

**Deorphanizing Human Cytochrome P450 Enzymes CYP4A22 and CYP4Z1 through
Mechanistic *in silico* modeling**

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy
of Freie Universität Berlin

by
David Machalz

Berlin
2021

The present thesis was prepared from November 2016 till July 2021 under the supervision of Prof. Dr. Gerhard Wolber at the Institute of Pharmacy of Freie Universität Berlin.

- | | |
|---------------------------|-------------------------------|
| 1. Reviewer/Gutachter*in: | Prof. Dr. Gerhard Wolber |
| 2. Reviewer/Gutachter*in: | Prof. Dr. Maria Kristina Parr |

Thesis defense on/Disputation am: 01. Dezember 2021

Give us insight, not numbers.

Charles Coulson, theoretical chemist, 1959

Acknowledgments

None of this research work would have been possible without the constant support from Prof. Gerhard Wolber. Thanks for all your advice and especially on science communication, molecular modeling and informatics. Thank you for putting your trust in me, which ultimately led to this piece of work.

Special thanks to Prof. Matthias Bureik from Tianjin University without whom my computational predictions would have been just dangling in the air. I am grateful for our continuous and truly collaborative work on cytochrome P450 and glucuronosyltransferase enzymes.

Furthermore, I would like to thank Prof. Maria Parr from the Freie Universität for multiple collaboration opportunities on steroid biosynthesis and action, as well as sulfotransferases. I also would like to thank her for co-supervising my doctoral studies and reviewing my thesis.

I would like to thank all members of the AG Wolber for fun, meaningful daily input and solidarity in tougher times. I want to especially mention Dr. Alexandra Nass, Dr. David Schaller, Szymon Pach, Dr. Dora Šribar, Theresa Noonan and Dr. Marcel Bermudez. I especially like to thank Szymon Pach, Dr. Dora Šribar, Theresa Noonan and Dr. Marcel Bermudez for carefully reading this thesis.

I would like to thank my working station N3 at the Institute of Pharmacy and the FU CURTA cluster for carrying out meaningful and some less meaningful molecular modeling calculations.

In parallel to my academic education, I would like to thank the late-night host Conan O'Brien for teaching me how to be silly, humorous, and intelligent at the same time as suitable for an academic setting.

Finally, I would like to thank my friends and family, especially my parents, for supporting me throughout this endeavor.

Abstract

Cytochrome P450 (CYP) enzymes are monooxygenases that catalyze the oxidation of structurally diverse substrates and are present in various lifeforms, including humans. Human CYPs catalyze the metabolism of xenobiotics including drugs and are involved in the essential biosynthesis of steroids, vitamins, and lipids. CYP-catalyzed metabolism and biosynthesis has been extensively studied recently, but several CYPs remain understudied despite their potential role in key biotransformation pathways. For these so-called 'orphaned CYPs', physiological function and structure are yet unknown, such as for CYP4A22 and 4Z1. CYP4A22 catalyzes the ω -hydroxylation of arachidonic acid to the angiogenic 20-hydroxyeicosatetraenoic acid. CYP4Z1 is overexpressed in breast cancer and other malignancies, which is correlated with tumor progression. Hence, CYP4Z1 is considered a promising breast cancer target that was not previously addressed by small molecule inhibitors. Here, we report our efforts to deorphanize CYP4A22 and 4Z1 together with our experimental partner Prof. Bureik. We were the first to predict the structure of CYP4A22 and 4Z1 by homology modeling and overcame the challenge of low-sequence similarity templates by incorporating substrate activities. We applied substrate docking and 3D pharmacophore modeling to rationalize how the binding site structure determines structure-activity relationships (SAR) trends. The well-known structural flexibility of CYPs was partially accounted for by molecular dynamics simulations. For the first time, enzyme-substrate interactions dynamics were analyzed with our novel dynamic pharmacophore approach, which led to the prediction of key residues. For CYP4A22, a residue influencing ω -hydroxylation (Phe320) and two binding residues (Arg96 and Arg233) were predicted. For CYP4Z1, the key role of Arg487 and assisting role of Asn381 for substrate binding were predicted, which was validated by *in vitro* mutational studies. The thereby validated CYP4Z1 model and substrate SAR were used in a virtual screening campaign resulting in a new potent and selective CYP4Z1 inhibitor (IC_{50} : 63 ± 19 nM). Taken together, we established an *in vitro/in silico* deorphanization protocol that shed light on the structure-function relationships of CYP4A22 and 4Z1. This enabled us to discover a potent inhibitor of CYP4Z1 that will allow further studies on the physiological and pathophysiological role of the enzyme and might be further

improved to target CYP4Z1 in a new therapeutical approach. Similar workflows could easily be applied to study other neglected enzymes in metabolism and other biotransformation pathways.

Zusammenfassung

Cytochrom P450 (CYP)-Enzyme sind Monooxygenasen, die die Oxidation strukturell diverser Substrate katalysieren und in verschiedenen Lebensformen, einschließlich des Menschen, vorkommen. Menschliche CYPs katalysieren den Metabolismus von Xenobiotika einschließlich Arzneistoffen und sind an der essenziellen Biosynthese von Steroiden, Vitaminen und Lipiden beteiligt. CYP-katalysierter Metabolismus und Biosynthese wurden in der Vergangenheit intensiv untersucht, aber einige CYPs sind trotz ihrer potenziellen Rolle in wichtigen Biotransformationswegen noch wenig erforscht. Für diese so genannten „orphaned“ oder „verwaisten“ CYPs, sind physiologische Funktion und Struktur noch unbekannt, wie z.B. CYP4A22 und 4Z1. CYP4A22 katalysiert die ω -Hydroxylierung von Arachidonsäure zu der angiogenen 20-Hydroxyeicosatetraensäure. CYP4Z1 wird bei Brustkrebs und anderen malignen Erkrankungen überexprimiert, was mit der Tumorprogression korreliert ist. Daher wird CYP4Z1 als ein vielversprechendes Brustkrebs-Target angesehen, das bisher nicht durch niedermolekulare Inhibitoren adressiert wurde. Hier berichten wir über unsere Bemühungen, CYP4A22 und 4Z1 zusammen mit unserem experimentellen Partner Prof. Bureik zu deorphanisieren. Wir waren die Ersten, die die Struktur von CYP4A22 und 4Z1 durch Homologiemodellierung vorhersagten und überwandten die Herausforderung der Templates mit geringer Sequenzähnlichkeit, indem wir Substrataktivitäten mit einbezogen. Wir wendeten Substrat-Docking und 3D-Pharmakophor-Modellierung an, um zu rationalisieren, wie die Struktur der Bindungstasche die Trends der Struktur-Aktivitäts-Beziehungen (SAR) bestimmt. Die bekannte strukturelle Flexibilität von CYPs wurde partiell durch Molekulardynamik-Simulationen berücksichtigt. Zum ersten Mal wurde die Dynamik der Enzym-Substrat-Interaktionen mit unserem neuartigen dynamischen Pharmakophor-Ansatz analysiert, was zur Vorhersage von wichtigen Aminosäuren führte. Für CYP4A22 wurde eine Aminosäure, die die ω -Hydroxylierung beeinflusst (Phe320) und zwei Bindungsaminosäuren (Arg96 und Arg233) vorhergesagt. Für CYP4Z1 wurde die Schlüsselrolle von Arg487 und die unterstützende Rolle von Asn381 für die Substratbindung vorhergesagt, welche durch *in vitro* Mutationsstudien validiert wurde. Das dadurch validierte CYP4Z1-Modell und die Substrat-SAR wurden in einer

virtuellen Screening-Kampagne verwendet, die zu einen neuen potenten und selektiven CYP4Z1-Inhibitor führte (IC_{50} : 63 ± 19 nM). Zusammengenommen haben wir ein *in vitro/in silico* Deorphanisierungsprotokoll etabliert, welches die Struktur-Funktionsbeziehungen von CYP4A22 und 4Z1 beleuchtet. Dies versetzte uns in die Lage einen potenten Inhibitor von CYP4Z1 zu entdecken, der weitere Studien über die physiologische und pathophysiologische Rolle des Enzyms ermöglichen wird und möglicherweise weiter verbessert werden kann, um CYP4Z1 in einem neuen therapeutischen Ansatz zu adressieren. Ähnliche Arbeitsabläufe könnte leicht angewendet werden, um andere vernachlässigte Enzyme im Metabolismus und anderen Biotransformationswegen zu untersuchen.

Table of Contents

Acknowledgments	vii
Abstract.....	ix
Zusammenfassung	xi
1 Introduction.....	1
1.1 The Human Cytochrome P450 Enzyme Family	1
1.2 The Chemistry of Human Cytochrome P450 Enzymes	4
1.3 The Structure of Human Cytochrome P450 Enzymes.....	8
1.4 The Relevance of Understudied Human Cytochrome P450 Enzymes	13
1.4.1 Structural Insights into Human Understudied Cytochrome P450 Enzymes	14
2 Research Aim	25
3 Computational Methods	27
3.1 Homology Modeling.....	27
3.2 Molecular Docking	32
3.3 Molecular Dynamics Simulations	34
3.4 Static and Dynamic 3D Pharmacophores and Virtual Screening	37
4 Results	41
4.1 Efficient Substrate Screening and Inhibitor Testing of Human CYP4Z1 Using Permeabilized Recombinant Fission Yeast	41
4.2 Functional Characterization and Mechanistic Modeling of the Human Cytochrome P450 Enzyme CYP4A22	56
4.3 Importance of Asparagine-381 and Arginine-487 for Substrate Recognition in CYP4Z1	74
4.4 Discovery of a novel potent cytochrome P450 CYP4Z1 inhibitor	86
5 Discussion.....	105
5.1 Characterization of CYP4Z1 <i>in vitro</i> and <i>in silico</i>	105

5.2	Characterization of CYP4A22 <i>in vitro</i> and <i>in silico</i>	107
5.3	Dissecting and Validating the CYP4Z1 Binding Site Model	110
5.4	Targeting CYP4Z1 by Rational Computational Drug Design.....	112
6	Conclusion.....	117
7	Bibliography.....	119
8	Appendix.....	133
8.1	List of Abbreviations.....	133
8.2	List of Figures.....	134
8.3	List of Tables.....	136
9	Publications	137
9.1	Publications included in this cumulative thesis.....	137
9.2	All peer-reviewed Publications	138
9.3	Conference Proceedings	140

1 Introduction

1.1 The Human Cytochrome P450 Enzyme Family

Cytochrome P450 (CYP or P450) enzymes are present in many lifeforms, such as plants, fungi, bacteria, and animals [1, 2]. CYP enzymes form a superfamily of monooxygenases that catalyze the insertion of a single oxygen into a substrate R (see Equation 1) [3]. The remaining oxygen atom of the dioxygen molecule is reduced to water. The reaction mechanism is discussed later in detail (see 1.2 The Chemistry of Human Cytochrome P450 Enzymes).



In order to bind the dioxygen for catalysis, CYPs use a prosthetic group named heme B (Figure 1). Prosthetic groups are cofactors that tightly bind to the enzyme. Heme B consists of an iron atom that is coordinated by a protoporphyrin IX moiety [4]. When the heme iron is at its ferrous state (Fe^{2+}) and binds carbon monoxide, an absorption maximum at 450 nm can be observed for the enzyme in the visible spectrum. This observation led to the names “pigment 450”, “P450”, and “P-450” for the enzyme family [5].

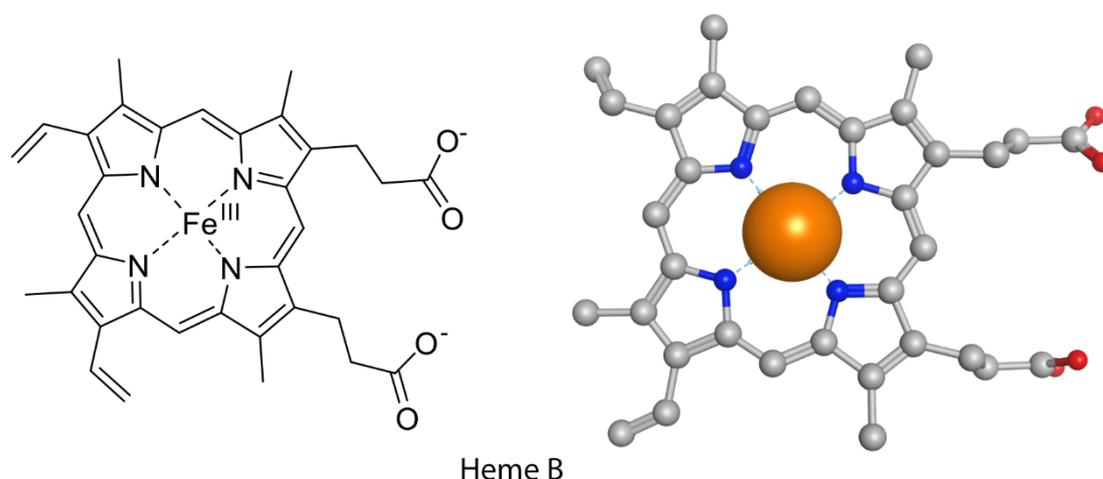


Figure 1. Heme B is the prosthetic group of cytochrome P450 enzymes. In 3D the heme is shown as ball and stick. Iron is shown as a large orange sphere, oxygen atoms are colored red, and nitrogens are colored, and carbons are light grey. Hydrogens have been omitted in both representations.

CYP genes are named after a general nomenclature that starts with the abbreviation CYP and is followed by a number that denotes the CYP family. The subsequent letter

indicates the subfamily and the last number in the name stands for the individual CYP isoform or isoenzyme. For example, the gene name CYP1A2 represents the second CYP isoform of the CYP1A subfamily in the CYP1 family. CYP families show at least 40% and subfamilies at least 55% sequence identity [2].

Table 1. Overview of the 57 human CYP enzymes. CYP families are indicated by background coloring. The following abbreviations are used DR drugs, ES eicosanoids, SS sex steroids, XB xenobiotics.

Gene name	Expression location [6]	Function(s) [7]
CYP1A1	Lung, several extrahepatic sites	metabolism of ES; XB
CYP1A2	Liver	metabolism of ES; DR and XB
CYP1B1	Several extrahepatic sites	metabolism of ES; XB
CYP2A6	Liver, lung, and several extrahepatic sites	metabolism of ES; DR and XB
CYP2A7	unknown	nitrosamine metabolism
CYP2A13	Nasal tissue	metabolism of ES; DR and XB
CYP2B6	Liver, Lung	metabolism of ES; DR and XB
CYP2C8	Liver	metabolism of ES; DR and XB
CYP2C9	Liver	metabolism of ES; DR and XB; contributes to warfarin metabolism
CYP2C18	Liver	metabolism of ES; DR and XB
CYP2C19	Liver	metabolism of ES; DR and XB
CYP2D6	Liver	metabolism of ES; DR and XB
CYP2E1	Liver, Lung, other tissues	metabolism of ES; DR and XB
CYP2F1	Lung	metabolism of ES; DR and XB
CYP2J2	Lung	metabolism of ES; DR and XB
CYP2R1	Liver	vitamin D 25-hydroxylase
CYP2S1	Lung	metabolism of ES; DR and XB; retinoids
CYP2U1	Thymus, brain	metabolism of ES; long-chain fatty acids; DR and XB
CYP2W1	Tumors	metabolism of ES; DR and XB
CYP3A4	Liver, small intestine	metabolism of ES, SS; greater than or equal to 60% of all clinically used DR; XB
CYP3A5	Liver, lung	metabolism of ES, SS; DR and XB
CYP3A7	Fetal liver	metabolism of ES, SS; DR and XB
CYP3A43	Brain, liver	metabolism of ES, SS; DR and XB
CYP4A11	Liver, kidney	metabolism of ES, medium- and long-chain fatty acids; DR and XB
CYP4A22	Liver, kidney	function(s) unknown
CYP4B1	Lung	metabolism of ES; DR and XB

Gene name	Expression location [6]	Function(s) [7]
CYP4F2	Liver	metabolism of ES; DR and XB; contributes to warfarin metabolism
CYP4F3	Neutrophils	metabolism of ES; DR and XB
CYP4F8	Seminal vesicles	metabolism of ES; DR and XB
CYP4F11	Liver	function(s) unknown
CYP4F12	Liver	metabolism of ES; DR and XB
CYP4F22	Liver	function(s) unknown
CYP4V2	Eye	function(s) unknown
CYP4X1	Liver, brain	function(s) unknown; anandamide metabolism in brain
CYP4Z1	Mammary gland	metabolism of ES, fatty acids; DR and XB
CYP5A1	Platelets	eicosanoid metabolism (thromboxane A2 synthase), participates in platelet aggregation (official gene name: TBXAS1)
CYP7A1	Liver	associated with elevated LDL-cholesterol levels; cholesterol 7 α -hydroxylase, biosynthesis of bile acids, metabolism of oxysterols
CYP7B1	Brain	oxysterol 7 α -hydroxylase, neurosteroid 7 α -hydroxylase
CYP8A1	Aorta, others	eicosanoid metabolism (prostacyclin (PGI ₂) synthase), participates in platelet disaggregation (official gene name: PTGIS)
CYP8B1	Liver	sterol 12 α -hydroxylase
CYP11A1	Adrenals, other steroidogenic tissue	cholesterol side-chain cleavage
CYP11B1	Adrenals	steroid 11 β -hydroxylase
CYP11B2	Adrenals	steroid 11 β - and 18-hydroxylase and 18-oxidase
CYP17A1	Steroidogenic tissue	steroid 17 α -hydroxylase, 17,20-lyase
CYP19A1	Steroidogenic tissue, adipose	androgen aromatase, oestrogen synthetase
CYP20A1	Liver, other tissue	function(s) unknown
CYP21A2	Steroidogenic tissue	steroid 21-hydroxylase
CYP24A1	Kidney	vitamin D 24-hydroxylase
CYP26A1	Several	retinoid acid inactivation (hydroxylase)
CYP26B1	Brain	retinoid acid inactivation (hydroxylase)
CYP26C1	unknown	retinoid acid inactivation (hydroxylase)
CYP27A1	Liver	bile acid biosynthesis, sterol 27-hydroxylase, vitamin D 25-hydroxylase
CYP27B1	Kidney	25-hydroxy-vitamin D 1 α -hydroxylase
CYP27C1	Liver	function(s) unknown
CYP39A1	Liver, other tissues	oxysterol 7 α -hydroxylase
CYP46A1	Brain	cholesterol 24-hydroxylase (brain)
CYP51A1	Liver, testes	lanosterol 14 α -demethylase

In humans, 57 unique CYP isoforms are known (Table 1) out of 13,000 sequenced CYPs in animals [8]. They are mainly expressed in the liver, but also in various other tissues. Human CYPs are essential biochemical tools that contribute to the biosynthesis of steroids, vitamins, and lipids [6, 7, 9-11]. CYP enzymes are known for their pivotal role in the metabolism of drugs and other chemicals foreign to the body, which are referred to as xenobiotics [6, 9, 12]. CYPs are involved in the metabolic fate of 75% of all drugs [6, 12]. Often, CYPs are not only involved in the metabolism of endogenous compounds or xenobiotics but also might act on both substrate groups [9], as apparent from Table 1. Some CYPs are also involved in genetic diseases and cancer [6, 7] or constitute drug targets themselves [6, 9, 13]. So far, research focused on drug metabolizing CYPs; hence, there are still CYP enzymes without known physiological function, which are also referred to as orphaned or neglected CYPs [14, 15]. Recently, all 57 human CYP isoforms have been expressed in their functional form and enzymatic activity was shown for the orphaned CYP2A7, CYP4A22, and CYP20A1 for the first time [16].

1.2 The Chemistry of Human Cytochrome P450 Enzymes

CYP enzymes are monooxygenases that facilitate the insertion of a single oxygen atom into the substrate as mentioned previously. Chemicals that are inert under physiological conditions, such as alkanes, can be oxygenated by the catalysis of CYPs at body temperature. Most CYP-catalyzed biotransformation reactions are carried out by the reactive iron-oxo species of CYPs named Compound I (Figure 2) [3, 17-21]. Compound I is a positively charged species that contains the heme, a cysteine as the fifth ligand, and an oxygen atom as the sixth ligand. The electron deficiency of Compound I makes it an ideal oxidant [3, 17-19]. Compound I was first described in 1976 by the laboratories of Groves and Coon [20, 21]. The production of Compound I is the focus of biochemical CYP activity. The impact of CYPs on the substrate conversion itself is mainly of steric nature [22]. The shape of the catalytic pocket of the CYP enzyme influences the steric accessibility of the substrate and thereby the stereo- and regioselectivity of the reaction [23] (see 1.3 The Structure of Human Cytochrome P450 Enzymes).

Most CYP-catalyzed oxidations follow a general catalytic reaction cycle with eight steps (Figure 2) [3, 24]. In the resting state (step 1) the heme iron is in its ferric (Fe^{3+}) state and is coordinated by a water molecule. The substrate binds, which initiates step 2; the first electron is transferred for the reduction of ferric to ferrous iron [25, 26]. This also acts as a protection mechanism against the formation of toxic superoxides and peroxides and waste of redox equivalents. For microsomal CYPs, the required electrons are in most cases provided by the redox partner cytochrome P reductase (POR, NADPH-P450 reductase) and in some cases by cytochrome b5. For some bacterial or mitochondrial CYPs, electrons can also originate from ferredoxin [24].

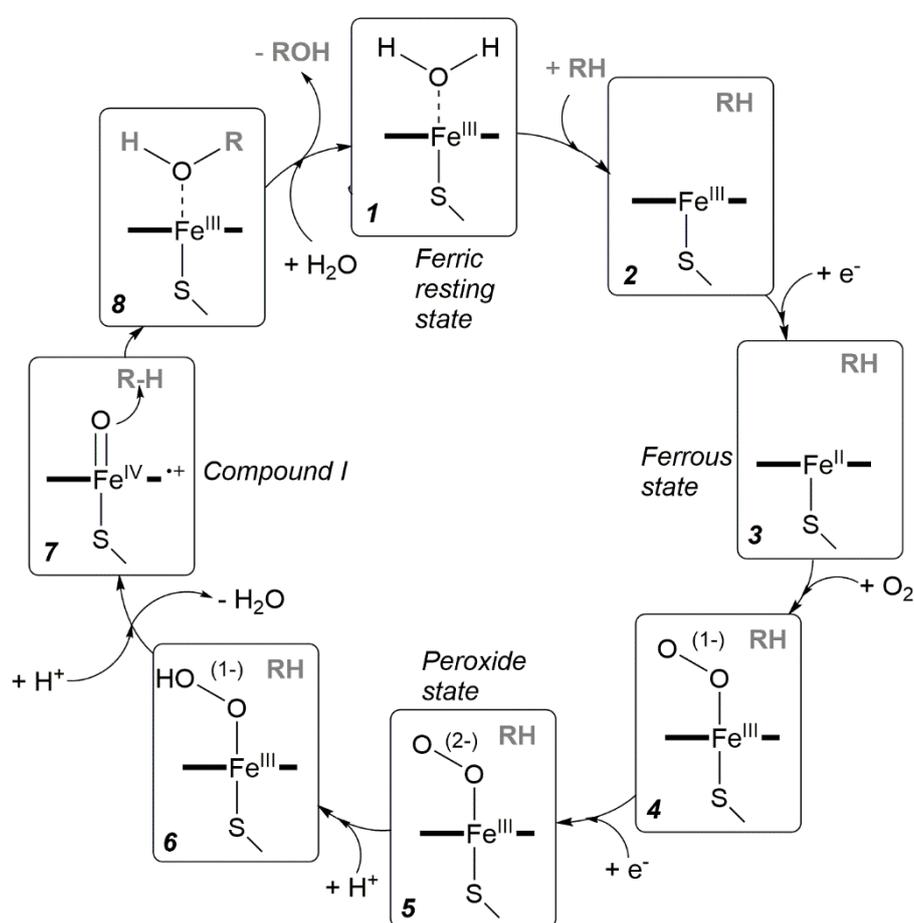


Figure 2. The catalytic reaction cycle of cytochrome P450 enzymes. Adapted from [27].

Substrate binding usually leads to the exit of active site waters including the water coordinating the heme iron. Consequently, the ferric iron switches from the low spin ($S=1/2$) hexacoordinated to the high spin ($S=5/2$) pentacoordinated state, which has an increased redox potential. The increased redox potential accelerates the reduction to ferrous iron (Fe^{2+}) in step 3. Now dioxygen can bind to the ferrous iron. The iron

transfers one electron to the dioxygen and returns to its ferric state. The dioxygen is reduced to superoxide in step 4. Another single electron is transferred and leads to the reduction of dioxygen to peroxide in step 5. Protonation of the distal oxygen leads to a hydroperoxide intermediate in step 6. In step 7 the distal oxygen undergoes a second protonation, which leads to the cleavage of the O-O bond. The distal oxygen atom leaves as a water molecule and a π -cationic ferryl-oxo intermediate is formed, which is also referred to as the previously described Compound I. An “oxygen rebound” mechanism was proposed as follows; Compound I cleaves off a hydrogen atom from the targeted C-H bond of the substrate R. This is the rate-limiting step of substrate oxidation and results in a hydroxyl radical. The hydroxyl radical is bound to the heme iron and subsequently reacts with the substrate’s carbon radical [20, 21]. The final product is an oxygenated metabolite in state 8 that can now leave the active site. The heme iron is again coordinated by a water molecule and the rest state 1 is reached [3, 24]. The CYP enzyme can now accommodate the next substrate for oxygenation.

The spectrum of biotransformation reactions catalyzed by CYPs can go beyond simple substrate oxygenation (Figure 3) [22, 24, 28, 29]. The most common biotransformation of xenobiotics is the hydroxylation of aliphatic and aromatic carbon atoms catalyzed by CYPs (Figure 3, reactions 1 and 6) [29]. During the hydroxylation reaction, an oxygen atom is inserted into a carbon-hydrogen bond, which results in a new hydroxy moiety. Aromatic hydroxylation produces an epoxide intermediate, which then opens and reorients to a phenolic moiety [22, 28]. The epoxide of an aromatic ring or a carbon-carbon double bond can also be the product in cases where epoxide opening is not possible (Figure 3, reaction 11). Dealkylation reactions can occur when a carbon atom that is bonded to a heteroatom, such as nitrogen, oxygen, or sulfur, is hydroxylated. In a subsequent reaction, the bond between heteroatom and hydroxylated carbon is broken. The final products are a protonated heteroatom moiety and an aldehyde group on the carbon. In amines or ethers, this biotransformation reaction is called N- or O-dealkylation [22, 28] (Figure 3, reactions 4 and 5). Nitrogen, sulfur, or phosphorus atoms can also be oxygenated by CYPs (Figure 3, reactions 7 to 10). The time-limiting step here is the transfer of a single electron from the heteroatom to Compound I, which often requires a higher activation energy than the hydrogen abstraction from carbon

atoms [22, 28]. Even the nitrogen of dihydropyridine can be oxidized by CYPs (Figure 3, reaction 12). Furthermore, uncommon reactions have been described for CYPs, such as the formation or expansion of rings, desaturation, and reductions [28].

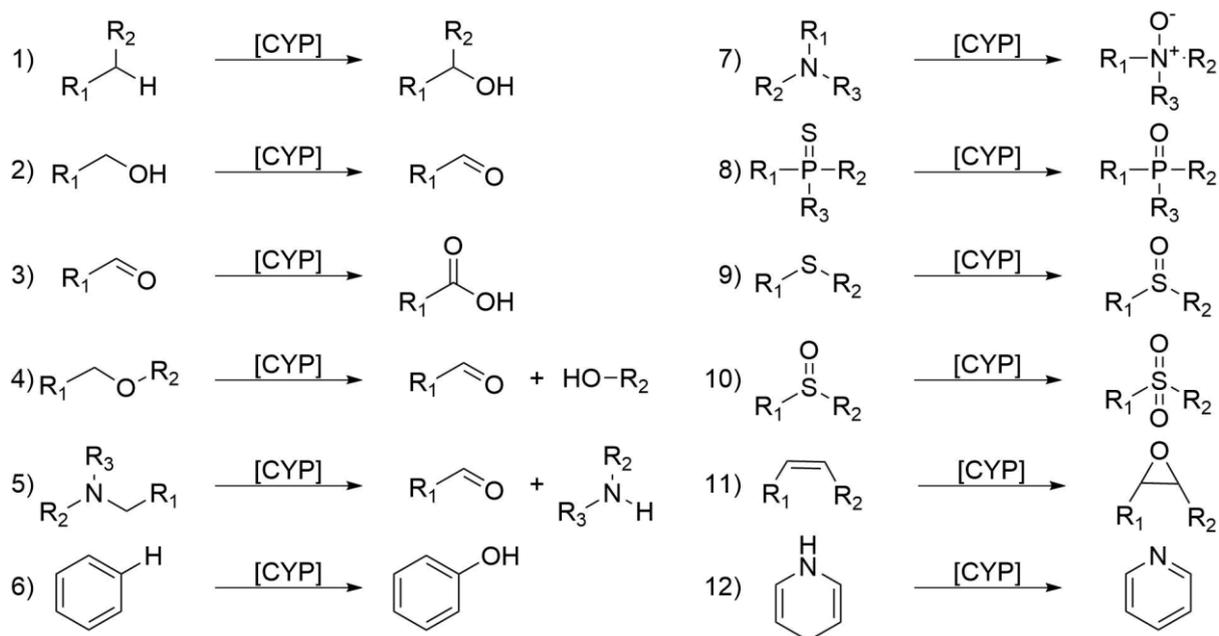


Figure 3. Examples of oxidation reactions catalyzed by Cytochrome P450 enzymes [22, 28]. (1) Hydroxylation of aliphatic carbon. (2) Oxidation of alcohol to aldehyde. (3) Oxidation of aldehyde to carboxylic acid. (4) O-dealkylation of ether. (5) N-dealkylation of amine. (6) Hydroxylation of aromatic carbon. (7, 8, 9, and 10) Heteroatom oxidation: N-oxidation, P-oxidation, and S-oxidation. (11) Epoxidation of carbon-carbon double bond. (12) Oxidation of dihydropyridine to pyridine. The reactions lead to the insertion of a single oxygen atom and for (4) and (5) to subsequent reactions.

Which substrate moiety will be biotransformed under CYP catalysis? This is mainly influenced by the electronic structure and shape of the substrate and the structure and flexibility of the CYP enzyme. As mentioned above, the rate-limiting step in CYP-catalyzed reactions is either the hydrogen abstraction or the transfer of a single electron from the heteroatom. Substrate moieties that are highly solvent-exposed and have low activation energy for the rate-limiting reaction step will be most reactive [23, 30]. According to quantum mechanical (QM) calculations, different CYP reaction mechanisms require different optimal substrate orientations [30-33]. The CYP enzyme controls the orientation of the substrate towards the heme moiety in the catalytic pocket [23, 24]. Highly reactive substrate moieties, such as secondary or tertiary carbons might be out of reach for oxygenation due to the shape of the catalytic pocket of the CYP enzyme. As a result, a less reactive substrate moiety might react instead such as ω -hydroxylation where a primary carbon in alkanes or fatty acids is

hydroxylated [34]. Structural flexibility of the CYP enzyme might make different substrate orientations possible. As a result, the same substrate might be biotransformed into several different metabolites, depending on the different substrate orientations in the catalytic pocket.

1.3 The Structure of Human Cytochrome P450 Enzymes

Human CYP enzymes comprise a membrane anchoring domain and a globular catalytic domain connected by a flexible loop (Figure 4). The globular domain can be separated into two sides by the plane of the catalytic heme moiety (Figure 5). The proximal side of the globular domain contains the cysteine that coordinates the heme moiety and interacts with the redox partner, such as POR. The distal side exhibits higher structural flexibility and contains the catalytic pocket that binds substrates [35]. CYP enzymes and their redox partner are located at the membrane of the endoplasmic reticulum or the inner mitochondrial membrane. Both CYPs and POR are anchored to the membrane via a transmembrane helix domain [36-38]. A rapid turnover is assured by the formation of a loose complex of the CYP enzyme and its redox partner protein and their rapid dissociation. An X-ray structure of the complex of the heme and Flavin mononucleotide binding domain of CYP102A1 of *Bacillus megaterium* shows an interface close to the cysteine ligand of the heme in the core of the globular catalytic domain (PDB ID: 1BVY) [39].

The CYP membrane anchoring domain is formed by 30 to 50 residues on the N-terminal end that precedes the globular catalytic domain in the sequence. Within the anchoring domain, a transmembrane helix is formed by a stretch of 20 hydrophobic residues roughly 30 Å in length spanning the full width of the hydrophobic core of the bilayer. The transmembrane helix is connected to the catalytic domain via a linker of 10 residues. The residues of the linker domain are often positively charged, which are likely to interact with the negative charge of the lipid headgroup of the bilayer. Variations in linker length likely induce alterations in the membrane placement of the catalytic domain [36]. In some CYPs, such as CYP51A1 (Figure 5), the membrane anchoring domain contains additional residues that precede the transmembrane helix [36, 40]. These additional residues form an amphiphilic helix that resides in the lipid head region on the distal side of the bilayer. CYP19A1 (aromatase) was expressed in

Escherichia coli without the anchoring domain retained its catalytic activity [41]. This is also the case for various other CYP isoforms, leading to the conclusion that the anchoring domain is not necessary for catalytic activity [36]. The majority of the crystallized CYPs were expressed without the anchoring domain as apparent from the CYP X-ray structure models deposited in the protein data bank (PDB, www.rcsb.org).

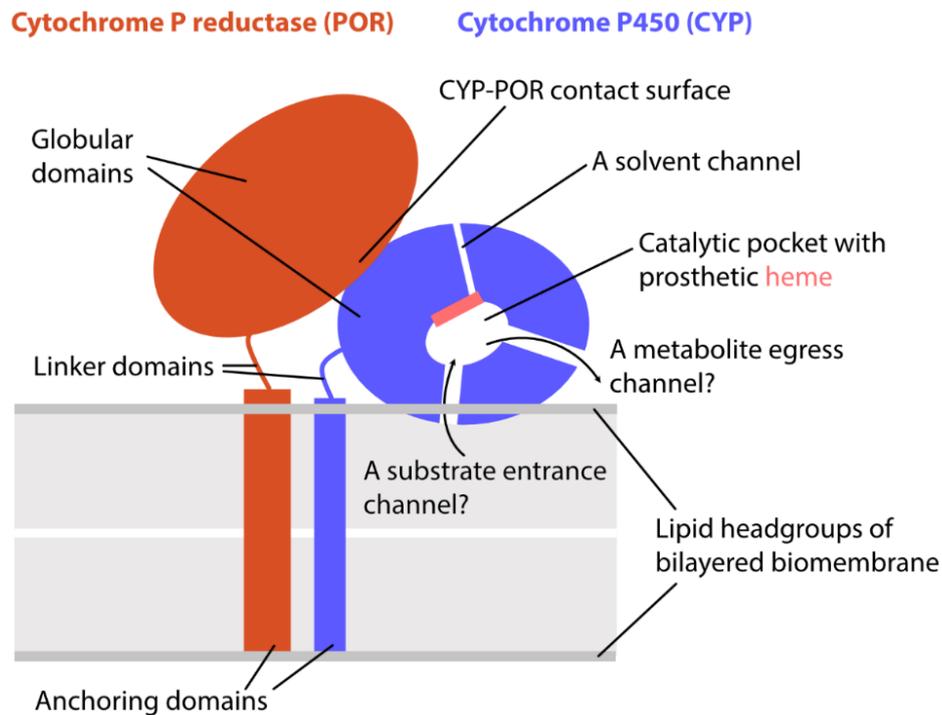


Figure 4. CYPs and their redox partner, such as cytochrome P reductase reside in a membrane. Different channels that connect the catalytic pocket and the enzyme exterior are surmised. Their function is still poorly understood as there is little consensus between CYP isoforms [42].

The general CYP fold of the catalytical globular domain is conserved and comprises 12 α -helices labeled A to L and four β -sheets named β 1-4 (Figure 5) [36, 37]. In some CYP enzymes loops between helices also contain inserted shorter helices, that are labeled with the letter of the helix closest in the sequence and a prime. Most notable is the F/G loop between helix F and G in mammalian CYPs, where two short helices F' and G' are inserted [37]. The F/G loop and parts of the B/C domain are reported to act as a lid above the catalytic pocket and the F/G loop dips into the membrane (Figure 5) [36, 38, 43]. Antibodies targeting the hydrophobic surface patch in the lid in CYP2B4 did not bind, which indicates that the surface patch of the F/G loop is buried in the membrane [44]. Variations in the F/G segment are likely to shift the orientation of the catalytic domain towards the bilayer and the extent of membrane insertion.

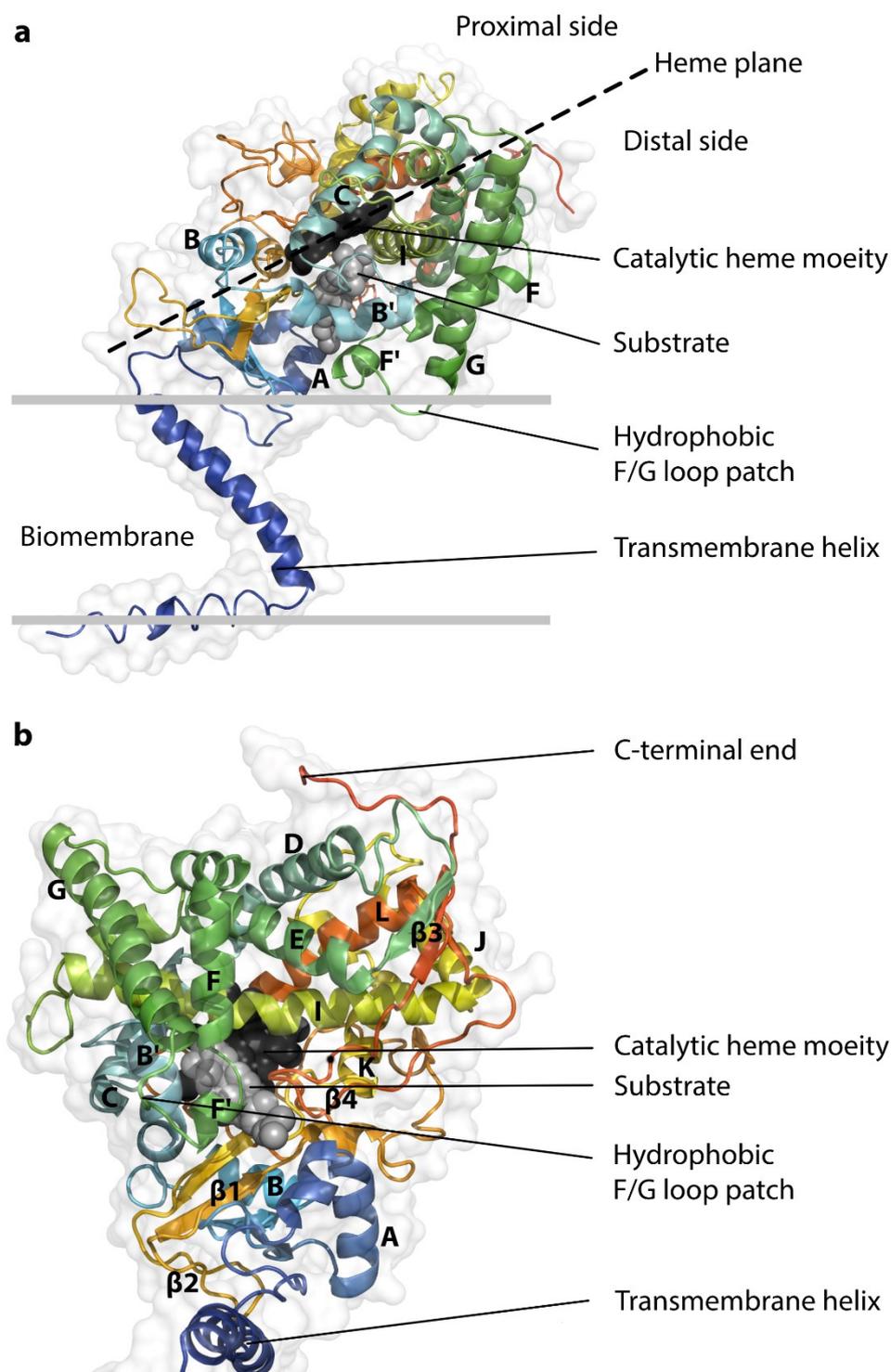


Figure 5. The general fold of Cytochrome P450 enzymes. (a) 3D structure of lanosterol 14- α demethylase (CYP51) the only CYP enzyme with an X-ray structure including the transmembrane helix in side view (PDB ID: 4LXJ) [40]. (b) CYP51 globular domain is shown from the bottom view where the plane of the heme is equal to the paper plane. 3D structure of the backbone is shown in cartoon representation and secondary structures are labelled.

The catalytic pocket of CYPs is buried in the center of their globular domain and accommodates the prosthetic heme moiety (Figure 5). The closer to the heme the more conserved are the structural elements, such as helices I and L, and the environment of the heme thiolate required for the catalytic CYP activity. Helix I is situated right below

the heme and adjacent to the heme (Figure 6).[37]. In CYP4B1, CYP4A11, CYP4A22, and CYP4V2 the heme is covalently bound to helix I via glutamate, which restricts the catalytic pocket laterally [34]. The heme iron is coordinated by a thiolate moiety of a cysteine residue that is situated in a highly conserved and rigid β -bulge [36]. The cysteine residue forms hydrogen bonds to two other residues assured by local rigidity. The first hydrogen bond is formed between the cysteine's thiolate chain and the backbone amide NH moiety of a second residue. The second hydrogen bond is formed between the cysteine's backbone amide NH moiety and the backbone oxygen amide moiety of a nearby phenylalanine residue. The redox potential is regulated with the help of these two hydrogen bonds [45]. The carboxylate moieties of the heme are anchored by a network of basic residues [37]. Helix I is distorted due to a threonine residue that uses its side chain hydroxy moiety for intrahelical hydrogen bonding instead of its backbone amide NH moiety (Figure 6). Although the threonine residue is not conserved throughout the whole CYP superfamily, it is not yet clearly understood how the distorted fold of the I helix is conserved, nevertheless. The distortion opens a small pocket in the active site between the heme and the I helix. The pocket might be already present in the holo form or open upon substrate binding and can accommodate a network of water molecules. The water network is believed to provide the protonation necessary for the heterolytic cleavage of the O-O bond in the dioxygen molecule, which is a crucial step in the catalyzed oxidation reaction (see 1.2 The Chemistry of Human Cytochrome P450 Enzymes) [3]. Binding of the redox partner putidaredoxin to CYP101A1 leads to the opening of the I helix cleft and a short channel connects the I helix cleft to bulk water that handles the dioxygen protonation [46].

In most X-ray structures of CYPs, the catalytic pocket near the heme thiolate moiety is closed off from the enzyme's environment. For substrates to enter and metabolites to exit the catalytic pocket, large-scale and concerted movements in the distal part of the globular CYP domain are required. X-ray structures of the open and closed state of the active site of several CYP isoforms, such as rabbit CYP2B4, indicate that several helices are involved in the opening of a channel towards the membrane including the F/G segment [47, 48]. Additionally, the F/G segment controls the opening and collapse of

various channels that connect the buried catalytic site with the distal enzyme surface. While some channels are quite narrow that only water can pass, others allow for substrate entrance and metabolite egress. The growing effort to characterize entry and egress channels shows that there is little consensus between CYP isoforms and that they might also contribute to substrate selectivity [42]. Most CYP substrates are rather hydrophobic and likely accumulate in the membrane below the CYP. Substrate access channels hence lead from the membrane to the catalytic pocket. After substrate oxygenation, the resulting metabolite can exit the catalytic pocket towards the solvent through an egress channel (Figure 4) [49].

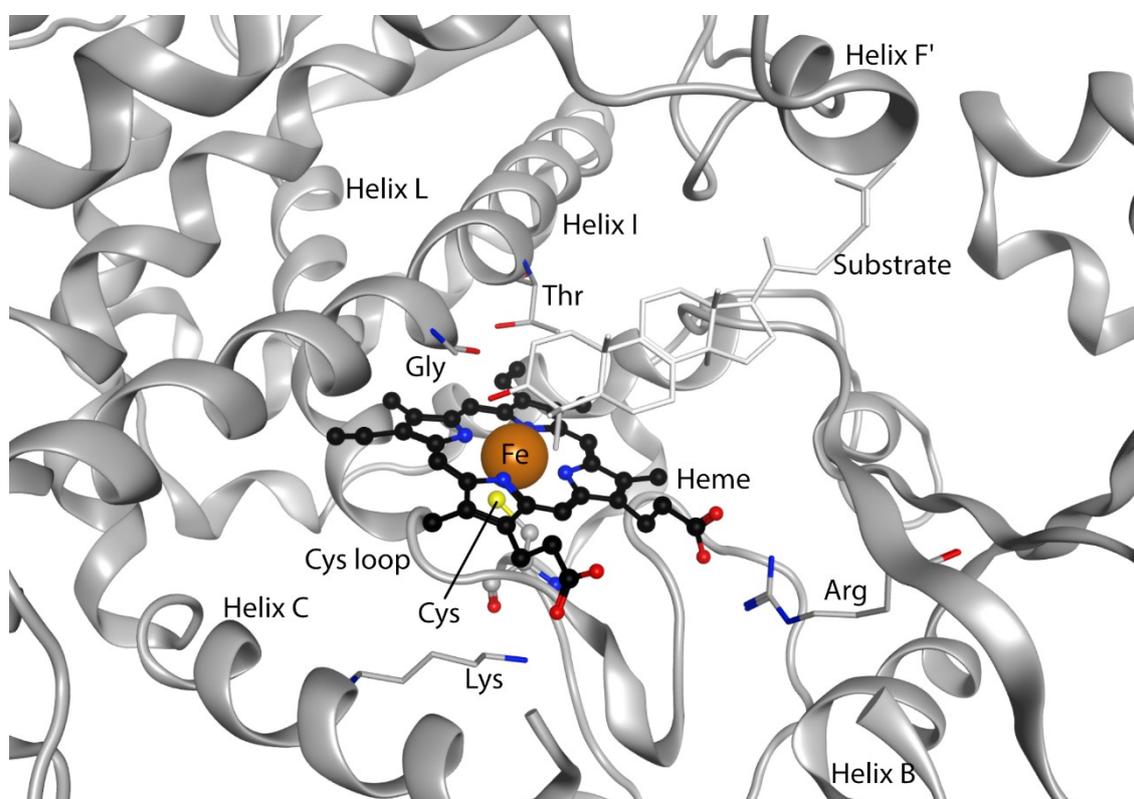


Figure 6. The catalytic pocket of Cytochrome P450 enzymes. The substrate (lanosterol) is situated above the prosthetic heme moiety ready for oxygenation in CYP51 (PDB ID: 4LXJ) [40]. The heme iron is coordinated by a cysteine residue that is situated in the Cys loop. Helix I is distorted by a hydrogen bond between the side chain of Thr and the backbone of the residue four positions away in the helix, here Gly. In some CYP enzymes Gly is mutated to Glu, which then forms a covalent bond to the heme. The heme moiety is held in place by salt bridges to basic residues (Arg and Lys). For sake of clarity helix B' is not shown. The protein backbone is displayed in the cartoon representation, Heme and Cys residue are shown in ball and stick. All other residues and the substrate are shown in stick representation. Heme carbon atoms are colored black, oxygens are red and nitrogens blue.

Despite the conserved general fold, CYP enzymes significantly vary in structural flexibility. Higher structural flexibility in the distal part of the globular domain allows CYP enzymes to adopt different shapes and sizes of the catalytic pocket. Diverse substrates can thereby be bound by some CYP isoforms, which are also referred to as

CYPs with high substrate promiscuity [35]. This is especially true for the famous example of the major drug-metabolizing CYP3A4 [50].

1.4 The Relevance of Understudied Human Cytochrome P450 Enzymes

The general mechanism for the formation of the reactive CYP structure, Compound I, is similar for all CYP isoforms [17, 22]. Hence, it is the CYP structure that determines active site shape and flexibility. The substrate spectrum and function of the individual CYP enzyme are determined by its structure [22]. The structure and function of CYPs in xenobiotic metabolism and biosynthesis of steroids and other endogenous compounds have been intensively studied in the recent past. Neglected CYP enzymes have been hardly investigated and in the case of orphaned CYPs not at all studied [14, 15]. Orphaned CYP4 family enzymes CYP4A22 and CYP4Z1 are overexpressed in several malignancies [51-55] and likely play a role in important metabolic pathways. The only activity described so far for CYP4A22 is the ω -hydroxylation of arachidonic acid to 20-hydroxyeicosatetraenoic acid (20-HETE), which induces angiogenesis [55]. CYP4A22 likely has other functions as this activity is not unique to this CYP enzyme [55, 56]. CYP4A22 is overexpressed in colon and ovary cancer [55]. CYP4Z1 is expressed in mammary tissue and overexpressed in breast cancer cells and other malignancies [51-54]. Neither the physiological nor the pathophysiological role of CYP4Z1 in the above-mentioned malignancies has been understood, although there is a clear correlation between CYP4Z1 expression and tumor progression [57]. However, the discovery of autoantibodies against CYP4Z1 suggests a pathophysiological role of CYP4Z1 in breast cancer [58, 59]. Hence, CYP4Z1 might be a promising drug target in breast cancer [15, 51, 60], which has not yet been targeted by small molecules.

1.4.1 Structural Insights into Human Understudied Cytochrome P450 Enzymes

David Machalz, Szymon Pach, Marcel Bermudez, Matthias Bureik, Gerhard Wolber

Drug Discovery Today (IF: 7.79), 26, Issue 10, (2021), 2456-2464

Copyright Elsevier

<https://doi.org/10.1016/j.drudis.2021.06.006>

In this review, we show the disbalance in the research of human CYP enzymes and focuses on recent insights into understudied human CYP isoforms. This review stresses the importance of investigating previously neglected or understudied CYP isoforms. Studies on these CYP isoforms are not only of interest to understand their physiological function, but some of those CYPs are also potential drug targets. We show how homology models allow to structurally access 13 understudied CYPs, including insights on CYP4 enzymes from the collaboration between the Wolber and Bureik lab.

Contribution (first author):

Conceptual design (80%)

Data analysis (100%)

Visualization (90%)

Manuscript preparation and writing (40%)

Pages 15 to 23 have been removed due to copyright.

2 Research Aim

To date, research has mainly focused on CYP enzymes that are involved in drug metabolism or steroid biosynthesis. However, there are several highly interesting CYPs that have been neglected so far [14, 15]. The understudied CYP4 family enzymes CYP4A22 and CYP4Z1 likely play a role in important metabolic pathways. Especially CYP4Z1 might be a promising drug target in breast cancer, but no potent inhibitor has been reported before this thesis was started [15, 51, 60]. Experimental information on substrates is sparse for CYP4Z1 and CYP4A22 and their 3D structure is not known. Therefore, this doctoral thesis aimed to shed light on structure and function of CYP4 family enzymes focusing on CYP4A22 and CYP4Z1 using the following steps:

- Investigate active site of both CYP enzymes by probe substrates *in vitro*,
- Predict the 3D structure of both CYP enzymes guided by probe substrate structure-activity relationships (SAR),
- Investigate CYP enzyme-substrate dynamics to predict binding mode and key binding residues in both CYPs,
- Validate key residues roles by *in vitro* mutagenesis studies for CYP4Z1,
- Develop predictive CYP4Z1 binding model for a virtual screening approach for new small molecule inhibitors, and
- Validate activity and selectivity for screening hits *in vitro*.

The computational methods that have been applied to achieve the research aim are described in detail in the following chapter. All *in vitro* experiments have been realized in the lab of Prof. Bureik at Tianjin University, China.

3 Computational Methods

Computational methods are frequently used in early-stage drug design and discovery in the pharmaceutical industry [61, 62]. Hillisch and colleagues stated that computational chemistry methods were involved in the design of 50% of phase I drug candidates [62]. In recent years, this percentage has most likely increased dramatically due to ever more powerful hardware and the rise of machine learning in drug discovery [63, 64]. Computer-aided drug design (CADD) is commonly divided into structure-based and ligand-based methods [65, 66]. Structure-based approaches explicitly consider the macromolecular target structure, typically proteins, deoxyribonucleic or ribonucleic acid (DNA, RNA), or a complex of these. Ligand-based approaches solely focus on ligands binding to the target of interest. Ligands can be either small molecules, peptides, or oligonucleotides. In the following paragraphs, structure-based drug design methods applied in this thesis are explained further.

3.1 Homology Modeling

An atomistic 3D model of the macromolecule is essential for the structure-based design of small-molecule ligands. Experimental methods used to elucidate the structure of a macromolecular target are X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, or cryogenic electron microscopy (Cryo-EM). Despite the growing number of experimental atomistic models in the PDB (www.rcsb.org) [67], the structure of a large portion of pharmacologically relevant targets has not yet been elucidated. In lack of this information, homology modeling can be used to construct a 3D model for a protein target without a known 3D structure. Available methods can be divided into comparative homology modeling and *de novo* homology modeling (also referred to as 'protein folding prediction'). The basis for comparative homology modeling is the observation that proteins with highly similar primary sequences have highly similar 3D structures [68, 69]. Comparative homology modeling predicts the target protein structure considering one or more experimental structures of related proteins as templates [70]. No structural template is required for *de novo* or *ab initio* homology modeling and is often used for targets without solved structures of close relatives [71]. Several reviews cover the methodology of homology modeling in depth and point out the pitfalls in comparative protein structure modeling [70-72].

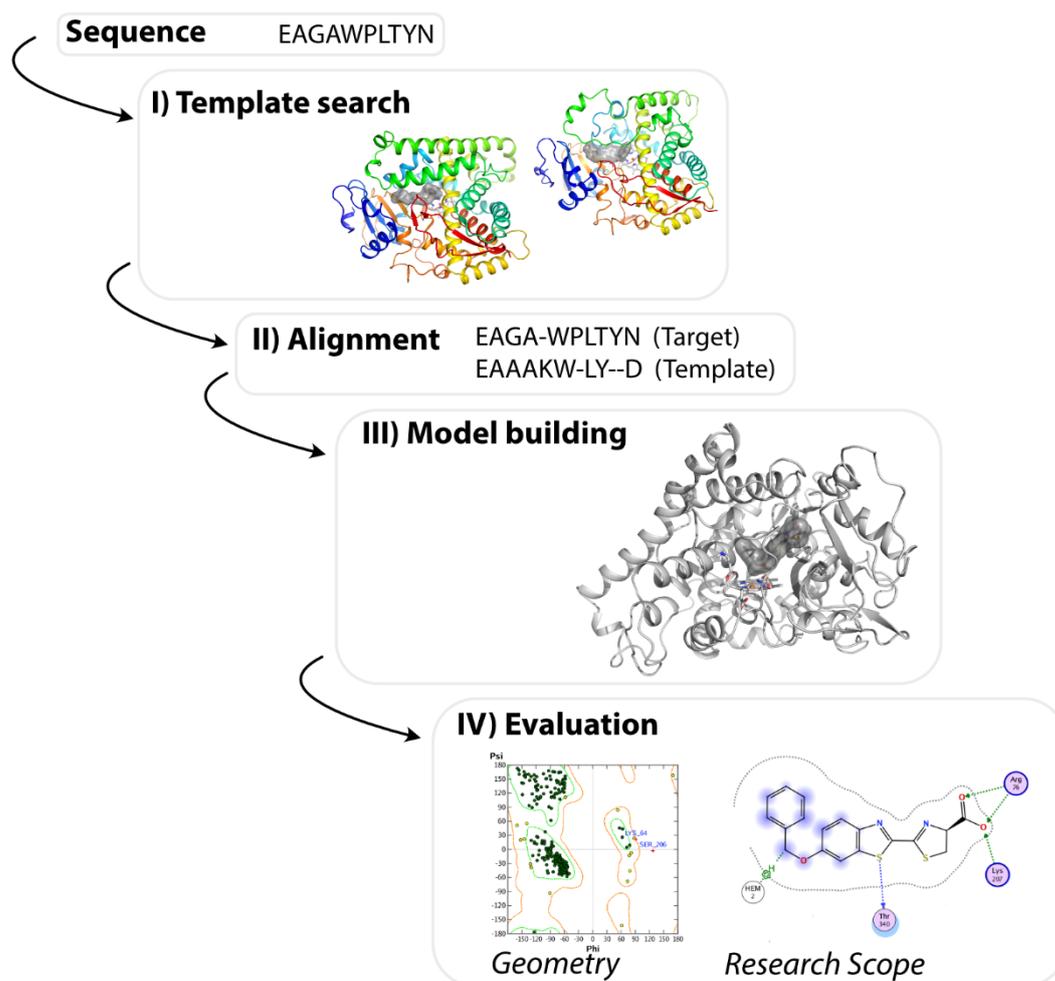


Figure 7. The four steps of comparative homology modeling.

In general, homology modeling for a protein of unknown structure, termed the “target”, consists of four steps (Figure 7). First, at least one similar protein with a known structure, termed the “template”, needs to be identified (step I). The primary sequences of both target and template need to be aligned accurately (step II). A 3D model of the target can then be built (step III) and evaluated for the fitness to the research question at hand and general model quality (step IV) [70].

In order to identify template structures that are similar to the target sequence, the basic local alignment search tool (BLAST) is commonly used. BLAST was designed to compare sequences in a rapid fashion, which makes it an optimal tool to quickly search large databases for similar sequences [73]. BLAST does a ‘quick and dirty’ sequence alignment that is not perfectly accurate, but good enough to estimate the sequence similarity. The target sequence can also be compared to several protein structures at once to increase sensitivity, which is also referred to as multiple sequence alignment

(MSA) and is frequently done with position-specific iterated BLAST (PSI-BLAST) [74]. Another approach to identifying suitable template structures is 3D template matching or threading. Template sequence and structure are used to recognize its fold. After the alignment of template and target, a scoring function predicts the probability that the target adopts the fold. For example, the local meta-threading-server (LOMETS) [75, 76] uses 11 different threading programs for fold recognition. The second-generation LOMETS2 tackles the problem that general MSA methods sometimes fail to provide an acceptable number of homolog structures, which results in poor alignment quality. LOMETS2 increases the number of non-redundant template sequences, also referred to as depth of the MSA. A deep MSA is generated in up to three search stages of increasing complexity using hidden Markov models [75]. LOMETS2 is an integral part of I-TASSER, a fully automated online homology modeling webserver [77, 78]. The LOMETS2 output comprises the best template prediction and sequence alignment for each of the programs based on the standard score (Z -score). The Z -score describes the deviation from the mean in a normalized distribution using the standard deviation (σ). For example, a Z -score of -2 means that the observed value is -2σ smaller than the mean. However, if possible, the selected template should be checked for an evolutionary relationship to the target and experimental quality. The close environment of target and template protein should be similar. Both proteins should be globular or transmembrane proteins, and have the same quaternary structure, such as consistently monomers or oligomers only. Subsequently, the template and target sequence must be aligned as accurately as possible, which is crucial for the homology modeling workflow. For pairwise alignment, the EMBOSS Needle-Stretcher algorithm [79] might be used, which makes use of the Needleman-Wunsch algorithm [80] best suited for this task. For multiple sequence alignment CLUSTAL Ω (or CLUSTAL O or CLUSTAL omega) [81] and SALIGN [82] are often used. Low sequence similarity is a severe problem in homology modeling. Cases with sequence identity below 25% are referred to as “the twilight area” [83].

After the identification and alignment of the template structure, the 3D model for the target protein can be built. High structural coverage is important for good model quality, which is the case when a large share of target protein residues have

corresponding residues in the template structure [72]. The workflow starts with the construction of the backbone of the target protein. The backbone model can be constructed by three different approaches: the assembly of rigid bodies, segment matching/coordinate reconstruction, or satisfaction of spatial restraints [70]. In the assembly of rigid bodies backbone atom positions of conserved core regions are taken from the most similar template. Backbone loop atoms are modeled using a database of all reported loop structures. The segment matching approach assumes that all hexapeptides can be classified into 100 structural classes. Templates are hence broken down into hexapeptide segments and reassembled based on conformations obtained from a stochastic search or an energy function-guided approach. Modeling by satisfaction of spatial restraints assumes that geometrical restraints are similar between template and target. General geometrical restraints from a force field (see 3.3 Molecular Dynamics Simulations) are also applied. The model is obtained after the minimization of restraint violations. Constraints or restraints derived from other methods such as NMR can also be included in this approach [70]. Less ordered domains of the target protein are built by loop modeling. Often, little information on the fold of the domain can be derived from the short sequence. Usually, the anchor regions, through which the loop is connected to the core regions, are used to guide the loop modeling. Loop modeling approaches can be divided into *ab initio*, database search, or a combination of these two methods. The *ab initio* methods search for conformations that are then evaluated by a scoring function, which is usually based on a force field. The database search scans databases for experimental loop structures with high sequence and anchor region similarity [70]. Once the backbone model is complete the side chains can be modeled. Backbone and side conformation are coupled and might make it necessary to address backbone flexibility while modeling sidechains. Between proteins with less than 30% sequence identity, the backbone fold might be conserved, but the sidechain interaction pattern is likely lost completely [83]. Side chain packing and solvent exposure drive correct modeling of side chain conformations [83, 84].

Finally, the constructed 3D model needs to be evaluated for general model quality and whether it is fit to model the research question at hand [70, 72]. Commonly checked general quality criteria are satisfactory geometry and the intra-protein hydrogen

bonding potential. Geometrical parameters usually include dihedral angles, atom distance clashes, bond angles, and lengths.

Several web services, such as VERIFY3D [85, 86], PROCHECK [87], and WHATCHECK [88], are available to check the quality parameters. The SAVES web server (saves.mbi.ucla.edu) comprises all the aforementioned web services. Furthermore, the final model should be examined by an expert with knowledge of the structure and function of the protein class. This is especially true for modeling in the twilight area where the built model likely needs further information from orthogonal experiments, such as mutagenesis studies or extended ligand structure-activity relationship (SAR) studies. The latter approach might also be called ligand-guided or ligand-steered homology modeling [72]. Generally, the model needs to be fit for purpose, which means that observed experimental trends must be explainable within the scope of the model. In the case of disagreement between experiment and model, the model is iteratively optimized. The approach and level of required optimization strongly depend on the complexity of the model and research question at hand [72]. The optimization strategy can range from a simple side chain rotation to elaborate techniques that consider the dynamic nature of macromolecules, such as molecular dynamics. In this doctoral thesis, homology models were used to explain and predict ligand binding and key binding residues.

There are various automated homology modeling web services, such as SWISS-MODEL [89], ROBETTA [90] RaptorX [91], and I-TASSER [75, 77, 78]. I-TASSER has been used excessively in the presented doctoral project, hence its methodology is discussed further below. The templates are fragmented and reassembled into initial models of the target. The models undergo stochastic conformational sampling. In detail, the models are subjected to multiple Monte Carlo simulations in parallel and the simulations can exchange atom positions, which is also referred to as replica exchange. The processed models are then clustered to find representative (centroid) model structures, which are then refined by two-step energy minimization to remove potential atom clashes. The top five models are chosen based on the confidence score and presented to the user. In the last step, COACH provides several predictions for

the top model, such as closest related template structure (by TM-align), function annotation, and binding site prediction.

3.2 Molecular Docking

How does a small molecular ligand bind to a macromolecular target, such as a protein, to exert its biological effect? This is one of the fundamental questions in drug discovery and is putatively addressed by molecular docking. Molecular docking suggests hypotheses for the binding of a ligand to a target (Figure 8) [92, 93]. The binding hypotheses of the ligand are also referred to as docking poses. The structure of the macromolecule can originate from homology modeling or experiments, such as X-ray crystallography, NMR, or Cryo-EM. Experimental atomistic 3D structures are deposited in the PDB [67]. The docking program faces two challenges to obtaining meaningful ligand binding hypotheses. First, the program needs to find plausible poses for the ligand at a predefined part of the target, such as the catalytic pocket of an enzyme. Subsequently, the found docking poses and the involved modeled macromolecule-ligand interactions must be evaluated and ranked by a scoring function. In general, the rotational and translational degrees of freedom of the rigid ligand and rotations around its rotatable bonds (conformations) define the ligand's flexibility space. The conformation of the macromolecular target is often kept rigid to reduce the complexity of the problem. Flexible docking approaches somewhat address target flexibility and allow for small changes in the target structure, such as side-chain rotation. Target flexibility can also be addressed by other, more qualified methods to sample the biologically relevant conformational target space, such as molecular dynamics (MD) or Monte Carlo simulations (see below) [92]. In ensemble docking, the ligand is docked to not only one, but multiple target conformations, which is widely used with highly flexible targets [94].

The ligand's flexibility is addressed by the search, or sampling, of ligand poses, which is handled differently by the various docking programs. The search techniques can be categorized into five sections: incremental construction approach (FlexX [95]), conformational search (MOE docker[96]), Monte Carlo search (Autodock [97]), systematic search techniques (Glide [98, 99]), and genetic algorithm (GOLD [100]). For the incremental construction approach, the ligand is broken down into fragments that

are docked individually. The program then attempts to reconstruct the ligand by ligating suitable fragment docking poses [95]. MOE docker samples ligand conformations first and then attempts to place ligand conformations via rigid-body docking into the designated target binding site[96]. In the early versions of AutoDock, the conformation of the ligand is changed randomly by a Monte Carlo search algorithm. The change in energy caused by the conformational change is evaluated by a scoring function.

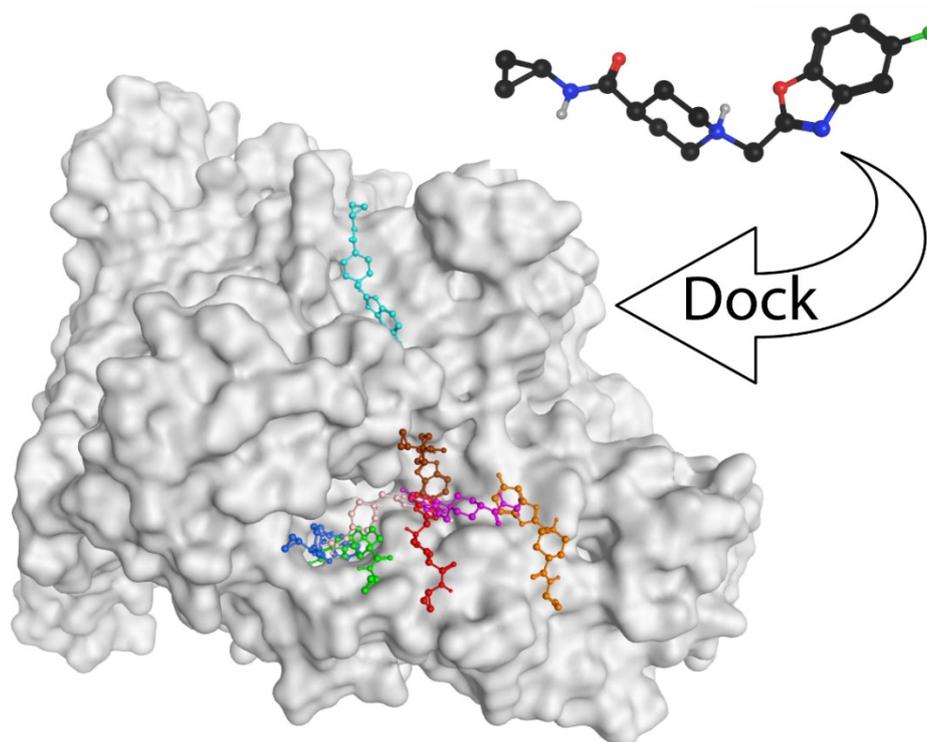


Figure 8. Molecular docking of a small molecule ligand to a macromolecular target. Docking poses of BO-I in the vicinity to the substrate funnel of choline trimethylamine-lyase (CutC, PDB ID: 5FAU [101]) obtained with GOLD are shown [102]. The protein's surface is shown in grey and ligand poses are colored individually and shown in ball and stick.

The Monte Carlo algorithm accepts conformational changes that are energetically more favorable. Changes that increase energy are not rejected right away, but the probability of rejection increases with the extent of the energy increase [97]. Glide searches for all possible combinations of ligand conformation and position. After a preselection of poses, further refinement is achieved by Monte Carlo sampling [98, 99]. The primarily used docking program in the presented doctoral studies is GOLD (Genetic algorithm for ligand docking) and will therefore be described more in depth further below. GOLD applies a genetic algorithm to identify plausible ligand poses

[100]. Genetic algorithms attempt to solve a given problem by simulating competition, or natural selection, between different suggested problem solutions as chromosomes, here docking poses. An initial set of chromosomes (docking poses) is generated. One chromosome holds the parameter details of one initial docking pose and its fitness is evaluated by the scoring function. The fittest chromosome (docking pose with the highest score) is selected and undergoes crossover (partial exchange in docking pose parameters) and/or mutation (random changes in individual parameters). The fitness of the new offspring chromosomes (new docking poses) is again evaluated by the scoring function. The least fit parent chromosomes (lowest scoring old docking poses) are replaced by the fittest offspring chromosomes. The genetic algorithm might now start again with chromosome manipulation until it either converges (parents = offspring) or reaches the predefined maximum number of runs. The final set of docking poses might now be re-scored by a different scoring function.

The scoring functions used to evaluate docking poses can be empirical (ChemScore [103]), derived from force fields (GOLD scoring function [104]), or knowledge-based (DrugScore [105]). The scoring function attempts to predict the free energy of binding ΔG for each docking pose and often fails with its prediction [106, 107]. Due to the limitations of scoring functions to identify pharmacologically relevant ligand conformations, the usage of more suitable metrics for the selection of a plausible binding hypothesis is encouraged [106]. Docking programs will always generate ligand poses in the defined pocket, also for inactive compounds – they produce false positives at high rates [106, 107]. For CYPs the distance of the heme iron to the small molecular moiety associated with inhibition or undergoing CYP-catalyzed biotransformation represents a simple but meaningful descriptor for binding mode plausibility. Additionally, 3D pharmacophore modeling (see section 3.4) can be applied to the docking poses and a reference ligand orientation to point to the presence or absence of protein-ligand interactions.

3.3 Molecular Dynamics Simulations

Every biomolecular system, such as a macromolecule-ligand complex, is in constant motion. Molecular dynamics (MD) simulations can model motions of biomolecular systems on time scales from nanoseconds to microseconds (Figure 9). MD simulations

can act as a “computational microscope” and potentially explain biological phenomena that are outside the scope of experimental methods [108]. The fundamentals and the ever-increasing role of MD simulations in drug design and protein structure modeling have been reviewed extensively [108-112]. AMBER [113], GROMACS [114], NAMD [115], OpenMM [116], and DESMOND [117] are frequently-used programs to carry out the MD simulations.

The biomolecular system of interest is usually modeled in water with physiological salt concentrations and if applicable within a bilayer membrane. This can easily result in a model with 100,000 or more atoms, which demands a good trade-off between an accurate enough model of the system and reasonable computational expense. Molecular mechanics (MM) are a good middle ground that apply classical Newtonian mechanics and represent atoms as spheres and bonds as springs [111, 118, 119]. In MM, each atom is described by its position, mass, and point charge. MD simulations start from the coordinates of an experimental (X-ray, Cryo-EM) or theoretical (homology model) 3D model. Atomic motion is calculated under the assumption that the forces acting on each atom are constant for the duration of a time step. Hence, the time step needs to be sufficiently small, approximately 1-4 fs, to account for the fast, vibrational movements of hydrogens. Initially, every atom is assigned a random motion – more precisely, a velocity taken from the Maxwell-Boltzmann distribution at the defined temperature. The potential energy of an atom dictates its next move and is obtained by calculating the atom’s bonded and non-bonded interactions (Figure 9). A set of formulas and corresponding parameters that define the shape of the functions are used here and are commonly referred to as a force field (FF). The most commonly used FFs for biomolecular systems are AMBER [120], CHARMM [121], GROMOS [122], and OPLS[123]. The bonded interactions are described by four formulas for bond-stretching, bond angle bending, torsion angle, and out-of-plane bending. The non-bonding forces between atoms are implemented by two formulas for van der Waals forces and electrostatic forces (Coulomb’s law) [111, 118, 119]. Some FFs have an additional cross term in the energy function that describes the dependency of bond length and angles [119]. Function parameters are tuned during FF parametrization to reproduce known properties from experiment or QM calculations. For example, bond

length can be obtained from an X-ray structure or other experiment, while non-equilibrium properties, such as a torsion angle profile, can be derived from QM calculations. Atom and bond types might be defined for molecular moieties and parameter assignment, which assumes that parameters are transferable [119].

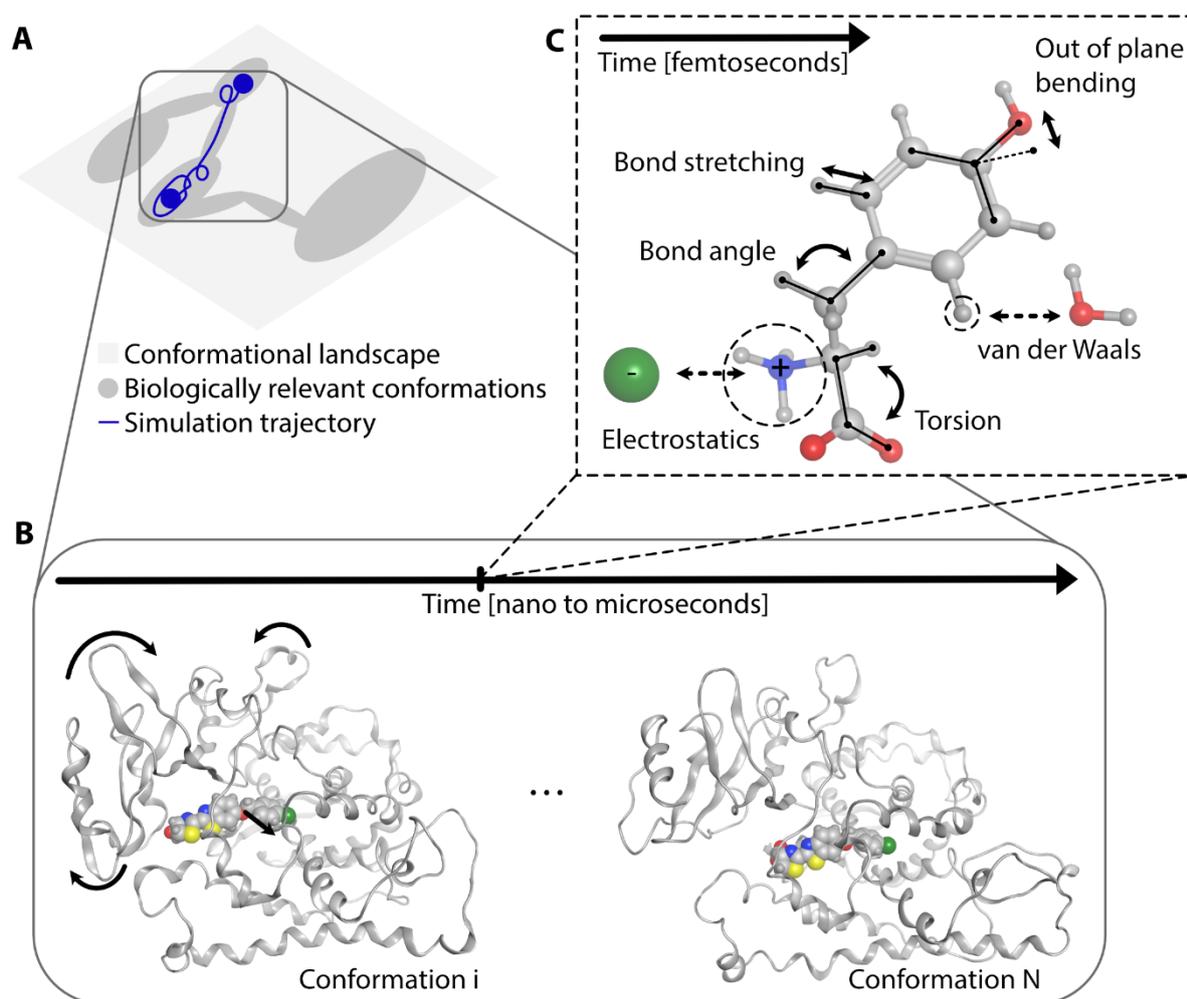


Figure 9. Molecular dynamics (MD) simulation of a biomolecular system. (A) MD simulations cover only a small part of the biological relevant conformational landscape of the biomolecular system. The conformational landscape is oversimplified as a 2D space, and the biologically relevant part of the conformational landscape is shown as darker patches. Conformations of the system during the MD simulation result in a trajectory and correspond to a path through the conformational landscape. (B) The biomolecular system undergoes conformational changes as indicated by curved arrows, which results in a simulation trajectory of conformations. (C) For each simulation step, the force field components are calculated that dictate the next conformational change. Bonded interactions include bond length, angle, torsion, and out of plane bending. Non-bonded interactions include electrostatic and van der Waals interactions.

Throughout the MD simulation, successive system conformations are obtained, which is also often referred to as a trajectory. The movement of parts of the system, such as the macromolecule, can be described by the root mean square deviation of heavy atom positions (RMSD) or the root mean square fluctuation (RMSF). With RMSD, the atom positions of two conformations of a molecule are compared. For MD simulation

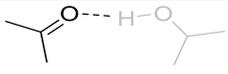
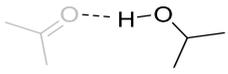
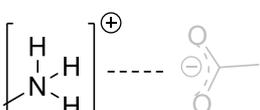
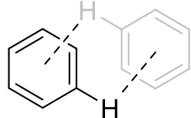
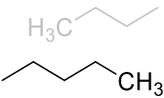
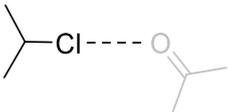
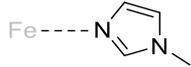
analysis each conformation is compared to one reference, which is usually the starting geometry. In RMSF calculations the average movement for each atom subgroup (e.g., protein residue) over time compared to a reference (usually starting geometry) is obtained. Monitoring of simple geometrical properties, such as distances, angles, dihedrals, or perimeters might be sufficient to link the simulation data to macroscopic experimental biological trends. The binding affinity of ligands to a macromolecule might be calculated by more advanced simulation protocols, such as free energy perturbation [124]. Specific interactions within the macromolecule or between ligand and macromolecule, such as hydrogen bonds, might be monitored over the MD trajectory by tools such as VMD [125] or MDAnalysis [126]. Interaction motifs might be analyzed by applying the concept of 3D pharmacophores to MD simulations as in the dynamic pharmacophores approach described in the following paragraph.

3.4 Static and Dynamic 3D Pharmacophores and Virtual Screening

A similar biological effect might be caused by a set of chemically diverse ligands binding to the same macromolecular target. Target-ligand interaction motifs might be derived from experimental or theoretical target-ligand complexes. The individual interaction motifs can be synthesized into a universal and simplistic 3D pharmacophore model. A 3D pharmacophore model comprises all macromolecule-ligand interactions in an easily visualized format ideal for communication with experimentalists. According to the IUPAC definition: “[a] pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supra-molecular interactions with a specific biological target structure and to trigger (or to block) its biological response.” [127]. The individual target-ligand interactions, such as hydrogen bonds, charge-charge interactions, halogen bonds, aromatic interactions, metal coordination, and hydrophobic contacts are encoded as pharmacophore features (Table 2) [128]. Each pharmacophore feature contains not only the chemical description but also its relative position and its weight in the model. For each feature, the allowed deviation from the position or interaction angle can be changed to get less strict or stricter geometrical tolerance. The weight of each feature in the 3D pharmacophore model can be adapted as well. In this fashion, the 3D pharmacophore

model can be fine-tuned to get the best description of the modeled system. The fine-tuned 3D pharmacophore model might then be used in virtual screening (see below).

Table 2. Overview of 3D pharmacophore features as defined in LigandScout [129-131]. Exemplary ligand moieties (black) and their corresponding interacting moiety (grey) are shown.

Feature	2D Depiction	Example
Hydrogen bond acceptor		
Hydrogen bond donor		
Negative ionizable		
Positive ionizable		
Aromatic interaction		
Hydrophobic contact		
Halogen bond		
Metal coordination		
Exclusion Volume		Any atoms of the target

3D pharmacophore models can either be derived via a ligand-based or a structure-based approach. For the ligand-based approach, a conformation of one or multiple aligned ligands is used to create a model of consensus 3D pharmacophore features. Forge [132] and Pharao [133] offer ligand-based 3D pharmacophore modeling. Structure-based 3D pharmacophores are created from a target-ligand complex, which has the advantage that interaction directionality and exclusion volumes are included. MOE (Molecular Operating Environment 2020.09; Chemical Computing Group ULC, Montreal, Canada), Catalyst [134], and LigandScout [129] are exemplary programs that

offer both structure- and ligand-based 3D pharmacophore modeling methods. Exclusion volumes are defined as a spherical space that cannot be occupied by the ligand since it would clash with target atoms at a similar position.

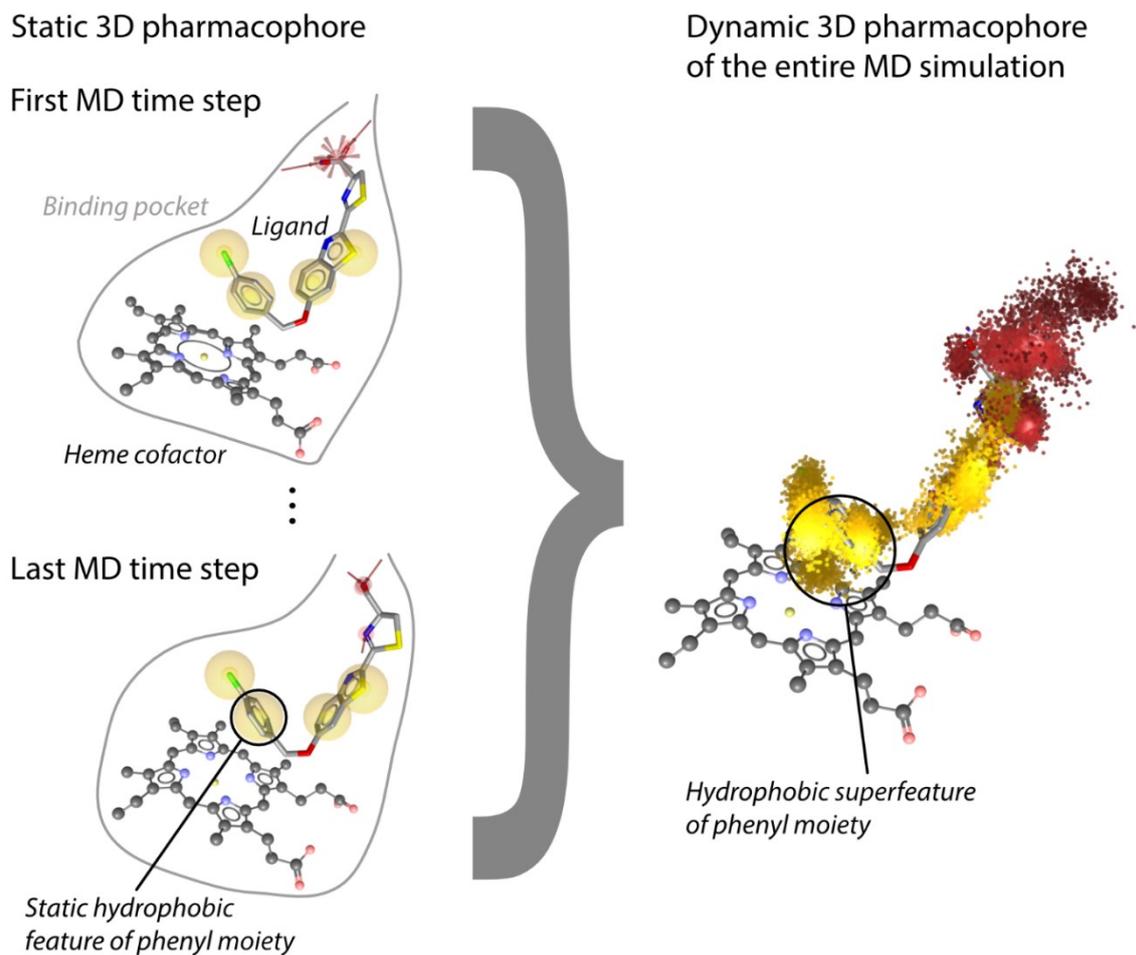


Figure 10. Dynamic 3D pharmacophore (dynophore) model from MD simulations. Besides the pharmacophore features only ligand and heme cofactor are depicted. The target is only depicted schematically as binding pocket shape.

3D pharmacophore-based virtual screening aims to identify active molecules that show the interactions encoded in the pharmacophore model [130, 135]. The 3D pharmacophore features abstract the chemical nature of ligand moieties, which allows scaffold hopping and matching of different bioisosteric moieties. A negative ionizable pharmacophore feature will not only match a carboxylate moiety, but also the bioisosteric tetrazole moiety. Virtual libraries of millions of compounds can be screened in an extremely rapid fashion using the simplistic 3D pharmacophore model(s) in software such as LigandScout [129-131], MOE, or Phase[136]. The 3D pharmacophore model is usually validated using reported active and inactive molecules against the target. If there are no known inactive molecules, molecules

(decoys) that are randomly selected and are likely to be inactive against the target can be used instead [137, 138]. Decoys are obtained by minimal structural changes to known active molecules while maintaining the physicochemical properties. After validation, precalculated conformations of the search library compounds are superposed on the 3D pharmacophore model. Compounds that satisfy the 3D pharmacophore model constraints are included in a virtual hit list. Virtual screening performance can be evaluated by several parameters including sensitivity, specificity, enrichment factor, and the receiver operating characteristic (ROC) curve [130]. The ROC curve visualizes the true positive rate, also referred to as sensitivity, and the false positive rate of the 3D pharmacophore model. The ROC curve shows the proportion of compounds selected by the 3D pharmacophore model with confirmed activity and those that are falsely predicted to be active. Recently, our lab has reviewed the latest advances in 3D pharmacophore modeling including virtual screening and other applications [139].

Target and ligand are in constant motion as described in the previous section (3.3 Molecular Dynamics Simulations). Hence it is only logical to include this motion into 3D pharmacophore modeling. Sydow and colleagues have therefore applied 3D pharmacophore modeling to MD simulations, which results in a dynamic pharmacophore model or 'dynophore' [140, 141]. For every time step of the MD simulation, each target-ligand interaction is captured as an individual interaction point. Related interaction points are grouped into superfeatures, such as all hydrophobic contacts the phenyl ring of a ligand forms with the target (Figure 10). The final dynamic 3D pharmacophore model is composed of multiple superfeatures. Superfeatures allow for spatial, temporal, and general statistical analysis. The spatial distribution of the interaction points gives insight into the stability of the modeled interactions and might identify substrates of binding. The temporal analysis with barcode plots, the occurrence of interactions over time, shows which interactions occur together and which are mutually exclusive. Statistical analysis, such as the frequency of occurrence of superfeatures can identify key interactions and discard minor interactions.

4 Results

4.1 Efficient Substrate Screening and Inhibitor Testing of Human CYP4Z1 Using Permeabilized Recombinant Fission Yeast

Qi Yan, David Machalz, Andy Zöllner, Erik J. Sorensen, Gerhard Wolber, Matthias Bureik

This article was published in *Biochemical Pharmacology* (IF: 5.29) 146 (2017) 174-187, Copyright Elsevier

<https://doi.org/10.1016/j.bcp.2017.09.011>

Cytochrome P450 4Z1 is overexpressed in breast cancer and might hence be a promising therapeutical target. In the following study, the lab of Prof. Bureik at Tianjin University established an expression method for CYP enzymes and probed the promising breast cancer target CYP4Z1 with luminogenic substrates and inhibitors. In order to rationalize the substrate structure-activity relationships (SAR) trends information on the enzyme structure is required. However, no experimentally resolved atomistic model of CYP4Z1 has been reported so far. Hence, our lab constructed a homology model of CYP4Z1 using the substrate SAR data. We docked the substrates into the active site of CYP4Z1 and predicted four binding residues with the help of 3D pharmacophore modeling.

Contribution (second author):

Conceptual design (20%)

Computational experiments (100%)

In vitro experiments (0%)

Visualization (60%)

Manuscript preparation and writing (20%)

Pages 42 to 55 have been removed due to copyright.

4.2 Functional Characterization and Mechanistic Modeling of the Human Cytochrome P450 Enzyme CYP4A22

Pradeepraj Durairaj, Linbing Fan, David Machalz, Gerhard Wolber, Matthias Bureik

This article was published in FEBS Letters (IF: 4.12) 593, Issue 16, (2019), 2214-2225, Copyright Federation of European Biochemical Society

<https://doi.org/10.1002/1873-3468.13489>

The Bureik lab at Tianjin University expressed the orphaned CYP4A22 in a functional form and reported activity on fatty acids and probe substrates. As no atomistic model of CYP4A22 is known, our lab followed the same protocol as previously reported for CYP4Z1 (see Section 4.1). We constructed a homology model for CYP4A22 incorporating the substrate SAR data. With the CYP4A22 model we could rationalize the effect of a mutation on the probable interface to the redox partner protein. Substrate docking and 3D pharmacophore modeling allowed us to rationalize activity trends. We used molecular dynamics simulations in combination with dynamic 3D pharmacophore (dynophore) modeling to study substrate recognition in CYP4A22 and predict binding residues.

Contribution (shared first author):

Conceptual design (30%)

Computational experiments (100%)

In vitro experiments (0%)

Visualization (50%)

Manuscript preparation and writing (40%)

Pages 57 to 73 have been removed due to copyright.

4.3 Importance of Asparagine-381 and Arginine-487 for Substrate Recognition in CYP4Z1

Wei Du, David Machalz, Qi Yan, Erik J. Sorensen, Gerhard Wolber, Matthias Bureik

This article was published in *Biochemical Pharmacology* (IF: 5.29) 174 (2020) 113850,
Copyright Elsevier

<https://doi.org/10.1016/j.bcp.2020.113850>

In a previous study of the promising anti-cancer target CYP4Z1 we predicted four residues for substrate recognition (see section 4.1). In this study, the Bureik lab at Tianjin University tested our prediction by *in vitro* mutagenesis studies. Surprisingly, the mutation led to enhanced substrate binding. Hence, our lab employed molecular dynamics simulations and dynamic pharmacophore modeling to study the interaction dynamics in CYP4Z1 substrate recognition. The analysis of the interaction dynamics suggested two new binding residues and provided a mechanistic explanation for enhanced activity CYP4Z1 mutants. New *in vitro* mutants expressed by the Bureik lab confirmed the key role of the new residues in substrate binding and recognition.

Contribution (shared first author):

Conceptual design (40%)

Computational experiments (100%)

In vitro experiments (0%)

Visualization (50%)

Manuscript preparation and writing (50%)

Pages 75 to 85 have been removed due to copyright.

4.4 Discovery of a novel potent cytochrome P450 CYP4Z1 inhibitor

David Machalz, Hongjie Li, Wei Du, Shishir Sharma, Sijie Liu, Matthias Bureik, Gerhard Wolber

This article was published in European Journal of Medicinal Chemistry (IF: 6.18) 215 (2021) 113255, Copyright Elsevier

<https://doi.org/10.1016/j.ejmech.2021.113255>

Previously, we elucidated the CYP4Z1 substrate binding mode and validated our CYP4Z1 model (see section 4.3). In this study, our lab moved forward in targeting the promising anti-cancer target CYP4Z1. As a starting point, we suggested a binding mode for a low micromolecular inhibitor. We encompassed the enzyme-inhibitor interactions into a 3D pharmacophore model which we enhanced with insights from previous CYP4Z1 mutagenesis studies. We used this 3D pharmacophore model to screen for new inhibitors of CYP4Z1. Structure activity relationships of the active screening hits were studied by MD simulations, dynamic pharmacophores, and quantum mechanics. Screening hits were evaluated *in vitro* for purity, activity, and selectivity by the Bureik lab at Tianjin University. We identified a low nanomolecular inhibitor that binds selectively to CYP4Z1 over the major drug metabolizing CYP enzymes.

Contribution (shared first author):

Conceptual design (90%)

Computational experiments (90%)

In vitro experiments (0%)

Visualization (60%)

Manuscript preparation and writing (50%)

Pages 87 to 104 have been removed due to copyright.

5 Discussion

The reaction mechanism is highly similar for all CYPs, and their common fold is largely preserved. However, structural differences between CYP enzymes might lead to distinct functions and substrate spectra. Hence, atomistic models of CYP4A22 and CYP4Z1 were necessary to follow the aim of this doctoral project: the elucidation of the structure and function of the two orphaned CYP enzymes. The first two studies focused on the development of ligand-guided homology models for CYP4Z1 (see section 4.1) and CYP4A22 (see section 4.2) by using probe substrate activities. CYP4Z1 was further investigated by elaborated *in silico* binding residue predictions and extensive mutagenesis investigations in a third study (see section 4.3). In the fourth study, CYP4Z1 was targeted in a virtual screening campaign and a new inhibitor was discovered with an IC_{50} value of 63 ± 19 nM (see section 4.4).

5.1 Characterization of CYP4Z1 *in vitro* and *in silico*

In the first study, we kickstarted the deorphanization of the promising anti-cancer target CYP4Z1 by further investigating its poorly understood activity (see section 4.1). The clear correlation between CYP4Z1 expression and tumor progression [57] makes the enzyme a promising anti-cancer target. Before this study, the only reported activity was the in-chain hydroxylation of lauric and myristic acid [142]. To escape the sparse ligand data situation the Bureik lab used the newly developed “enzyme bags” assay to probe CYP4Z1 with 15 different pro-luciferin substrates and six inhibitors.

In the absence of an experimentally determined structural model, a ligand-based model of the target’s active site can be developed based on the structure of reported ligands. However, the number of reported substrates of CYP4Z1 was too small for any ligand-based approach. Hence, the development of a homology model for CYP4Z1 was the most feasible way forward to investigate the structure-function relationship. Homology modeling suffers from increasing uncertainty with decreasing sequence identity between target and the used template protein structure (see section 3.1). At under 25% sequence identity, correct alignment of the target and template sequences becomes exceedingly difficult. This is also referred to as the “the twilight area” of homology modeling and can result in incorrect models [83]. CYP domains of high

sequence similarity can adapt different conformations for different ligands, which adds another challenge to the model development. This is especially true for CYP3A4 where bulky ligands ketoconazole and erythromycin require large-scale changes to the enzyme's holo-structure resulting in an increase of active site volume by more than 80% [50]. Hence, Guengerich et al. stated that "[...] a structure of a P450 in the absence of a ligand is probably not particularly useful in predicting what a P450 substrate complex will look like"[143]. Furthermore, it is critical to understand the chemistry of CYPs and their substrates in order to include the steric requirements of the heme moiety and arrive at plausible models of CYP enzymes, which can yield mechanistic insights. However, a first homology model of CYP4Z1 was reported in a review on the enzyme in early 2017 [60], which was based on the human X-ray structure CYP3A4 (PDB ID: 1TQN). Human CYP4Z1 and CYP3A4 are in the twilight area with a sequence identity of 25.4%. Fortunately, the first X-ray structure of a CYP4 family enzyme was published later in 2017, the X-ray structure of rabbit CYP4Z1 (PDB ID: 5T6Q) [34]. Human CYP4Z1 and rabbit CYP4B1 have a sequence identity of 44.0% and are more closely related than CYP4Z1 and CYP3A4. Hence, we could derive a first homology model of human CYP4Z1 based on the more suitable template of rabbit CYP4B1. The sequence identity below 50% nevertheless demanded additional orthogonal data for a reliable model. Here, ligand-based homology modeling can be applied, which includes indirect information on the binding site from ligand activity data [72]. Structure-activity relationships (SAR) from the pro-luciferin substrates were used to enhance the homology model. The smaller active site of CYP4B1 compared to CYP4Z1 posed an additional challenge. The co-crystallized ligand n-octane in CYP4B1 is somewhat smaller than the substrates tested in CYP4Z1. Hence, it is safe to assume that CYP4B1 is a suitable template in terms of backbone geometry and binding site differences are mainly caused by side chains. In CYP4B1 the binding site is restricted by a covalently bound heme, which is not the case in CYP4Z1. In the CYP4Z1 model, side chains had to be manually adapted in a ligand-guided fashion to allow for a slightly larger binding site than in CYP4B1.

Taken together, we augmented the homology modeling for CYP4Z1 with probe substrate data to shed first light on the active site of CYP4Z1. We predicted Ser113,

Ser222, Asn381, and Ser383 to be important for substrate binding in CYP4Z1. The CYP4Z1 model could now be validated *in vitro* in the next step and subsequently, be used for a structure-based drug design campaign targeting the enzyme. The incorporated probe substrate data was key to overcome the sparse data situation that hampers the homology modeling of orphaned CYPs and resulted in the first reported structural model yielding mechanistic insights into CYP4Z1 functionality.

5.2 Characterization of CYP4A22 *in vitro* and *in silico*

In this second study, we employed an *in vitro* and *in silico* protocol similar to the first study to investigate the activity of CYP4A22 (see section 4.2). CYP4A22 activity was poorly characterized before and only ω -hydroxylation of AA to 20-HETE was known. The Bureik lab expressed two variants of CYP4A22-WT and CYP4A22.1 that differ by the Tyr152Asn mutation. CYP4A22.1 was previously reported as the reference standard sequence [144]. However, bioinformatic analyses done for this study revealed that CYP4A22-WT is the most frequent variant. Due to an even sparser ligand data situation than for CYP4Z1, CYP4A22 was probed by 11 probe substrates, three fatty acids, and four inhibitors. CYP4A22 showed activity for 8 probe substrates and ω -hydroxylated lauric and myristic acid, which is the first report of enzymatic activity of CYP4A22.

Our lab built the first homology model of CYP4A22 informed by substrate SAR to understand the structural prerequisites that allow for the rather different reaction mechanisms. To the best of our knowledge, this was the first-ever reported homology model of CYP4A22. The CYP4A22 model most closely resembled the rabbit CYP4B1 structure [34]. With a sequence identity of 51.3%, CYP4A22 and CYP4B1 are more closely related than CYP4Z1 and CYP4B1 with 44.0%. Furthermore, the covalently bound heme moiety in CYP4A22 and CYP4B1 makes their active sites more similar and tunes them towards ω -hydroxylation activity. We surmised from the CYP4A22 model that the Tyr152Asn mutation enhances the interaction with the redox partner cytochrome P reductase and hence CYP4A22.1 activity.

Once again, a direct implementation of ligand data helped to fine-tune the atomistic model of the active site of CYP4A22. In particular, the position of the side chain of

Phe320 is worth noting here. The initial model featured the inwards rotamer of Phe320 similar to the position of Phe309 in CYP4B1. This CYP4A22 binding site shape was, however, not fully in line with the determined activity profile. Addressing this issue, we found that the outwards rotamer allows for the biotransformation of bulkier substrates as found *in vitro*. We surmised that the Phe320 rotation is only possible in CYP4A22, as CYP4B1 is more densely packed in the second layer behind the Phe309 (see section 4.2, Fig. 3A). The suggested increased residual mobility of Phe320 in CYP4A22 is likely the cause for the decreased ω -hydroxylation efficiency. The importance of Phe rotamers in CYP active sites has been reported before, for example for the drug-metabolizing CYP2D6. An *in silico* study found that CYP2D6 can adapt to its substrates with two different rotamers of a Phe residue in the active site [145]. This investigation started from a validated 3D model of a CYP2D6 X-ray structure, which is in clear contrast to our study of an unvalidated CYP4A22 model. Although we couldn't investigate the transition between the two Phe rotamers for CYP4A22 by MD simulations as done in the CYP2D6 study within the framework of this thesis due to the lack of a validated CYP4A22 model, this would certainly be an interesting question for future investigations.

We used MD simulations to describe the dynamics of interactions between CYP4A22.1 and the most active probe substrate Luciferin-BE. Here, the hypothesized salt bridge between the substrate's carboxylate moiety and Arg96 and Arg233 was of particular interest. The suggested role of these two Arg residues was further supported by MD simulations analyzed by dynamic pharmacophores. Dynamic pharmacophores describe how macromolecule-ligand interactions evolve over time during an MD simulation creating a 3D pharmacophore model for each simulation time step. Dynamic pharmacophores allow us not only to view the occurrence of individual intermolecular interactions, but also which interactions occur together and those that are mutually exclusive. This way we can dissect complex interaction profiles to suggest which interactions are important and how subtle changes influence binding affinity. The simulation results of CYP4A22 and Luciferin-BE suggested stable interaction with Arg96 and less frequent interaction with Arg233. Hence, sole interaction with Arg96 might suffice for substrate recognition.

MD simulations allow us to follow the movements of a biomolecular system over time, such as a CYP enzyme-substrate complex. In this way, MD simulations can sample an ensemble of conformational states of the enzyme-substrate complex. MD simulations work under the assumption of the ergodic hypothesis, which implies that a system can reach every possible state no matter the initial state the system was in and that the time average equals the ensemble average [119]. Given infinite simulation time and assuming the ergodic hypothesis, an MD simulation will visit all conformational states of the complex. In reality, MD simulation time is finite and limited by the available hardware. The complexity of the simulated system and time scale of the investigated event, therefore, determine how long the MD simulation needs to be to sample the relevant conformational space. Movements of protein domains can be well in the microsecond time scale. Since we were particularly interested in small changes to the enzyme-substrate complex related to biological activity, simulations in the hundreds of nanoseconds are sufficient and feasible. Nevertheless, MD simulations can get trapped in low-energy conformational states, which leads to local oversampling and hampers sufficient conformational sampling. To identify or even circumvent artifacts, such as local oversampling, it is common practice to run multiple MD simulation replicas of the system. Conclusions drawn from a single long MD simulation as opposed to multiple shorter simulation replicas might be false positive [146]. Hence, we ran the MD simulation of CYP4A22 and Luciferin-BE in three replicas.

MD simulations apply Molecular Mechanics (MM), a reductionistic representation of the system. In MM-based force fields, charge-charge interactions decay more quickly with increasing distance than *in vivo*. Beyond 9 Å charge-charge interactions are described implicitly in the default settings of the DESMOND simulation engine. Hence, the cancellation of the salt bridge between CYP4A22 and Luciferin-BE might be an artifact; once the guanidine moiety of the Arg233 residue moved further away than 9 Å from the substrate carboxylate the interaction was canceled out, which is unlikely in reality. A way to investigate this potential artifact would be to increase the calculated interaction range, which might however slow the MD simulation down severely and was outside of the scope of this study. Ideally, the effects of these

hypothesized salt bridges would be investigated by *in vitro* mutagenesis studies first before investigating CYP4A22 more in depth *in silico*.

Taken together, we reported the first functional expression and homology model of CYP4A22, which was investigated in atomistic detail in this study. We predicted two key Arg residues, described the binding site including covalent heme binding, and rationalized the ω -hydroxylation activity of CYP4A22. With the help of this CYP4A22 model, we rationalized the activity shift between the two most common CYP4A22 variants. These findings will benefit further studies on the function of the still orphaned CYP4A22.

5.3 Dissecting and Validating the CYP4Z1 Binding Site Model

In the third study, we continued to characterize CYP4Z1 in terms of the binding site and substrate recognition (see section 4.3) based on the findings of the first study (see section 4.1). Previously, we predicted Ser113, Ser222, Asn381, and Ser383 to be major drivers of ligand binding in CYP4Z1. Mutagenesis experiments done by the Bureik lab in this study could not confirm the importance of these residues. In fact, the Ser113Ala and Ser383Ala mutants increased CYP4Z1 activity. We predicted that the CYP4Z1 mutants Ser113Ala and Ser383Ala will hamper substrate recognition and thereby reduce biotransformation activity. However, the two mutants led to increased CYP4Z1 activity *in vitro*, which clearly shows that a mutation can have unforeseen effects in complex interaction networks and can benefit from rationalization through MD simulations. This is especially true for poorly understood enzymes such as CYP4Z1.

Previously, MD simulations could provide the necessary explanation for complex mutagenesis data of CYP enzymes [147-149]. Our MD simulation analysis protocol was inspired by a study on the regioselectivity of macrocycle oxidation by CYPBM3 by MD simulations by Petrovic et al [147]. Petrovic and colleagues monitored atom distances in MD simulations and used them to describe the conformational space of the substrate in the active site of CYPBM3. This methodology heavily inspired our analysis of MD simulations of CYP4Z1 and Luciferin-4F12. But how do MD simulations relate to *in vitro* trends? A common approach is to focus on the second step in the catalytic cycle where the substrate has already entered the active site of the CYP

enzyme, but the reaction has not yet taken place. Here, the distance between the heme iron that will bind the oxygen for reaction and the substrate atom undergoing oxidation, the site of oxidation (SoO), is measured. Catalytically competent binding is assumed at an Fe-SoO distance smaller than 6 Å [145]. In the complex of CYP4B1 and n-octane (PDB ID: 5T6Q), the Fe-SoO distance is 3.4 Å. In CYP21A2, the endogenic substrate progesterone is 4.0 Å away from the heme (PDB ID: 4Y8W). However, as mentioned above, X-ray structures can also provide a false picture of catalytically non-competent binding. In the X-ray structure of the well-established CYP model system CYP102A1 of *Bacillus megaterium* (CYPBM3) the substrate N-palmitoylglycine is bound in a catalytically non-competent pose (PDB ID: 1JPZ). We measured an Fe-SoO distance of 7.5 Å in this structure. Here MD simulations with advanced sampling were able to produce a catalytically competent pose of N-palmitoylglycine with an Fe-SoO distance of 3.6 Å [150]. These examples establish the Fe-SoO-distance as a good surrogate parameter for catalytic competence of a CYP enzyme-substrate conformation from X-ray structures or MD simulations.

Hence, we went back to the CYP4Z1 model and employed MD simulations to study substrate binding dynamics as previously done for CYP4A22 (see section 4.2). We found the previously unreported binding residue Arg487 that forms a salt bridge to the carboxylate group from MD simulations of CYP4Z1 and probe substrate Luciferin-4F12. The importance of this salt bridge was validated *in vitro*. Substrate recognition via long-range charge-charge interactions is a common mechanism in enzymes [151]. CYP4 enzymes are mostly fatty acid hydroxylases, and it seems plausible that positively charged amino acids recognize the carboxylate moiety of fatty acids. This notion is further supported by the previous finding that probe substrates without a free carboxylate moiety showed drastically lowered activity in the first study of CYP4Z1 (see section 4.1). However, it is worth noting that, due to the convenience of the luminescence assay, CYP4Z1 was investigated with Luciferin-4F12, which is not its physiological substrate. Apart from the carboxylic acid moiety that Luciferin-4F12 and fatty acids share, fatty acids have far more rotatable bonds and therefore can adapt more conformations. For CYP4A22 we surmised that fatty acids could adapt different catalytically competent conformations (see section 4.2).

In the second stage of MD simulations, we could successfully explain the activity shifts for three mutants: Ser113Ala, Ser383Ala, and Arg487Met. Catalytically competent substrate binding was less frequent for Arg487Met due to the loss of the salt bridge to the substrate carboxylate. In the gain of function mutants Ser113Ala and Ser383Ala, a subpocket in the binding site closed, which led to more frequent catalytically competent substrate binding. It is not clear whether the gain of function mutations might have also positively affected active site solvation or access channel dynamics. Furthermore, we dissected the interaction dynamics for CYP4Z1.1 into substrate's and residue's contributions. We found that besides the key residue Arg487, Asn381 plays a supporting role in substrate binding, which was confirmed by a double mutant *in vitro* (see section 4.3).

Taken together, we were able to validate our CYP4Z1 model *in vitro* and study substrate binding thoroughly *in silico*. We reported two key residues, predicted *in silico* and validated *in vitro*. These detailed insights into CYP4Z1 substrate binding are crucial to investigate the still poorly understood physiological role of CYP4Z1 further. Furthermore, the validated CYP4Z1 model is perfectly apt to be used in a structure-based drug design campaign to find new small molecules targeting CYP4Z1.

5.4 Targeting CYP4Z1 by Rational Computational Drug Design

In this fourth study, we used our validated CYP4Z1 model to discover a new potent inhibitor of CYP4Z1, which was validated in two different assays *in vitro* (see section 4.4). We used the previously determined substrate SAR (see sections 4.1 and 4.3) to augment a 3D pharmacophore model of the most potent reported non-covalent CYP4Z1 inhibitor 1-benzylimidazole. This augmented 3D pharmacophore model was used in a virtual screening campaign and yielded 8 hits after a careful post-screening pipeline.

A 3D pharmacophore model that is used for virtual screening should be validated with active and inactive molecules. As CYP4Z1 is an orphaned enzyme very few active inhibitors are known. The reported inhibitors of CYP4Z1 are structurally diverse, which hampers the development of a consensus 3D pharmacophore model. Hence, we decided to augment the pharmacophore model with validated substrate SAR data

instead. Another potential pitfall in 3D pharmacophore modeling is the consideration of the binding site shape, which can be described by exclusion volumes that don't allow inhibitor placement. It can be very cumbersome to completely describe the binding site with exclusion spheres. Hence, we used docking to account for steric limitations of the binding site and to identify implausible inhibitor poses.

Five of the virtual screening hits were tested by the Bureik lab in primary screening. Surprisingly, the marketed drug ozagrel was amongst the validated hits, and the *in vivo* implication of this activity needs to be studied further. For this study, the Bureik lab developed a new CYP4Z1 inhibition assay by transfecting MCF-7 breast cancer cells to overexpress CYP4Z1. Under these more *in vivo* like conditions, the most potent hit (inhibitor 9) showed an IC₅₀ value of 63 ± 19 nM, which corresponds to 100-fold increased potency compared to the starting point of 1-benzylimidazole. For the approval of new drugs, authorities, such as the Food and Drug Administration (FDA) or the European Medical Agency (EMA), routinely request data on the inhibition of the six major drug-metabolizing CYPs (1A2, 2C9, 2C19, 2D6, 2E1, and 3A4). The nanomolar activity of inhibitor 9 is especially astonishing as hits from virtual screening do not often show high affinity in contrast to lead compounds that already underwent multiple optimization cycles. Furthermore, Inhibitor 9 showed no activity except in CYP2D6, where it showed only minor inhibition. This selectivity profile makes inhibitor 9 a hit worthy of being optimized towards a lead structure.

To understand the SAR of inhibitor 9, we employed MD simulations including dynamophore analysis. However, the MM force fields showed a severe limitation here; the interaction between inhibitor and iron was sufficiently represented only after the introduction of a virtual covalent bond. Hence it was necessary to explicitly model valence electrons with the quantum mechanical (QM) method density functional theory to correctly model the key interaction. QM methods have been frequently used to describe CYP enzyme functionality and inhibition [23, 152-156]. SMARTCyp used previously performed QM calculations to predict CYP-related metabolism solely based on the substrate structure [152]. The ω-hydroxylation of vitamin K₁ by CYP4F2 has been described by QM methods [153]. More importantly, the inhibition of the heme by nitrogen-containing heterocycles has been thoroughly studied by QM methods

[154, 155] and has even been used to discover inhibitors of prostate cancer target CYP17A1 [156]. Since QM methods are computationally more expensive, it is only feasible to perform calculations on small systems. Hence, CYPs are often reduced to an iron atom, the heme moiety or a porphyrin ring, the cysteine ligand or a methylmercaptide and some truncated model of the ligand. Possible electronic and steric effects on imidazole moiety of the inhibitor by the rest of the molecule or the CYP enzyme are neglected this way. Due to the explicit modeling of electrons, molecular interactions become very sensitive to directionality. Hence, we used the most frequent conformation coordinates from our MD simulation to account for steric effects instead of an optimized structure as done in other QM studies on CYPs mentioned above.

Shortly before the publication of our CYP4Z1 study, a new inhibitor, 8-[(1H-benzotriazol-1-yl)amino]octanoic acid, was reported, which covalently binds to the heme of CYP41 [157]. Kowalski and colleagues started from the fragment 1-aminobenzotriazole, which can covalently bind to the heme, and extended it with fatty acids of different lengths. The two most potent inhibitors showed selective binding to CYP4Z1 over seven other CYP4 enzymes. The Kowalski inhibitor did not inhibit seven major drug-metabolizing CYPs *in vitro*, similar to our inhibitor 9. However, their study did not propose a binding mode, which could hamper future rational structure-based optimization. The IC_{50} value of the Kowalski inhibitor was reported as 5900 ± 900 nM in contrast to our inhibitor with an IC_{50} value of 63 ± 19 nM. However, it must be noted that IC_{50} values originating from different assays can hardly be compared directly, especially when the mode of action is different. As mentioned above, the Kowalski inhibitor is a mechanism-based inhibitor similar to the marketed aromatase inhibitor exemestane. Our inhibitor 9 is a non-covalent inhibitor, which is more similar to the antifungal drug ketoconazole inhibiting CYP51A1 or the prostate cancer drug abiraterone inhibiting CYP17A1. Both compounds are drug-like in terms of predicted standard physicochemical parameters, such as total polar surface area, molecular weight, and clog.

Hence, the Kowalski inhibitor and our inhibitor 9 could be merged into a new inhibitor in a future study. The new inhibitor would carry our more drug-like including the

carboxylate crucial for high affinity. The imidazole would be replaced the 1-aminobenzotriazole of the Kowalski inhibitor for covalent and selective CYP4Z1 binding. Furthermore, the scaffold's amide moiety could be reversed to validated, if it forms hydrogen bonds contributing to the high binding affinity.

Taken together, we identified three new inhibitors by our carefully developed virtual screening workflow and validated inhibitor potency *in vitro*. The most potent hit, inhibitor 9, was active in yeast-based and breast cancer cell-based CYP4Z1 inhibition assays and inactive on six drug-metabolizing CYPs. This new potent inhibitor of CYP4Z1 is a valuable pharmacological tool and could be used for lead structure discovery for a novel therapeutic approach targeting CYP4Z1.

6 Conclusion

Human CYP enzymes are involved in many essential biotransformation pathways ranging from xenobiotic metabolism to the biosynthesis of steroids, vitamins, and fatty acids. The majority of CYPs involved in these processes have been studied extensively in the past, while other CYPs have been neglected or are completely orphaned. CYP4A22 and CYP4Z1 are two of the orphaned CYPs whose physiological function and structure are not known. The ω -hydroxylation activity of CYP4A22 produces 20-HETE, which triggers angiogenesis. Furthermore, CYP4Z1 is a promising breast cancer target that was not previously targeted by small molecules.

This doctoral study aimed to contribute to the deorphanization of CYP enzymes CYP4A22 and CYP4Z1. We have established a protocol to escape the sparse ligand data situation for orphaned enzymes where increasingly popular ligand data-driven methods, such as machine learning cannot be applied. Our experimental partner Prof. Bureik has established rapid testing with the described “enzyme bags” in combination with an efficient luminescence assay, which is also suitable for inhibitor screening. The experimental work of the Bureik lab enabled our ligand-driven homology modeling of CYP4Z1 and CYP4A22. We reported the first CYP4A22 model and a validated model of CYP4Z1. This highlights the importance of supporting homology modeling with activity data and modeling that emulates the mechanistics of the investigated biological system. However, besides the active site, the overall CYP4Z1 structure remains elusive. One open question is the role of access channels in substrate selectivity, which remains sparsely understood in all CYPs but has been recently tackled for CYP2D6 [158]. Nevertheless, with the help of these 3D models, we could show how structure determines CYP functionality. Small changes, such as the flip of a phenylalanine residue in CYP4A22 or the absence of a covalently bound heme in CYP4Z1, influence functionality. Recently, the deorphanization protocol has been used for other understudied CYPs [159-161] and has even been extended to understudied phase II metabolism enzymes, such as Sulfotransferases [162] and Glucuronosyltransferases [163]. In the advent of deep learning-enabled structure prediction, such as AlphaFold2 [164], large-scale structure-based enzyme deorphanization and selectivity modeling might come within reach.

Throughout this doctoral project, we have made use of different scales of molecular modeling appropriate for the research question at hand. We used 3D pharmacophore modeling as abstract molecular representation to rapidly screen millions of molecules to find new inhibitors of CYP4Z1. We used molecular mechanics, which model atoms and bonds explicitly, in molecular dynamics simulations and homology model refinement. We could show that taking the plastic structure of CYPs into account with MD simulations could lead to valuable insights. This involved the discovery of a new and important binding residue Arg487 and the dissection of the substrate recognition network in CYP4Z1. Here, dynamic pharmacophores have proven to be an invaluable tool to explain trends in MD simulations of protein-ligand complexes and detect even very subtle trends. We also demonstrated that complex descriptors are not always needed and a simple distance between two atoms (Fe-SoO) can link a simulation to reality. We explicitly modeled electrons for the description of the non-trivial coordination of the heme iron by our newly discovered CYP4Z1 inhibitor. However, absolute proof for the binding mode in CYP4Z1 can only be achieved through *in vitro* experiments. Especially for X-ray diffraction inhibitor 9 is a useful tool as only a potent inhibitor can establish the stable complex that is required for this technique.

The novel CYP4Z1 inhibitor we discovered is a new valuable pharmacological tool to further study the physiological and pathophysiological role of CYP4Z1 and is an ideal starting point for new therapeutics targeting CYP4Z1. Nevertheless, it is important to reiterate that CYP4Z1 is not a validated therapeutic target. *In vivo* experiments, such as CYP4Z1 knockout mice are still required. Inhibitor 9 or derivatives thereof could be applied in CYP4Z1 overexpressing mice to study the effect on survival.

Altogether, we have demonstrated how molecular modeling methods can drive the understanding of enzyme functionality and the discovery of new potent small molecules - never alone, always together with experimentalists.

7 Bibliography

- [1] D.R. Nelson, J.V. Goldstone, J.J. Stegeman, The cytochrome P450 genesis locus: the origin and evolution of animal cytochrome P450s, *Philos Trans R Soc Lond B Biol Sci* 368(1612) (2013) 20120474.
- [2] D.R. Nelson, L. Koymans, T. Kamataki, J.J. Stegeman, R. Feyereisen, D.J. Waxman, M.R. Waterman, O. Gotoh, M.J. Coon, R.W. Estabrook, I.C. Gunsalus, D.W. Nebert, P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature, *Pharmacogenetics* 6(1) (1996) 1-42.
- [3] I.G. Denisov, S.G. Sligar, Activation of molecular oxygen in cytochromes P450, in: P.R. Ortiz de Montellano (Ed.), *Cytochrome P450*, Springer International Publishing, Cham, 2015, pp. 69-109.
- [4] T. Omura, R. Sato, The carbon monoxide-binding pigment of liver microsomes. I. evidence for its hemoprotein nature, *J Biol Chem* 239(7) (1964) 2370-8.
- [5] T. Omura, R. Sato, A new cytochrome in liver microsomes, *J Biol Chem* 237 (1962) 1375-6.
- [6] F.P. Guengerich, Human cytochrome P450 enzymes, in: P.R. Ortiz de Montellano (Ed.), *Cytochrome P450*, Springer International Publishing, Cham, 2015, pp. 523-785.
- [7] D.W. Nebert, K. Wikvall, W.L. Miller, Human cytochromes P450 in health and disease, *Philos Trans R Soc Lond B Biol Sci* 368(1612) (2013) 20120431.
- [8] D.R. Nelson, Cytochrome P450 diversity in the tree of life, *Biochim Biophys Acta Proteins Proteom* 1866(1) (2018) 141-154.
- [9] F.P. Guengerich, Intersection of the roles of cytochrome P450 enzymes with xenobiotic and endogenous substrates: relevance to toxicity and drug interactions, *Chem Res Toxicol* 30(1) (2017) 2-12.
- [10] M.L. Edin, J. Cheng, A. Gruzdev, S.L. Hoopes, D.C. Zeldin, P450 enzymes in lipid oxidation, in: P.R. Ortiz de Montellano (Ed.), *Cytochrome P450*, Springer International Publishing, Cham, 2015, pp. 881-905.
- [11] R.J. Auchus, W.L. Miller, P450 enzymes in steroid processing, in: P.R. Ortiz de Montellano (Ed.), *Cytochrome P450*, Springer International Publishing, Cham, 2015, pp. 851-879.
- [12] S. Rendic, F.P. Guengerich, Survey of human oxidoreductases and cytochrome P450 enzymes involved in the metabolism of xenobiotic and natural chemicals, *Chem Res Toxicol* 28(1) (2015) 38-42.
- [13] S.P. Rendic, F. Peter Guengerich, Human cytochrome P450 enzymes 5-51 as targets of drugs and natural and environmental compounds: mechanisms, induction, and inhibition - toxic effects and benefits, *Drug Metab Rev* 50(3) (2018) 256-342.

- [14] F.P. Guengerich, Q. Cheng, Orphans in the human cytochrome P450 superfamily: approaches to discovering functions and relevance in pharmacology, *Pharmacol Rev* 63(3) (2011) 684-99.
- [15] B. Licznerska, W. Baer-Dubowska, What might the presence of 'orphan' CYP450 isoforms in breast epithelial cells mean for the future of targeted therapeutics?, *Expert Opin Drug Metab Toxicol* 17(2) (2021) 135-137.
- [16] P. Durairaj, L. Fan, W. Du, S. Ahmad, D. Mebrahtu, S. Sharma, R.A. Ashraf, J. Liu, Q. Liu, M. Bureik, Functional expression and activity screening of all human cytochrome P450 enzymes in fission yeast, *FEBS Lett* 593(12) (2019) 1372-1380.
- [17] R. Lonsdale, J. Olah, A.J. Mulholland, J.N. Harvey, Does compound I vary significantly between isoforms of cytochrome P450?, *J Am Chem Soc* 133(39) (2011) 15464-74.
- [18] S.P. de Visser, F. Ogliaro, P.K. Sharma, S. Shaik, Hydrogen bonding modulates the selectivity of enzymatic oxidation by P450: chameleon oxidant behavior by compound I, *Angew Chem Int Ed Engl* 41(11) (2002) 1947-51.
- [19] F. Ogliaro, S. Cohen, S.P. de Visser, S. Shaik, Medium polarization and hydrogen bonding effects on compound I of cytochrome P450: what kind of a radical is it really?, *J Am Chem Soc* 122(51) (2000) 12892-12893.
- [20] J.T. Groves, G.A. McClusky, Aliphatic hydroxylation by highly purified liver microsomal cytochrome P-450. evidence for a carbon radical intermediate, *Biochem Biophys Res Commun* 81(1) (1978) 154-60.
- [21] J.T. Groves, G.A. McClusky, Aliphatic hydroxylation via oxygen rebound. oxygen transfer catalyzed by iron, *J Am Chem Soc* 98(3) (1976) 859-861.
- [22] P.R. Ortiz de Montellano, Substrate oxidation by cytochrome P450 enzymes, in: P.R. Ortiz de Montellano (Ed.), *Cytochrome P450*, Springer International Publishing, Cham, 2015, pp. 111-176.
- [23] J. Kirchmair, A.H. Goller, D. Lang, J. Kunze, B. Testa, I.D. Wilson, R.C. Glen, G. Schneider, Predicting drug metabolism: experiment and/or computation?, *Nat Rev Drug Discov* 14(6) (2015) 387-404.
- [24] F.P. Guengerich, Mechanisms of cytochrome P450-catalyzed oxidations, *ACS Catal* 8(12) (2018) 10964-10976.
- [25] S.G. Sligar, Coupling of spin, substrate, and redox equilibriums in cytochrome P450, *J Biochemistry* 15(24) (1976) 5399-5406.
- [26] S.N. Daff, S.K. Chapman, K.L. Turner, R.A. Holt, S. Govindaraj, T.L. Poulos, A.W. Munro, Redox control of the catalytic cycle of flavocytochrome P-450 BM3, *Biochemistry* 36(45) (1997) 13816-23.

- [27] I.G. Denisov, T.M. Makris, S.G. Sligar, I. Schlichting, Structure and chemistry of cytochrome P450, *Chem Rev* 105(6) (2005) 2253-77.
- [28] F.P. Guengerich, Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity, *Chem Res Toxicol* 14(6) (2001) 611-50.
- [29] B. Testa, A. Pedretti, G. Vistoli, Reactions and enzymes in the metabolism of drugs and other xenobiotics, *Drug Discov Today* 17(11-12) (2012) 549-60.
- [30] P. Rydberg, U. Ryde, L. Olsen, Sulfoxide, sulfur, and nitrogen oxidation and dealkylation by cytochrome P450, *J Chem Theory Comput* 4(8) (2008) 1369-77.
- [31] S.T. Seger, P. Rydberg, L. Olsen, Mechanism of the N-hydroxylation of primary and secondary amines by cytochrome P450, *Chem Res Toxicol* 28(4) (2015) 597-603.
- [32] P. Rydberg, L. Olsen, Do two different reaction mechanisms contribute to the hydroxylation of primary amines by cytochrome P450?, *J Chem Theory Comput* 7(10) (2011) 3399-404.
- [33] J. Olah, A.J. Mulholland, J.N. Harvey, Understanding the determinants of selectivity in drug metabolism through modeling of dextromethorphan oxidation by cytochrome P450, *Proc Natl Acad Sci U S A* 108(15) (2011) 6050-5.
- [34] M.H. Hsu, B.R. Baer, A.E. Rettie, E.F. Johnson, The crystal structure of cytochrome P450 4B1 (CYP4B1) monooxygenase complexed with octane discloses several structural adaptations for omega-hydroxylation, *J Biol Chem* 292(13) (2017) 5610-5621.
- [35] M. Otyepka, K. Berka, P. Anzenbacher, Is there a relationship between the substrate preferences and structural flexibility of cytochromes P450?, *Curr Drug Metab* 13(2) (2012) 130-42.
- [36] T.L. Poulos, E.F. Johnson, Structures of cytochrome P450 enzymes, in: P.R. Ortiz de Montellano (Ed.), *Cytochrome P450*, Springer International Publishing, Cham, 2015, pp. 3-32.
- [37] M. Otyepka, J. Skopalik, E. Anzenbacherova, P. Anzenbacher, What common structural features and variations of mammalian P450s are known to date?, *Biochim Biophys Acta* 1770(3) (2007) 376-89.
- [38] M. Srejber, V. Navratilova, M. Paloncyova, V. Bazgier, K. Berka, P. Anzenbacher, M. Otyepka, Membrane-attached mammalian cytochromes P450: an overview of the membrane's effects on structure, drug binding, and interactions with redox partners, *J Inorg Biochem* 183 (2018) 117-136.
- [39] I.F. Sevrioukova, H. Li, H. Zhang, J.A. Peterson, T.L. Poulos, Structure of a cytochrome P450-redox partner electron-transfer complex, *Proc Natl Acad Sci U S A* 96(5) (1999) 1863-8.

- [40] B.C. Monk, T.M. Tomasiak, M.V. Keniya, F.U. Huschmann, J.D. Tyndall, J.D. O'Connell, 3rd, R.D. Cannon, J.G. McDonald, A. Rodriguez, J.S. Finer-Moore, R.M. Stroud, Architecture of a single membrane spanning cytochrome P450 suggests constraints that orient the catalytic domain relative to a bilayer, *Proc Natl Acad Sci U S A* 111(10) (2014) 3865-70.
- [41] J. Lo, G. Di Nardo, J. Griswold, C. Egbuta, W. Jiang, G. Gilardi, D. Ghosh, Structural basis for the functional roles of critical residues in human cytochrome p450 aromatase, *Biochemistry* 52(34) (2013) 5821-9.
- [42] P. Urban, T. Lautier, D. Pompon, G. Truan, Ligand access channels in cytochrome P450 enzymes: a review, *Int J Mol Sci* 19(6) (2018).
- [43] P.A. Williams, J. Cosme, V. Sridhar, E.F. Johnson, D.E. McRee, Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity, *Mol Cell* 5(1) (2000) 121-31.
- [44] C. De Lemos-Chiarandini, A.B. Frey, D.D. Sabatini, G. Kreibich, Determination of the membrane topology of the phenobarbital-inducible rat liver cytochrome P-450 isoenzyme PB-4 using site-specific antibodies, *J Cell Biol* 104(2) (1987) 209-19.
- [45] N. Ueyama, N. Nishikawa, Y. Yamada, T. Okamura, A. Nakamura, Cytochrome P-450 model (porphinato)(thiolato)iron(III) complexes with single and double NH \cdots S hydrogen bonds at the thiolate site, *J Am Chem Soc* 118(50) (1996) 12826-12827.
- [46] S. Tripathi, H. Li, T.L. Poulos, Structural basis for effector control and redox partner recognition in cytochrome P450, *Science* 340(6137) (2013) 1227-30.
- [47] S.C. Gay, M.B. Shah, J.C. Talakad, K. Maekawa, A.G. Roberts, P.R. Wilderman, L. Sun, J.Y. Yang, S.C. Huelga, W.X. Hong, Q. Zhang, C.D. Stout, J.R. Halpert, Crystal structure of a cytochrome P450 2B6 genetic variant in complex with the inhibitor 4-(4-chlorophenyl)imidazole at 2.0-Å resolution, *Mol Pharmacol* 77(4) (2010) 529-38.
- [48] M.B. Shah, P.R. Wilderman, J. Pascual, Q. Zhang, C.D. Stout, J.R. Halpert, Conformational adaptation of human cytochrome P450 2B6 and rabbit cytochrome P450 2B4 revealed upon binding multiple amlodipine molecules, *Biochemistry* 51(37) (2012) 7225-38.
- [49] K. Berka, M. Paloncyova, P. Anzenbacher, M. Otyepka, Behavior of human cytochromes P450 on lipid membranes, *J Phys Chem B* 117(39) (2013) 11556-64.
- [50] M. Ekroos, T. Sjogren, Structural basis for ligand promiscuity in cytochrome P450 3A4, *Proc Natl Acad Sci U S A* 103(37) (2006) 13682-7.
- [51] M.A. Rieger, R. Ebner, D.R. Bell, A. Kiessling, J. Rohayem, M. Schmitz, A. Temme, E.P. Rieber, B. Weigle, Identification of a novel mammary-restricted cytochrome P450, CYP4Z1, with overexpression in breast carcinoma, *Cancer Res.* 64(7) (2004) 2357-64.

- [52] D. Downie, M.C. McFadyen, P.H. Rooney, M.E. Cruickshank, D.E. Parkin, I.D. Miller, C. Telfer, W.T. Melvin, G.I. Murray, Profiling cytochrome P450 expression in ovarian cancer: identification of prognostic markers, *Clinical cancer research : an official journal of the American Association for Cancer Research* 11(20) (2005) 7369-75.
- [53] D.T. Ross, U. Scherf, M.B. Eisen, C.M. Perou, C. Rees, P. Spellman, V. Iyer, S.S. Jeffrey, M. Van de Rijn, M. Waltham, A. Pergamenschikov, J.C. Lee, D. Lashkari, D. Shalon, T.G. Myers, J.N. Weinstein, D. Botstein, P.O. Brown, Systematic variation in gene expression patterns in human cancer cell lines, *Nat Genet* 24(3) (2000) 227-35.
- [54] X.J. Ma, Z. Wang, P.D. Ryan, S.J. Isakoff, A. Barmettler, A. Fuller, B. Muir, G. Mohapatra, R. Salunga, J.T. Tuggle, Y. Tran, D. Tran, A. Tassin, P. Amon, W. Wang, W. Wang, E. Enright, K. Stecker, E. Estepa-Sabal, B. Smith, J. Younger, U. Balis, J. Michaelson, A. Bhan, K. Habin, T.M. Baer, J. Brugge, D.A. Haber, M.G. Erlander, D.C. Sgroi, A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen, *Cancer Cell* 5(6) (2004) 607-16.
- [55] A. Alexanian, B. Miller, R.J. Roman, A. Sorokin, 20-HETE-producing enzymes are up-regulated in human cancers, *Cancer Genomics Proteomics* 9(4) (2012) 163-9.
- [56] A. Alexanian, A. Sorokin, Targeting 20-HETE producing enzymes in cancer - rationale, pharmacology, and clinical potential, *Onco Targets Ther* 6 (2013) 243-55.
- [57] G.I. Murray, S. Patimalla, K.N. Stewart, I.D. Miller, S.D. Heys, Profiling the expression of cytochrome P450 in breast cancer, *Histopathology* 57(2) (2010) 202-11.
- [58] V. Nunna, N. Jalal, M. Bureik, Anti-CYP4Z1 autoantibodies detected in breast cancer patients, *Cell Mol Immunol* 14(6) (2017) 572-574.
- [59] C. Khayeka-Wandabwa, X. Ma, X. Cao, V. Nunna, J.L. Pathak, R. Bernhardt, P. Cai, M. Bureik, Plasma membrane localization of CYP4Z1 and CYP19A1 and the detection of anti-CYP19A1 autoantibodies in humans, *Int Immunopharmacol* 73 (2019) 64-71.
- [60] X. Yang, M. Hutter, W.W.B. Goh, M. Bureik, CYP4Z1 - A human cytochrome P450 enzyme that might hold the key to curing breast cancer, *Curr Pharm Des* 23(14) (2017) 2060-2064.
- [61] I. Muegge, A. Bergner, J.M. Kriegel, Computer-aided drug design at Boehringer Ingelheim, *J Comput Aided Mol Des* 31(3) (2017) 275-285.
- [62] A. Hillisch, N. Heinrich, H. Wild, Computational chemistry in the pharmaceutical industry: from childhood to adolescence, *ChemMedChem* 10(12) (2015) 1958-62.
- [63] A. Bender, I. Cortes-Ciriano, Artificial intelligence in drug discovery: what is realistic, what are illusions? part 1: ways to make an impact, and why we are not there yet, *Drug Discov Today* 26(2) (2021) 511-524.

- [64] A.H. Goller, L. Kuhnke, F. Montanari, A. Bonin, S. Schneckener, A. Ter Laak, J. Wichard, M. Lobell, A. Hillisch, Bayer's in silico ADMET platform: a journey of machine learning over the past two decades, *Drug Discov Today* 25(9) (2020) 1702-1709.
- [65] P. Sledz, A. Caflisch, Protein structure-based drug design: from docking to molecular dynamics, *Curr Opin Struct Biol* 48 (2018) 93-102.
- [66] G. Sliwoski, S. Kothiwale, J. Meiler, E.W. Lowe, Jr., Computational methods in drug discovery, *Pharmacol Rev* 66(1) (2014) 334-95.
- [67] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, *Nucleic Acids Res* 28(1) (2000) 235-42.
- [68] S.H. Bryant, C.E. Lawrence, The frequency of ion-pair substructures in proteins is quantitatively related to electrostatic potential: a statistical model for nonbonded interactions, *Proteins* 9(2) (1991) 108-19.
- [69] C. Chothia, A.M. Lesk, The relation between the divergence of sequence and structure in proteins, *EMBO J* 5(4) (1986) 823-6.
- [70] M.A. Marti-Renom, A.C. Stuart, A. Fiser, R. Sanchez, F. Melo, A. Sali, Comparative protein structure modeling of genes and genomes, *Annu Rev Biophys Biomol Struct* 29(1) (2000) 291-325.
- [71] B. Kuhlman, P. Bradley, Advances in protein structure prediction and design, *Nat Rev Mol Cell Biol* 20(11) (2019) 681-697.
- [72] T. Schmidt, A. Bergner, T. Schwede, Modelling three-dimensional protein structures for applications in drug design, *Drug Discov Today* 19(7) (2014) 890-7.
- [73] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, *J Mol Biol* 215(3) (1990) 403-10.
- [74] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res* 25(17) (1997) 3389-402.
- [75] W. Zheng, C. Zhang, Q. Wuyun, R. Pearce, Y. Li, Y. Zhang, LOMETS2: improved meta-threading server for fold-recognition and structure-based function annotation for distant-homology proteins, *Nucleic Acids Res* 47(W1) (2019) W429-W436.
- [76] S. Wu, Y. Zhang, LOMETS: a local meta-threading-server for protein structure prediction, *Nucleic Acids Res* 35(10) (2007) 3375-82.
- [77] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, The I-TASSER Suite: protein structure and function prediction, *Nat Methods* 12(1) (2015) 7-8.

- [78] A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction, *Nat Protoc* 5(4) (2010) 725-38.
- [79] F. Madeira, Y.M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A.R.N. Tivey, S.C. Potter, R.D. Finn, R. Lopez, The EMBL-EBI search and sequence analysis tools APIs in 2019, *Nucleic Acids Res* 47(W1) (2019) W636-W641.
- [80] S.B. Needleman, C.D. Wunsch, A general method applicable to the search for similarities in the amino acid sequence of two proteins, *J Mol Biol* 48(3) (1970) 443-53.
- [81] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol Syst Biol* 7 (2011) 539.
- [82] H. Braberg, B.M. Webb, E. Tjioe, U. Pieper, A. Sali, M.S. Madhusudhan, SALIGN: a web server for alignment of multiple protein sequences and structures, *Bioinformatics* 28(15) (2012) 2072-3.
- [83] S.Y. Chung, S. Subbiah, A structural explanation for the twilight zone of protein sequence homology, *Structure* 4(10) (1996) 1123-7.
- [84] R.J. Petrella, T. Lazaridis, M. Karplus, Protein sidechain conformer prediction: a test of the energy function, *Fold Des* 3(5) (1998) 353-77.
- [85] J.U. Bowie, R. Luthy, D. Eisenberg, A method to identify protein sequences that fold into a known three-dimensional structure, *Science* 253(5016) (1991) 164-70.
- [86] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers* 22(12) (1983) 2577-637.
- [87] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, *J Appl Crystallogr* 26(2) (1993) 283-291.
- [88] G. Vriend, WHAT IF: a molecular modeling and drug design program, *J Mol Graph* 8(1) (1990) 52-56.
- [89] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F.T. Heer, T.A.P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: homology modelling of protein structures and complexes, *Nucleic Acids Res* 46(W1) (2018) W296-W303.
- [90] Y. Song, F. DiMaio, R.Y. Wang, D. Kim, C. Miles, T. Brunette, J. Thompson, D. Baker, High-resolution comparative modeling with RosettaCM, *Structure* 21(10) (2013) 1735-42.

- [91] M. Kallberg, H. Wang, S. Wang, J. Peng, Z. Wang, H. Lu, J. Xu, Template-based protein structure modeling using the RaptorX web server, *Nat Protoc* 7(8) (2012) 1511-22.
- [92] N.S. Pagadala, K. Syed, J. Tuszynski, Software for molecular docking: a review, *Biophys Rev* 9(2) (2017) 91-102.
- [93] D.B. Kitchen, H. Decornez, J.R. Furr, J. Bajorath, Docking and scoring in virtual screening for drug discovery: methods and applications, *Nat Rev Drug Discov* 3(11) (2004) 935-49.
- [94] R.M. Knegtel, I.D. Kuntz, C.M. Oshiro, Molecular docking to ensembles of protein structures, *J Mol Biol* 266(2) (1997) 424-40.
- [95] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, A fast flexible docking method using an incremental construction algorithm, *J Mol Biol* 261(3) (1996) 470-89.
- [96] C.R. Corbeil, C.I. Williams, P. Labute, Variability in docking success rates due to dataset preparation, *J Comput Aided Mol Des* 26(6) (2012) 775-86.
- [97] D.S. Goodsell, H. Lauble, C.D. Stout, A.J. Olson, Automated docking in crystallography: analysis of the substrates of aconitase, *Proteins* 17(1) (1993) 1-10.
- [98] T.A. Halgren, R.B. Murphy, R.A. Friesner, H.S. Beard, L.L. Frye, W.T. Pollard, J.L. Banks, Glide: a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening, *J Med Chem* 47(7) (2004) 1750-9.
- [99] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy, *J Med Chem* 47(7) (2004) 1739-49.
- [100] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, *J Mol Biol* 267(3) (1997) 727-48.
- [101] S. Bodea, M.A. Funk, E.P. Balskus, C.L. Drennan, Molecular basis of C-N bond cleavage by the glycy radical enzyme choline trimethylamine-lyase, *Cell Chem Biol* 23(10) (2016) 1206-1216.
- [102] M.T. Gabr, D. Machalz, S. Pach, G. Wolber, A benzoxazole derivative as an inhibitor of anaerobic choline metabolism by human gut microbiota, *Rsc Med Chem* 11(12) (2020) 1402-1412.
- [103] M.D. Eldridge, C.W. Murray, T.R. Auton, G.V. Paolini, R.P. Mee, Empirical scoring functions: I. the development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes, *J Comput Aided Mol Des* 11(5) (1997) 425-45.
- [104] M.L. Verdonk, J.C. Cole, M.J. Hartshorn, C.W. Murray, R.D. Taylor, Improved protein-ligand docking using GOLD, *Proteins* 52(4) (2003) 609-23.

- [105] H. Gohlke, M. Hendlich, G. Klebe, Knowledge-based scoring function to predict protein-ligand interactions, *J Mol Biol* 295(2) (2000) 337-56.
- [106] G.L. Warren, C.W. Andrews, A.M. Capelli, B. Clarke, J. LaLonde, M.H. Lambert, M. Lindvall, N. Nevins, S.F. Semus, S. Senger, G. Tedesco, I.D. Wall, J.M. Woolven, C.E. Peishoff, M.S. Head, A critical assessment of docking programs and scoring functions, *J Med Chem* 49(20) (2006) 5912-31.
- [107] A. Fischer, M. Smiesko, M. Sellner, M.A. Lill, Decision making in structure-based drug discovery: visual inspection of docking results, *J Med Chem* 64(5) (2021) 2489-2500.
- [108] R.O. Dror, R.M. Dirks, J.P. Grossman, H. Xu, D.E. Shaw, Biomolecular simulation: a computational microscope for molecular biology, *Annu Rev Biophys* 41 (2012) 429-52.
- [109] J. Mortier, C. Rakers, M. Bermudez, M.S. Murgueitio, S. Riniker, G. Wolber, The impact of molecular dynamics on drug design: applications for the characterization of ligand-macromolecule complexes, *Drug Discov Today* 20(6) (2015) 686-702.
- [110] M. Bermudez, J. Mortier, C. Rakers, D. Sydow, G. Wolber, More than a look into a crystal ball: protein structure elucidation guided by molecular dynamics simulations, *Drug Discov Today* 21(11) (2016) 1799-1805.
- [111] M. De Vivo, M. Masetti, G. Bottegoni, A. Cavalli, Role of molecular dynamics and related methods in drug discovery, *J Med Chem* 59(9) (2016) 4035-61.
- [112] W.F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D.P. Geerke, A. Glattli, P.H. Hunenberger, M.A. Kastholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N.F. van der Vegt, H.B. Yu, Biomolecular modeling: goals, problems, perspectives, *Angew Chem Int Ed Engl* 45(25) (2006) 4064-92.
- [113] D.A. Case, T.E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K.M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs, *J Comput Chem* 26(16) (2005) 1668-88.
- [114] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen, GROMACS: fast, flexible, and free, *J Comput Chem* 26(16) (2005) 1701-18.
- [115] J.C. Phillips, D.J. Hardy, J.D.C. Maia, J.E. Stone, J.V. Ribeiro, R.C. Bernardi, R. Buch, G. Fiorin, J. Henin, W. Jiang, R. McGreevy, M.C.R. Melo, B.K. Radak, R.D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L.V. Kale, K. Schulten, C. Chipot, E. Tajkhorshid, Scalable molecular dynamics on CPU and GPU architectures with NAMD, *J Chem Phys* 153(4) (2020) 044130.
- [116] P. Eastman, J. Swails, J.D. Chodera, R.T. McGibbon, Y. Zhao, K.A. Beauchamp, L.P. Wang, A.C. Simmonett, M.P. Harrigan, C.D. Stern, R.P. Wiewiora, B.R. Brooks,

V.S. Pande, OpenMM 7: rapid development of high performance algorithms for molecular dynamics, *PLoS Comput Biol* 13(7) (2017) e1005659.

[117] K. Bowers, E. Chow, H. Xu, R. Dror, M. Eastwood, B. Gregersen, J. Klepeis, I. Kolossvary, M. Moraes, F. Sacerdoti, J. Salmon, Y. Shan, D.E. Shaw, Scalable algorithms for molecular dynamics simulations on commodity clusters, *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing* (2006) 43--43.

[118] S. Riniker, Fixed-charge atomistic force fields for molecular dynamics simulations in the condensed phase: an overview, *J Chem Inf Model* 58(3) (2018) 565-578.

[119] F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons 2017.

[120] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T.s. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *J Am Chem Soc* 117(19) (2002) 5179-5197.

[121] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J Phys Chem B* 102(18) (1998) 3586-616.

[122] W.R.P. Scott, P.H. Hünenberger, I.G. Tironi, A.E. Mark, S.R. Billeter, J. Fennen, A.E. Torda, T. Huber, P. Krüger, W.F. van Gunsteren, The GROMOS biomolecular simulation program package, *J Phys Chem A* 103(19) (1999) 3596-3607.

[123] W.L. Jorgensen, J. Tirado-Rives, The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin, *J Am Chem Soc* 110(6) (1988) 1657-66.

[124] R.W. Zwanzig, High-temperature equation of state by a perturbation method. I. nonpolar gases, *J Chem Phys* 22(8) (1954) 1420-1426.

[125] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, *J Mol Graph* 14(1) (1996) 33-8, 27-8.

[126] N. Michaud-Agrawal, E.J. Denning, T.B. Woolf, O. Beckstein, MDAnalysis: a toolkit for the analysis of molecular dynamics simulations, *J Comput Chem* 32(10) (2011) 2319-27.

[127] C.G. Wermuth, C.R. Ganellin, P. Lindberg, L.A. Mitscher, Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998), *Pure and Applied Chemistry* 70(5) (1998) 1129-1143.

[128] T. Seidel, G. Wolber, M.S. Murgueitio, Pharmacophore perception and applications *Applied Chemoinformatics* 2018, pp. 259-282.

- [129] G. Wolber, T. Langer, LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters, *J Chem Inf Model* 45(1) (2005) 160-9.
- [130] T. Seidel, G. Ibis, F. Bendix, G. Wolber, Strategies for 3D pharmacophore-based virtual screening, *Drug Discov Today Technol* 7(4) (2010) e203-70.
- [131] G. Wolber, W. Sippl, Pharmacophore identification and pseudo-receptor modelling, in: C.G. Wermuth, D. Rognan (Eds.), *The Practice of Medicinal Chemistry* (4th edition), Elsevier Ltd, Philadelphia, PA, USA, 2015, pp. 489-507.
- [132] T.J. Cheeseright, M.D. Mackey, R.A. Scoffin, High content pharmacophores from molecular fields: a biologically relevant method for comparing and understanding ligands, *Curr Comput Aided Drug Des* 7(3) (2011) 190-205.
- [133] J. Taminau, G. Thijs, H. De Winter, Pharao: pharmacophore alignment and optimization, *J Mol Graph Model* 27(2) (2008) 161-9.
- [134] D. Barnum, J. Greene, A. Smellie, P. Sprague, Identification of common functional configurations among molecules, *J Chem Inf Comput Sci* 36(3) (1996) 563-71.
- [135] B.K. Shoichet, Virtual screening of chemical libraries, *Nature* 432(7019) (2004) 862-5.
- [136] S.L. Dixon, A.M. Smondryev, E.H. Knoll, S.N. Rao, D.E. Shaw, R.A. Friesner, PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. methodology and preliminary results, *J Comput Aided Mol Des* 20(10-11) (2006) 647-71.
- [137] M.M. Mysinger, M. Carchia, J.J. Irwin, B.K. Shoichet, Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking, *J Med Chem* 55(14) (2012) 6582-94.
- [138] N. Huang, B.K. Shoichet, J.J. Irwin, Benchmarking sets for molecular docking, *J Med Chem* 49(23) (2006) 6789-801.
- [139] D. Schaller, D. Šribar, T. Noonan, L. Deng, T.N. Nguyen, S. Pach, D. Machalz, M. Bermudez, G. Wolber, Next generation 3D pharmacophore modeling, *WIREs Computational Molecular Science* 10(4) (2020).
- [140] D. Sydow, Dynophores: novel dynamic pharmacophores, Humboldt-Universität zu Berlin, Lebenswissenschaftliche Fakultät, 2015.
- [141] M. Bermudez, A. Bock, F. Krebs, U. Holzgrabe, K. Mohr, M.J. Lohse, G. Wolber, Ligand-specific restriction of extracellular conformational dynamics constrains signaling of the M2 muscarinic receptor, *ACS Chem Biol* 12(7) (2017) 1743-1748.

- [142] A. Zollner, C.A. Dragan, D. Pistorius, R. Muller, H.B. Bode, F.T. Peters, H.H. Maurer, M. Bureik, Human CYP4Z1 catalyzes the in-chain hydroxylation of lauric acid and myristic acid, *Biol Chem* 390(4) (2009) 313-7.
- [143] F.P. Guengerich, M.R. Waterman, M. Egli, Recent structural insights into cytochrome P450 function, *Trends Pharmacol Sci* 37(8) (2016) 625-640.
- [144] X. Cao, P. Durairaj, F. Yang, M. Bureik, A comprehensive overview of common polymorphic variants that cause missense mutations in human CYPs and UGTs, *Biomed Pharmacother* 111 (2019) 983-992.
- [145] J. Hritz, A. de Rooter, C. Oostenbrink, Impact of plasticity and flexibility on docking results for cytochrome P450 2D6: a combined approach of molecular dynamics and ligand docking, *J Med Chem* 51(23) (2008) 7469-77.
- [146] B. Knapp, L. Ospina, C.M. Deane, Avoiding false positive conclusions in molecular simulation: the importance of replicas, *J Chem Theory Comput* 14(12) (2018) 6127-6138.
- [147] D. Petrovic, A. Bokel, M. Allan, V.B. Urlacher, B. Strodel, Simulation-guided design of cytochrome P450 for chemo- and regioselective macrocyclic oxidation, *J Chem Inf Model* 58(4) (2018) 848-858.
- [148] L. Capoferri, R. Leth, E. ter Haar, A.K. Mohanty, P.D. Grootenhuis, E. Vottero, J.N. Commandeur, N.P. Vermeulen, F.S. Jorgensen, L. Olsen, D.P. Geerke, Insights into regioselective metabolism of mefenamic acid by cytochrome P450 BM3 mutants through crystallography, docking, molecular dynamics, and free energy calculations, *Proteins* 84(3) (2016) 383-96.
- [149] P.W. de Waal, K.F. Sunden, L.L. Furge, Molecular dynamics of CYP2D6 polymorphisms in the absence and presence of a mechanism-based inactivator reveals changes in local flexibility and dominant substrate access channels, *PLoS One* 9(10) (2014) e108607.
- [150] K.P. Ravindranathan, E. Gallicchio, R.A. Friesner, A.E. McDermott, R.M. Levy, Conformational equilibrium of cytochrome P450 BM-3 complexed with N-palmitoylglycine: a replica exchange molecular dynamics study, *J Am Chem Soc* 128(17) (2006) 5786-91.
- [151] B. Honig, A. Nicholls, Classical electrostatics in biology and chemistry, *Science* 268(5214) (1995) 1144-9.
- [152] P. Rydberg, D.E. Gloriam, J. Zaretski, C. Breneman, L. Olsen, SMARTCyp: a 2D method for prediction of cytochrome P450-mediated drug metabolism, *ACS Med Chem Lett* 1(3) (2010) 96-100.
- [153] J. Li, H. Zhang, G. Liu, Y. Tang, Y. Tu, W. Li, Computational insight into vitamin K1 omega-hydroxylation by cytochrome P450 4F2, *Front Pharmacol* 9 (2018) 1065.

- [154] P. Rydberg, F.S. Jorgensen, L. Olsen, Use of density functional theory in drug metabolism studies, *Expert Opin Drug Metab Toxicol* 10(2) (2014) 215-27.
- [155] A.G. Leach, N.J. Kidley, Quantitatively interpreted enhanced inhibition of cytochrome P450s by heteroaromatic rings containing nitrogen, *J Chem Inf Model* 51(5) (2011) 1048-63.
- [156] S. Bonomo, C.H. Hansen, E.M. Petrunak, E.E. Scott, B. Styrisshave, F.S. Jorgensen, L. Olsen, Promising tools in prostate cancer research: selective non-steroidal cytochrome P450 17A1 inhibitors, *Sci Rep* 6 (2016) 29468.
- [157] J.P. Kowalski, M.G. McDonald, R.D. Pelletier, H. Hanenberg, C. Wiek, A.E. Rettie, Design and characterization of the first Selective and potent mechanism-based Inhibitor of cytochrome P450 4Z1, *J Med Chem* 63(9) (2020) 4824-4836.
- [158] A. Fischer, M. Smiesko, Spontaneous ligand access events to membrane-bound cytochrome P450 2D6 sampled at atomic resolution, *Sci Rep* 9(1) (2019) 16411.
- [159] J. Liu, D. Machalz, G. Wolber, E.J. Sorensen, M. Bureik, New pro-luciferin substrates for human CYP4 family enzymes, *Appl Biochem Biotechnol* 193(1) (2021) 218-237.
- [160] S. Sharma, J. Liu, X. Zhang, S.S. Sharma, E.J. Sorensen, M. Bureik, New luciferin-based probe substrates for human CYP26A1, *Biochem Biophys Rep* 24 (2020) 100861.
- [161] P. Durairaj, L. Fan, S.S. Sharma, Z. Jie, M. Bureik, Identification of new probe substrates for human CYP20A1, *Biol Chem* 401(3) (2020) 361-365.
- [162] Y. Sun, D. Machalz, G. Wolber, M.K. Parr, M. Bureik, Functional expression of all human sulfotransferases in fission yeast, assay development, and structural models for isoforms SULT4A1 and SULT6B1, *Biomolecules* 10(11) (2020).
- [163] F. Yang, D. Machalz, S. Wang, Z. Li, G. Wolber, M. Bureik, A common polymorphic variant of UGT1A5 displays increased activity due to optimized cofactor binding, *FEBS Lett* 592(11) (2018) 1837-1846.
- [164] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohli, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, *Nature* (2021).

8 Appendix

8.1 List of Abbreviations

20-HETE	20-hydroxyeicosatetraenoic acid
BLAST	basic local alignment search tool
CADD	computer-aided drug design
CASP	critical assessment of techniques for protein structure prediction
Cryo-EM	cryogenic electron microscopy.
CYP	cytochrome P450
DNA	deoxyribonucleic acid
IC ₅₀	half maximal inhibitory concentration
FF	force field
GOLD	genetic optimisation for ligand docking
IUPAC	international union of pure and applied chemistry
LOMETS	local meta-threading-server
MD	molecular dynamics
MM	molecular mechanics
MOE	molecular operating environment
MSA	multiple sequence alignment
NMR	nuclear magnetic resonance
OPLS	optimized potentials for liquid simulations
P450	cytochrome P450
PDB	protein data bank
POR	cytochrome P reductase
PSI-BLAST	position-specific iterated basic local alignment search tool
ROC	receiver operator curve
RMSD	root mean square deviation
RMSF	root mean square fluctuation
RNA	ribonucleic acid
SAR	structure activity relationship
VS	virtual screening
QM	quantum mechanics

8.2 List of Figures

- Figure 1. Heme B is the prosthetic group of cytochrome P450 enzymes. In 3D the heme is shown as ball and stick. Iron is shown as a large orange sphere, oxygen atoms are colored red, and nitrogens are colored, and carbons are light grey. Hydrogens have been omitted in both representations. 1
- Figure 2. The catalytic reaction cycle of cytochrome P450 enzymes. Adapted from [27].
..... 5
- Figure 3. Examples of oxidation reactions catalyzed by Cytochrome P450 enzymes [22, 28]. (1) Hydroxylation of aliphatic carbon. (2) Oxidation of alcohol to aldehyde. (3) Oxidation of aldehyde to carbonic acid. (4) O-dealkylation of ether. (5) N-dealkylation of amine. (6) Hydroxylation of aromatic carbon. (7, 8, 9, and 10) Heteroatom oxidation: N-oxidation, P-oxidation, and S-oxidation. (11) Epoxidation of carbon-carbon double bond. (12) Oxidation of dihydropyridine to pyridine. The reactions lead to the insertion of a single oxygen atom and for (4) and (5) to subsequent reactions. 7
- Figure 4. CYPs and their redox partner, such as cytochrome P reductase reside in a membrane. Different channels that connect the catalytic pocket and the enzyme exterior are surmised. Their function is still poorly understood as there is little consensus between CYP isoforms [42]. 9
- Figure 5. The general fold of Cytochrome P450 enzymes. (a) 3D structure of lanosterol 14- α demethylase (CYP51) the only CYP enzyme with an X-ray structure including the transmembrane helix in side view (PDB ID: 4LXJ) [40]. (b) CYP51 globular domain is shown from the bottom view where the plane of the heme is equal to the paper plane. 3D structure of the backbone is shown in cartoon representation and secondary structures are labelled. 10
- Figure 6. The catalytic pocket of Cytochrome P450 enzymes. The substrate (lanosterol) is situated above the prosthetic heme moiety ready for oxygenation in CYP51 (PDB ID: 4LXJ) [40]. The heme iron is coordinated by a cysteine residue that is situated in the Cys loop. Helix I is distorted by a hydrogen bond between the side chain of Thr and the backbone of the residue four positions away in the helix, here Gly. In some CYP enzymes Gly is mutated to Glu, which then forms a covalent bond to the

- heme. The heme moiety is held in place by salt bridges to basic residues (Arg and Lys). For sake of clarity helix B' is not shown. The protein backbone is displayed in the cartoon representation, Heme and Cys residue are shown in ball and stick. All other residues and the substrate are shown in stick representation. Heme carbon atoms are colored black, oxygens are red and nitrogens blue. 12
- Figure 7. The four steps of comparative homology modeling..... 28
- Figure 8. Molecular docking of a small molecule ligand to a macromolecular target. Docking poses of BO-I in the vicinity to the substrate funnel of choline trimethylamine-lyase (CutC, PDB ID: 5FAU [101]) obtained with GOLD are shown [102]. The protein's surface is shown in grey and ligand poses are colored individually and shown in ball and stick..... 33
- Figure 9. Molecular dynamics (MD) simulation of a biomolecular system. **(A)** MD simulations cover only a small part of the biological relevant conformational landscape of the biomolecular system. The conformational landscape is oversimplified as a 2D space, and the biologically relevant part of the conformational landscape is shown as darker patches. Conformations of the system during the MD simulation result in a trajectory and correspond to a path through the conformational landscape. **(B)** The biomolecular system undergoes conformational changes as indicated by curved arrows, which results in a simulation trajectory of conformations. **(C)** For each simulation step, the force field components are calculated that dictate the next conformational change. Bonded interactions include bond length, angle, torsion, and out of plane bending. Non-bonded interactions include electrostatic and van der Waals interactions..... 36
- Figure 10. Dynamic 3D pharmacophore (dynophore) model from MD simulations. Besides the pharmacophore features only ligand and heme cofactor are depicted. The target is only depicted schematically as binding pocket shape. 39

8.3 List of Tables

Table 1. Overview of the 57 human CYP enzymes. CYP families are indicated by background coloring. The following abbreviations are used DR drugs, ES eicosanoids, SS sex steroids, XB xenobiotics.....	2
Table 2. Overview of 3D pharmacophore features as defined in LigandScout [129-131]. Exemplary ligand moieties (black) and their corresponding interacting moiety (grey) are shown.	38

9 Publications

9.1 Publications included in this cumulative thesis

1) First author, Impact factor: 7.79

Machalz, D., Pach, S., Bermudez, M., Bureik, M., & Wolber, G. (2021). Structural insights into understudied human cytochrome P450 enzymes. *Drug Discovery Today*, 26, Issue 10, 2456-2464.

<https://doi.org/10.1016/j.drudis.2021.06.006>

2) Second author, Impact factor: 5.29

Yan, Q., Machalz, D., Zöllner, A., Sorensen, E. J., Wolber, G., Bureik, M. (2017). Efficient substrate screening and inhibitor testing of human CYP4Z1 using permeabilized recombinant fission yeast. *Biochemical Pharmacology*, 146, 174-87.

<https://doi.org/10.1016/j.bcp.2017.09.011>

3) Shared first author, Impact factor: 4.12

Durairaj, P., Fan, L., Machalz, D., Wolber, G., Bureik, M. (2019). Functional characterization and mechanistic modeling of the human cytochrome P450 enzyme CYP4A22. *FEBS Letters*, 593, Issue 16, 2214-2225.

<https://doi.org/10.1002/1873-3468.13489>

4) Shared first author, Impact factor: 5.29

Du, W., Machalz, D., Yan, Q., Sorensen, E. J., Wolber, G., Bureik, M. (2020). Importance of asparagine-381 and arginine-487 for substrate recognition in CYP4Z1. *Biochemical Pharmacology*, 174, 113850.

<https://doi.org/10.1016/j.bcp.2020.113850>

5) Shared first author, Impact factor: 6.18

Machalz, D., Li, H., Du, W., Sharma, S., Liu, S., Bureik, M., Wolber, G. (2021). Discovery of a novel potent cytochrome P450 CYP4Z1 inhibitor. *European Journal of Medicinal Chemistry*, 215, 113255.

<https://doi.org/10.1016/j.ejmech.2021.113255>

9.2 All peer-reviewed Publications

Machalz, D., Pach, S., Bermudez, M., Bureik, M., & Wolber, G. (2021). Structural insights into understudied human cytochrome P450 enzymes. *Drug Discovery Today*, 26, Issue 10, 2456-2464.

- 1) <https://doi.org/10.1016/j.drudis.2021.06.006>
- 2) Loke, S., Stoll, A., Machalz, D., Botrè, F., Wolber, G., Bureik, M., & Parr, M. K. (2021). Corticosteroid Biosynthesis Revisited: No Direct Hydroxylation of Pregnenolone by Steroid 21-Hydroxylase. *Frontiers in Endocrinology*, 12(629). <https://doi.org/10.3389/fendo.2021.633785>
- 3) Machalz, D., Li, H., Du, W., Sharma, S., Liu, S., Bureik, M., Wolber, G. (2021). Discovery of a novel potent cytochrome P450 CYP4Z1 inhibitor. *European Journal of Medicinal Chemistry*, 215, 113255. <https://doi.org/10.1016/j.ejmech.2021.113255>
- 4) Liu, J., Machalz, D., Wolber, G., Sorensen, E. J., Bureik, M. (2021). New ProLuciferin Substrates for Human CYP4 Family Enzymes. *Applied Biochemistry and Biotechnology*, 193(1), 218–237. <https://doi.org/10.1007/s12010-020-03388-6>
- 5) Sun, Y., Machalz, D., Wolber, G., Parr, M. K., Bureik, M. (2020). Functional Expression of All Human Sulfotransferases in Fission Yeast, Assay Development, and Structural Models for Isoforms SULT4A1 and SULT6B1. *Biomolecules*, 10(11), 1517. <https://doi.org/10.3390/biom10111517>
- 6) Gabr, M. T., Machalz, D., Pach, S., Wolber, G. (2020). A benzoxazole derivative as an inhibitor of anaerobic choline metabolism by human gut microbiota. *RSC Medicinal Chemistry*. <https://doi.org/10.1039/D0MD00218F>
- 7) Schaller, D., Šribar, D., Noonan, T., Deng, L., Nguyen, T. N., Pach, S., Machalz, D., Bermudez, M., Wolber, G. (2020). Next generation 3D pharmacophore modeling. *WIREs Computational Molecular Science*. <https://doi.org/10.1002/wcms.1468>
- 8) Du, W., Machalz, D., Yan, Q., Sorensen, E. J., Wolber, G., Bureik, M. (2020). Importance of asparagine-381 and arginine-487 for substrate recognition in CYP4Z1. *Biochemical Pharmacology*, 174, 113850. <https://doi.org/10.1016/j.bcp.2020.113850>

-
- 9) Stoll, A., Loke, S., Joseph, J. F., Machalz, D., de la Torre, X., Botrè, F., Wolber, G., Bureik, M., Parr, M. K. (2019). Fine-mapping of the substrate specificity of human steroid 21-hydroxylase (CYP21A2). *The Journal of Steroid Biochemistry and Molecular Biology*, 105446. <https://doi.org/10.1016/j.jsbmb.2019.105446>
 - 10) Durairaj, P., Fan, L., Machalz, D., Wolber, G., Bureik, M. (2019). Functional characterization and mechanistic modeling of the human cytochrome P450 enzyme CYP4A22. *FEBS Letters*, 593, Issue 16, 2214-2225. <https://doi.org/10.1002/1873-3468.13489>
 - 11) Zierau, O., Kolodziejczyk, A., Vollmer, G., Machalz, D., Wolber, G., Thieme, D., Keiler, A. M. (2019). Comparison of the three SARMs RAD-140, GLPG0492 and GSK-2881078 in two different in vitro bioassays, and in an in silico androgen receptor binding assay. *Journal of Steroid Biochemistry and Molecular Biology*, 189. <https://doi.org/10.1016/j.jsbmb.2019.02.014>
 - 12) Yang, F., Machalz, D., Wang, S., Li, Z., Wolber, G., Bureik, M. (2018). A common polymorphic variant of UGT1A5 displays increased activity due to optimized cofactor binding. *FEBS Letters*. <https://doi.org/10.1002/1873-3468.13072>
 - 13) Keiler, A. M., Zierau, O., Wolf, S., Diel, P., Schänzer, W., Vollmer, G., Machalz, D., Wolber, G., Parr, M. K. (2018). Androgen- and estrogen-receptor mediated activities of 4-hydroxytestosterone, 4-hydroxyandrostenedione and their human metabolites in yeast based assays. *Toxicology Letters*. <https://doi.org/10.1016/j.toxlet.2018.04.026>
 - 14) Yan, Q., Machalz, D., Zöllner, A., Sorensen, E. J., Wolber, G., Bureik, M. (2017). Efficient substrate screening and inhibitor testing of human CYP4Z1 using permeabilized recombinant fission yeast. *Biochemical Pharmacology*, 146, 174-87. <https://doi.org/10.1016/j.bcp.2017.09.011>

9.3 Conference Proceedings

1) **EUROPIN Meeting**

2020, online

Talk: Novel Potent Inhibitor for Promising Anticancer Target Cytochrome P450 4Z1

2) **EUROPIN Summer School on Drug Design**

2019, Vienna, Austria

Talk: Targeting Cytochrome P450 4Z1 as Novel Cancer Target

3) **14th German Conference on Chemoinformatics (GCC)**

2018, Mainz, Germany

Poster: Holistic CYP4Z1 modeling towards tissue specific prodrug activation

4) **11th International Conference on Chemical Structures (ICCS)**

2018, Noordwijkerhout, the Netherlands

Poster: Molecular nature of the increased activity of the Uridine 5'-diphosphoglucuronosyltransferase (UGT) nine-fold mutant 1A5*8

5) **EUROPIN Summer School on Drug Design**

2017, Vienna, Austria

Poster: Screening for substrates and inhibitors for Cytochrome P450 4Z1 - a promising breast cancer target

Talk: Mechanistic in silico profiling of xenobiotic metabolism

Declaration of independent work

I hereby declare that I have written this dissertation independently and without any unauthorized help. I have only used the indicated sources and aids. No doctoral thesis on this topic and in this form has been submitted yet and any extraneous ideas or work has been marked as such.

Berlin, den 18.08.2021

David Machalz