| **CYP2D6** | 53.5 (2.30) |
| **CYP2E1** | 56.1 (3.86) |
| **CYP3A4** | 57.8 (1.12) |
| **CYP3A5** | 53.9 (1.26) |

[a] Mean and SEM; in °.

**Table S3:** Residue numbers of the F/G-regions in the investigated human CYP enzymes and average energies per residue due to rigid contacts.

| Isoform | F/G-region[a] | $\bar{E}_{FG,CNA}$[b] |
|---|---|---|
| CYP1A1 | 202-266 | -23.04 (0.148) |
| CYP1A2 | 213-274 | -28.89 (0.101) |
| CYP1B1 | 220-282 | -26.04 (0.157) |
| CYP2A6 | 194-251 | -32.62 (0.117) |
| CYP2B6 | 193-250 | -30.95 (0.395) |
| CYP2C8 | 189-251 | -27.73 (0.164) |
| CYP2C9 | 191-254 | -27.08 (0.157) |
| CYP2C19 | 192-253 | -27.85 (0.133) |
| CYP2D6 | 199-259 | -22.16 (0.163) |
| CYP2E1 | 193-255 | -31.72 (0.109) |
| CYP3A4 | 202-266 | -20.32 (0.112) |
| CYP3A5 | 202-266 | -23.06 (0.200) |

[a] The numbers refer to residue positions in the respective sequences.

[b] Average of the chemical potential of rigid contacts (eq. S1) between the residues in the F/G region (eq. S2) and SEM considering one trajectory as one sample; in kcal $mol^{-1}$.

**Table S4:** Promiscuity index ($I_{cat}$) calculated with $K_M$ and $k_{cat}$ from ref. [30].

| Isoform | $I_{cat}$[a] |
|---|---|
| CYP3A4 | 0.66 |
| CYP3A5 | 0.67 |
| CYP2D6 | 0.61 |
| CYP2E1 | 0.26 |
| CYP2C8 | 0.44 |
| CYP2C9 | 0.22 |
| CYP2C19 | 0.48 |
| CYP1A2 | 0.38 |

[a] Calculated with eq. 4 from ref. [28].

**Table S5:** Average energies per residue due to rigid contacts.

| Isoform | $\bar{E}_{\mathrm{all,CNA}}$[a] |
|---|---|
| CYP1A1 | -23.04 (0.148) |
| CYP1A2 | -28.89 (0.101) |
| CYP1B1 | -26.04 (0.157) |
| CYP2A6 | -32.62 (0.117) |
| CYP2B6 | -30.95 (0.395) |
| CYP2C8 | -27.73 (0.164) |
| CYP2C9 | -27.08 (0.157) |
| CYP2C19 | -27.85 (0.133) |
| CYP2D6 | -22.16 (0.163) |
| CYP2E1 | -31.72 (0.109) |
| CYP3A4 | -20.32 (0.112) |
| CYP3A5 | -23.06 (0.200) |

[a] Average of the chemical potential of rigid contacts (eq. S1) between all residues (eq. S2) and SEM considering one trajectory as one sample; in kcal $mol^{-1}$.

**Table S6:** Residue numbers of the I-helix in the investigated human CYP enzymes and average energies per residue due to rigid contacts.

| Isoform | I-region[a] | $\bar{E}_{\mathrm{I,CNA}}$[b] |
|---|---|---|
| CYP1A1 | 307-335 | -37.36 (0.219) |
| CYP1A2 | 304-335 | -36.79 (0.185) |
| CYP1B1 | 317-348 | -40.36 (0.141) |
| CYP2A6 | 288-319 | -43.37 (0.130) |
| CYP2B6 | 288-316 | -39.52 (0.623) |
| CYP2C8 | 284-315 | -34.10 (0.273) |
| CYP2C9 | 284-314 | -37.38 (0.318) |
| CYP2C19 | 284-315 | -38.46 (0.112) |
| CYP2D6 | 292-323 | -39.80 (0.336) |
| CYP2E1 | 286-317 | -42.25 (0.132) |
| CYP3A4 | 292-323 | -35.86 (0.093) |
| CYP3A5 | 292-323 | -39.21 (0.193) |

[a] The numbers refer to residue positions in the respective sequences.

[b] Average of the chemical potential of rigid contacts (eq. S1) between the residues in the I region (eq. S2) and SEM considering one trajectory as one sample; in kcal $mol^{-1}$.

**Table S7:** Residue numbers of the B/C-region in the investigated human CYP enzymes and average energies per residue due to rigid contacts.

| Isoform | B/C-region[a] | $\bar{E}_{BC,CNA}$[b] |
|---|---|---|
| CYP1A1 | 89-147 | -27.68 (0.080) |
| CYP1A2 | 91-149 | -28.37 (0.113) |
| CYP1B1 | 100-155 | -29.91 (0.133) |
| CYP2A6 | 84-138 | -28.08 (0.131) |
| CYP2B6 | 81-131 | -26.85 (0.327) |
| CYP2C8 | 80-131 | -25.36 (0.144) |
| CYP2C9 | 80-131 | -24.81 (0.179) |
| CYP2C19 | 83-131 | -27.86 (0.135) |
| CYP2D6 | 82-143 | -31.44 (0.184) |
| CYP2E1 | 83-136 | -30.42 (0.175) |
| CYP3A4 | 87-137 | -27.62 (0.086) |
| CYP3A5 | 87-132 | -24.17 (0.110) |

[a] The numbers refer to residue positions in the respective sequences.

[b] Average of the chemical potential of rigid contacts (eq. S1) between the residues in the B/C region (eq. S2) and SEM considering one trajectory as one sample; in kcal $mol^{-1}$.

## 9. Supplemental Figures

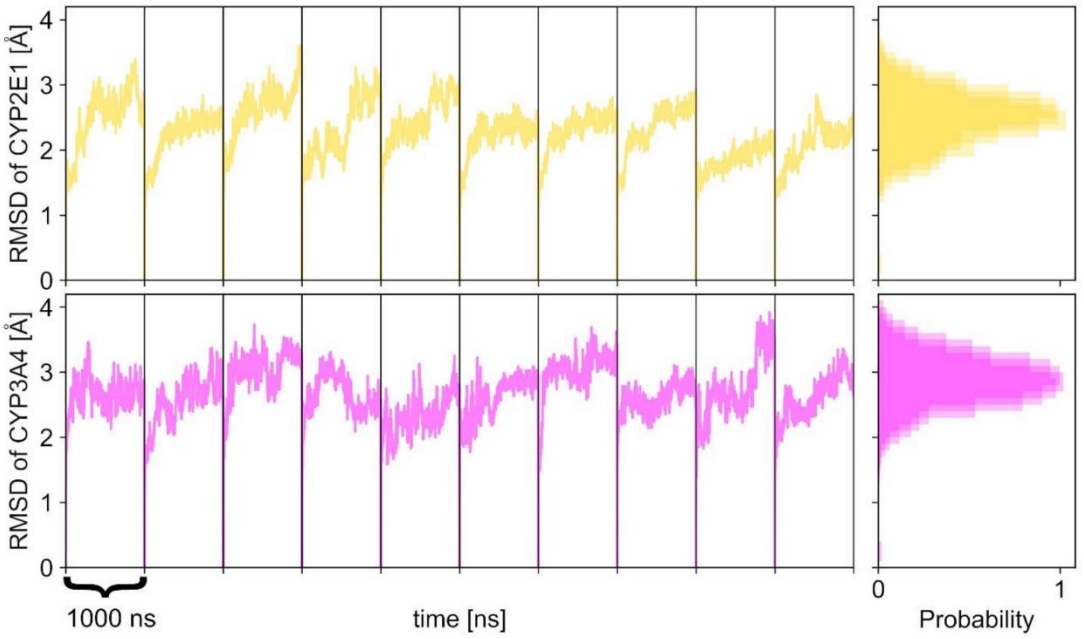



**Figure S1: Root mean-square deviation (RMSD) of the backbone of CYP2E1 and CYP3A4 along 10 MD trajectories each and aggregate probability distributions over all trajectories.** The RMSD is calculated with respect to the structure in the first frame after thermalization. The first 30 residues were not considered in the RMSD calculations because the TM helix can move independently in the membrane. Analyses were performed with CPPTRAJ.

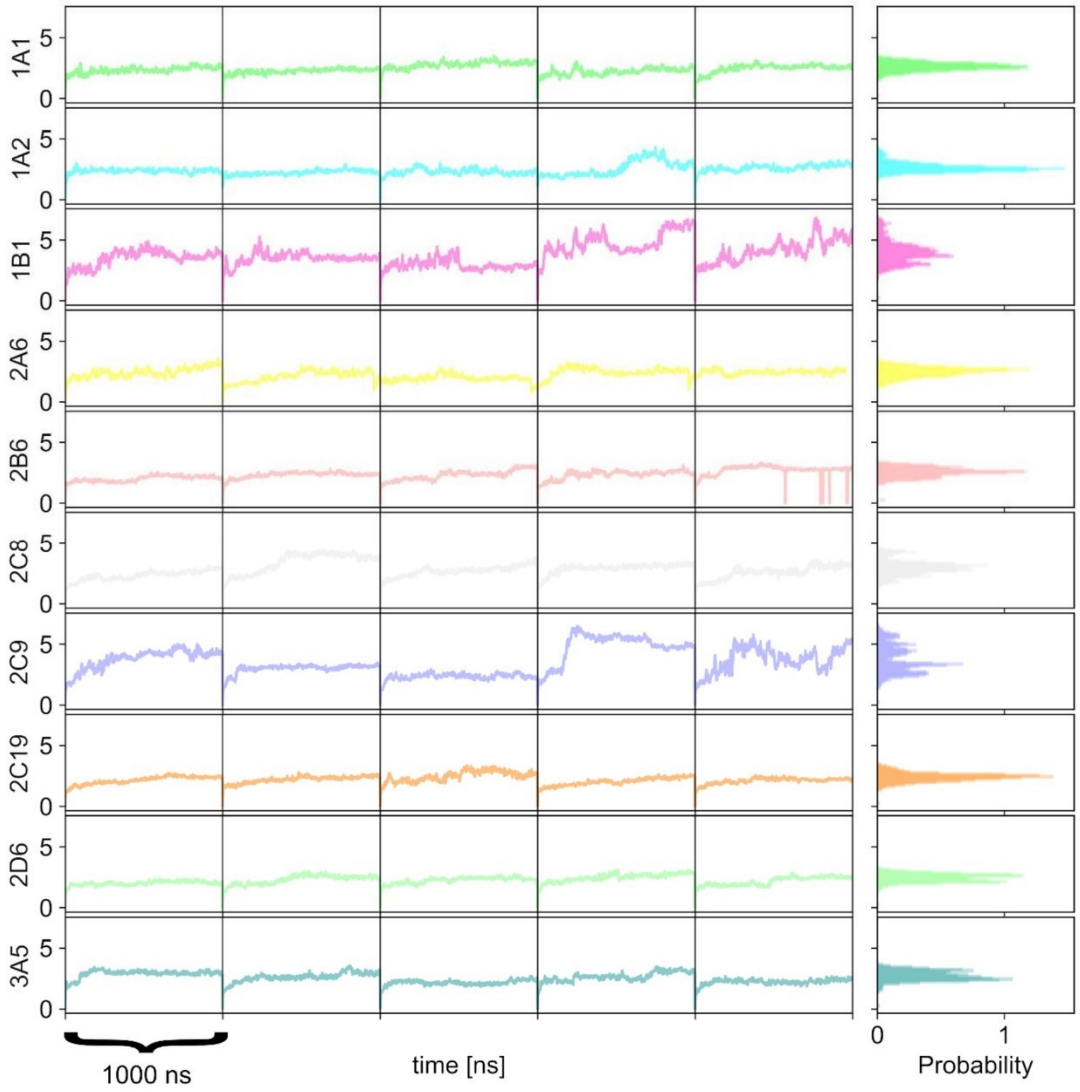



**Figure S2: Root mean-square deviation (RMSD) of the backbone of the ten other investigated CYP isoforms along five MD trajectories each and aggregate probability distributions.** The RMSD is calculated with respect to the structure in the first frame after equilibration. The first 30 residues were not considered in the RMSD calculations because the TM helix can move independently in the membrane. Analyses were performed with CPPTRAJ. Values per trajectory are separated by black vertical lines. Analyses were performed with CPPTRAJ.

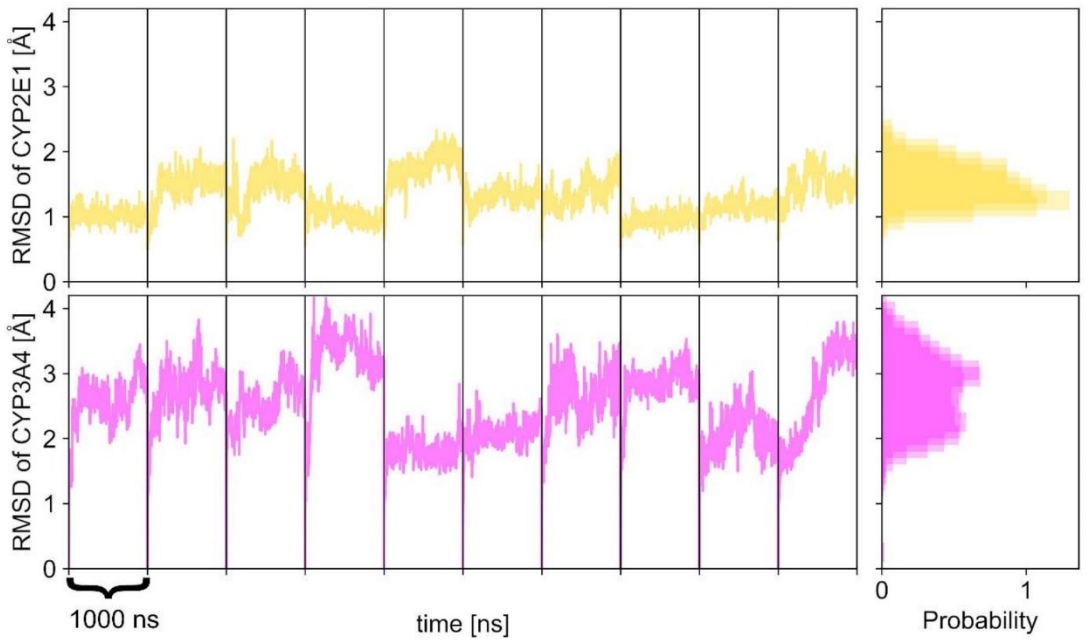



**Figure S3: Root mean-square deviation (RMSD) of the backbone of F/G-regions of CYP2E1 and CYP3A4 along 10 MD trajectories each and aggregate probability distributions over all trajectories.** The RMSD is calculated with respect to the structure in the first frame after thermalization. Analyses were performed with pyTRAJ.

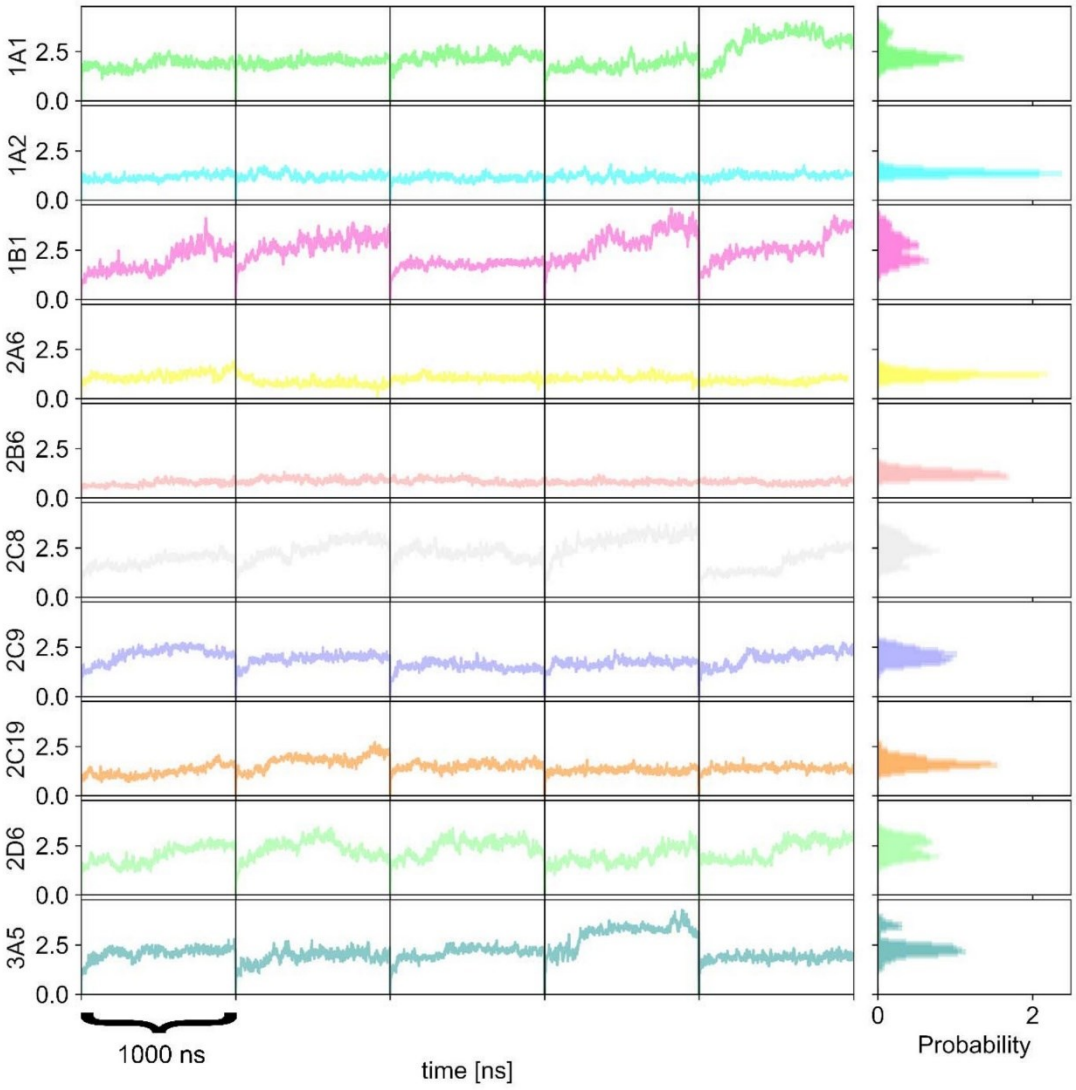



**Figure S4: Root mean-square deviation (RMSD) of the backbone of F/G-regions of the ten other investigated CYP isoforms along five MD trajectories each and aggregate probability distributions.** The RMSD is calculated with respect to the structure in the first frame after equilibration. Analyses were performed with pyTRAJ.

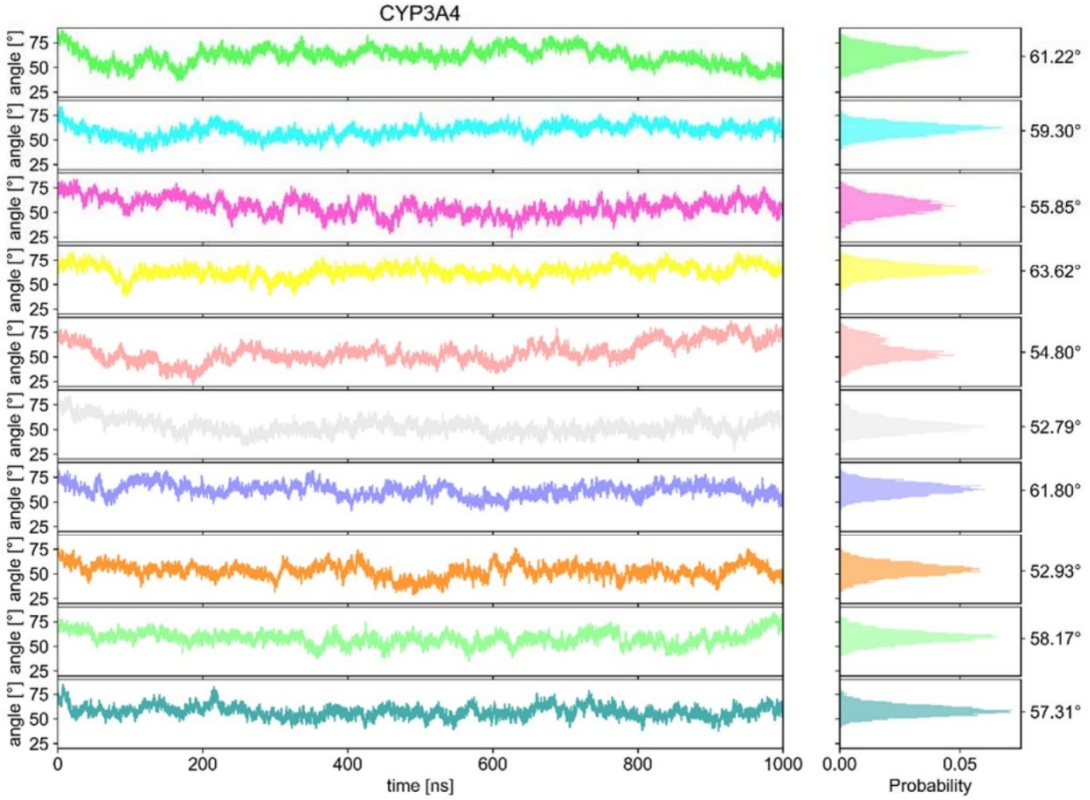



**Figure S5: Angle of the porphyrin ring system to the z-axis of the simulation box of CYP3A4 over 10 MD trajectories and probability distributions.** The values on the right denote the mean.

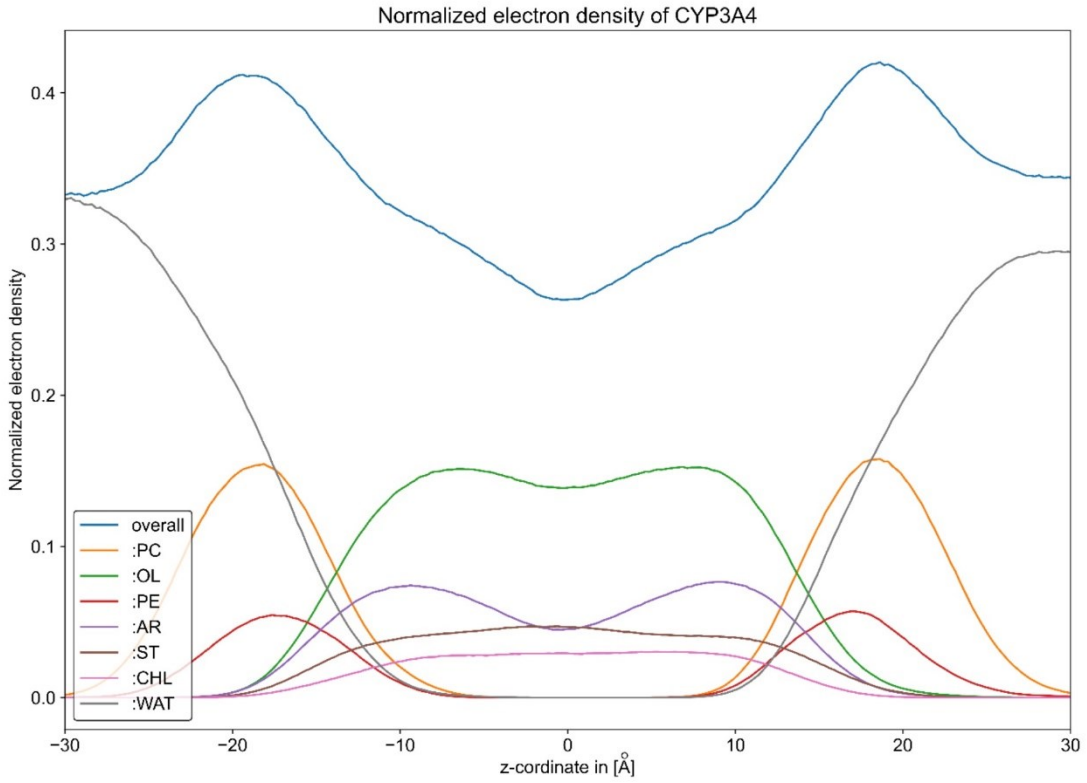



**Figure S6: Electron density plot of membrane components based on one MD simulation of CYP3A4.** The electron density is plotted against the z-coordinate of the simulation box for PC: phosphocholine, OL: oleic acid, PE: phosphoethanolamine, AR: arachidonic acid, ST: stearic acid, CHL: cholesterol, and WAT: water. The asymmetric curves result because of the embedded CYP enzyme. Analyses were performed with CPPTRAJ.

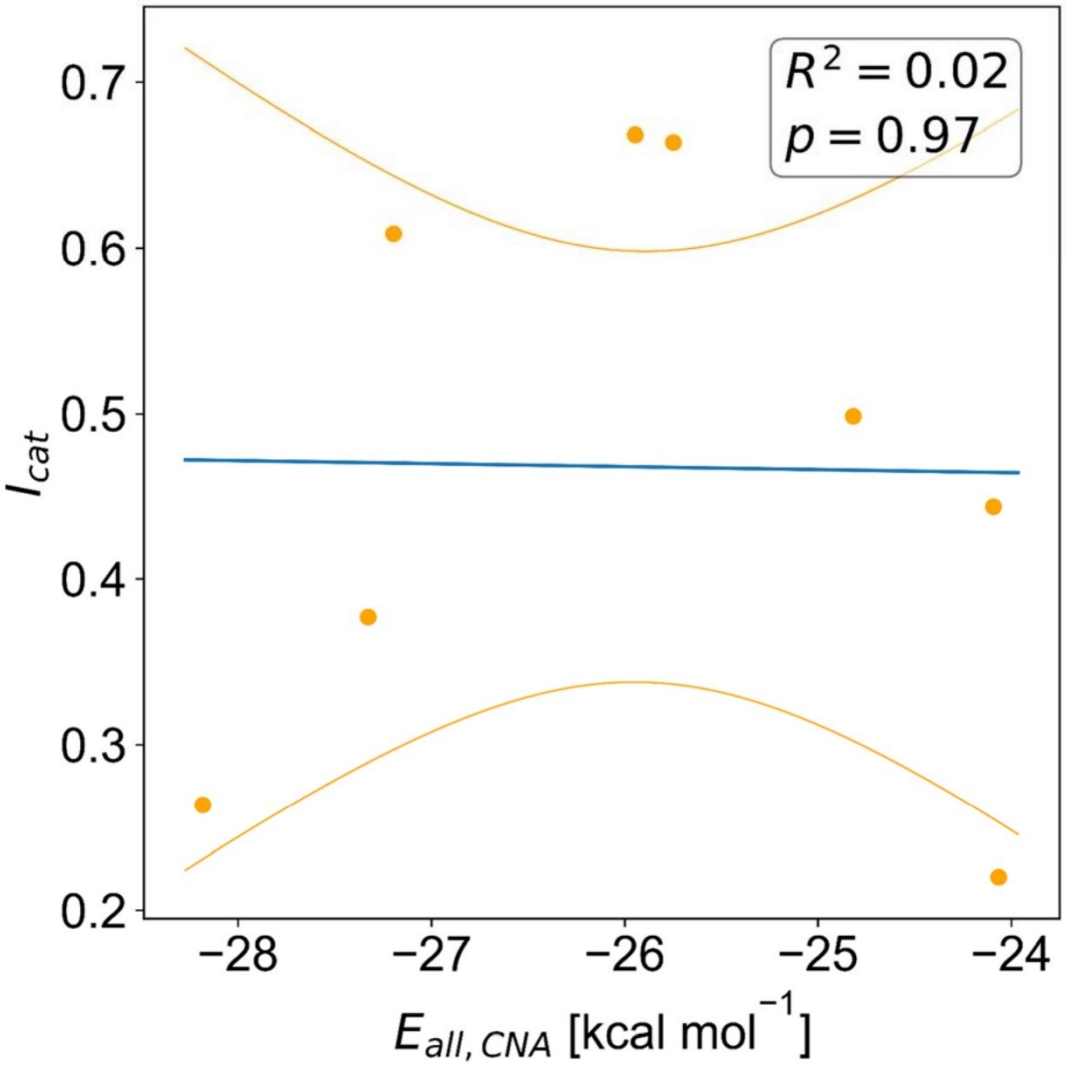



**Figure S7: The promiscuity of CYP isoforms is not correlated to the overall structural rigidity of the enzymes.** There is no significant correlation between the promiscuity index ($I_{cat}$) and $\bar{E}_{\mathrm{all,CNA}}$ (eq. S2) for CYP1A2, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP3A4, CYP3A5.

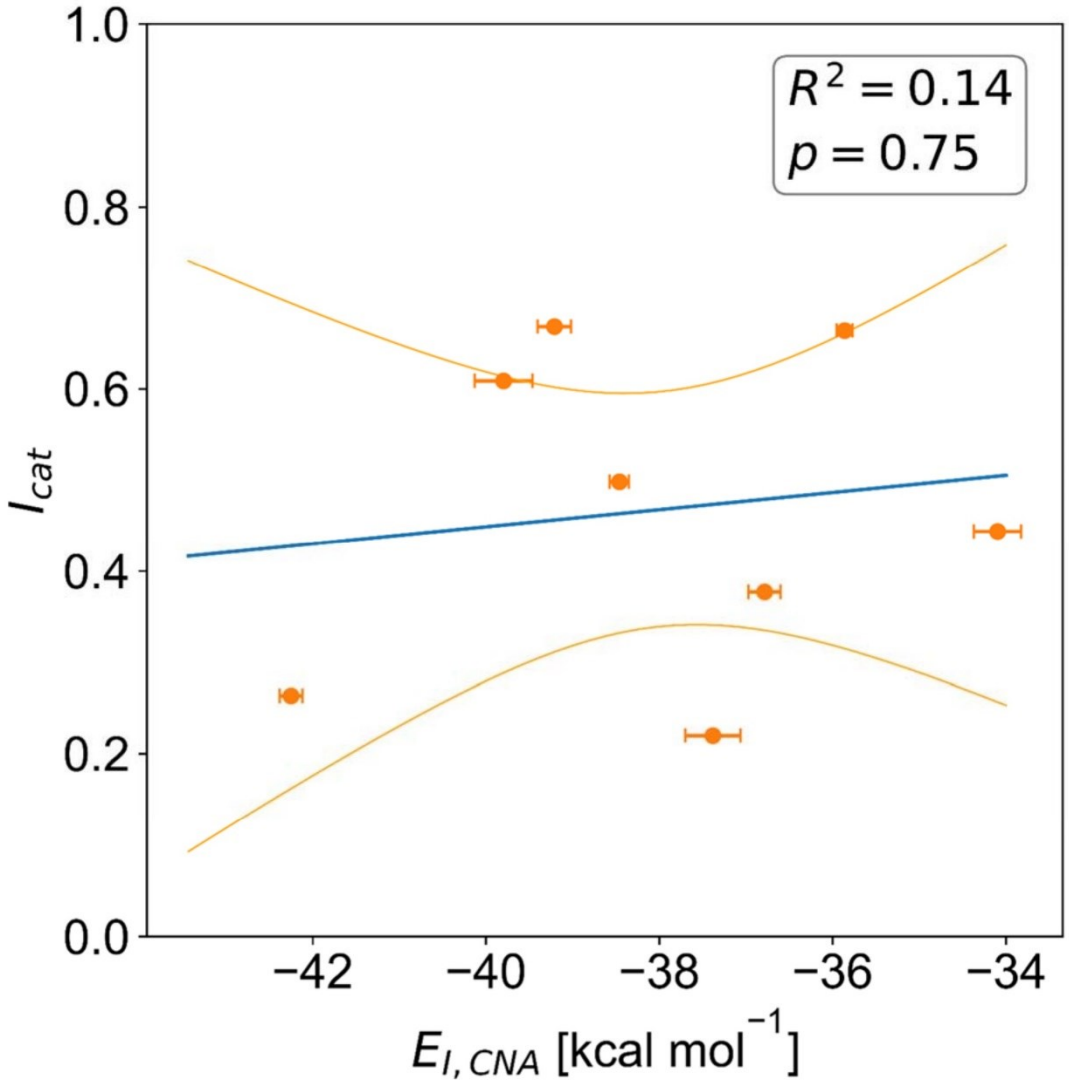



**Figure S8: The promiscuity of CYP isoforms is not correlated to the structural rigidity of the I-helix.** There is no significant correlation between the promiscuity index ($I_{cat}$) and $\bar{E}_{\mathrm{I,CNA}}$ (eq. S2) for CYP1A2, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP3A4, CYP3A5.

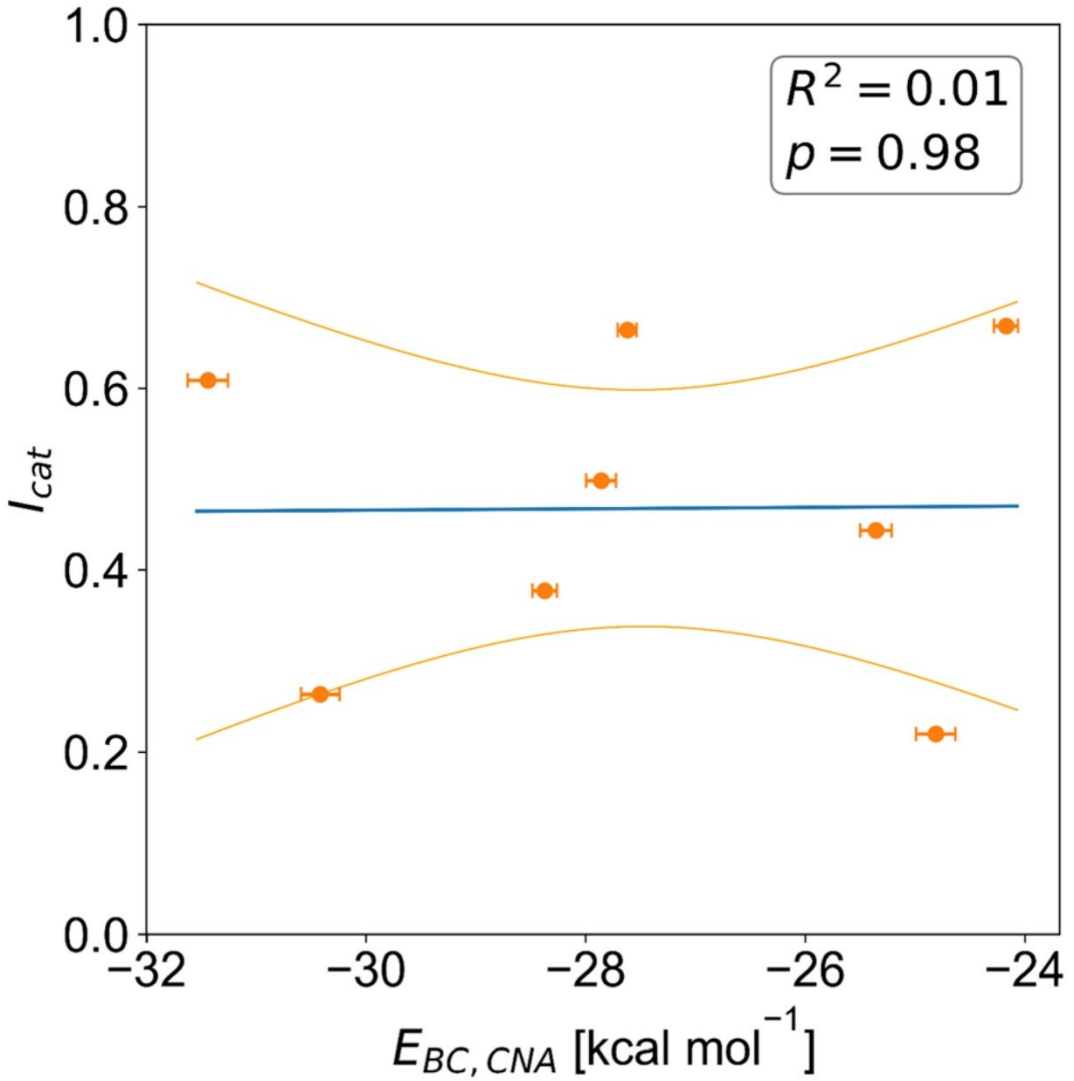



**Figure S9: The promiscuity of CYP isoforms is not correlated to the structural rigidity of their B/C-region.** There is no significant correlation between the promiscuity index ($I_{cat}$) and $\bar{E}_{\mathrm{BC,CNA}}$ (eq. S2) for CYP1A2, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP3A4, CYP3A5 does not show any trend.

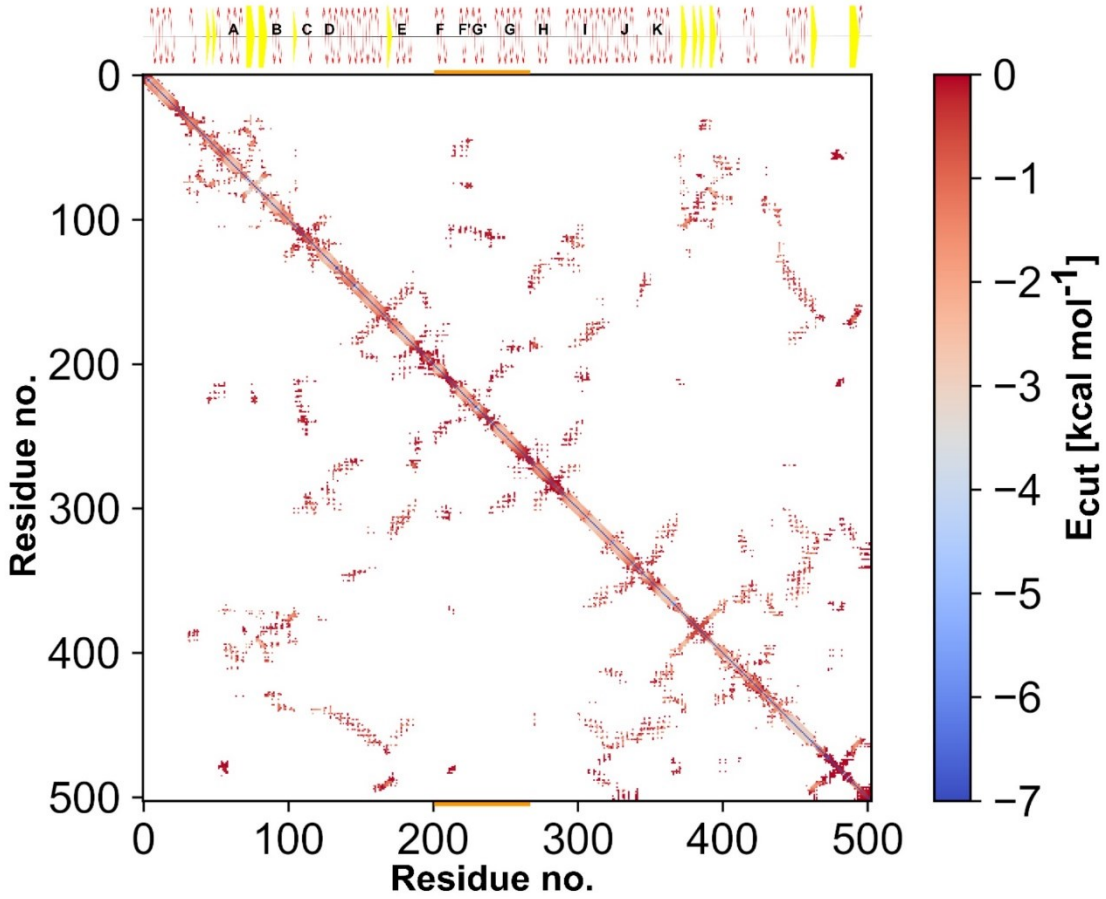



**Figure S10: Average neighbor stability map of CYP3A4.** Red (blue) colors indicate pairs of residues where a weak (strong) rigid contact exists. White color indicates pairs of residues more than 5 Å away from each other. Secondary structure elements identified with DSSP[31] are colored in red for helices, yellow for beta-sheets, and green for loops and given for helices and beta-sheets above the plot. Helices are named according to Williams *et al.*[32] The F/G-region is marked in orange.

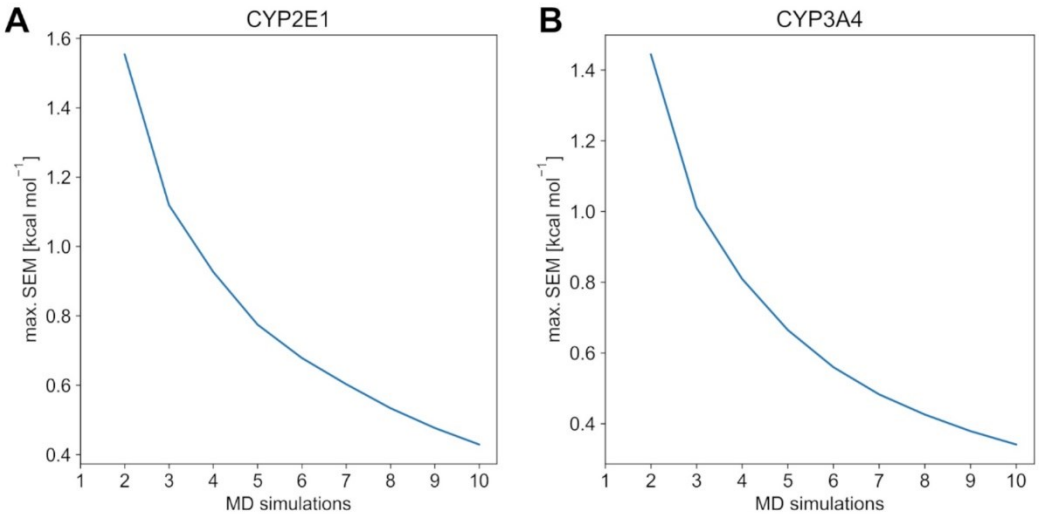



**Figure S11: Results of bootstrapping for CYP2E1 and CYP3AA.** Bootstrapping was performed by calculating the maximal *SEM* for all possible combinations (sampling without replacement) for different numbers of simulations. The maximal *SEM* decreases with the number of MD simulations considered in the sampling. Beyond five independent MD simulations, the maximal *SEM* was < 0.8 kcal mol$^{-1}$. This is almost equivalent to thermal energy at 300 K. **A:** Bootstrapping of CYP2E1. **B:** Bootstrapping of CYP3A4.

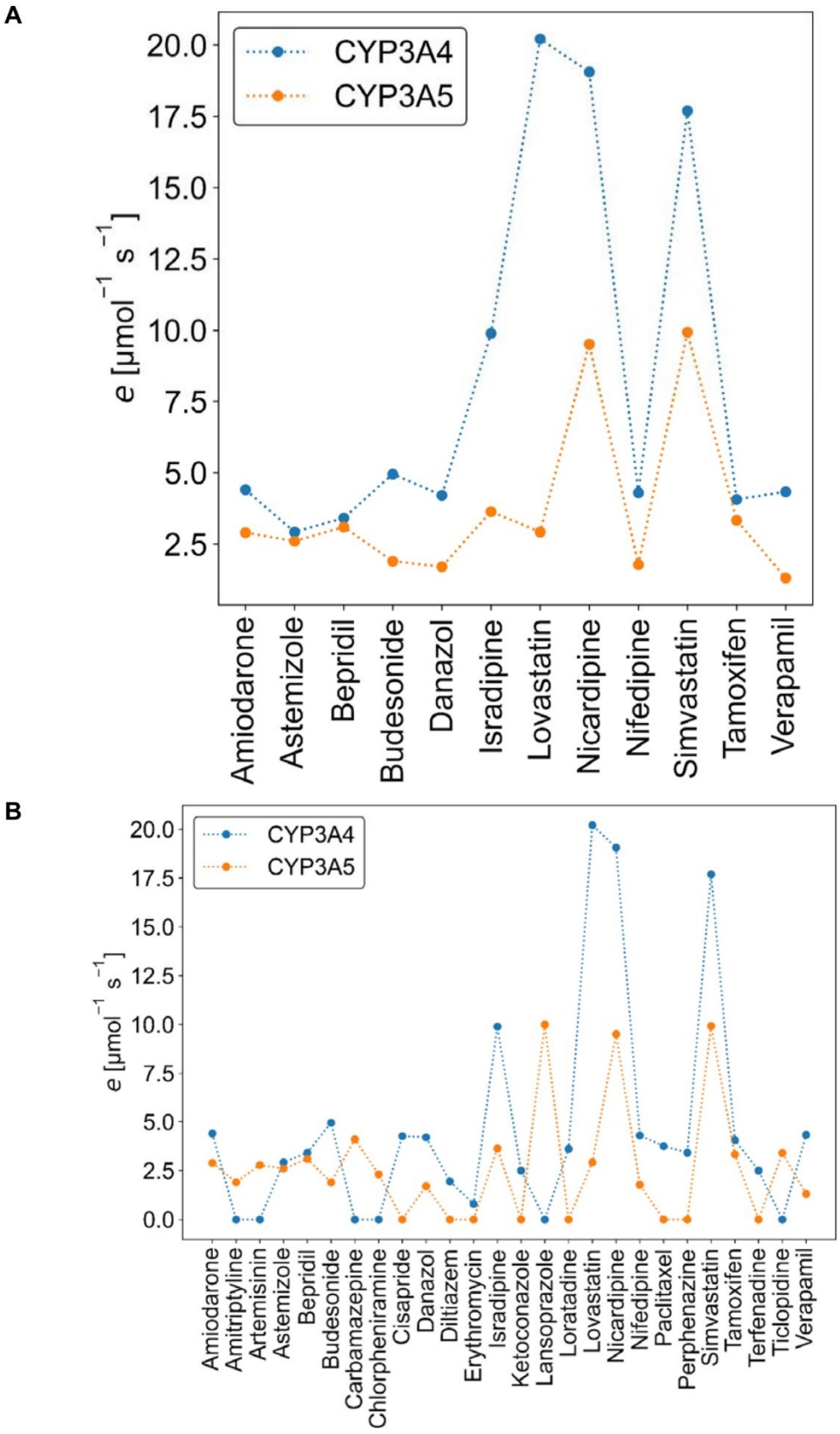

A
CYP3A4
CYP3A5
e [µmol⁻¹ s⁻¹]
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
Amiodarone
Astemizole
Bepridil
Budesonide
Danazol
Isradipine
Lovastatin
Nicardipine
Nifedipine
Simvastatin
Tamoxifen
Verapamil
B
CYP3A4
CYP3A5
e [µmol⁻¹ s⁻¹]
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Amiodarone
Amitriptyline
Artemisinin
Astemizole
Bepridil
Budesonide
Carbamazepine
Chlorpheniramine
Cisapride
Danazol
Diltiazem
Erythromycin
Isradipine
Ketoconazole
Lansoprazole
Loratadine
Lovastatin
Nicardipine
Nifedipine
Paclitaxel
Perphenazine
Simvastatin
Tamoxifen
Terfenadine
Ticlopidine
Verapamil

**Figure S12: Catalytic efficiencies of CYP3A4 versus CYP3A5. A:** Catalytic efficiencies of CYP3A4 are higher than CYP3A5 for every drug oxidized by both isoforms. Only drugs are shown where reactions are detected for both isoforms. **B:** Catalytic efficiencies of CYP3A4 are mostly (20 of 26) higher than CYP3A5 for drugs oxidized by at least one isoform. Catalytic efficiency values are calculated based on ref. [30]. Values of CYP3A4 are shown in blue, values of CYP3A5 in orange.

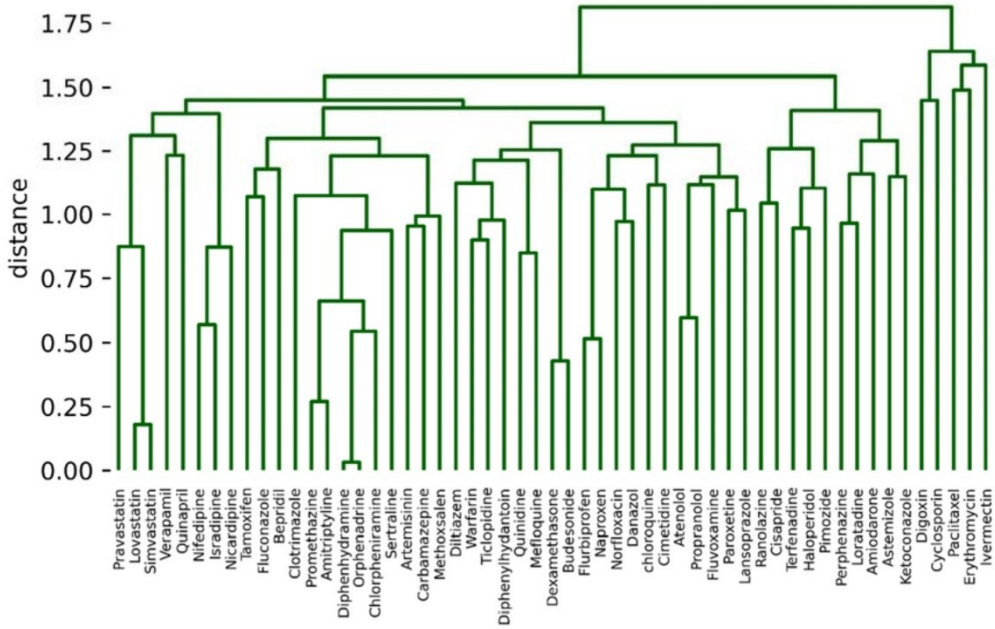



**Figure S13: Hierarchical clustering of substrates from ref. [30].** The complete linkage clustering method implemented in SciPy[25] was used to define clusters using the all-vs.-all pairwise distance matrix calculated for all 55 compounds derived from the respective matrix of maximum pairwise Tanimoto-Combo distance scores $\delta_{ij}$. The maximal distance is 2. At a distance of 1.0, 35 clusters exist with on average 1.6 substrates.

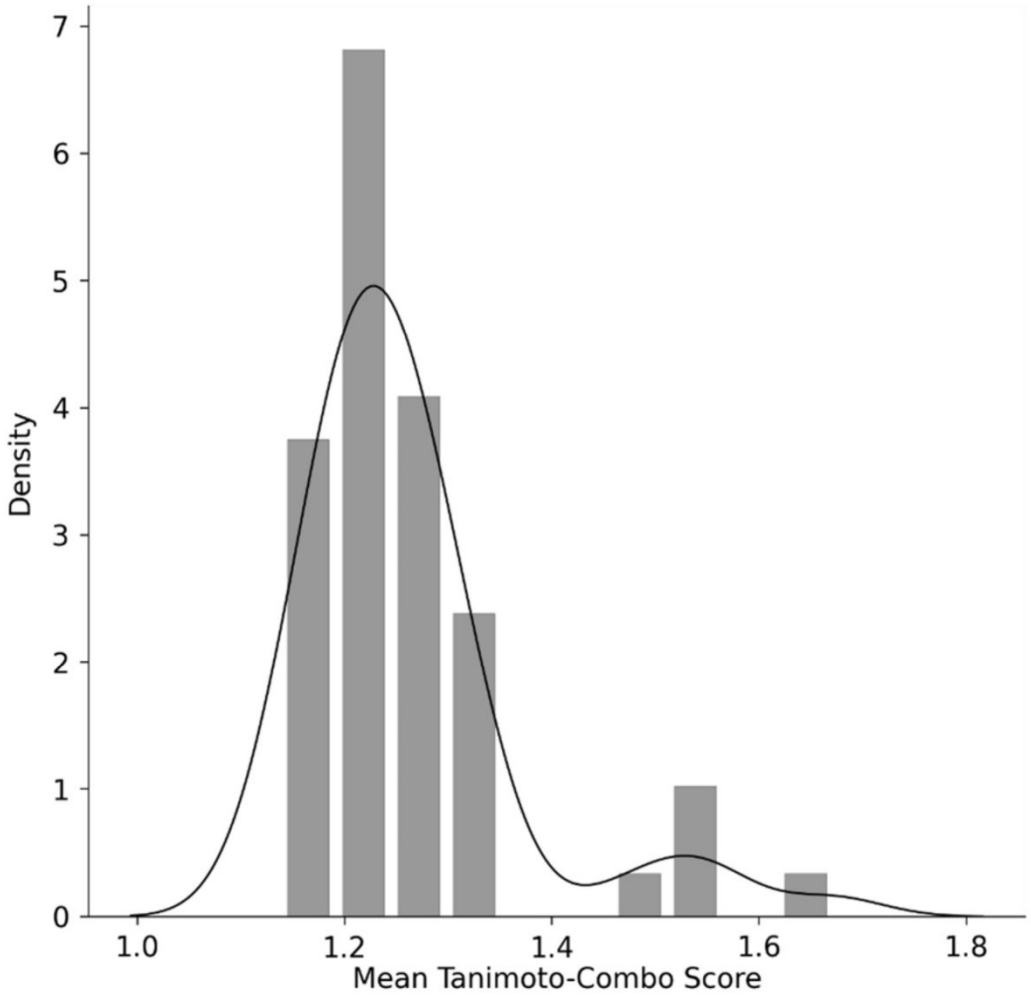



**Figure S14: Distribution of the mean Tanimoto-Combo distance scores $\delta_i$.** The mean scores were calculated for each of the 55 substrates from ref. [30] as the respective row average in the all-vs.-all matrix of maximum pairwise Tanimoto-Combo distance scores $\delta_{ij}$; $\delta_{ij}$ is bounded between 0 (identical) and 2 (dissimilar). The distribution peaks at ~1.25.

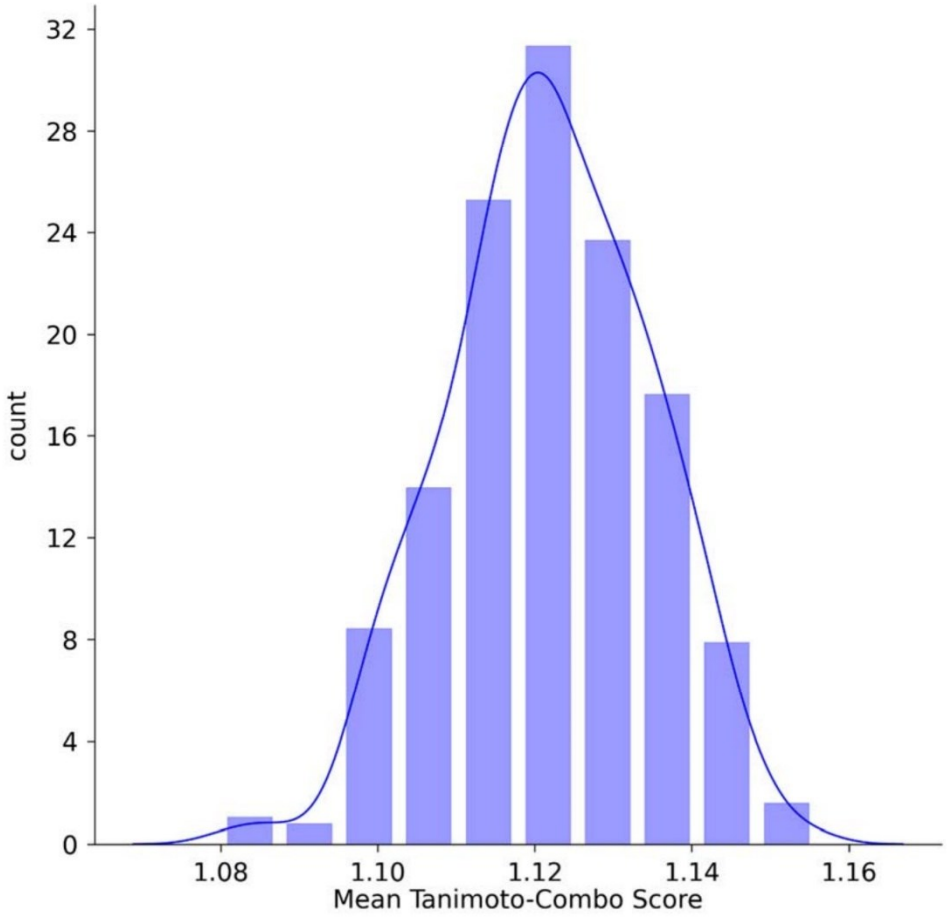



**Figure S15: Distribution of the means of mean Tanimoto-Combo distance scores $\delta_i$ from 500 randomly drawn samples of the ZINC drug-like database[33].** The means of mean Tanimoto-Combo distance scores $\delta_i$ for each randomly drawn sample range between 1.08 and 1.16. To obtain these values, we drew 500 times randomly 55 out of 981,259,785 substances from the ZINC drug-like database and computed the mean Tanimoto-Combo distance scores for each compound of the sample. Afterwards, we calculated the mean of $\delta_i$ per sample. Further information about the random drawing and which substances were sampled can be found in the supporting repository.

## 10. Supplemental References


1. Schott-Verdugo, S.; Gohlke, H., PACKMOL-Memgen: A Simple-To-Use, Generalized Workflow for Membrane-Protein-Lipid-Bilayer System Building. *J. Chem. Inf. Model.* **2019**, 59, 2522-2528.
2. Yeagle, P., *The membranes of cells*. Third edition. ed.; Elsevier/AP: Amsterdam ; Boston, 2016; p xii, 439 pages.
3. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, 9, 3878-3888.
4. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, I., T.E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Giambasu, G.; Giese, T.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Huang, Y.; Izadi, S.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wilson, L.; Wolf, R. M.; Wu, X.; Xiao, L.; Xiong, Y.; York, D. M.; Kollman, P. A. *AMBER 2019*, University of California, San Francisco, 2019.
5. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, 11, 3696-3713.
6. Dickson, C. J.; Madej, B. D.; Skjevik, A. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C., Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014**, 10, 865-879.
7. Izadi, S.; Anandakrishnan, R.; Onufriev, A. V., Building Water Models: A Different Approach. *The Journal of Physical Chemistry Letters* **2014**, 5, 3863-3871.
8. Shahrokh, K.; Orendt, A.; Yost, G. S.; Cheatham, T. E., 3rd, Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. *J. Comput. Chem.* **2012**, 33, 119-133.
9. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, 23, 327-341.
10. Schott-Verdugo, S.; Muller, L.; Classen, E.; Gohlke, H.; Groth, G., Structural Model of the ETR1 Ethylene Receptor Transmembrane Sensor Domain. *Sci Rep* **2019**, 9, 8869.
11. Quigley, D.; Probert, M. I. J., Langevin dynamics in constant pressure extended systems. *The Journal of Chemical Physics* **2004**, 120, 11432-11441.
12. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984**, 81, 3684-3690.
13. Berka, K.; Paloncýová, M.; Anzenbacher, P.; Otyepka, M., Behavior of Human Cytochromes P450 on Lipid Membranes. *The Journal of Physical Chemistry B* **2013**, 117, 11556-11564.
14. Nguyen, H.; Roe, D. R.; Swails, J.; Case, D. A. *PYTRAJ: Interactive data analysis for molecular dynamics simulations*, pytraj GitHub repository: 2016.
15. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, 9, 3084-3095.
16. Šrejber, M.; Navrátilová, V.; Paloncýová, M.; Bazgier, V.; Berka, K.; Anzenbacher, P.; Otyepka, M., Membrane-attached mammalian cytochromes P450: An overview of the membrane's effects on structure, drug binding, and interactions with redox partners. *J. Inorg. Biochem.* **2018**, 183, 117-136.
17. Headlam, M. J.; Wilce, M. C. J.; Tuckey, R. C., The F–G loop region of cytochrome P450scc (CYP11A1) interacts with the phospholipid membrane. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2003**, 1617, 96-108.
18. Pfleger, C.; Rathi, P. C.; Klein, D. L.; Radestock, S.; Gohlke, H., Constraint Network Analysis (CNA): A Python Software Package for Efficiently Linking Biomacromolecular Structure, Flexibility, (Thermo-)Stability, and Function. *J. Chem. Inf. Model.* **2013**, 53, 1007-1015.
19. Hermans, S. M. A.; Pfleger, C.; Nutschel, C.; Hanke, C. A.; Gohlke, H., Rigidity theory for biomolecules: concepts, software, and applications. *WIREs Computational Molecular Science* **2017**, 7, e1311.

20. Jacobs, D. J.; Rader, A. J.; Kuhn, L. A.; Thorpe, M. F., Protein flexibility predictions using graph theory. *Proteins* **2001**, 44, 150-165.
21. Dahiyat, B. I.; Benjamin Gordon, D.; Mayo, S. L., Automated design of the surface positions of protein helices. *Protein Sci.* **1997**, 6, 1333-1337.
22. Pfleger, C.; Radestock, S.; Schmidt, E.; Gohlke, H., Global and local indices for characterizing biomolecular flexibility and rigidity. *J. Comput. Chem.* **2013**, 34, 220-233.
23. Rathi, P. C.; Mulnaes, D.; Gohlke, H., VisualCNA: a GUI for interactive constraint network analysis and protein engineering for improving thermostability. *Bioinformatics* **2015**, 31, 2394-6.
24. Walt, S. v. d.; Colbert, S. C.; Varoquaux, G., The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering* **2011**, 13, 22-30.
25. Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Vijaykumar, A.; Bardelli, A. P.; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G.-L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; de Miranda Cardoso, J. V.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmerer, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y.; SciPy, C., SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, 17, 261-272.
26. Katoh, K.; Rozewicki, J.; Yamada, K. D., MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* **2017**, 20, 1160-1166.
27. Waterhouse, A. M.; Procter, J. B.; Martin, D. M. A.; Clamp, M.; Barton, G. J., Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, 25, 1189-1191.
28. Nath, A.; Atkins, W. M., A Quantitative Index of Substrate Promiscuity. *Biochemistry* **2008**, 47, 157-166.
29. Hawkins, P. C.; Skillman, A. G.; Nicholls, A., Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* **2007**, 50, 74-82.
30. Foti, R. S.; Honaker, M.; Nath, A.; Pearson, J. T.; Buttrick, B.; Isoherranen, N.; Atkins, W. M., Catalytic versus Inhibitory Promiscuity in Cytochrome P450s: Implications for Evolution of New Function. *Biochemistry* **2011**, 50, 2387-2393.
31. Kabsch, W.; Sander, C., Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, 22, 2577-2637.
32. Williams, P. A.; Cosme, J.; Vinković, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H., Crystal Structures of Human Cytochrome P450 3A4 Bound to Metyrapone and Progesterone. *Science* **2004**, 305, 683.
33. Sterling, T.; Irwin, J. J., ZINC 15 – Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, 55, 2324-2337.

# Molecular mechanisms underlying single nucleotide polymorphism-induced reactivity decrease in CYP2D6

Daniel Becker[a], Prassad V. Bharatam [b], Holger Gohlke[a,c,1]

[a]Mathematisch-Naturwissenschaftliche Fakultät, Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany

[b] Department of Medicinal Chemistry, National Institute of Pharmaceutical Education and Research (NIPER), Sector 67, S. A. S. Nagar, Mohali, 160 062 Punjab, India

[c]John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre (JSC), and Institute of Bio- and Geosciences (IBG-4: Bioinformatics), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

[1] To whom correspondence should be addressed.

Running title: SNP-induced reactivity decrease in CYP2D6

Keywords: Cytochrome P450, drug metabolism, CYP2D6, SNPs, mutations

[*]Address: Universitätsstr. 1, 40225 Düsseldorf, Germany.

Phone: (+49) 211 81 13662; Fax: (+49) 211 81 13847

E-mail: gohlke@uni-duesseldorf.de

## Abstract

Cytochrome P450 2D6 (CYP2D6) is one of the most important enzymes involved in drug metabolism. Genetic polymorphism can influence drug metabolism by CYP2D6 such that a therapy is seriously affected by under- or overdosing of drugs. However, a general explanation at the atomistic level for poor activity is missing so far. Here we show for the 20 most common single nucleotide polymorphisms (SNPs) of CYP2D6 that poor metabolism is driven by four mechanisms. We found in extensive all-atom molecular dynamic simulations that the rigidity of the I-helix (central helix), distance between central phenylalanines (stabilizing bound substrate), availability of basic residues on the surface of CYP2D6 (binding of Cytochrome P450 reductase), and position of arginine 132 (electron transfer to heme) are essential for an extensive function of the enzyme. These results were applied to SNPs with unknown effects and potential SNPs that may lead to poor drug metabolism were identified. The revealed molecular mechanisms might be important for other drug-metabolizing Cytochrome P450 enzymes.

## Introduction

The isoenzyme cytochrome P450 (CYP) 2D6 plays a crucial role in human drug metabolism.[1] It catalyzes the oxidation of especially lipophilic drugs that takes place as phase-1 biotransformation before excretion. Already in 1977, Mahgoub *et al.* described that differences in the hydroxylation rate of debrisoquine, an antihypertensive drug, depend on a single autosomal gene,[2] called later CYP2D6 according to its position in the CYP gene superfamily.[3] Since then, CYP2D6 polymorphism has raised increasing awareness during the development of small-molecule drugs because the polymorphism can lead to underdosing – leading to a failure of therapy or intoxication by drug metabolites – or overdosing – leading to intoxication by the drug itself.[4]

Differential mechanisms drive the differences in CYP2D6-related metabolism. The "very extensive" or "ultra-metabolizer" phenotype is mostly driven by gene duplication.[5] This duplication results in the overexpression of the CYP2D6 enzyme. Around 7% of Caucasians show an "ultra-metabolizer" phenotype.[6] Although an explanation at the genetic level for the "ultra-metabolizer" phenotype is often appropriate, mutations in the CYP2D6 gene can also lead to higher enzyme activity.[7] The "poor-metabolizer" and "non-metabolizer" phenotypes are established either at the genetic level, probably related to regulatory factors[8], or protein level, in which a combination of single nucleotide polymorphisms (SNPs) changes the enzymatic activity[9]. CYP2D6 is a highly polymorphic protein: So far, 165 genetic variants (alleles) have been discovered.[10] Deciphering the molecular mechanisms underlying the SNP-induced reactivity decrease of CYP2D6 is of utmost interest from a fundamental and drug development point of view.

CYP2D6, like all human CYP enzymes, is a membrane-associated and heme-containing protein, where the heme is buried and only accessible via channels.[11, 12] The oxidation reaction catalyzed by CYP enzymes occurs in multiple steps, starting from a heme with Fe(III) configuration (Figure 1A):[13] First, the ligand usually binds to the heme iron by displacing a distal water molecule. Afterward, an electron is transferred to the heme from Cytochrome P450 Reductase (CPR), a membrane-bound enzyme required for electron transfer from NADPH to CYP.[14, 15] In mammalian CYPs, the electron is transferred with arginine (in CYP2D6: Arg132), acting as a bridge between the flavine mononucleotide (FMN) of CPR and the heme iron.[16] Fe(II) of heme now binds an oxygen molecule, and a second electron is transferred, yielding - after protonation - a ferric hydroperoxyl complex that rapidly separates one water molecule and results in a ferryl-coupled porphyrin radical cation. Finally, the highly reactive radical cation oxidizes the substrate, and after the product egress, the cycle is completed.[17]

CPR undergoes a complex mechanism consisting of open and closed states to provide electrons to CYP. First, the closing of the FMN domain establishes an interaction between this domain and the flavine dinucleotide (FAD) domain.[18] Due to this interaction, electrons of FAD, originating from NADPH, are transferred to FMN. Second, CPR reopens and forms a complex with CYP, mainly driven by polar interactions. The surface of the CYP enzyme is positively charged in the interaction region by several arginine and lysine residues (in the case of CYP2D6: Arg129, Arg133, Arg140, Lys146, Lys429, Arg440, and Arg450), while the surface of the FMN domain is negatively charged due to aspartic acid and glutamic acid residues.[19]

Effects on CYP2D6 reactivity related to different alleles, including combinations of mutations, were investigated previously in wet lab experiments and computational studies.[9, 20-22] Due to the many naturally occurring alleles, only investigations of a subgroup of alleles have been possible. In these studies, different aspects of how mutations can affect the catalysis steps have been pointed out. I) Decreased activity can be correlated with an instability of the expressed enzyme variant.[23] II) Mutations of Phe120 and Phe483 affect the substrate binding.[24-26] III) As to the arginines and lysines playing a crucial role in the interaction between CYP and CPR, the allele CYP2D6*31 (combining SNPs R296C, R440H, and S486T; listed in the Pharmacogene Variation Consortium database) has a mutation in one of these positions, likely underlying that this CYP2D6 variant has no enzymatic activity *in vivo* and *in vitro*.[27] IV) The bridging arginine between FMN of CPR that supports the electron transfer [16] is another crucial residue; as a substitution probably prevents electron transfer to the heme, i.e., leads to inactivity, no substitution has been documented.

In this work, we aim to elucidate the effect of mutations found in naturally occurring alleles on CYP2D6 activity on an atomistic level. Due to the large number of natural mutants, our study focuses on the twenty most common SNPs in the global population (Figure 1B). We performed extensive all-atom MD simulations with a cumulated simulation time of over 200 μs. On this basis, we pursue a mechanism- and hypothesis-driven evaluation to detect if and how an SNP probably leads to a decreased reactivity of the mutant.

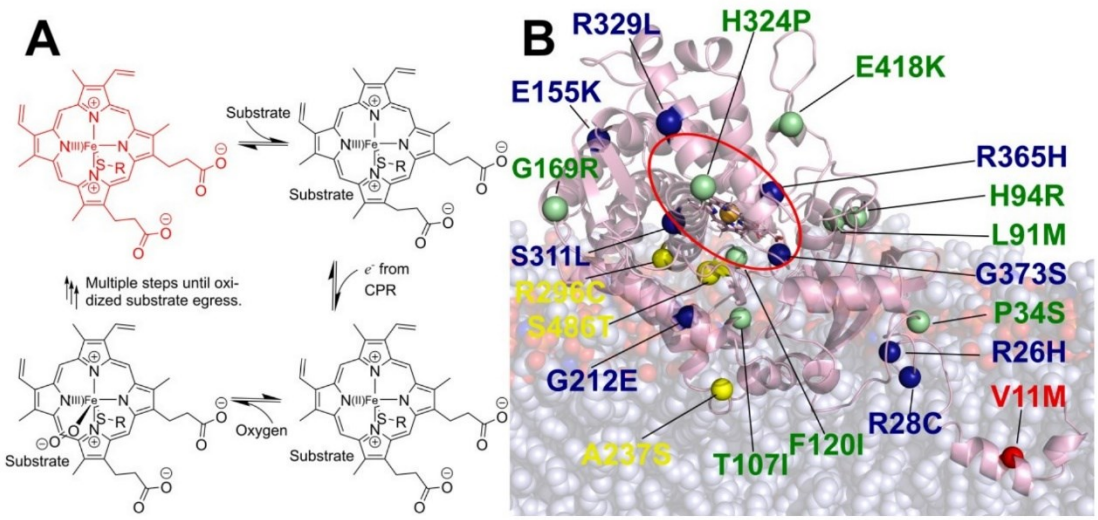



Figure 1**: CYP2D6 is a heme-containing, membrane-associated enzyme. A:** Parts of the reaction cycle of CYP enzymes underline the importance of substrate binding as well as electron transfer from CPR. The scheme was adapted from ref.[13]. **B:** Overview of the 20 most frequent SNPs mapped onto CYP2D6 (pink). Mutations that lead to poor metabolism are colored in green, variants with no effect are colored in yellow, V11M leading to an ultra-rapid metabolization is colored in red, and variants with unknown effects are colored in blue. Cα atoms of the positions are shown as spheres. The orange sphere in the center indicates the heme iron in the center of the porphyrin ring, which is shown as sticks.

# Materials and methods

## Structural models of CYP2D6 mutants

We used the equilibrated structural model of CYP2D6 wildtype generated in our previous work by modeling the globular part and transmembrane helix separately with TopModel[28] and docking both structures with spatial restraints as a basis[29]. Repeating the modeling of CYP2D6 wildtype by AlphaFold v2[30] yields a root-mean-square deviation (RMSD) < 2 Å, confirming the accuracy of our previous modeling. The selected mutations were introduced into the structural model using PyMOL, thereby paying attention that the rotamer with the least clashes was chosen.[31] The protonation state of the protein residues was estimated for pH 7.4 with PROPKA3[32] using HTMD.[33] The generated models were embedded by PACKMOL-Memgen[34] into a bilayer membrane with a composition reflecting the main lipid components of the human endoplasmic reticulum[35] (CHL:DOPC:DSPC:DAPC:DOPE 10:22:13:19:21). Potassium chloride was added with 0.15 M concentration. The orientation of the structural model in the membrane was determined with MEMEMBED.[36]

## Molecular dynamics simulations

The GPU particle mesh Ewald implementation[37] of the AMBER18[38] molecular simulations suite was used with ff14SB parameters[39] for the protein, Lipid17 parameters[40] for the membrane, and OPC as the water model[41] and Li/Merz parameters for the ions[42]. Parameters for the heme and cystine residue forming the S-Fe bridge between heme and the protein were taken from Shahrokh *et al.*[43]. Because covalent bonds to hydrogens were constrained with the SHAKE algorithm[44], a time step of 2 fs was used. The cutoff for nonbonded interactions was set to 10 Å.

Production runs were prepared as described by Schott-Verdugo et al.[45], but the minimization without restraints was performed for 40,000 steps. Langevin dynamics[46] with a friction coefficient of 1 $ps^{-1}$ was used to keep the temperature at 310 K. Initially, the minimized systems were heated by gradually increasing the temperature from 0 K to 310 K for 280 ps under NVT conditions. The system density was adjusted using NPT conditions at 1 bar for 2 ns. Production runs were then performed using the same NPT conditions. The pressure was controlled with the Berendsen barostat[47] with semi-isotropic pressure scaling, coupling the membrane (x,y) plane.

Geometric analyses were performed with cpptraj[48] from the AmberTools suite[49].

## Conformational sampling

To improve the robustness of the analyses and quantify the statistical uncertainty of the results, we carried out all analyses on ensembles of conformations generated from ten MD trajectories of 1 μs length for each of the enzymes. The independence of the replicas was ensured by using heating processes with a different random seed for each replica. The structures remain structurally stable with a backbone RMSD < 3.5 Å compared to the starting structure, considering the protein part without the transmembrane helix, which moves independently from the globular part.

## Constraint network analysis

We analyzed static properties, i.e., structural rigidity and its opposite flexibility,[50] of the 20 CYP2D6 mutants and the wildtype. Therefore, the enzymes were represented as constraint networks, where atoms are the nodes, and covalent and noncovalent bonds constitute constraints in between.[51] Noncovalent interactions such as hydrogen bonds, salt bridges, hydrophobic tethers, and stacking interactions contribute most to biomolecular stability. Using an empirical energy function allows for quantifying the strength of hydrogen bonds and salt bridges.[52] By gradually removing these polar noncovalent constraints from an initial network representation of a biomolecule according to a cutoff energy $E_{\mathrm{cut}}$, a succession of network states $\sigma$ is generated that forms a 'constraint dilution trajectory'.[53, 54] To this end, hydrogen bonds and salt bridges are removed in the order of increasing strength such that for network state $\sigma$, only those hydrogen bonds are kept that have an energy $E_{\mathrm{HB}} \leq E_{\mathrm{cut}}(\sigma)$. Performing rigidity analysis[55] on such a trajectory reveals a hierarchy of structural stability that reflects the modular structure of biomolecules in terms of secondary, tertiary, and super-tertiary structure.

By this, we obtained for each system a neighbor stability map ($rc_{ij,\mathrm{neighbor}}(E_{\mathrm{cut}}(\sigma))$) that contains information accumulated over all network states $\sigma$ along the trajectory [56, 57] in that it monitors the persistence of rigid contacts for pairs of residues during a constraint dilution process. From the stability maps, we calculated $E_{i,\mathrm{CNA}}$ (eq. 1), the sum of energies associated with rigid contacts between residue $i$ to all other residues.

$$E_{\mathrm{i,CNA}} = \sum_{j=1}^{n} rc_{ij,neighbor} \qquad \textbf{(eq. 1)}$$

We also calculated $\bar{E}_{\mathrm{region,CNA}}$ (eq. 2), where "region" is either the entire protein, the F/G region, or the I-helix, as the average of the chemical potential of rigid contacts between the $n$ residues within a region.

$$\bar{E}_{\text{region,CNA}} = \frac{1}{n} \sum_{i=region_{start}}^{region_{end}} E_{\text{i,CNA}} \qquad \textbf{(eq. 2)}$$

The computations were done with the Constraint Network Analysis (CNA) program (version 4.0) developed by us.[54] For further details, see our previous work[29].

## Generation of a structural model of the CYP2D6-CPR complex

The structural model of the complex of CYP2D6 and the FMN domain of CPR (Uniprot ID: P16435) was obtained through protein-protein docking using the HADDOCK webserver with default settings[58, 59] and is used for visualization purposes. For the FMN domain, residues 142, 144, 147, 179, and 209 were considered involved in the interaction. For CYP2D6, residues 129, 133, 140, 146, 429, and 440 were considered involved. The involved residues were chosen based on literature[15] and serve as putative interacting residues during the docking.

# Results

## Identification of relevant SNPs

The most common SNPs were identified based on the Uniprot entry (Uniprot-ID: P10635),[60] which reports on 50 SNPs. We restricted ourselves to the 20 most frequent mutations with a relative frequency of at least $10^{-4}$ in the global population. This frequency was taken from the dbSNP database.[61] Data from ClinVar,[62] PharmVar,[27] and the Human Gene Mutation Databank (HGMD)[63] were used to assess the clinical relevance of the mutations (see Table 1). For variants leading to decreased reactivity, we used the term "poor", for the wildtype and variants with no change in reactivity, the term "extensive" was used, the variant associated with ultrarapid metabolism was labeled as "ultrarapid", and the variants with unknown effects on reactivity were labeled as "unknown".

**Table 1:** Classification of clinical effects of CYP2D6 SNPs investigated here and source of the classification.

| SNP | ClinVar[62] | HGMD[63] | PharmVar[27] | Class |
|---|---|---|---|---|
| WT | | | | extensive |
| V11M | likely benign | ultrarapid metabolizer | | ultrarapid |
| A237S | likely benign | | normal | extensive |
| R296C | likely benign | | | extensive |
| S486T | likely benign | | normal | extensive |
| P34S | poor metabolism | poor metabolizer | | poor |
| L91M | | poor metabolizer | | poor |
| H94R | | poor metabolizer | | poor |
| T107I | likely benign | poor metabolizer | | poor |
| F120I | | poor metabolizer | | poor |
| G169R | poor metabolism | poor metabolizer | | poor |
| H324P | likely benign | poor metabolizer | no function | poor |
| E418K | | reduced enzyme activity | | poor |
| R26H | | | | unknown |
| R28C | | | | unknown |
| E155K | | | | unknown |
| G212E | | | | unknown |
| S311L | | | | unknown |
| R329L | | | | unknown |
| G373S | | | | unknown |
| R365H | | | | unknown |

## General strategy to identify atomistic mechanisms

For this study, we assumed that mutations documented as extensive and ultrarapid can be used as a comparison group (CG) to identify the atomistic mechanisms leading to decreased activity in the other variants. The variants were investigated as to different aspects relevant to the catalyzed reaction, i.e., enzyme stability, substrate binding, binding of CPR, electron transfer, and substrate access as detailed below. We then defined criteria for determining whether an aspect is significantly different from the CG, which could also be applied to characterize other variants not included in this dataset.

To do this, for a mechanism $A$, we identified the upper or lower limit (depending on the mechanistic relation) of a criterion $u_{\mathrm{A,limit}}$ (eq. 3) that shows a significant difference for a given significance level $p$ with the average criterion of the control group $\bar{u}_{A,\mathrm{CG}}$ with standard deviation $\sigma_{A,CG}$ (eq. 4). To obtain the most robust $u_{\mathrm{A,limit}}$ estimate according to our dataset, we respectively considered the largest standard deviation $\sigma_{A,nonCG,max}$ incurred for any of the variants that are not in the control group (nonCG).

$$u_{A,limit} = \min\left\{u_{test} \middle| T\left(\bar{u}_{A,CG}, \sigma_{A,CG}, u_{test}, \sigma_{A,nonCG,max}\right) < p \wedge u_{test} > \bar{u}_{A,CG}\right\}$$

$$\text{or} \qquad \textbf{(eq 3).}$$

$$u_{A,limit} = \max\left\{u_{test} \middle| T\left(\bar{u}_{A,CG}, \sigma_{A,CG}, u_{test}, \sigma_{A,nonCG,max}\right) < p \wedge u_{test} < \bar{u}_{A,CG}\right\}$$

where $p$ was set to 0.05 and $T$ is the $t$-test function implemented in SciPy v1.7.3.[64]

$$\bar{u}_{A,CG} = \frac{1}{5}\sum_{V \in CG}\frac{1}{10}\sum_{i=1}^{10} u_{A,V,i} \qquad \textbf{(eq. 4)}$$

## Structural instability of the I-helix

The first molecular mechanism relates to a loss of rigid contacts from the I-helix (residues L293-L323) to other parts of CYP2D6 as deduced from CNA. The I-helix is the central and one of the least mobile helices within the enzyme (Figure 2A, S1A,B). The loss of rigid contacts is indicated by a decreased average chemical potential (eq. 2) in this region, which is significant only for the substitution H324P with respect to CG and all other nonCG ($p < 0.01$, two-sided $t$-test) (Figure 2B).

From a structural viewpoint, proline is known for disrupting the secondary structure,[65] which may be intensified by the proline at position 325 already being present in the wild type. In the H324P variant, the $\phi$ dihedral angle ($C_{i-1}$-$N_i$-$C\alpha_i$-$C_i$) of this residue changes to ~ -75°

(loop conformation) from ~ -120° (β-sheet-like conformation) in the wildtype (Figure 2C). Due to this change, the loop between I-helix and J-helix is elongated by one residue (WT: L323-P325; H324P: L323-D326), which may contribute to destabilizing the I-helix. Furthermore, the replacement of the histidine sidechain leads to more space for the C-terminal loop of CYP2D6, which also contains C443 interacting with the heme group (Figure S2A,B,C). With H324, the C-terminal loop formed more rigid contacts to the I-helix than with a proline at this position, which led to the decrease in structural stability of the I-helix (Figure 2B).

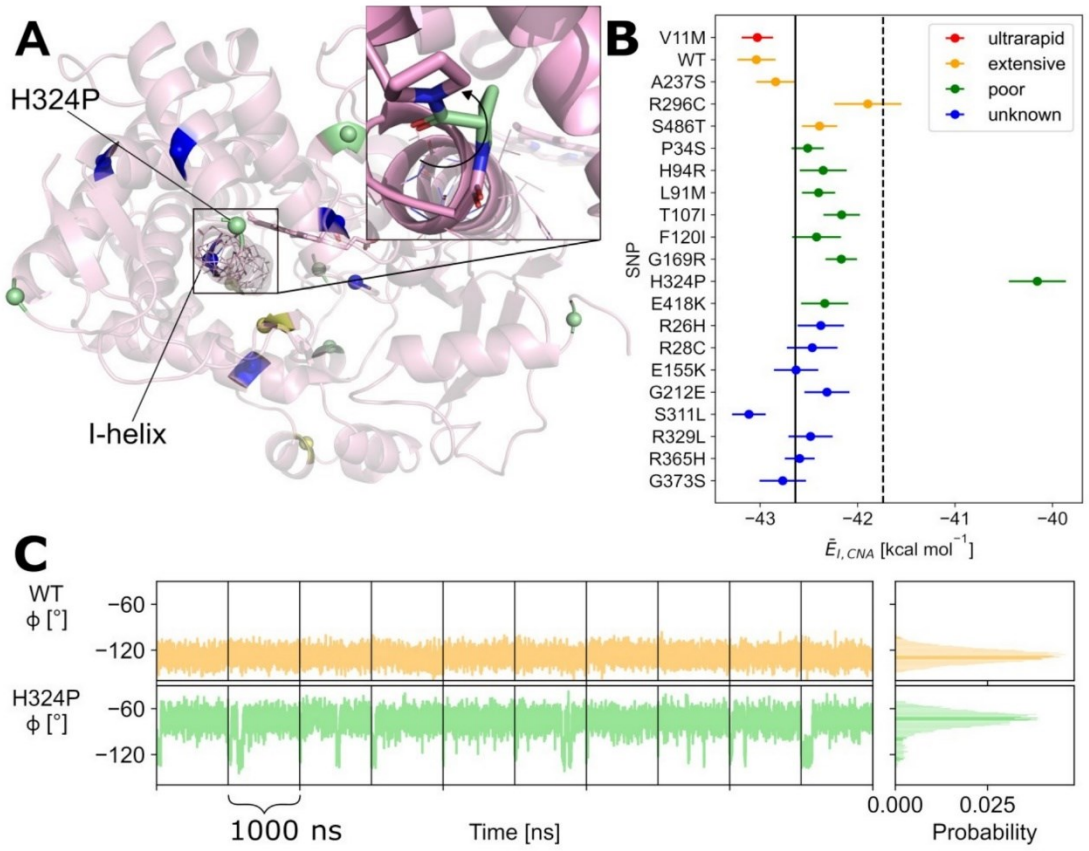


Figure 2: **Structural instability of the I-helix leads to poor metabolism of CYP2D6 H324P. A:** Structure of the globular part of CYP2D6 (pink) shown with the I-helix marked with backbone lines and positions of SNPs as shown in Figure 1. The heme group is shown as sticks. In the blowup, the H324P variant is shown with the measured ϕ angle indicated. **B:** Average chemical potential of rigid contacts of I-helix residues ($\bar{E}_{I,CNA}$) shown for each investigated SNP. For all variants except H324P, the value is approximately between -43 and -42 kcal mol$^{-1}$. Only H324P differs significantly ($p < 0.01$, two-sided t-test) from the comparison group, leading to a decreased structural stability of the I-helix. The continuous vertical line denotes the mean of $\bar{E}_{I,CNA}$ of the comparison group (eq. 2); the error bars denote the SEM determined by error propagation along 10 independent simulations. The dashed vertical line denotes the limit for a significantly different value ($p < 0.01$, two-sided *t*-test, eq. 3). **C:** The ϕ angle, i.e., dihedral angle $C_{i-1}$-$N_i$- $C\alpha_i$ -$C_i$, of residue 324 in the wildtype and the H324P variant is shown over the simulation time (left) and as frequency distribution (right). Vertical lines separate the 10 replicas. In each replica of H324P, a shift to ~-75° is visible, while the dihedral remains at ~ -120° in all replicas of the wild type.

## Distortion of the binding site integrity

F120 and F483 are important in substrate binding, especially for aromatic substrates (Figure 3A).[24-26] The two phenylalanines build a hydrophobic dome that narrows down the binding pocket and builds a hydrophobic environment close to the heme necessary for the correct binding of the substrate.[66] In our MD simulations, we saw large and significant differences ($p < 0.01$, two-sided *t*-test) in the distance between both phenylalanines in the variants T107I and G169R compared to the CG (Figure 3B). This suggests that substrate binding could be disfavorably impacted.

For T107I, the change in the distance is probably caused by a conformational change in the C'-helix where a hydrogen bond from the threonine hydroxyl group to the V104 backbone oxygen cannot be established anymore. The missing interaction leads to a higher variance of the ψ dihedral angle ($N_i$-$C\alpha_i$-$C_i$-$N_{i+1}$) at P105, with significantly more positive values in T107I than in the wildtype (Figure S3) ($p < 0.01$, two-sided *t*-test). The C'-helix is linked to the B-C-loop, which comprises F120 (Figure 3C).

G169R is part of the D-E-loop that interacts with the C-terminal loop by a hydrogen bond between F172 in the D-E-loop and L492 one residue before the β3-2-strand (Figure S4A,B,C). In the G169R variant, we observed a decreased distance between the $C_\alpha$ of R169 and the $C_\alpha$ of P496, which is the last residue of the β3-2-strand caused by electrostatic interactions of the guanidino group of R169 with the S168 hydroxyl group and the backbone nitrogen of P496 (Figure S4D-H). Due to the interaction between both loops, the shift in the D-E-loop is transferred to the C-terminal loop, which comprises F483.

In the F120I variant, the binding of aromatic substrates is likely hindered due to the missing π-interactions; at least one aromatic ring occurs in 99% of a database containing more than 3,500 compounds published by the medicinal chemistry departments of AstraZeneca, Pfizer, and GlaxoSmithKline.[67] Thus, the variants T107I, G169R, and F120I are probably poor metabolizing enzymes due to changes in the substrate binding strength. If we apply the criterion to the group of unknown SNPs, R26H, R28C, and S311L lead to significantly higher distances between F120 and F483, which suggests that substrate binding is also decreased for these variants.

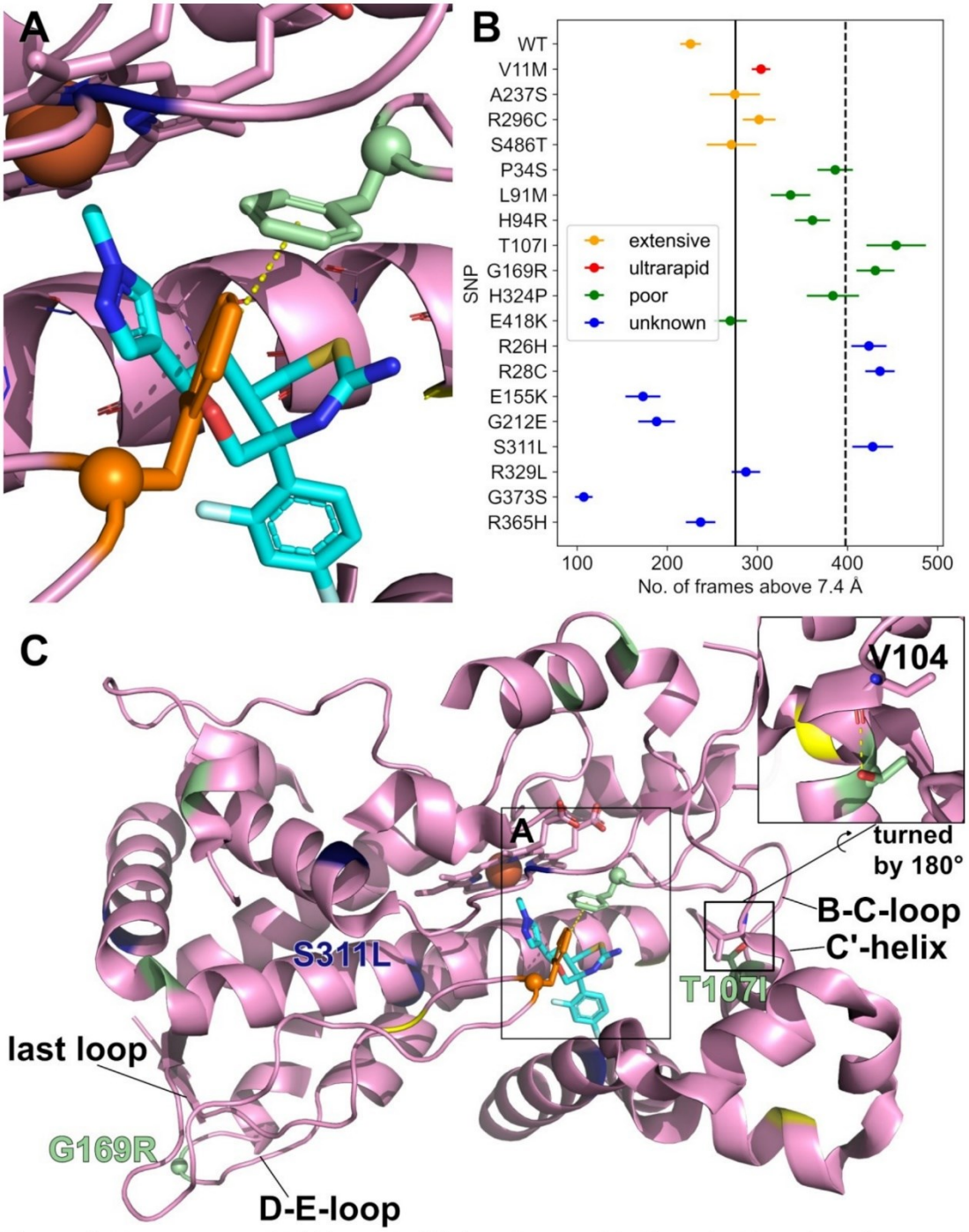



**Figure 3: Importance of F120 and F483 in substrate binding. A:** Zoom into the crystal structure of the active site of CYP2D6 (pink) binding inhibitor BACE-1 (cyan) (PDB-ID: 4XRY). The heme and ligand are shown as sticks; F120 is marked in green and F483 in orange. The iron within the heme is shown as a sphere. **B:** The number of frames with a distance between F120 and F483 above 7.4 Å is depicted for all investigated variants, except F120I. The distance was measured between the phenyl rings. The continuous vertical line denotes the mean distance between F120 and F483 of the comparison group (eq. 3); the error bars denote the SEM determined by error propagation along 10 independent simulations. The dashed vertical line denotes the limit for a significantly different value ($p < 0.01$, two-sided t-test, eq. 3). **C:** Crystal

structure of CYP2D6 bound to BACE-1. Locations of mutations T107I and G169R are labeled. The influence on the distance between F120 and F483 could be explained by the vicinity of both mutations to the loops F120 (green) and F483 (orange) are part of. T107 is stabilizing the C' helix via hydrogen bonding with the backbone oxygen of V104 shown in the blowup. Breaking this by an exchange of threonine by isoleucine leads to conformational changes of the B-C-loop, of which F120 is part. In the G169R variant, a small glycine is mutated to a voluminous arginine in the D-E-loop, which is close to the last loop in the sequence behind the L-helix, of which F483 is part.

## Reduced interaction with Cytochrome P450 reductase

For the reaction cycle of CYP enzymes, the transfer of an electron from Cytochrome P450 reductase (CPR) is mandatory. CPR itself undergoes a complex cycle that requires binding and unbinding from the CYP enzyme.[68] The binding between CYP and CPR is mediated by basic residues (positive charges) on the CYP side and acidic residues (negative charges) on the CPR side.[15, 69] For CYP2D6, the basic residues are: R129, R133, R140, K146, K429, R440, and R450 (Figure 4A,B). For successful binding, the availability of these residues plays a crucial role. All of them do not form intramolecular interactions during the MD simulations except R450, which forms a salt bridge with E150. This salt bridge occurs in all investigated variants, but more often in L91M (significant: $p < 0.01$, t-test) and G373S (Figure 4C). This is also reflected by the average chemical potential of the rigid contact between R450 and E150 ($\bar{E}_{\text{R450-E150,CNA}}$; eq. 2) (Figure 4D). Hence, the formation of this salt bridge may lead to reduced binding of CYP2D6 L91M and CYP2D6 G373S to CPR and a lack of electrons for the metabolizing reaction.

L91M influences the salt bridge prevalence between R450 and E150 via the K''-L-loop of CYP2D6 (Figure 4E). L91M is part of the B-helix, which is close to the K''-L-loop. In MD simulations of the L91M variant, we observed a significantly decreased distance between the K'-L-loop ($C_\alpha$ of S437, G439, and R441) and the β1-3-strand ($C_\alpha$ of L395) compared to the CG, which is located below the B-helix (Figure S5A-E) ($p < 0.01$, two-sided *t*-test). The conformation of the K''-L-loop itself does not change (Figure S5F). The movement of the K-L-loop leads to a change in the relative position of R450, which is close to the L-helix, which increases the prevalence of salt bridge formation with E150.

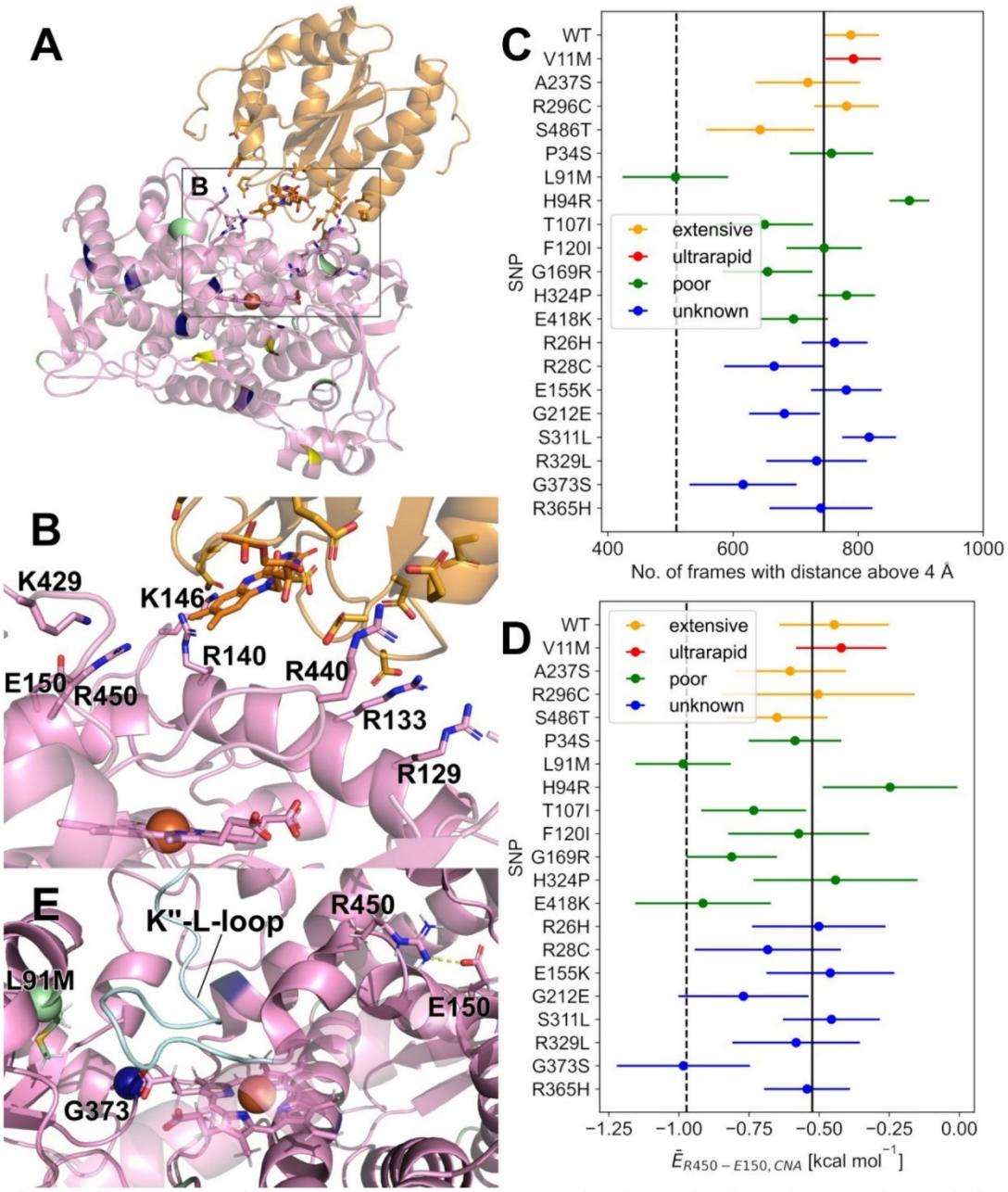



**Figure 4: Arginine 450 plays an important role in CPR binding. A:** Interface of CPR (yellow) and CYP2D6 (pink) derived by docking using the HADDOCK web server [58]. E150 does not take part in the interaction with CPR but can form a salt bridge with R450. Heme and charged residues within the interface are shown as sticks. The iron within the heme is shown as a sphere. The residues are colored according to their effect on metabolism. **B:** The close-up view reveals the importance of charged residues for the interaction between CPR and CYP2D6. While the interface of CYP2D6 is mainly positively charged, the interface of CPR is negatively charged due to many glutamic and aspartic acids. **C:** The number of frames with a distance between R450 and E150 above 4 Å is depicted for all investigated variants. The distance was measured as the closest distance between all combinations of sidechain N's of arginine and sidechain O's of glutamic acid. Above 4 Å, the salt bridge is not considered formed[70]. Then, R450 is free to interact with CPR. The continuous vertical line denotes the mean distance between R450 and E150 of the CG; the error bars denote the SEM determined by error

propagation along 10 independent simulations. The dashed vertical line denotes the limit for a significantly different value ($p < 0.01$, two-sided t-test). **D:** Average chemical potential of rigid contacts ($\bar{E}_{\text{R450-E150,CNA}}$) between R450 and E150. The salt bridge in the L91M and the G373S variants is significantly more stable than in all other variants. The continuous vertical line denotes the mean of $\bar{E}_{\text{R450-E150,CNA}}$ of the CG; the error bars denote the SEM determined by error propagation along 10 independent simulations. The dashed vertical line denotes the limit for a significantly different value ($p < 0.01$, two-sided t-test). **E:** The close-up view on CYP2D6 L91M shows that L91M but also G373S may influence the salt bridge stability between R450 and E150 by a shift of the K''-L-loop (cyan) due to larger space requirements of the substitutions compared to the wildtype residues. The locations of L91M and G373 are shown as spheres. The M91 side chain is shown as sticks, as are R450 and E150.

## Reduced electron transfer to heme mediated by R132

For the electron transfer from the flavin mononucleotide of CPR to the heme group of CYP, R132 of CYP, located between both cofactors, is needed (Figure 5A,B).[16] R132 is part of the C-helix and is on average 7.8 Å away from the heme iron. For the three variants E418K, H94R, and P34S with decreased reactivity, we found that the distance between R132 and the iron of the heme group is on average significantly increased compared to the CG (Figure 5C). Thus, these variants are probably poor metabolizing enzymes due to the lower probability of an electron transfer via R132. The same criterion is fulfilled by R26H, R329L, and G373S of the unknown group.

E418K is part of the K''-helix and at least 31.9 Å away from R132. Its sidechain points to the protein's surface (Figure S6A). Our MD simulations revealed a minor conformational change at the K''-helix, due to which the neighboring K429 is shifted towards the K''-L-loop (Figure S6B). Consequently, the distance to A449 is significantly decreased by 0.2 Å ($p < 0.01$, two-sided *t*-test) (Figure S6C). This shift is then transmitted from A449 to E446 (Figure S6D) and from E446 to K140 (Figure S6E), which is part of the C-helix and has a constant distance to R132 (Figure S6F). These changes lead to a significantly higher distance between R132 and heme iron (Figure S6G).

H94R is part of the B-helix, which is spatially located between the β2-1-strand and the K''-L-loop (Figure S7A-C). Our MD simulations revealed that the distance between the B-helix and the β2-1-strand (measured as the distance between Cα 94 and Cα 383 as the center of the β2-1-strand) is significantly increased by 0.7 Å ($p < 0.01$, two-sided *t*-test) (Figure S7D). Due to this shift, the interaction between the C-terminus of the B-helix and the C'-C-loop is affected, i.e., the hydrogen bond between E96 and Y124 is less often formed in H94R than the wildtype ($p < 0.01$, two-sided *t*-test) (Figure S7E). This leads to an overall higher displacement of the R132-containing C-helix in relation to the heme (Figure S7F).

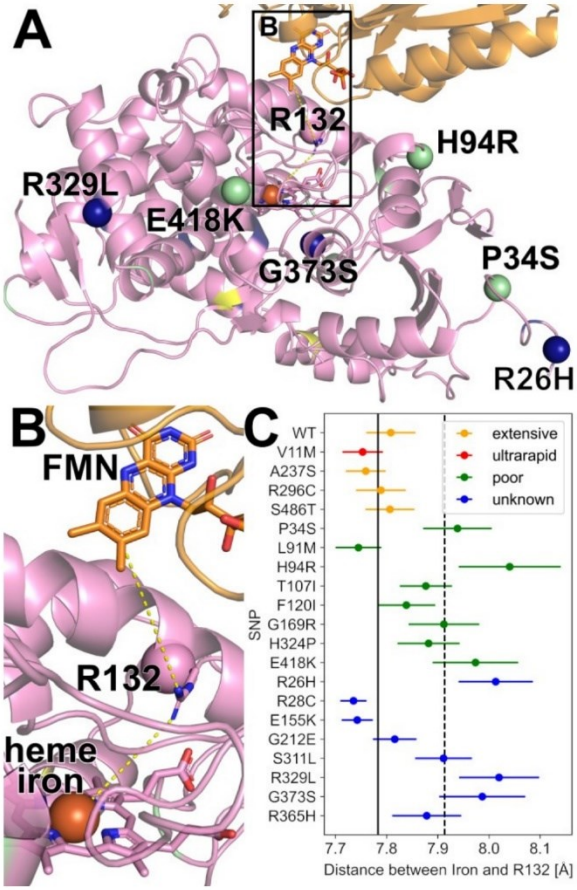



**Figure 5: R132 plays a crucial role in electron transfer from CPR to CYP2D6. A:** Interface of CPR (yellow) and CYP2D6 (pink) derived by docking using the HADDOCK web server [58]. Heme and flavin mononucleotide are shown as sticks as is R132. The iron within the heme is shown as a sphere. All Cα-atoms of substitutions in variants that show a significantly higher distance between iron and R132 are shown as a sphere. **B:** Close-up view of the interface between CYP2D6 and CPR. The route of the electron transfer according to ref.[16] is indicated with dashed yellow lines. Thus, the distance between R132 and the heme iron is an important factor for the likelihood of an electron transfer. **C:** Average distance between R132 and the heme iron. The continuous vertical line denotes the mean of the distance of the CG; the error bars denote the SEM determined by error propagation along 10 independent simulations. The dashed vertical line denotes the limit for a significantly different value ($p < 0.01$, two-sided *t*-test). The distance is measured as the closest distance between the terminal nitrogen atoms of arginine and iron.

P34S is part of the loop between the TM-helix and A-helix. Together with Y33, it stabilizes the β2-sheet by holding Y33 in between V68 (β1-1-strand) and F387 (β2-2-strand) (Figure S8A-C). In our MD simulations of P34S, we observed that Y33 moves away from this position ($p < 0.01$, two-sided *t*-test), which leads to a reorientation in the β2-2-strand, and the aromatic ring of F387 adopts the position of Y33 (Figure S8D). The consequence of the different conformation of the β2-2-strand is a shift of the B-helix, which interacts with the β2-strand as

described above. The shift of the B-helix is transmitted via the interaction of E96 and Y124 to the C-helix, of which R132 is part.

## Unchanged stability of the F/G-Region

In our previous work, we showed that the structural stability of the F/G-region (residues P200 – M260), containing the F- and G-helices next to the substrate entrance channel (Figure 6A), determines the substrate promiscuity of drug-metabolizing human CYP enzymes.[29] This is probably caused by the lid effect of this region. The stability of the F/G-region was determined by the average chemical potential due to rigid contacts in this region ($\bar{E}_{\mathrm{FG,CNA}}$, eq. 2).

Here, we hypothesized that the overall reactivity of CYP2D6 can be decreased by substitutions that increase the structural stability of the F/G-region, as this would hamper substrate access due to a less mobile lid function. We, thus, computed $\bar{E}_{\mathrm{FG,CNA}}$ for all variants with the same approach used before.[29] However, differences between any of the variants and the WT were insignificant (Figure 6B). As a corollary of this finding, we expect that the substrate promiscuity of all investigated variants is similar to the wildtype.

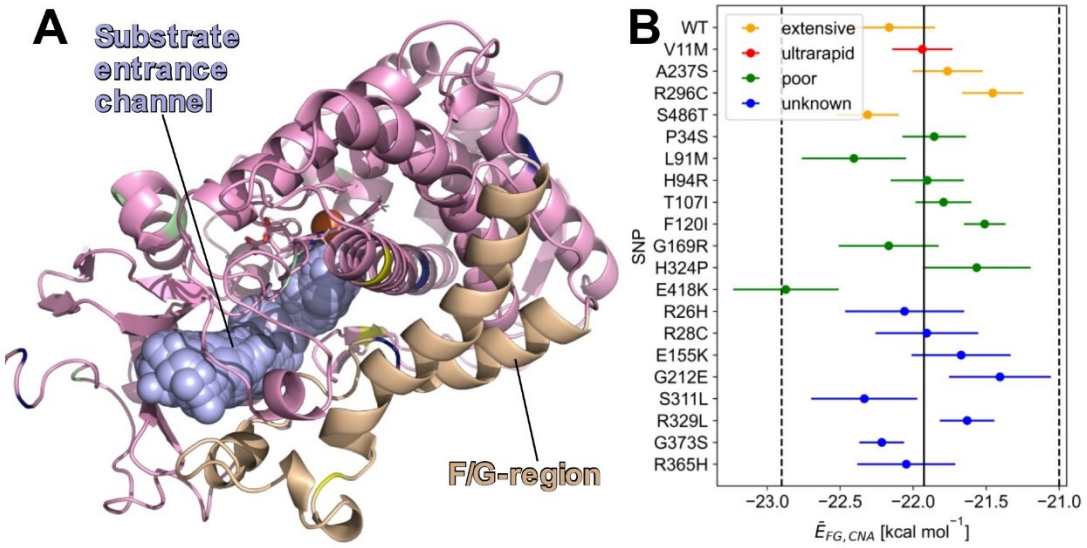



**Figure 6: The structural stability of the F/G-region does not differ significantly between the different variants. A:** The position of the substrate entrance channel (cyan) in CYP2D6 is next to the F/G region(beige). The rigidity of the F/G-region is correlated with substrate promiscuity in human CYP enzymes.[29] Other parts of CYP2D6 are colored pink. Substitutions that lead to poor metabolism are colored in green, substitutions leading to an extensive metabolism are colored in yellow, V11M leading to an ultra-rapid metabolization is colored in red, and substitutions with unknown effects are colored in blue. **B:** To estimate the substrate promiscuity of the different isoforms, the structural stability of the F/G region ($\bar{E}_{FG,CNA}$) was calculated[29]. The continuous vertical line denotes the mean of $\bar{E}_{FG,CNA}$ of the CG; the error bars denote the SEM determined by error propagation along 10 independent simulations. The dashed vertical line denotes the limit for a significantly different value ($p < 0.01$, two-sided *t*-test). Since no substitution leads to a significant difference, the substrate promiscuity of the variants is likely similar to the wild type.

## Decision tree to predict substitutions leading to poor metabolism

The above-analyzed aspects relevant to the CYP2D6-catalyzed reaction can be arranged in a hierarchy with associated criteria for when an aspect is significantly different from the CG. The resulting decision tree (Figure 7A) shall allow us to predict when a CYP2D6 substitution leads to poor metabolism.

The importance of the I-helix as the central helix of the enzyme entails that if the enzyme is destabilized in this region, it loses its function.[71] Thus, we use the structural stability of the I-helix as the most relevant aspect to determine if CYP2D6 activity is reduced. Compared to the CG of five variants with extensive or ultrarapid activity, an average chemical potential (eq. 2) in this region $\bar{E}_{I,CNA}^{lim.}$ = -41.57 kcal mol$^{-1}$ is significantly different even when considering the maximum standard deviation of the nonCG ($p < 0.01$, two-sided t-test) (Figure 2B). Thus, variants with $E_{I,CNA} > E_{I,CNA}^{lim.}$ are considered to lead to poor metabolism.

The substrate binding is a crucial step in the reaction process, which we consider the second most relevant aspect to determine if CYP2D6 activity is reduced. In the crystal structure of CYP2D6 bound to BACE1 (PDB ID: 4XRY), F120 and F483 stabilize the aromatic substrate (Figure 3A), and the distance between the ring systems of both residues is 7.4 Å. Assuming that this distance is relevant for substrate binding, we investigated the persistence with which this distance is above 7.4 Å during our MD trajectories. Compared to the CG, a persistence of 4% is significantly different ($p < 0.01$, two-sided t-test) (Figure 3B). Thus, variants with a persistence > 4% are considered to lead to poor metabolism.

The transfer of an electron from CPR is mandatory for the reaction cycle of CYP enzymes. Binding between both proteins is a prerequisite, which we consider the third most relevant aspect to determine if CYP2D6 activity is reduced. The unavailability of R450 of CPY2D6 as an interaction partner in the interface due to intramolecular salt bridge formation with E150 was analyzed in terms of the average chemical potential (eq. 2) of this interaction, $\bar{E}_{\text{R450-E150,CNA}}$. Compared to the CG, $\bar{E}_{\text{R450-E150,CNA}}^{lim.} = -1.0$ kcal mol$^{-1}$ is significantly different ($p < 0.01$, two-sided t-test) (Figure 4D). Thus, variants with $E_{\text{R450-E150,CNA}} < E_{\text{R450-E150,CNA}}^{lim.}$ are considered to lead to poor metabolism.

We considered the transfer of an electron mediated by R132 as the last aspect to determine if CYP2D6 activity is reduced. As a criterion, the distance between iron and R132 was evaluated. Compared to the CG, a distance of 7.91 Å is significantly different ($p < 0.01$, two-sided t-test) (Figure 5C). Thus, variants with a distance > 7.91 Å are considered to lead to poor metabolism.

Following this decision tree, all variants of the CG and all variants that lead to known poor metabolism are correctly classified, with the first to fourth aspect leading to a decision in 1, 3, 1, and 3 cases for the latter group, respectively (Figure 7B). Application of this decision tree to variants with unknown effects leads to the predictions that variants R26H, R28C, S311L, R329L, and G373S result in poor metabolism.

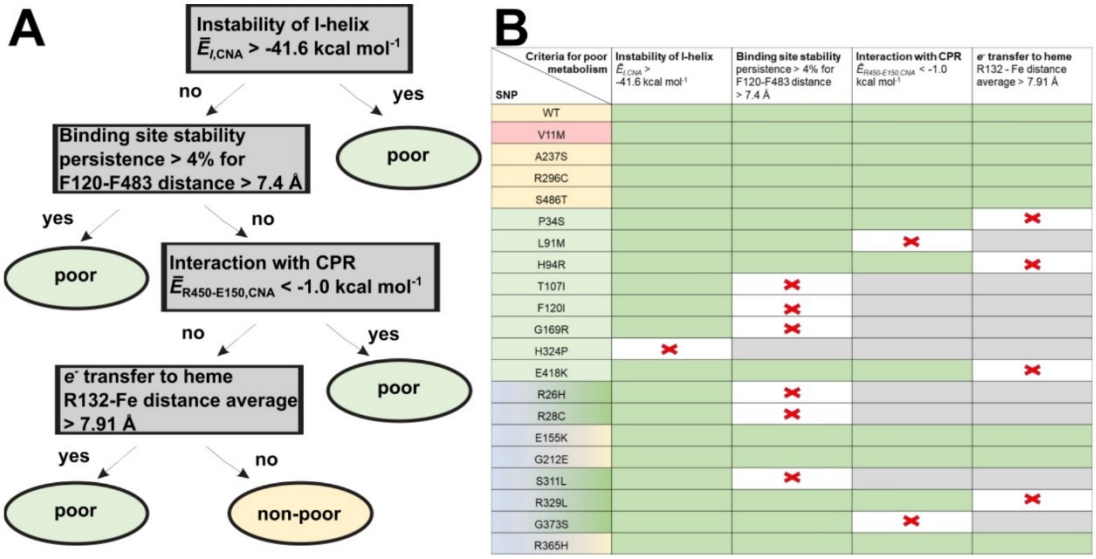



**Figure 7: Decision tree for application of the derived criteria to the variants with unknown effect. A:** Decision tree to distinguish poorly metabolizing variants from the CG. **B:** Aspects that lead to the prediction of a poor metabolism are marked with a red cross. Aspects that were not met are visualized with a green background, and aspects that were not checked due to the hierarchy of the decision tree are marked with a grey background. As a result, R26H, R28C, and S311L are predicted to lead to poor metabolism due to diminished binding site stability, whereas G373S is predicted to lead to poor metabolism due to diminished interaction with CPR. The first column is colored according to the (predicted) effect on the metabolism (poor metabolism: green; variants with no effect: yellow; V11M leading to an ultra-rapid metabolization: red; variants with unknown effects are colored with a color gradient from blue to the predicted effect on metabolism (green: poor; yellow: non-poor).

## Possible molecular explanations for variants predicted to result in poor metabolism

### Distortion of binding site integrity

R26H and R28C are part of the loop between the TM-helix and the A-helix and lead to a destabilization of the binding site as measured by the persistence with which the distance between F120 and F483 is above 7.4 Å during our MD trajectories. As residues 26 and 28 are at least 21.2 Å away from residues 120 and 483, this involves combined structural changes. F120 is pulled from the binding site due to a B'-helix (residues 105-109) movement (Figure S9A,B). The B'-helix moves due to a shift of the F'-helix. The shift is identified based on the constant distance between the B'- and F'-helices – measured as the distance between V223 (chosen as a reference point because V223 is the most central residue in the F'-helix) and Q108 (Figure S9C). The movement of the F'-helix is induced by a shift of the β-turn at W75. W75 is

part of a β-turn between β1-1 strand and β1-2 strand, and the shift lets the F'-helix move toward the TM-helix (Figure S9D). The β-turn shifts due to a - in both variants -more frequently occurring hydrogen bond between the sidechain oxygen of Q27 and the indol nitrogen of W75 (Figure S9E,F).

S311L is part of the I-helix (Figure S10A) and leads to a destabilization of the binding site as measured by the persistence with which the distance between F120 and F483 is above 7.4 Å during our MD trajectories. F483 is moving away from the binding site because of a shift of the C-terminal loop, which contains F483. The shift is induced by W316 in the I-helix, which forms CH-π-stacking interactions with P487 in the C-terminal loop (Figure S10B,C). W316 is shifted by the increased spatial requirement of leucine compared to serine at position 311. The S311L substitution does not impact the I-helix itself, e.g., the distance between $C_\alpha$ 311 and $C_\alpha$ 316 does not change.

**Reduced interaction with Cytochrome P450 reductase**

G373S is part of the K-K'-loop, located between the K-helix and β1-4-strand. The K-K'-loop is close to the carboxy groups of heme (Figure S11A). The substitution is predicted to lead to a reduced interaction of the variant with CPR as indicated by $\bar{E}_{\text{R450-E150,CNA}} < -1.0$ kcal mol$^{-1}$.

Due to a shift of E446 towards S373 in G373S, R450 is shifted closer to E150. The shift of E446 is induced by a shift of the K'-L-loop due to a decreased interaction between S437 and the heme carboxy group (Figure S11B). The distance between the S437 hydroxyl group, located in the K'-L-loop, and the heme carboxy group is significantly decreased by 0.3 Å compared to the CG (Figure S11C) ($p < 0.01$, two-sided *t*-test). The interaction between S437 and the heme carboxy group is decreased due to the frequently formed hydrogen bond between S373 in G373S, which cannot be formed in the GC.

**Reduced electron transfer to heme mediated by R132**

R329L is part of the J-helix and points to the protein surface. In the R329L variant, we observed a increase of $\bar{E}_{\text{J,CNA}}$ (eq. 2) of 0.6 kcal mol$^{-1}$. Thus, the J-helix forms less stable rigid contacts to the surrounding structural elements. This leads to higher values of $\bar{E}_{\text{J-K'-loop,CNA}}$ and $\bar{E}_{\text{C,CNA}}$, which explains the higher mobility of R132 in the C-helix.

## Discussion

In this study, we intended to elucidate the effect of mutations found in naturally occurring alleles on CYP2D6 activity on an atomistic level. Our results demonstrate that SNPs correlated to changes in enzymatic activity led to changes in four crucial aspects of the CYP-catalyzed reaction, the stability of the main helix I-helix, substrate binding, binding of CPR, and electron transfer from CPR to heme iron. By contrast, no significant impacts on the substrate uptake channels due to the investigated SNPs were found.

We identified the four mechanisms by the most extensive MD simulations on CYP2D6 variants performed so far. Overall, we performed full atomistic MD simulations of membrane-bound CYP2D6 wildtype and variants of a cumulated time of 210 µs. This value exceeds previous studies[22, 72, 73] by two orders of magnitude. The generated data allowed us to identify significant changes, even if they are small, among the different isoforms and versus the wildtype and to predict the behavior of unknown SNPs. For all four mechanisms, we identified parameters that can be used to distinguish between substitutions leading to a change in activity or not as well as are connected to the underlying molecular mechanisms. Also, the criteria match the current knowledge of the mechanism CYP enzymes undergo while oxidizing a substrate[12, 13, 74]

To overcome the challenge of 165 known natural variants of CYP2D6, we investigated the 20 most frequent SNPs. This approach provided detailed insights into their molecular mechanisms but precludes to detect combinations of SNPs that lead to a larger structural change, e.g., CYP2D6*53, that combines F120I and A122S and does not show a change in activity[75], even if F120I in CYP2D6*49 is associated with a decreased activity,[76] and F120I was predicted as a mutation that decreases activity by HGMD[77].

Interestingly, according to our analyses, the F/G-region stability and, by this, the substrate promiscuity is not changed by any of the investigated SNPs. This fits with the clinical practice because if a patient is identified as a poor metabolizer, metabolism will be poor for any CYP2D6 substrate.[78] Thus, the drug regime will need to be modified once for all substrates and not for all substrates separately.

The derived mechanisms and the criteria associated with the mechanisms can be used to identify the effects of unknown SNPs and to scrutinize all known variants as to the molecualr origins that lead to the clinical effect. This can especially be useful for clinical effects with low incidence.

In summary, we identified four distinct mechanisms that lead to poor metabolism of CYP2D6 variants with the help of extensive MD simulations and Constraint Network Analysis.

Our model shall allow us to predict if newly identified SNPs of CYP2D6 will lead to poor metabolism, which may be used for recommendations to modify drug regimes.

## Acknowledgments

We are grateful for computational support and infrastructure provided by the "Zentrum für Informations- und Medientechnologie" (ZIM) at the Heinrich Heine University Düsseldorf and the computing time provided by the John von Neumann Institute for Computing (NIC) on the supercomputer JUWELS at Jülich Supercomputing Centre (JSC) (user ID: HKF7, CYP450). The study was supported by Bundesministerium für Bildung und Forschung, Germany (BMBF) and the Department of Biotechnology, India (DBT) through project ReMetaDrug (funding number: 01DQ19002).

The FP7 WeNMR (project# 261572), H2020 West-Life (project# 675858), EOSC-hub (project# 777536), and the EGI-ACE (project# 101017567) European e-Infrastructure projects are acknowledged for the use of their web portals, which make use of the EGI infrastructure with the dedicated support of CESNET-MCC, INFN-PADOVA-STACK, INFN-LNL-2, NCG-INGRID-PT, TW-NCHC, CESGA, IFCA-LCG2, UA-BITP, SURFsara and NIKHEF, and the additional support of the national GRID Initiatives of Belgium, France, Italy, Germany, the Netherlands, Poland, Portugal, Spain, UK, Taiwan, and the US Open Science Grid.

## Author contributions

HG designed the study; DB performed computations; DB and HG analyzed results; DB and HG wrote the manuscript; PVB revised the manuscript; HG and PVB secured funding.

## Supporting information

The Supporting Information is available free of charge at:

Supporting materials and methods, supporting figures, and supporting references (PDF). A repository with used scripts, MD simulation data, and CNA results are available from researchdata.hhu.de via below provided DOI.

## Data and software availability

The CNA software is available under academic licenses from http://cpclab.uni-duesseldorf.de/index.php/Software. The CNA web server is accessible at http://cpclab.uni-duesseldorf.de/cna/.

## References

1. Rendic, S.; Guengerich, F. P., Survey of Human Oxidoreductases and Cytochrome P450 Enzymes Involved in the Metabolism of Xenobiotic and Natural Chemicals. *Chem. Res. Toxicol.* **2015**, 28, 38-42.
2. Mahgoub, A.; Dring, L. G.; Idle, J. R.; Lancaster, R.; Smith, R. L., POLYMORPHIC HYDROXYLATION OF DEBRISOQUINE IN MAN. *The Lancet* **1977**, 310, 584-586.
3. Kimura, S.; Umeno, M.; Skoda, R. C.; Meyer, U. A.; Gonzalez, F. J., The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and identification of the polymorphic CYP2D6 gene, a related gene, and a pseudogene. *Am. J. Hum. Genet.* **1989**, 45, 889-904.
4. Nebert, D. W.; Russell, D. W., Clinical importance of the cytochromes P450. *Lancet* **2002**, 360, 1155-1162.
5. Dahl, M. L.; Johansson, I.; Bertilsson, L.; Ingelman-Sundberg, M.; Sjöqvist, F., Ultrarapid hydroxylation of debrisoquine in a Swedish population. Analysis of the molecular genetic basis. *J. Pharmacol. Exp. Ther.* **1995**, 274, 516-20.
6. Løvlie, R.; Daly, A. K.; Molven, A.; Idle, J. R.; Steen, V. M., Ultrarapid metabolizers of debrisoquine: Characterization and PCR-based detection of alleles with duplication of the CYP2D6 gene. *FEBS Lett.* **1996**, 392, 30-34.
7. Johansson, I.; Lundqvist, E.; Bertilsson, L.; Dahl, M. L.; Sjöqvist, F.; Ingelman-Sundberg, M., Inherited amplification of an active gene in the cytochrome P450 CYP2D locus as a cause of ultrarapid metabolism of debrisoquine. *Proceedings of the National Academy of Sciences* **1993**, 90, 11825-11829.
8. Chiba, K.; Kato, M.; Ito, T.; Suwa, T.; Sugiyama, Y., Inter-individual variability of in vivo CYP2D6 activity in different genotypes. *Drug Metabolism and Pharmacokinetics* **2012**, advpub, 1201160353-1201160353.
9. Yu, A.; Kneller, B. M.; Rettie, A. E.; Haining, R. L., 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.
10. Guengerich, F. P. Human Cytochrome P450 Enzymes. In *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Ortiz de Montellano, P. R., Ed.; Springer International Publishing: Cham, 2015, pp 523-785.
11. Black, S. D., Membrane topology of the mammalian P450 cytochromes. *FASEB J.* **1992**, 6, 680-685.
12. Cojocaru, V.; Winn, P. J.; Wade, R. C., The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta, Gen. Subj.* **2007**, 1770, 390-401.
13. Schlichting, I.; Berendzen, J.; Chu, K.; Stock Ann, M.; Maves Shelley, A.; Benson David, E.; Sweet Robert, M.; Ringe, D.; Petsko Gregory, A.; Sligar Stephen, G., The Catalytic Pathway of Cytochrome P450cam at Atomic Resolution. *Science* **2000**, 287, 1615-1622.
14. Masters, B. S. S.; Okita, R. T., The history, properties, and function of NADPH-cytochrome P-450 reductase. *Pharmacol. Ther.* **1980**, 9, 227-244.
15. Hamdane, D.; Xia, C.; Im, S.-C.; Zhang, H.; Kim, J.-J. P.; Waskell, L., Structure and Function of an NADPH-Cytochrome P450 Oxidoreductase in an Open Conformation Capable of Reducing Cytochrome P450 *<sup> </sup>. *J. Biol. Chem.* **2009**, 284, 11374-11384.
16. Prade, E.; Mahajan, M.; Im, S.-C.; Zhang, M.; Gentry, K. A.; Anantharamaiah, G. M.; Waskell, L.; Ramamoorthy, A., A Minimal Functional Complex of Cytochrome P450 and FBD of Cytochrome P450 Reductase in Nanodiscs. *Angew. Chem. Int. Ed.* **2018**, 57, 8458-8462.

17. Ortiz de Montellano, P. R. Substrate Oxidation by Cytochrome P450 Enzymes. In *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Ortiz de Montellano, P. R., Ed.; Springer International Publishing: Cham, 2015, pp 111-176.
18. Xia, C.; Hamdane, D.; Shen, A. L.; Choi, V.; Kasper, C. B.; Pearl, N. M.; Zhang, H.; Im, S.-C.; Waskell, L.; Kim, J.-J. P., Conformational Changes of NADPH-Cytochrome P450 Oxidoreductase Are Essential for Catalysis and Cofactor Binding * <sup>
</sup>. *J. Biol. Chem.* **2011**, 286, 16246-16260.
19. Waskell, L.; Kim, J.-J. P. Electron Transfer Partners of Cytochrome P450. In *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Ortiz de Montellano, P. R., Ed.; Springer International Publishing: Cham, 2015, pp 33-68.
20. Dai, D.-P.; Geng, P.-W.; Wang, S.-H.; Cai, J.; Hu, L.-M.; Nie, J.-J.; Hu, J.-H.; Hu, G.-X.; Cai, J.-P., In vitro functional assessment of 22 newly identified CYP2D6 allelic variants in the Chinese population. *Basic Clin Pharmacol Toxicol* **2015**, 117, 39-43.
21. de Groot, M. J.; Wakenhut, F.; Whitlock, G.; Hyland, R., Understanding CYP2D6 interactions. *Drug Discov. Today* **2009**, 14, 964-972.
22. Huff, H. C.; Vasan, A.; Roy, P.; Kaul, A.; Tajkhorshid, E.; Das, A., Differential Interactions of Selected Phytocannabinoids with Human CYP2D6 Polymorphisms. *Biochemistry* **2021**, 60, 2749-2760.
23. Zanger, U. M.; Fischer, J.; Raimundo, S.; Stüven, T.; Evert, B. O.; Schwab, M.; Eichelbaum, M., Comprehensive analysis of the genetic factors determining expression and function of hepatic CYP2D6. *Pharmacogenetics and Genomics* **2001**, 11.
24. Flanagan, J. U.; MarÉChal, J.-D.; Ward, R.; Kemp, C. A.; McLaughlin, L. A.; Sutcliffe, M. J.; Roberts, G. C. K.; Paine, M. J. I.; Wolf, C. R., 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.
25. Keizers, P. H. J.; Lussenburg, B. M. A.; de Graaf, C.; Mentink, L. M.; Vermeulen, N. P. E.; Commandeur, J. N. M., Influence of phenylalanine 120 on cytochrome P450 2D6 catalytic selectivity and regiospecificity: crucial role in 7-methoxy-4-(aminomethyl)-coumarin metabolism. *Biochem. Pharmacol.* **2004**, 68, 2263-2271.
26. Lussenburg, B. M. A.; Keizers, P. H. J.; de Graaf, C.; Hidestrand, M.; Ingelman-Sundberg, M.; Vermeulen, N. P. E.; Commandeur, J. N. M., The role of phenylalanine 483 in cytochrome P450 2D6 is strongly substrate dependent. *Biochem. Pharmacol.* **2005**, 70, 1253-1261.
27. Gaedigk, A.; Ingelman-Sundberg, M.; Miller, N. A.; Leeder, J. S.; Whirl-Carrillo, M.; Klein, T. E.; the PharmVar Steering, C., The Pharmacogene Variation (PharmVar) Consortium: Incorporation of the Human Cytochrome P450 (CYP) Allele Nomenclature Database. *Clinical Pharmacology & Therapeutics* **2018**, 103, 399-401.
28. Mulnaes, D.; Porta, N.; Clemens, R.; Apanasenko, I.; Reiners, J.; Gremer, L.; Neudecker, P.; Smits, S. H. J.; Gohlke, H., TopModel: Template-Based Protein Structure Prediction at Low Sequence Identity Using Top-Down Consensus and Deep Neural Networks. *J. Chem. Theory Comput.* **2020**, 16, 1953-1967.
29. Becker, D.; Bharatam, P. V.; Gohlke, H., F/G Region Rigidity is Inversely Correlated to Substrate Promiscuity of Human CYP Isoforms Involved in Metabolism. *J. Chem. Inf. Model.* **2021**, 61, 4023-4030.
30. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli,

P.; Hassabis, D., Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, 596, 583-589.
31. System, T. P. M. G. *The PyMOL Molecular Graphics System*, Version 2.0; Schrödinger, LLC.
32. Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H., Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **2011**, 7, 2284-2295.
33. Doerr, S.; Harvey, M. J.; Noé, F.; De Fabritiis, G., HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *J. Chem. Theory Comput.* **2016**, 12, 1845-1852.
34. Schott-Verdugo, S.; Gohlke, H., PACKMOL-Memgen: A Simple-To-Use, Generalized Workflow for Membrane-Protein-Lipid-Bilayer System Building. *J. Chem. Inf. Model.* **2019**, 59, 2522-2528.
35. Yeagle, P., *The membranes of cells*. Third edition. ed.; Elsevier/AP: Amsterdam ; Boston, 2016; p xii, 439 pages.
36. Nugent, T.; Jones, D. T., Membrane protein orientation and refinement using a knowledge-based statistical potential. *BMC Bioinformatics* **2013**, 14, 276.
37. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, 9, 3878-3888.
38. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, I., T.E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Giambasu, G.; Giese, T.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Huang, Y.; Izadi, S.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wilson, L.; Wolf, R. M.; Wu, X.; Xiao, L.; Xiong, Y.; York, D. M.; Kollman, P. A. *AMBER 2019*, University of California, San Francisco, 2019.
39. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, 11, 3696-3713.
40. Dickson, C. J.; Madej, B. D.; Skjevik, A. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C., Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014**, 10, 865-879.
41. Izadi, S.; Anandakrishnan, R.; Onufriev, A. V., Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, 5, 3863-3871.
42. Sengupta, A.; Li, Z.; Song, L. F.; Li, P.; Merz, K. M., Jr., Parameterization of Monovalent Ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB Water Models. *J. Chem. Inf. Model.* **2021**, 61, 869-880.
43. Shahrokh, K.; Orendt, A.; Yost, G. S.; Cheatham, T. E., 3rd, Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. *J. Comput. Chem.* **2012**, 33, 119-133.
44. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, 23, 327-341.
45. Schott-Verdugo, S.; Muller, L.; Classen, E.; Gohlke, H.; Groth, G., Structural Model of the ETR1 Ethylene Receptor Transmembrane Sensor Domain. *Sci Rep* **2019**, 9, 8869.
46. Quigley, D.; Probert, M. I. J., Langevin dynamics in constant pressure extended systems. *J. Chem. Phys.* **2004**, 120, 11432-11441.
47. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, 81, 3684-3690.

48. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, 9, 3084-3095.
49. Case, D. A.; Aktulga, H. M.; Belfon, K.; Cerutti, D. S.; Cisneros, G. A.; Cruzeiro, V. W. D.; Forouzesh, N.; Giese, T. J.; Götz, A. W.; Gohlke, H.; Izadi, S.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kurtzman, T.; Lee, T.-S.; Li, P.; Liu, J.; Luchko, T.; Luo, R.; Manathunga, M.; Machado, M. R.; Nguyen, H. M.; O'Hearn, K. A.; Onufriev, A. V.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Risheh, A.; Schott-Verdugo, S.; Shajan, A.; Swails, J.; Wang, J.; Wei, H.; Wu, X.; Wu, Y.; Zhang, S.; Zhao, S.; Zhu, Q.; Cheatham, T. E., III; Roe, D. R.; Roitberg, A.; Simmerling, C.; York, D. M.; Nagan, M. C.; Merz, K. M., Jr., AmberTools. *J. Chem. Inf. Model.* **2023**.
50. Hermans, S. M. A.; Pfleger, C.; Nutschel, C.; Hanke, C. A.; Gohlke, H., Rigidity theory for biomolecules: concepts, software, and applications. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2017**, 7, e1311.
51. Jacobs, D. J.; Rader, A. J.; Kuhn, L. A.; Thorpe, M. F., Protein flexibility predictions using graph theory. *Proteins* **2001**, 44, 150-165.
52. Dahiyat, B. I.; Benjamin Gordon, D.; Mayo, S. L., Automated design of the surface positions of protein helices. *Protein Sci.* **1997**, 6, 1333-1337.
53. Radestock, S.; Gohlke, H., Exploiting the Link between Protein Rigidity and Thermostability for Data-Driven Protein Engineering. *Eng. Life Sci.* **2008**, 8, 507-522.
54. Pfleger, C.; Rathi, P. C.; Klein, D. L.; Radestock, S.; Gohlke, H., Constraint Network Analysis (CNA): A Python Software Package for Efficiently Linking Biomacromolecular Structure, Flexibility, (Thermo-)Stability, and Function. *J. Chem. Inf. Model.* **2013**, 53, 1007-1015.
55. Jacobs, D. J.; Thorpe, M. F., Generic Rigidity Percolation: The Pebble Game. *Phys. Rev. Lett.* **1995**, 75, 4051-4054.
56. Pfleger, C.; Radestock, S.; Schmidt, E.; Gohlke, H., Global and local indices for characterizing biomolecular flexibility and rigidity. *J. Comput. Chem.* **2013**, 34, 220-233.
57. Pfleger, C.; Minges, A.; Boehm, M.; McClendon, C. L.; Torella, R.; Gohlke, H., Ensemble- and Rigidity Theory-Based Perturbation Approach To Analyze Dynamic Allostery. *J. Chem. Theory Comput.* **2017**, 13, 6343-6357.
58. van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A. M. J. J., The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, 428, 720-725.
59. Honorato, R. V.; Koukos, P. I.; Jiménez-García, B.; Tsaregorodtsev, A.; Verlato, M.; Giachetti, A.; Rosato, A.; Bonvin, A. M. J. J., Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem. *Frontiers in Molecular Biosciences* **2021**, 8.
60. UniProt, C., UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, 47, D506-D515.
61. Sherry, S. T.; Ward, M. H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E. M.; Sirotkin, K., dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **2001**, 29, 308-11.
62. Landrum, M. J.; Chitipiralla, S.; Brown, G. R.; Chen, C.; Gu, B.; Hart, J.; Hoffman, D.; Jang, W.; Kaur, K.; Liu, C.; Lyoshin, V.; Maddipatla, Z.; Maiti, R.; Mitchell, J.; O'Leary, N.; Riley, G. R.; Shi, W.; Zhou, G.; Schneider, V.; Maglott, D.; Holmes, J. B.; Kattman, B. L., ClinVar: improvements to accessing data. *Nucleic Acids Res.* **2020**, 48, D835-D844.
63. HGMD (human gene mutation database). In *Encyclopedia of Genetics, Genomics, Proteomics and Informatics*, Rédei, G. P., Ed.; Springer Netherlands: Dordrecht, 2008, pp 873-873.

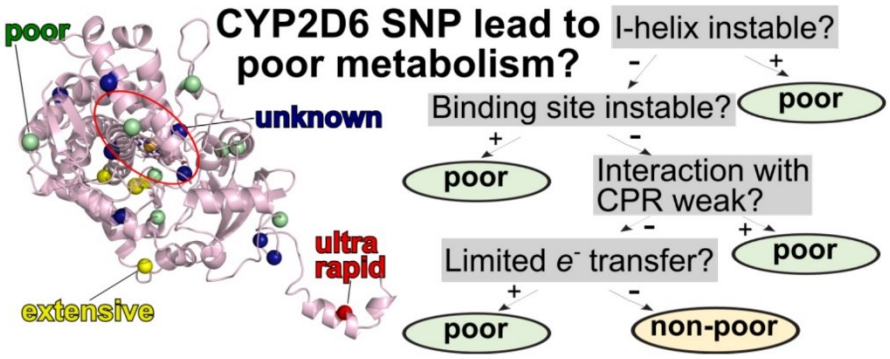
64. Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Vijaykumar, A.; Bardelli, A. P.; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G.-L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; de Miranda Cardoso, J. V.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmerer, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y.; SciPy, C., SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, 17, 261-272.
65. Morris, A. L.; MacArthur, M. W.; Hutchinson, E. G.; Thornton, J. M., Stereochemical quality of protein structure coordinates. *Proteins* **1992**, 12, 345-364.
66. Ramesh, M.; Bharatam, P. V., Importance of hydrophobic parameters in identifying appropriate pose of CYP substrates in cytochromes. *Eur. J. Med. Chem.* **2014**, 71, 15-23.
67. Roughley, S. D.; Jordan, A. M., The Medicinal Chemist's Toolbox: An Analysis of Reactions Used in the Pursuit of Drug Candidates. *J. Med. Chem.* **2011**, 54, 3451-3479.
68. Im, S.-C.; Waskell, L., The interaction of microsomal cytochrome P450 2B4 with its redox partners, cytochrome P450 reductase and cytochrome b5. *Arch. Biochem. Biophys.* **2011**, 507, 144-153.
69. Bridges, A.; Gruenke, L.; Chang, Y.-T.; Vakser, I. A.; Loew, G.; Waskell, L., Identification of the Binding Site on Cytochrome P450 2B4 for Cytochrome b <sub>5</sub> and Cytochrome P450 Reductase *. *J. Biol. Chem.* **1998**, 273, 17036-17049.
70. Kumar, S.; Nussinov, R., Close-Range Electrostatic Interactions in Proteins. *ChemBioChem* **2002**, 3, 604-617.
71. Evert, B.; Griese, E.-U.; Eichelbaum, M., A missense mutation in exon 6 of the CYP2D6 gene leading to a histidine 324 to proline exchange is associated with the poor metabolizer phenotype of sparteine. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **1994**, 350, 434-439.
72. Fischer, A.; Don, C. G.; 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.
73. Dorne, J. L. C. M.; Cirlini, M.; Louisse, J.; Pedroni, L.; Galaverna, G.; Dellafiora, L., A Computational Understanding of Inter-Individual Variability in CYP2D6 Activity to Investigate the Impact of Missense Mutations on Ochratoxin A Metabolism. **2022**, 14, 207.
74. Shaik, S.; Dubey, K. D., The catalytic cycle of cytochrome P450: a fascinating choreography. *Trends in Chemistry* **2021**, 3, 1027-1044.
75. Ebisawa, A.; Hiratsuka, M.; Sakuyama, K.; Konno, Y.; Sasaki, T.; Mizugaki, M., Two novel single nucleotide polymorphisms (SNPs) of the CYP2D6 gene in Japanese individuals. *Drug Metab Pharmacokinet* **2005**, 20, 294-9.
76. Soyama, A.; Kubo, T.; Miyajima, A.; Saito, Y.; Shiseki, K.; Komamura, K.; Ueno, K.; Kamakura, S.; Kitakaze, M.; Tomoike, H.; Ozawa, S.; Sawada, J., Novel nonsynonymous

single nucleotide polymorphisms in the CYP2D6 gene. *Drug Metab Pharmacokinet* **2004**, 19, 313-9.

77. Masayuki, M.; Hiroshi, Y.; Kazuma, K.; Shunsuke, I.; Jyunji, S.; Kazuko, N.; Akiko, S.; Shogo, O.; Jun-Ichi, S.; Eiji, K.; Moritoshi, K.; Tetsuya, K., Two Novel <em>CYP2D6*10</em> Haplotypes As Possible Causes of a Poor Metabolic Phenotype in Japanese. *Drug Metab. Dispos.* **2009**, 37, 699.

78. Crews, K. R.; Monte, A. A.; Huddart, R.; Caudle, K. E.; Kharasch, E. D.; Gaedigk, A.; Dunnenberger, H. M.; Leeder, J. S.; Callaghan, J. T.; Samer, C. F.; Klein, T. E.; Haidar, C. E.; Van Driest, S. L.; Ruano, G.; Sangkuhl, K.; Cavallari, L. H.; Müller, D. J.; Prows, C. A.; Nagy, M.; Somogyi, A. A.; Skaar, T. C., Clinical Pharmacogenetics Implementation Consortium Guideline for CYP2D6, OPRM1, and COMT Genotypes and Select Opioid Therapy. *Clinical Pharmacology & Therapeutics* **2021**, 110, 888-896.

**Table of Contents Graphic**



# Supporting Information
# Molecular mechanisms underlying single nucleotide polymorphism-induced reactivity decrease in CYP2D6

Daniel Becker[a], Prassad V. Bharatam [b], Holger Gohlke[a,c,1]

[a]Mathematisch-Naturwissenschaftliche Fakultät, Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany

[b] Department of Medicinal Chemistry, National Institute of Pharmaceutical Education and Research (NIPER), Sector 67, S. A. S. Nagar, Mohali, 160 062 Punjab, India

[c]John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre (JSC), and Institute of Bio- and Geosciences (IBG-4: Bioinformatics), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

[1] To whom correspondence should be addressed.

[*]Address: Universitätsstr. 1, 40225 Düsseldorf, Germany.
Phone: (+49) 211 81 13662; Fax: (+49) 211 81 13847
E-mail: gohlke@uni-duesseldorf.de

# Supporting Methods

## Statistical Analysis and Data Analysis of Computational Results

The standard error of the mean (SEM) for the computed $\bar{E}_{region,\mathrm{CNA}}$ values was calculated by considering each trajectory as an independent sample. Data of CNA analyses were visualized by visualCNA.[1]

The SEM was propagated by eq. S1:

$$SEM = \frac{1}{10}\sqrt{\sum_{i=1}^{10} SEM_i^{\,2}} \qquad \textbf{(eq. S1)}$$

All calculations were performed with NumPy.[2] Curve fitting was performed with the SciPy module stats.[3]

## Linear Interaction Energy Analysis

In the analysis of the G169R variant, the linear interaction energy (LIE) was calculated by the *lie* function implemented in pytraj.[4] We[5] and others[6, 7] showed that LIE is an efficient method to obtain good affinity predictions. The electrostatic and van der Waals interactions of R169 were calculated to the surrounding residues S168 and P496. The cutoff for interactions was set to 12.0 Å. The ligand mask was set to the atoms of the guanidino group, and the surrounding mask was set to S168 and P496.

## Supplemental Figures

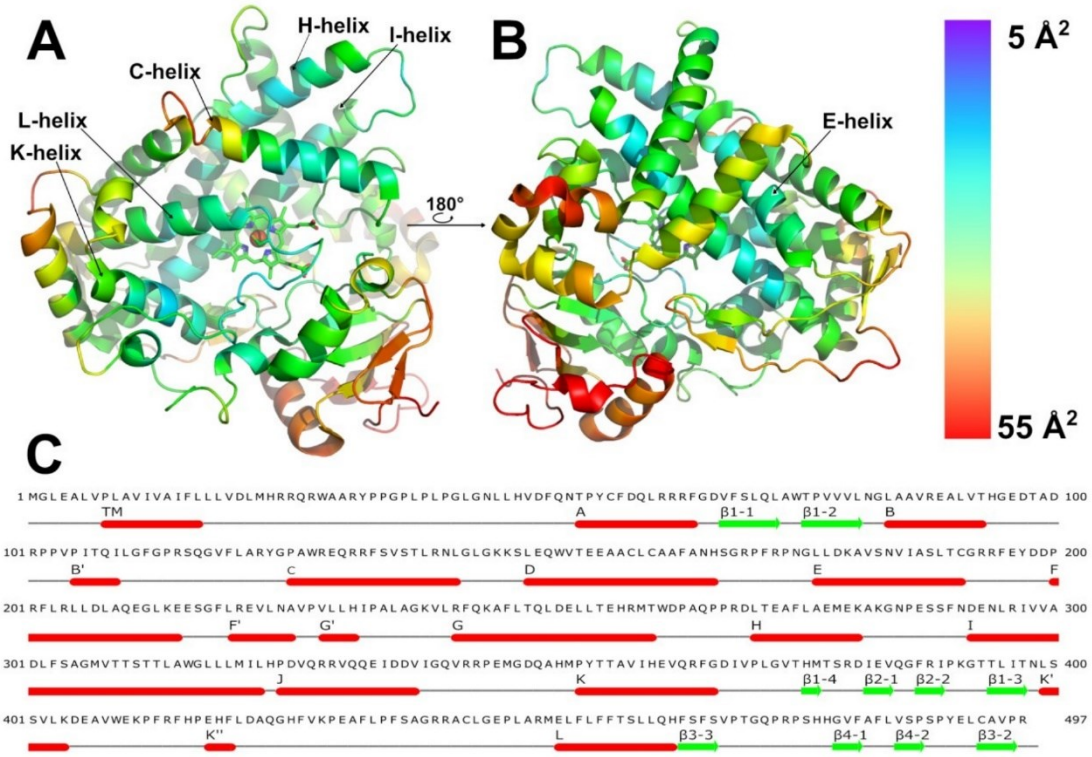



**Figure S1:** Crystal structure (PDB ID: 3TBG) of the globular part of CYP2D6 colored by B-factor (red: most mobile, blue: least mobile). **A:** The helices E, H, I, L, and K are the least mobile helices in CYP2D6. The heme ring system is shown as sticks, while the heme iron is shown as a sphere. **B:** Structure of CYP2D6 rotated by 180°. The color bar on the right denotes the B-factor. **C:** Sequence of CYP2D6 with marked secondary structure elements and the nomenclature.

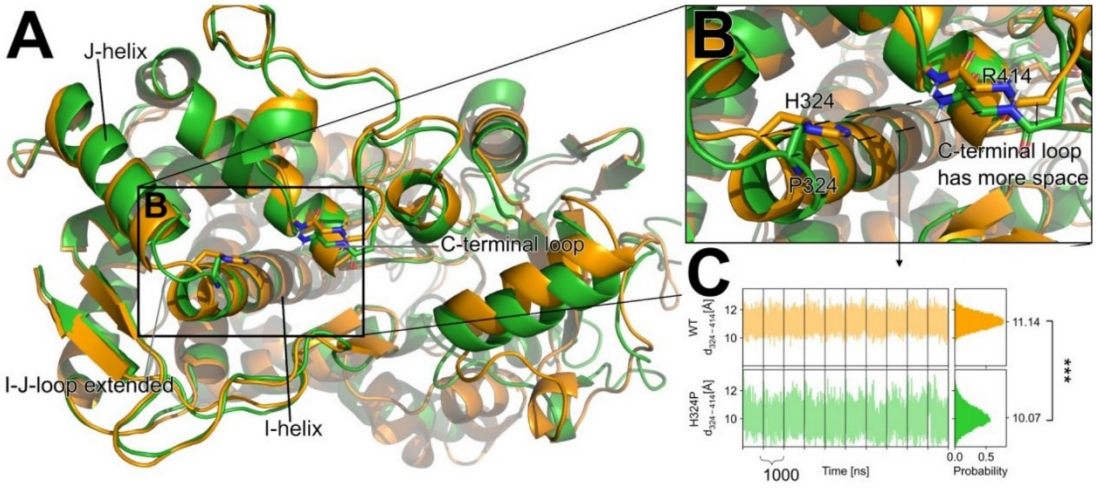



**Figure S2:** Effect of the H324P mutation on CYP2D6. **A:** The mutation in the variant H324P (lime-green) is located in the loop between I-helix and J-helix. Compared to the wildtype (orange), the I-J-loop is extended by one residue. **B:** In variant H324P, the C-terminal loop (marked with an arrow) is shifted towards the I-helix. Sidechains at position 324 are shown as sticks, and the backbone of R414 is shown as sticks. Dashed lines indicate distances measured in panel C. **C:** The $C_\alpha$ distance between residues 324 and 414 is significantly decreased for the variant H324P compared to wildtype. Left: time series of the distance, values per trajectory are separated by black vertical lines; right: aggregate probability distributions. Statistical analysis was performed using the two-sided t-test (* $p < 0.01$; ** $p < 0.001$; *** $p < 0.0001$; n.s.: $p > 0.01$).

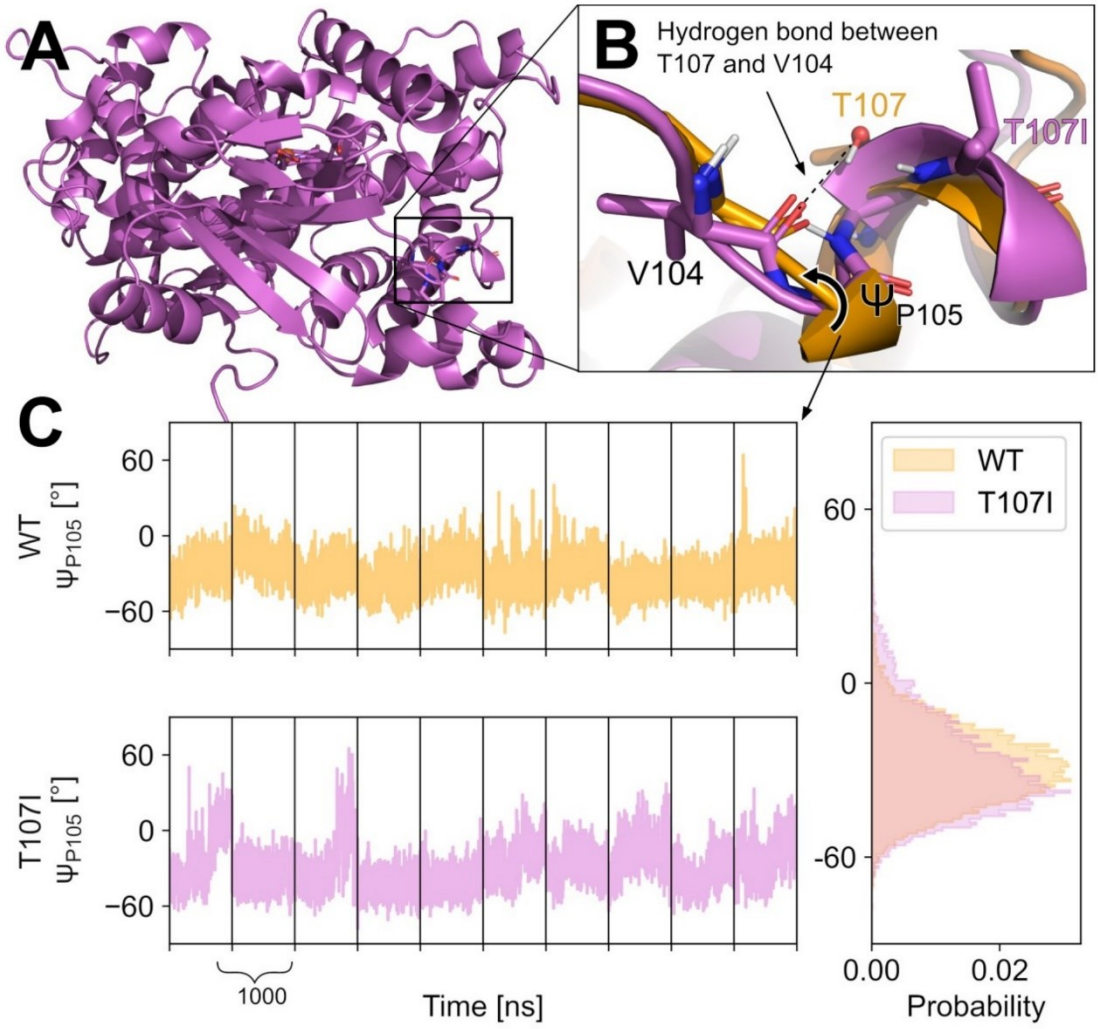



**Figure S3:** Effect of the T107I mutation on CYP2D6. **A:** The mutation in the variant T107I (orchid) is located in the C'-helix. The heme ring system is shown as sticks, while the heme iron is shown as a sphere. **B:** The missing interaction between V104 and T107 in the T107I variant destabilizes the C'-helix. The backbone of positions 104-107 is shown as sticks, as is the sidechain of position 107. The Ψ torsion at P105 is indicated with an arrow. **C:** The Ψ torsion has significantly more often positive values ($p < 0.01$, two-sided *t*-test) in T107I compared to the wildtype. Left: time series of the Ψ torsion, values per trajectory are separated by black vertical lines; right: aggregate probability distributions.

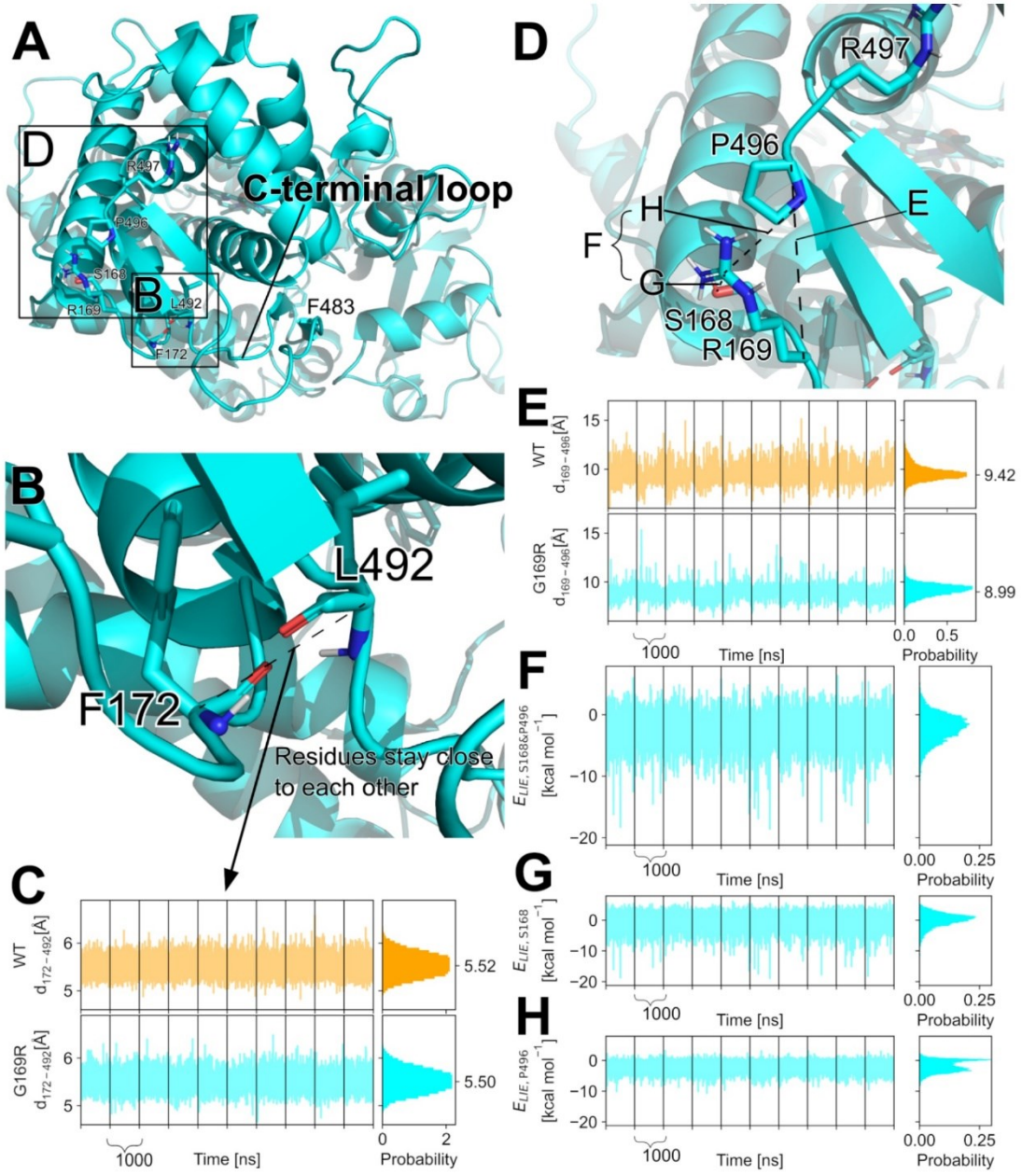



**Figure S4:** Effect of the G169R mutation on CYP2D6. **A:** The mutation in the variant G169R (cyan) is located in the D-E-loop. The heme ring system is shown as sticks, while the heme iron is shown as a sphere. **B:** F172 and L492 form backbone hydrogen atoms between the D-E-loop and the C-terminal loop. Amide hydrogens of both residues point to the carbonyl oxygens of the opposite residues. The backbone of both residues is shown as sticks. A dashed line indicates the distance measured in panel C. **C:** The $C_\alpha$ distance between F172 and L492 is highly constant and similar in the wild type (orange) and the variant G169R. Left: time series of the distance, values per trajectory are separated by black vertical lines; right: aggregate probability distributions. **D:** R169 in the variant G169R interacts with S168 and P496.

Interactions evaluated in panels F-H are indicated with dashed lines and the measured distance between $C_\alpha$ 169 and $C_\alpha$ 496 (panel E). **E:** The distance between $C_\alpha$ 169 and $C_\alpha$ 496 is significantly decreased in the variant G169R compared to wild type ($p < 0.01$, two-sided *t*-test). Left: time series of the distance, values per trajectory are separated by black vertical lines; right: aggregate probability distributions. **F-H:** Electrostatic and van der Waals energies computed according to the Linear Interaction Energy approach between the R169 guanidino group and S168 and P496 (F), R169 guanidino group and S168 (G), and R169 guanidino group and P496 (H) in the G169R variant. Left: time series of the electrostatic and van der Waals energies, values per trajectory are separated by black vertical lines; right: aggregate probability distributions. (F) In most frames, R169 interacts favorably with S168 *and* P496. In comparison to panels G and H, both residues interact with R169 in the investigated G169R variant. (G) In most frames, R169 shows interactions with S168. (H) In most frames, R169 shows interactions with P496.

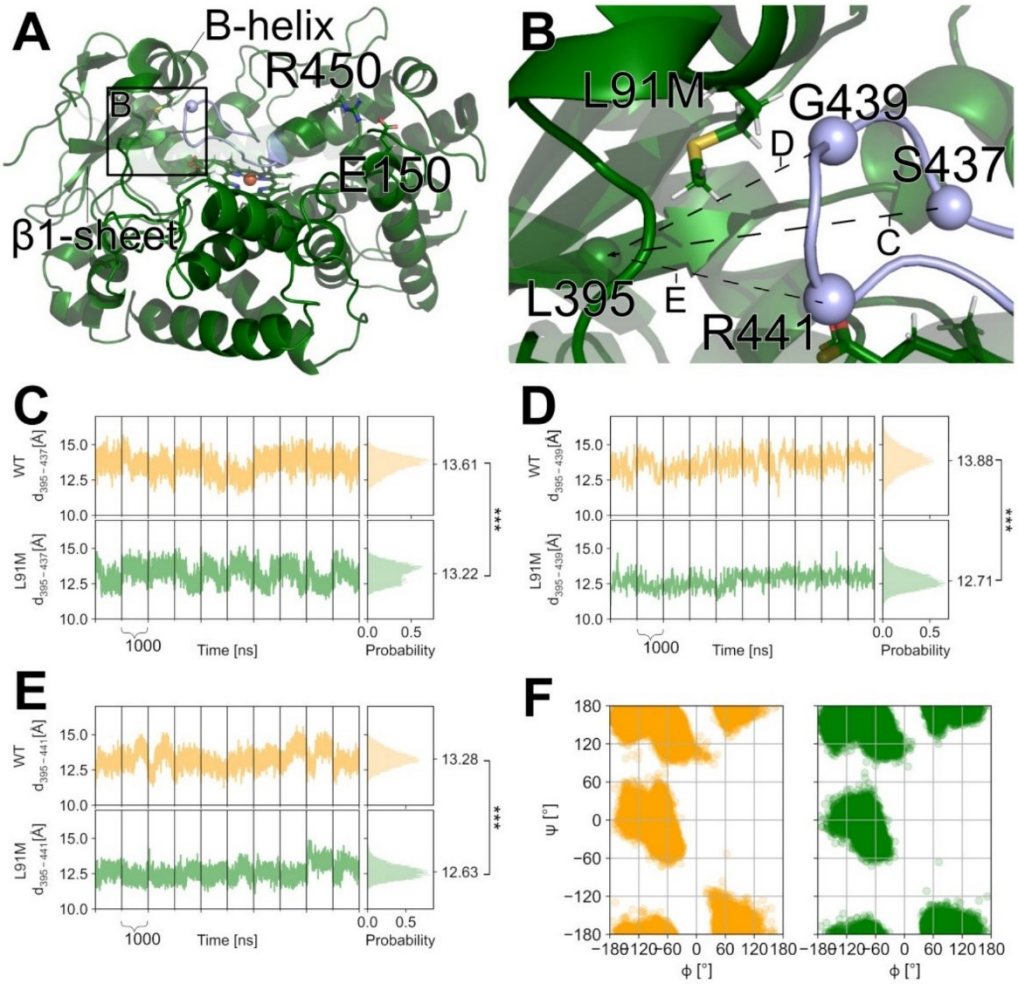



**Figure S5:** Effect of the L91M mutation on CYP2D6. **A:** The mutation L91M (green) is located on the B-helix. The heme ring is shown as sticks, while the heme iron and important $C_\alpha$ atoms are shown as spheres. The C-helix is shown with high transparency for visualization purposes. **B:** The missing second $C_\gamma$ in the variant leads to a closer distance between the K-L-loop (purple) and the β1 strand. $C_\alpha$ of L395 and G439 are shown as spheres. A dashed line indicates the measured distance. **C:** The distance between $C_\alpha$ atoms of L395 and S437 is significantly reduced in the L91M variant compared to wild type ($p < 0.01$, two-sided *t*-test). **D:** The distance between $C_\alpha$ atoms of L395 and G439 is significantly reduced in the L91M variant compared to wild type ($p < 0.01$, two-sided *t*-test). **E:** The distance between $C_\alpha$ atoms of L395 and R441 is significantly reduced in the L91M variant compared to wild type ($p < 0.01$, two-sided *t*-test). C-E: Left: time series of the distance, black vertical lines separate values per trajectory; right: aggregate probability distributions. **F:** Left: Ramachandran plot of the K-L-loop (residues 436-444) of the wild type. Right: Ramachandran plot of the K-L-loop (residues 436-444) of the L91M. Each data point represents one residue in one snapshot over 10 x 1 µs of simulation

time. No major conformational change is visible between wild type and variant. Statistical analysis was performed using the two-sided t-test (* $p < 0.01$; ** $p < 0.001$; *** $p < 0.0001$; n.s.: $p > 0.01$).

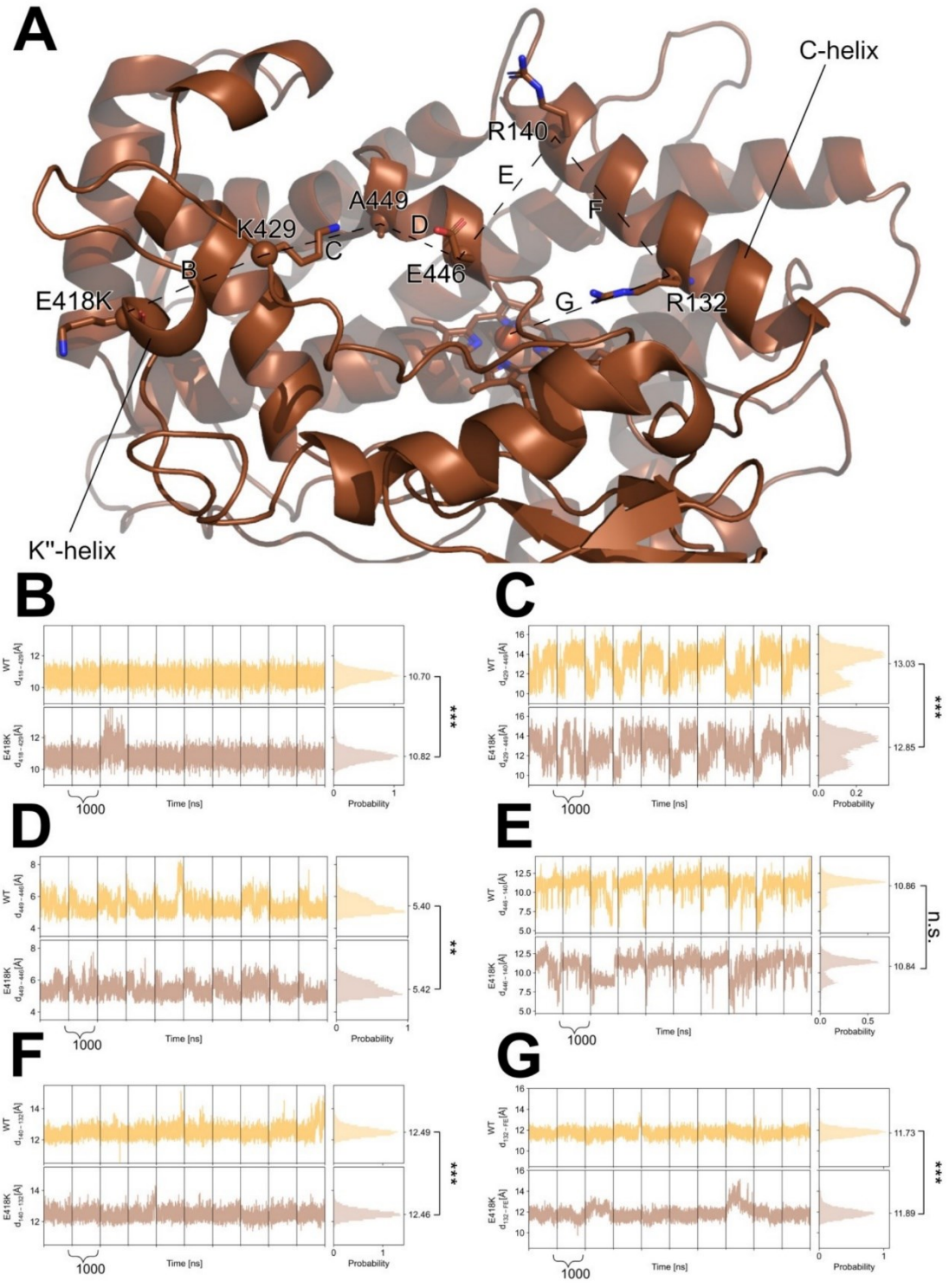



**Figure S6:** Effect of the E418K mutation on CYP2D6. **A:** The mutation in E418K is located on the K''-helix, while R132 is part of the C-helix. The $C_\alpha$ atoms of residues between a distance measured are shown as spheres. Dashed lines indicate measured distances. The heme ring system is shown as sticks, while the heme iron is shown as a sphere. **B:** The distance between

$C_\alpha$ of residue E418K and $C_\alpha$ of residue K429 is significantly increased in E418K ($p < 0.01$, two-sided *t*-test). **C:** The distance between $C_\alpha$ of residue K429 and $C_\alpha$ of residue A449 is significantly decreased in E418K ($p < 0.01$, two-sided *t*-test). **D:** The distance between $C_\alpha$ of residue E446 and $C_\alpha$ of residue A449 differs on average only by 0.02 Å. **E:** The distance between $C_\alpha$ of residue R140 and $C_\alpha$ of residue E446 does not differ significantly between WT and E418K. **F:** The distance between $C_\alpha$ of residue R132 and $C_\alpha$ of residue R140 differs on average only by 0.03 Å. **G:** The distance between $C_\alpha$ of residue R132 and iron in heme is significantly increased in E418K ($p < 0.01$, two-sided *t*-test). B-G: Left: time series of the distance, black vertical lines separate values per trajectory; right: aggregate probability distributions. Statistical analysis was performed using the two-sided t-test (* $p < 0.01$; ** $p < 0.001$; *** $p < 0.0001$; n.s.: $p > 0.01$).

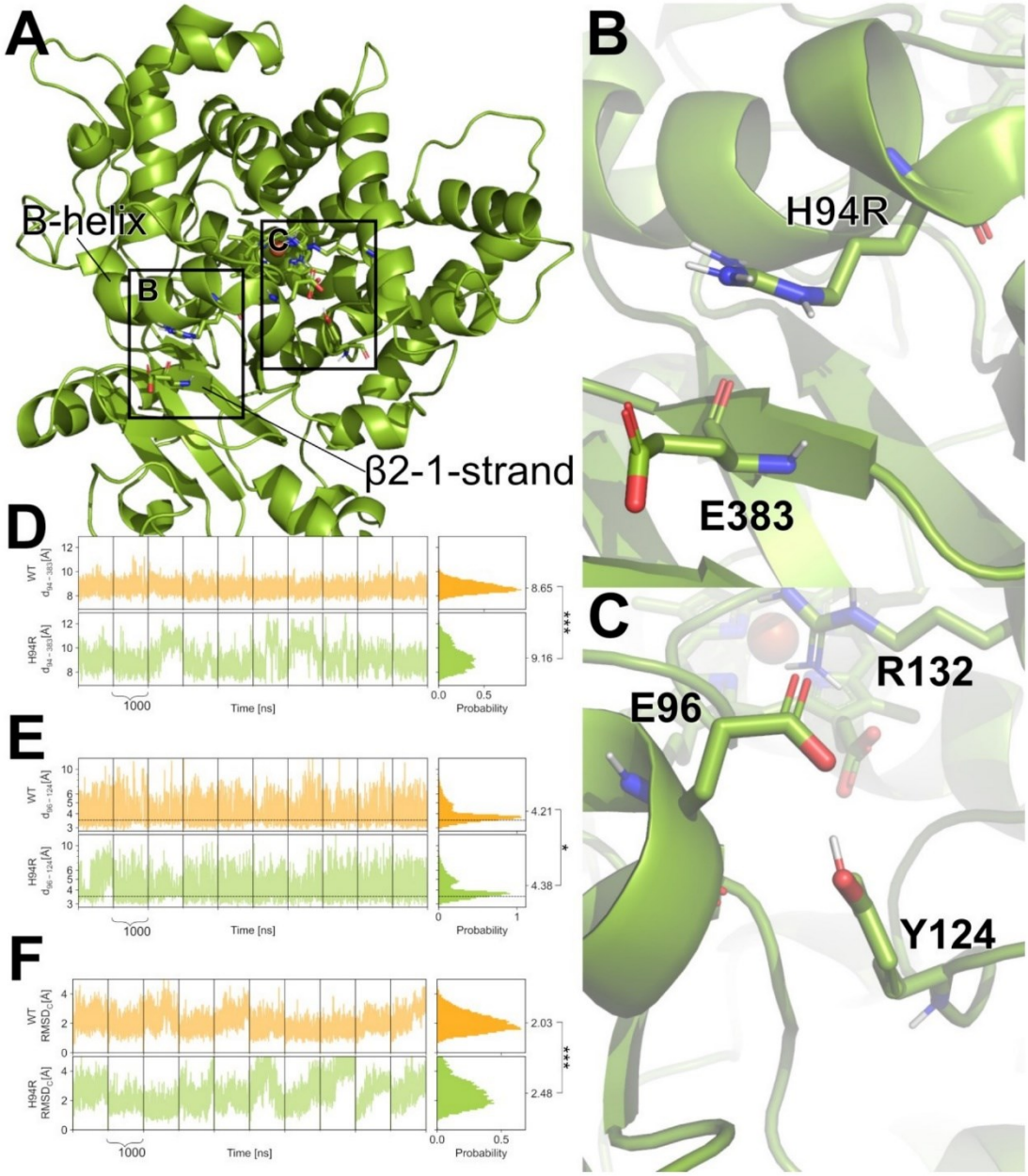



**Figure S7:** Effect of the H94R mutation on CYP2D6. **A:** The mutation in H94R is located on the B-helix. The heme ring and important residues are shown as sticks, while the heme iron is shown as a sphere. **B:** Due to the larger space requirement of R94, the distance between the $C_\alpha$ atom of R94 and the $C_\alpha$ atom of E383 is increased. E383 is part of the β2-1 strand. **C:** The charge-assisted hydrogen bond between E96 and Y124 is more often formed in WT than in H94R due to a shift of the B-helix. **D:** Distance between the $C_\alpha$ atom of residue H94/R94 and the $C_\alpha$ atom of residue E383 is significantly increased ($p < 0.01$, two-sided *t*-test). **E:** The hydrogen bond between E96 and Y124 is significantly less often formed in H94R than in the wild type ($p < 0.01$, two-sided *t*-test). The dashed line at 3.5 Å indicates the formation of a

hydrogen bond. **F:** The $C_\alpha$-RMSD (residues 126-143) of the C-helix after fitting on the heme group is significantly increased. D-F: Left: time series of the distance or the $C_\alpha$-RMSD, values per trajectory are separated by black vertical lines; right: aggregate probability distributions. Statistical analysis was performed using the two-sided t-test (* $p < 0.01$;** $p < 0.001$; *** $p < 0.0001$; n.s.: $p > 0.01$).

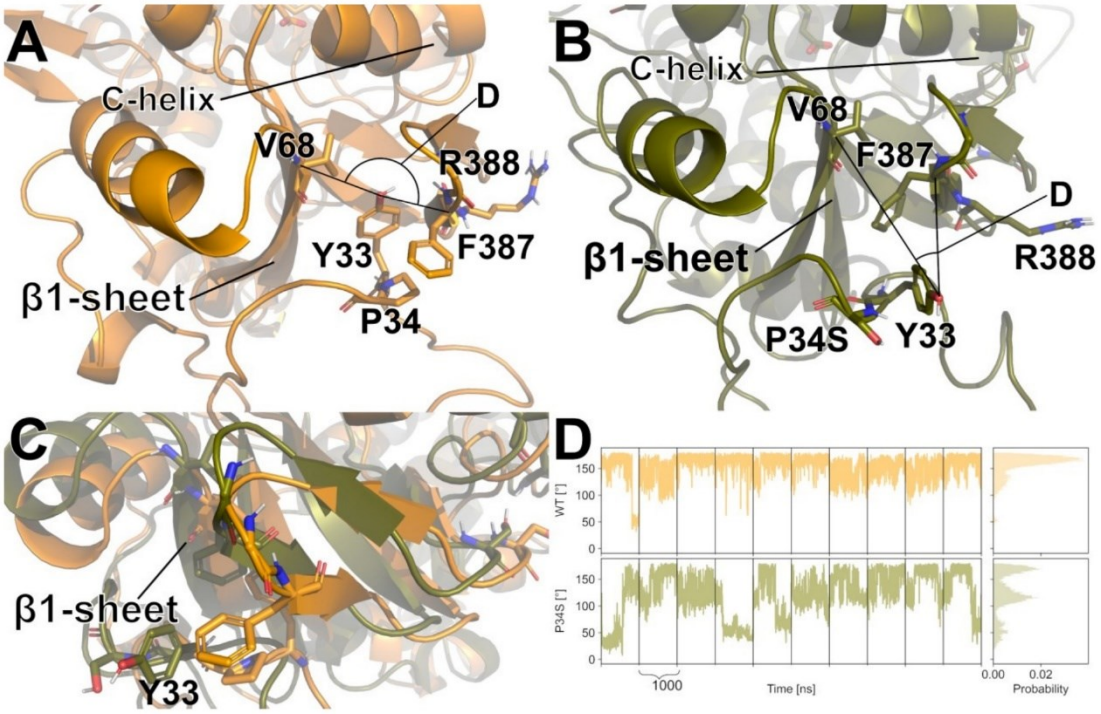



**Figure S8:** Effect of the P34S mutation on CYP2D6. **A:** P34 in CYP2D6 wild type stabilizes the position of Y33 by rigidifying the loop Y33 belongs to between V68 and F387 (see also panel D). Important residues are shown as sticks. Black lines indicate the angle measured in panel D. **B:** P34S in the CYP2D6 variant does not stabilize the position of Y33 between V68 and F387 (see also panel D). Y33 is shifted away from the β1 strand, and F387 is moving towards V69. Important residues are shown as sticks. Black lines indicate the angle measured in panel D. **C:** Overlay of CYP2D6 wild type (orange) and the P34S variant (olive) reveals the drastic change in the β1 strand conformation upon the Y33 shift. The relative orientation of the β1 strand leads to changes in the interaction with the C-helix. **D:** Y33 flips out of its position more often in P34S than in the wild type. Left: time series of the angle ($C_\alpha$ of V68, oxygen of the hydroxy group of Y33, $C_\alpha$ of F387), values per trajectory are separated by black vertical lines; right: aggregate probability distributions.

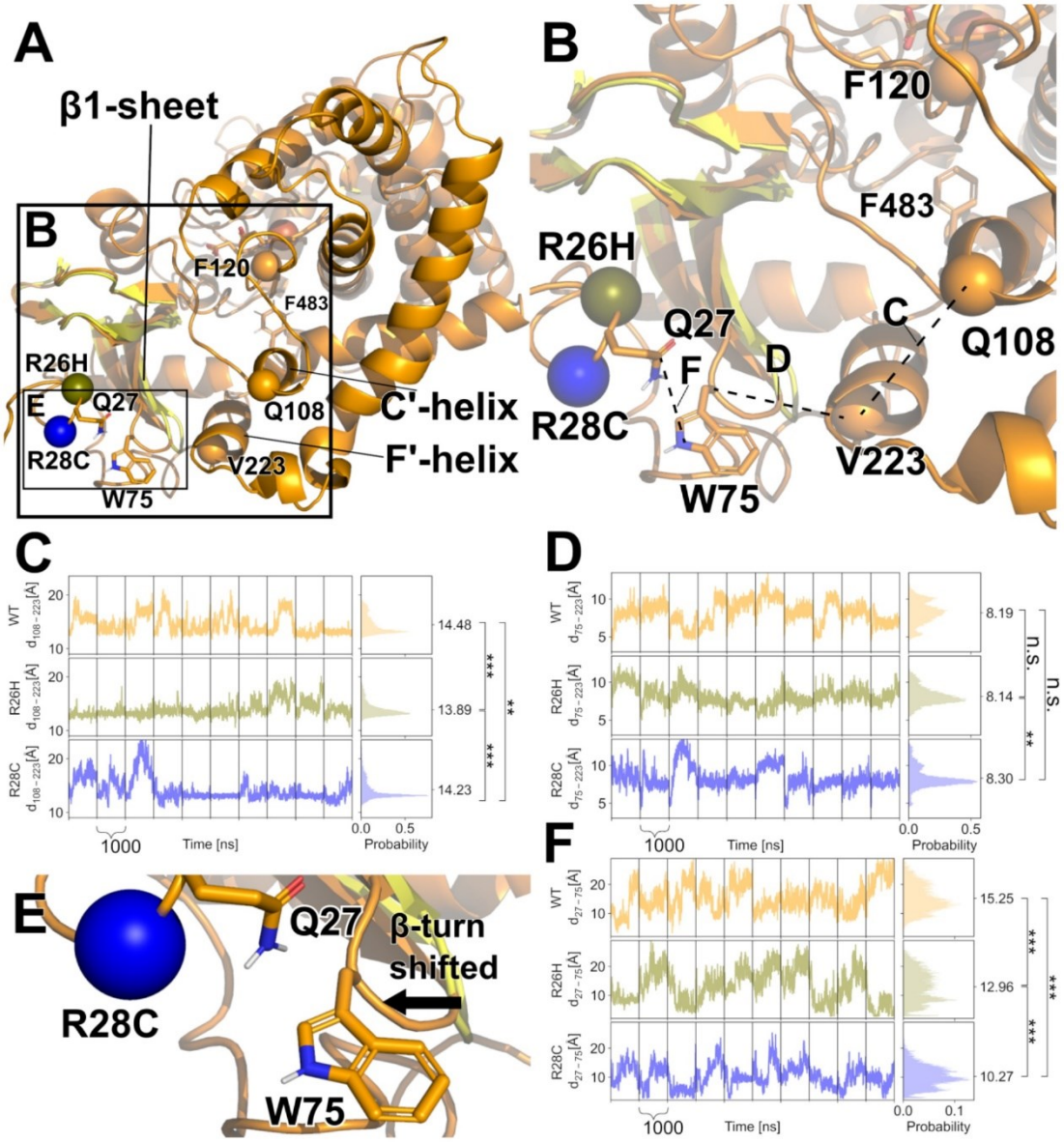



**Figure S9:** Effect of the R26H and R28C mutations on CYP2D6. **A:** The mutations R26H (olive) and R28C (blue) are located in the loop between the TM-helix and β1-sheet. The heme ring system is shown as sticks, while the heme iron is shown as a sphere. **B:** Mutations in the loop between the TM-helix and strand β1 influence the F'-helix and C'-helix. Q27 and W75 are shown as sticks; the $C_\alpha$ atoms of residues between a distance was measured are shown as spheres. Dashed lines indicate measured distances. **C:** The distance between $C_\alpha$ of residue Q108 and $C_\alpha$ of residue V223 is unaffected by either variant. **D:** The distance between $C_\alpha$ of residue W75 and $C_\alpha$ of residue V223 is significantly decreased in R26H and R28C ($p < 0.01$, two-sided *t*-test). **E:** Q27 interacts with W75 and leads to a shift of the β turn. **F:** The distance between $C_\alpha$ of residue Q27 and the aromatic nitrogen of W75 is significantly decreased in R26H and R28C

($p < 0.01$, two-sided *t*-test). C, D, F: Left: time series of the distance, values per trajectory are separated by black vertical lines; right: aggregate probability distributions. Statistical analysis was performed using the two-sided t-test (* $p < 0.01$;** $p < 0.001$; *** $p < 0.0001$; n.s.: $p > 0.01$)

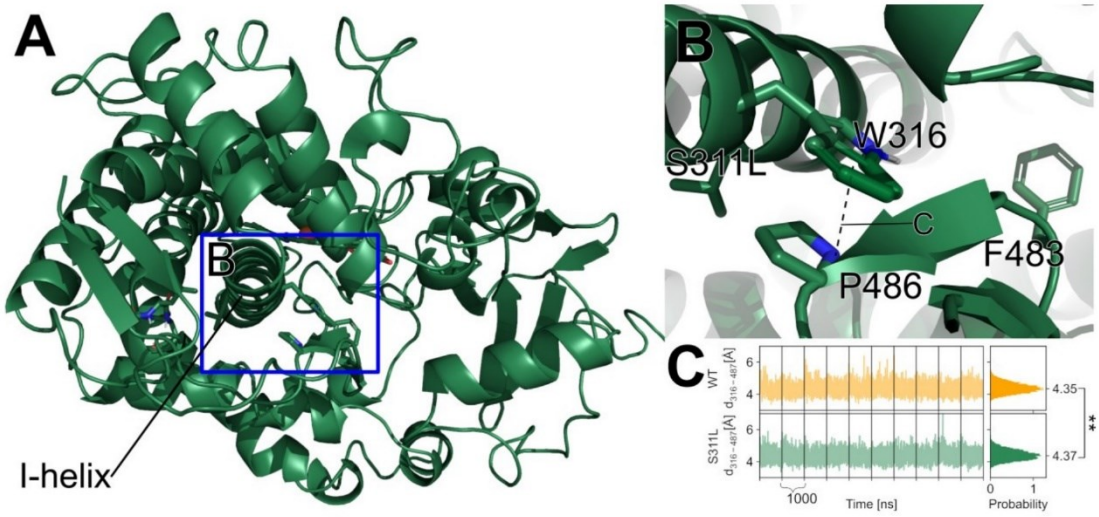



**Figure S10:** Effect of the S311L mutation on CYP2D6. **A:** The mutation S311L (sea green) is located in the I-helix. The heme ring system is shown as sticks, while the heme iron is shown as a sphere. **B:** W316 and P486 have a "stacked-like" arrangement.[8] The close contact pushes the C-terminal loop to the β1-sheet in the S311L variant due to space requirements. Residues S311L, W316, F483, and P486 are shown as sticks. Dashed lines indicate the measured distance. **C:** The distance between $C_\alpha$ of residue W316 and $C_\alpha$ of residue P486 differs on average only by 0.02 Å.. Left: time series of the distance, values per trajectory are separated by black vertical lines; right: aggregate probability distributions. Statistical analysis was performed using the two-sided t-test (* $p < 0.01$;** $p < 0.001$; *** $p < 0.0001$; n.s.: $p > 0.01$)

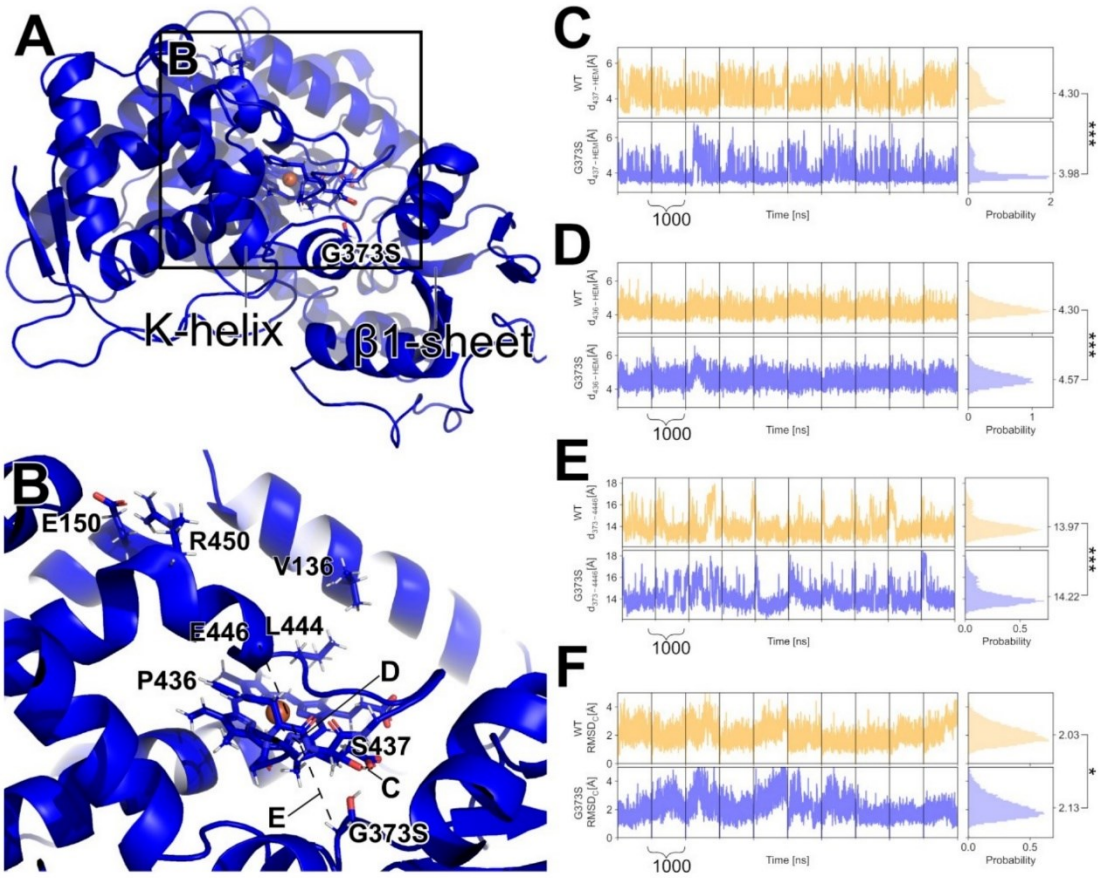



**Figure S11:** Effect of the G373S mutation on CYP2D6. **A:** The mutation in G373S is located in the K-K'-loop. The heme ring system is shown as sticks, while the heme iron and important $C_\alpha$ atoms are shown as spheres. **B:** The OH group in the variant stabilizes the position of the close-by carbonyl group in the heme moiety. This stabilizes the position of S437, which lifts the K-L-loop relative to the heme. This lifting leads to a higher distance between the $C_\alpha$ atom of residue G373/S373 (see panel E) and the $C_\alpha$ atom of residue E446 because of the interactions between P436 and the backbone of E446. This also increases the hydrophobic interaction between L444 and V136. Dashed lines indicate measured distances. **C:** The interaction between S437-OH and the heme carboxy group of ring D is significantly increased in G373S. **D:** The $C_\alpha$ atom of P436 shows an increased distance to the aromatic carbon 17 in ring D of heme. **E:** The distance between the $C_\alpha$ atom of residue G373/S373 and the $C_\alpha$ atom of residue E446 is significantly increased ($p < 0.01$, two-sided *t*-test). C-E: Left: time series of the distance, black vertical lines separate values per trajectory; right: aggregate probability distributions. **F:** The $C_\alpha$-RMSD of the C-helix is significantly increased in G373S. Left: time series of the $C_\alpha$-RMSD, values per trajectory are separated by black vertical lines; right: aggregate probability

distributions. Statistical analysis was performed using the two-sided t-test (* $p < 0.01$;** $p < 0.001$; *** $p < 0.0001$; n.s.: $p > 0.01$).

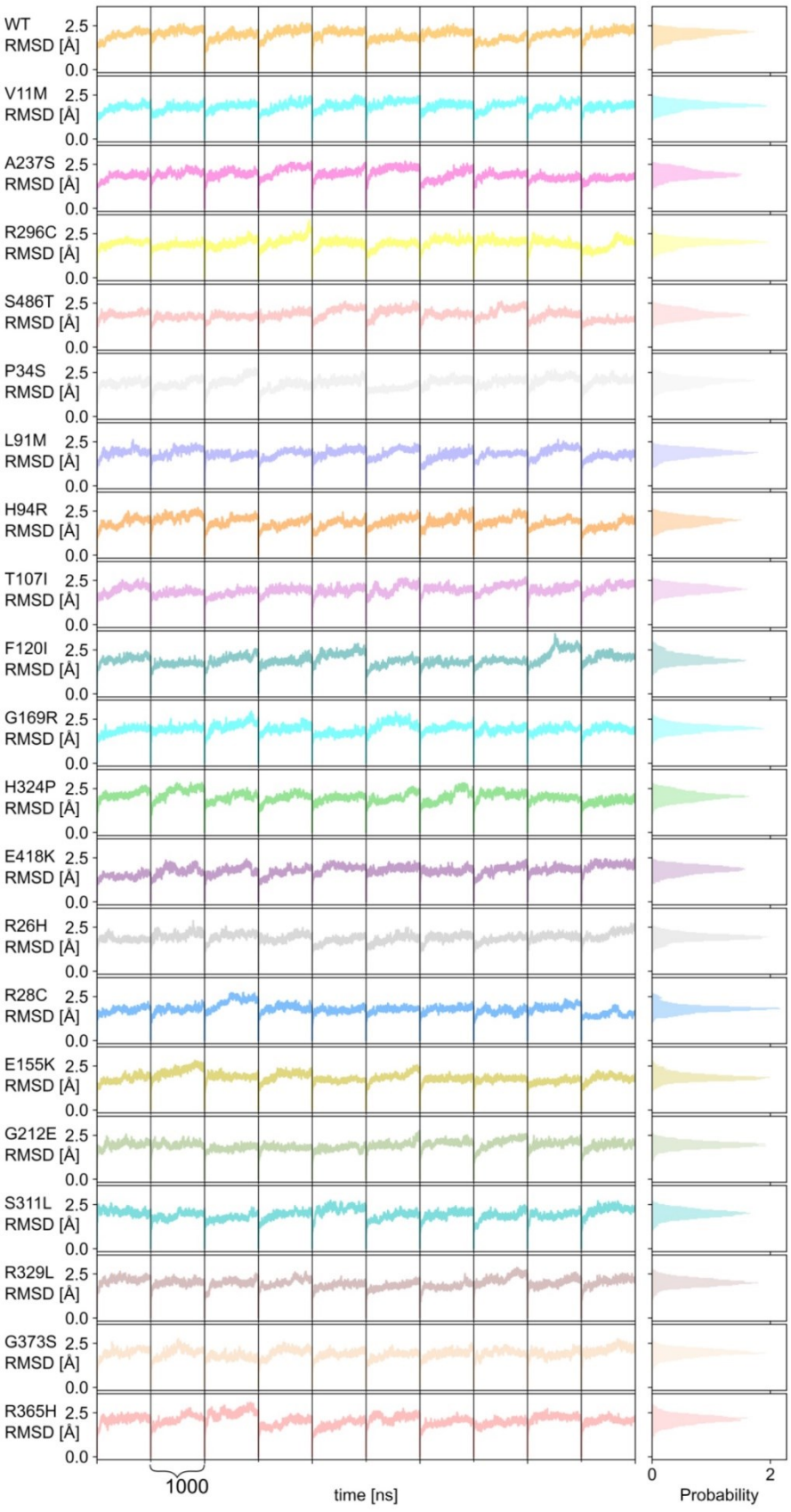

WT
V11M
A237S
R296C
S486T
P34S
L91M
H94R
T107I
F120I
G169R
H324P
E418K
R26H
R28C
E155K
G212E
S311L
R329L
G373S
R365H
RMSD [Å]
2.5
0.0
1000
time [ns]
0
2
Probability

**Figure S12: Root-mean-square deviation (RMSD) of the backbone of the CYP2D6 wild type and variants along ten MD trajectories each and aggregate probability distributions.** The RMSD is calculated with respect to the structure in the first frame after equilibration. The first 50 residues were not considered in the RMSD calculations because the TM-helix can move independently in the membrane.

## Supplemental References

1. Rathi, P. C.; Mulnaes, D.; Gohlke, H., VisualCNA: a GUI for interactive constraint network analysis and protein engineering for improving thermostability. *Bioinformatics* **2015**, 31, 2394-6.
2. Walt, S. v. d.; Colbert, S. C.; Varoquaux, G., The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering* **2011**, 13, 22-30.
3. Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Vijaykumar, A.; Bardelli, A. P.; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G.-L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; de Miranda Cardoso, J. V.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmerer, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y.; SciPy, C., SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, 17, 261-272.
4. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, 9, 3084-3095.
5. Homeyer, N.; Gohlke, H., FEW: A workflow tool for free energy calculations of ligand binding. *J. Comput. Chem.* **2013**, 34, 965-973.
6. Åqvist, J.; Luzhkov, V. B.; Brandsdal, B. O., Ligand Binding Affinities from MD Simulations. *Acc. Chem. Res.* **2002**, 35, 358-365.
7. Brandsdal, B. O.; Österberg, F.; Almlöf, M.; Feierberg, I.; Luzhkov, V. B.; Åqvist, J. Free Energy Calculations and Ligand Binding. In *Adv. Protein Chem.*; Academic Press: 2003; Vol. 66, pp 123-158.
8. Biedermannova, L.; E. Riley, K.; Berka, K.; Hobza, P.; Vondrasek, J., Another role of proline: stabilization interactions in proteins and protein complexes concerning proline and tryptophane. *Phys. Chem. Chem. Phys.* **2008**, 10, 6350-6359.