universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

In silico investigation of cytochrome P450 inhibition by peptides

verfasst von / submitted by

Iris Hauser

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magistra pharmaciae (Mag.pharm.)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 605
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Pharmazie
Betreut von / Supervisor:	Assoz. Prof. Mag. Dr. Johannes Kirchmair

Acknowledments

First of all, I would like to thank my supervisor Assoz. Prof. Mag. Dr. Johannes Kirchmair for letting me be part of his research group and thus giving me the opportunity to gain insight into the field of computational approaches in drug discovery for one semester.
Prof. Kirchmair supported me from the beginning of my work and was immediately on hand with solutions and new ideas when problems occurred.
I would also like to thank Anya Ya Chen and Steffen Hirte who made me feel very welcome in the group and assisted me with any issues I had.

Furthermore, I would like to thank especially my parents for their mental and financial support that made this study possible.

I would also like to thank my friends and fellow students for supporting each other and making this time an unforgettable period of my life.

Finally, my special thanks go to my partner who has been with me from the beginning of my studies, and without him, I would not be where I am today. Thank you for your daily support and your patience in every situation.

Zusammenfassung

Wir sind täglich einer Vielzahl von Stoffen ausgesetzt, wie Umweltgifte, Kosmetika, Lebensmittel oder Arzneistoffen, sogenannten Xenobiotika. Die Abbaumechanismen, insbesondere der Abbau von Arzneistoffen sind wichtige Stoffwechselvorgänge und schützen den Körper vor toxischen Effekten. Die wichtigste Enzymklasse, die für diesen Metabolismus verantwortlich sind, sind die in der Leber gebildeten Cytochrom P450 Enzyme.
CYPs verstoffwechseln eine Reihe von Xenobiotika und wichtigen Arzneistoffen wie Rifampicin, Barbiturate oder Phenytoin, was zu schwerwiegenden Arzneimittel Interaktion oder toxischen Plasmakonzentrationen führen kann, wenn diese Enzyme gehemmt werden. Peptide und Proteine gehören zu den wichtigsten Makronährstoffe und spielen eine wesentliche Rolle bei Vorgängen wie der Immunabwehr, Übertragung von Nervensignalen, Zellstoffwechsel oder Metabolismus. Sie werden grundsätzlich nicht durch CYP Enzyme abgebaut, sondern unterliegen einer Proteolyse durch Proteasen. Dennoch wurde mit Cyclosporin bereits ein starker CYP3A4 Inhibitor bestimmt, der in der Lage ist, schwerwiegende Wechselwirkungen mit anderen wichtigen Arzneistoffen wie Lovastatin oder Atorvastatin einzugehen und unterstreicht die Wichtigkeit der Untersuchung einer möglichen CYP Hemmung durch Peptide.
In dieser Arbeit werden mithilfe von computerunterstützen Anwendungen drei Ansätze verfolgt, um eventuelle Interaktionen von Peptiden mit CYP Enzymen zu beobachten. Alle Schritte zur Datenverarbeitung wurde mit dem Data-Mining Programms KNIME durchgeführt. 1) Zunächst wurden physikochemischen Eigenschaften für Peptide und CYP Inhibitoren berechnet und auf Ähnlichkeiten untersucht, um festzustellen ob sie einen Einfluss auf eine mögliche CYP Hemmung haben. Unter der Verwendung von DataWarrior zeigte eine PCA die Verteilung von bekannten Inhibitoren, nicht Inhibitoren, zugelassene Arzneistoffe und Peptide im chemischen Raum entsprechend ihrer Eigenschaften. 2) Als nächstes wurde im Schrödinger Maestro alle verfügbaren Inhibitor Strukturen aus der PDB auf ihre Eigenschaften und Merkmale untersucht, die sie zu einem starken Hemmstoff machen. 3) Zuletzt wurden alle verfügbaren Peptide aus der PepTherDia in kleinere Fragmente gespalten und gemeinsam mit allen bekannten CYP Inhibitoren in Glide von Schrödinger Maestro Docking Simulationen unterzogen und anschließend auf Basis der ermittelten docking scores ROC Kurven erstellt.

Die Ergebnisse der Arbeit zeigen, dass Peptide im Großen und Ganzen nicht in der Lage sind CYP Enyzme zu hemmen, da sie in ihren Eigenschaften gravierende Abweichungen von bekannten Inhibitoren zeigen. Wenn sie allerdings in kleineren Fragmenten vorliegen, besteht durchaus die Möglichkeit, dass sie in ähnlicher Weise wir Inhibitoren agieren können und hier Bindungen eingegangen werden.

Abstract

Every day we are exposed to a large variety of substances, such as environmental toxins, cosmetics, foods or drugs, and so-called xenobiotics. The degradation mechanisms, especially the degradation of drugs, are important metabolic processes and protect the body from toxic effects. The most important enzyme class responsible for this metabolism is the cytochrome P450 enzymes, which are produced in the liver.

CYPs metabolize the majority of all xenobiotics and important drugs such as rifampicin, barbiturates or phenytoin, which can lead to serious drug interactions or toxic plasma concentrations if these enzymes are inhibited.

Peptides and proteins are among the most important macronutrients and play an essential role in processes such as immune defense, transmission of nerve signals, cell metabolism or metabolism. They are usually not degraded by CYP enzymes but are subject to proteolysis by proteases. Nevertheless, with cyclosporine one strong CYP3A4 inhibitor has already been identified. Via its inhibitory activity on CYP3A4, cyclosporine can cause serious drug-drug interactions when applied together with other drugs such as lovastatin or atorvastatin. This underlines the importance of investigating possible CYP inhibition by peptides.

In this work, three approaches are followed using computational applications to identify potential interactions of peptides with CYP enzymes. All data processing steps were performed using the data-mining program KNIME. 1) Physicochemical properties of peptides and CYP inhibitors were calculated and analyzed for similarities to determine if these properties had an impact on potential CYP inhibition. A PCA created with DataWarrior, showed the distribution of known inhibitors, non-inhibitors, approved drugs, and peptides in chemical space according to their properties. 2) Then, Schrödinger Maestro was used to screen all available inhibitor structures from the PDB for their properties and characteristics that make them a potent inhibitor. 3) Finally, all available peptides from PepTherDia, a free accessible database with all currently approved peptide drugs, were dissected into smaller fragments and subjected to docking simulation using Schrödinger Maestro together with all known CYP inhibitors in Glide, and then ROC curves were generated based on the determined docking scores.

The results of this work show that, by and large, peptides are not able to inhibit CYP enzymes because their properties are very different from those of known inhibitors. However, if they are present in smaller fragments, it is quite possible that they can act similarly to inhibitors and form bonds here.

Table of contents

List of abbreviations

CYP P450: Cytochrom P450

ROC: Rapid Overlay of Chemical Structures

PDB: Protein Data Bank

ANF: a- napthoflavone

PCA: Principal Component Analysis

NADPH: nicotinamide adenine dinucleotide phosphate

FDA: Food and Drug Administration

GLP-1: glucagon-like-peptide-1

DPP4: dipeptidyl-peptidase-4

NEP: neutral endopeptidase

PEG: polyethylene glycol

ADMET: absorption, distribution, metabolism and excretion

QSAR: Quantitative structure-activity relationship model

EMA: European medicines agency

CCD: Chemical Component Dictionary

PC: Principal Components

DprE1: decaprenylphosphoryl-b-d-ribofuranose 2-oxidase)

AUC: Area Under the Curve

1. Introduction

1.1. Metabolism of Xenobiotics

Xenobiotics are substances that do not occur naturally in the human body, but with which we are confronted in many different ways. These include food, environmental toxins, cosmetics, alcohol, but most importantly the medications we take. When such xenobiotics enter the body, they undergo a degradation process that helps to eliminate them and prevent toxic accumulation. Especially with drugs, it is important to understand the metabolism, as they must not be degraded too quickly or too slowly. Prodrugs use exactly this metabolism to form an active metabolite from an inactive form and thus ensure the release of the active ingredient over a longer period of time without harmful effects. The fundamental role of metabolism is to convert substances from a lipophilic to a hydrophilic state in various organs and tissues such as the gastrointestinal tract, lungs, and liver before being eliminated via the kidneys. We can distinguish between phase I and phase II reactions, although these do not have to occur in this order and may even take place independently or only as phase II.
In phase I reactions, enzyme-mediated oxidations, reductions, or hydrolyses alter, add, or remove functional groups that make substances more polar. In this phase, cytochromes P450 enzymes form the most important enzyme family. In phase II conjugation reactions, on the other hand, strongly hydrophilic groups are attached, and these water-soluble metabolites are subsequently excreted in the urine. Here, reactions such as glucuronidation, sulfation, methylation or acetylation take place, mediated by UDP-glucuronosyltransferases, sulfotransferases or acetyltransferases. [1] [2]

1.2. Cytochrome P450 function and structure

Cytochrome P450 is a superfamily of heme-containing enzymes found predominantly in the smooth endoplasmic reticulum, liver cells, and intestine. There are currently 57 human CYP isozymes that differ in form, binding pocket size, and flexibility, resulting in varied characteristics and substrate binding. CYP enzymes take over a number of important catalytic reactions in the body such as drug metabolism of more than 80% of clinically used drugs, steroid biosynthesis or oxidation, and metabolism of unsaturated fatty acids and fat-soluble vitamins. [3] [4] As mentioned above, they are mostly involved in the hydroxylations of phase I reactions, resulting in the conversion of lipophilic compounds to more hydrophilic, water-soluble compounds for better excretion and further metabolism. [2] Six CYP isoenzymes: 3A4,

1A2, 2D6, 2C9 and 2C19 are particularly important and responsible for the metabolism of most approved drugs.

Cytochrome P450 are monooxygenases and consist of about 500 amino acids that form a polypeptide chain. [3] Most enzymes are basic and share a common overall fold and topology with a large α- helix and a smaller β-sheet domain. Nevertheless, the isozymes differ significantly in their properties and show a sequence identity of less than 20%. [5] The catalytic center is composed of three parallel (D, L, I) and one antiparallel (E) Helix with iron protoporphyrin IX as the prosthetic group enclosed between the I-Helix and the L – Helix.[5] [3] The protoporphyrin center is a dianion with a large, flat aromatic surface filled with four coordination sites of iron, the fifth is a cysteine anion and the sixth an exchangeable water molecule.[6]

1.2.1. Catalytic cycle

The prosthetic group iron protoporphyrin IX linked with the cysteine ligand forms the catalytic center and is responsible for the hydroxylation of substrates in most cases. When cytochromes bind their substrates via the thiolate sulfur of a cysteine residue, the catalytic cycle starts with a change in the oxidation potential of the iron center, either with or without, the loss of the water molecule.[7] NADPH-P450 reductase transfers an electron and reduces the active site from Fe^{3+} to Fe^{2+}. Then oxygen is bonded, followed by another electron transfer and reduction. After the attachment of a proton and cleavage of the O-O bond, water is released, and a "perferry-oxygen complex" (FeO_3^+) remains. The substrate is hydroxylated to this complex, released, the active site returns to its original state, and the cycle begins again. Whether the substrate-binding is reversible or irreversible depends mainly on the FeIII state with water, i.e. whether it is displaced from the active site or not.[8] [9]

1.2.2. CYP inhibition

Patients taking multiple medications at the same time are particularly at risk for drug interactions caused by inhibition or induction of CYP P450 enzymes. [10] Inhibition of cytochromes can lead to an increase in the plasma concentration of administered drugs, as they can no longer be biotransformed by CYP, which can result in various side effects and toxic reactions. Because of these adverse side effects, some drugs have already been

withdrawn from the market.[11] Inhibitors are structurally similar to substrates are capable of inhibiting the enzyme by two different mechanisms: 1) reversible inhibitors which do not replace the water molecule in the active site are called type I inhibitors. Here one can distinguish between competitive, noncompetitive, or uncompetitive where substrate and inhibitor either compete for the binding site or bind at different sites. Very often, however, mixed forms are observed. 2) type II inhibitors or mechanism-based inhibitors, which bind directly to the heme iron by displacement of water and thus irreversibly inhibit CYP P450 enzymes by forming a stable complex. Some substrates also exhibit typical type II inhibitor binding and act as both substrate and inhibitor. [11] [8]

Normally generated metabolites are different in structure and unstable for an inhibitory effect. But with some it is quite possible, as for example with the antihypertensive diltiazem. It has been observed that the metabolites N- desmethyl diltiazem and N,N-didesmethyl diltiazem are more powerful inhibitors of CYP3A4 than the parent compound diltiazem. [10]

1.2.3. CYP induction

Xenobiotics such as alcohol, drugs or environmental factors can not only inhibit but also induce the metabolic processes of Cytochromes. Principally, this stimulation of metabolism is a protective reaction of the body for the cells as increased detoxification activity. However, it should be noted that this can also activate drugs, leading to increased metabolism, reducing plasma concentrations and potentially ineffective pharmacotherapy for patients. This occurs through the activation of key transcription factors in the cell such as pregnane X receptor (PXR), constitutive androstane receptor (CAR), and aryl hydrocarbon receptor (AhR). The result is increased transcription of CYPs and thus increased mRNA synthesis. An example would be the induction of CYP3A4 by St. John's wort with co-administration of cyclosporine. Both the metabolism and clearance of cyclosporine decrease, and the plasma concentration falls below therapeutic levels.

Other problems may include an imbalance between intoxication and detoxification, but also a decrease or increase in toxicity as more reactive metabolites are formed.

For example, induction of rifampicin can lead to hepatotoxic reactions with acetaminophen, through increased production of toxic metabolites. [12] [13]

1.3. Peptides

Proteins and peptides are linear polymer molecules composed of amino acids covalently linked by peptide bonds. In general, proteins are composed of primary, secondary, tertiary, and quaternary structures. The primary structure describes the linear sequence of amino acids, the secondary structure includes alpha helices and beta sheets. The tertiary structure relates to the 3D arrangement of the protein and the quaternary structure to the 3D structure of proteins consisting of more than one chain. Proteins are part of the most important macronutrients and play a key role in the body's immune defense, metabolism, nerve signal transmission, and cell control, regulation, and coordination.[14]

There is still disagreement in the scientific community about the peptide definition and when "peptide" ends and "protein" begins.[15] For example, the FDA[16] defines peptides as "any polymer composed of 40 or fewer amino acids,"[17] while the International Union of Pure and Applied Chemistry (IUPAC) defines peptides as "amides derived from two or more amino carboxylic acid molecules (the same or different) by formation of a covalent bond from the carbonyl carbon of one to the nitrogen atom of another with formal loss of water." [18]

1.3.1. Metabolism

In the development of new drugs, it is particularly important to know the metabolism of these since it has a significant influence on the pharmacology and safety of the therapeutic drugs. Peptides are not degraded by CYP enzymes, which convert lipophilic compounds into hydrophilic ones, as is the case with most low-molecular-weight compounds but are especially susceptible to proteases.

Peptidases degrade peptides to amino acids depending on their catalytic site. We can divide the peptidases into two groups: endopeptidases and exopeptidases. Endopeptidases break the bond inside the polypeptide chain; among the exopeptidases, a distinction is made between aminopeptidases which cut peptides at the N- terminus, and the carboxypeptidases where the cleavage takes at the C-terminus. Depending on whether a peptide drug is absorbed orally or parenterally, it passes through different pathways during degradation. However, metabolization mainly occurs in tissues such as the gastrointestinal tract, liver, kidney, and blood. Glomerular filtration and renal reabsorption are the primary mechanisms for excretion in the proximal tubule of the kidney. [19] [20]

1.3.2. Peptide drugs

Peptide drugs already achieved outstanding success at the beginning of the 20th century with therapeutics such as penicillin or the polypeptide insulin. Despite good efficacy and low toxicity, they were nevertheless considered unsuitable for pharmacotherapy for a very long time and less good than small molecules because of their susceptibility to degradation by proteases and their poor bioavailability. Thanks to the great technological progress made in recent years, especially in simplifying the analysis, peptides are being used as both diagnostics and therapeutics and are gaining more and more interest and importance. Peptides account for about 1.5% of the global drug market, with about 85% of peptide drugs administered primarily as parenteral injections. Drug formulations range from vaccines to antimicrobial peptides that serve as alternatives to antibiotics to which resistance has developed, to peptides bound to biopolymers or polyethylenes.[21]

They are especially important for the treatment of type II diabetes, where they target the receptor for glucagon-like peptide 1 (GLP-1). Numerous drugs are already on the market for this purpose, including metformin, somapacitan, liraglutide or semaglutide.

Liraglutide, for example, is a long-acting GLP-1 agonist with 97% amino acid homology to native GLP1, marketed under the trade name Victoza®.

Due to the acylation of a lysine residue and an amino acid substitution from lysine to arginine, it has a greatly prolonged plasma half-life making it suitable for once-daily subcutaneous injections. In contrast to GLP-1, which is rapidly degraded by DPP4 and NEP, liraglutide has a slowed release due to high degree of binding to the human serum albumin. Liraglutide enhances meal-induced insulin release for effective lowering of blood glucose, appetite and body weight combined with prolonged duration of action.[22]

However, despite their numerous advantages, peptides continue to be a challenge for research due to their low metabolic activity, high molecular weight, low bioavailability, short half-life, and rapid enzymatic degradation by proteases. [20]

1.3.3. Strategies to improve the metabolic stability of peptides

In recent years, some methods have been developed to overcome these problems, in particular, to increase the stability against proteolysis and to extend the half-life: [20] [19]

1. Dipeptidyl- peptidase IV (DPP-IV) is an important enzyme in tissues for the peptide degradation of GLP1 and GLP1 analogs like Liraglutide. A modification or

substitution of the N-and/or C-terminal sites such as N-methylation or a C-terminal extension can help to improve the resistance

2. D-amino acids or unnatural amino acids also exhibited better stability to proteolytic degradation compared to L- amino acids
3. Conversion of linear peptides and peptide backbone into bicyclic structures, as well as backbone modification, e.g., replacement of beta residues by alpha residues, stabilize and show a prolonged half-life
4. Another method represents the formulation of nanoparticles: peptides are stimulated to self-assemble by adding amino acids like glutamine and asparagine to the C- or N-terminus, changing their lipophilic properties. These nanostructures are found in lytic peptides, which are resistant to degradation.

A further problem is rapid excretion via the kidneys. Molecules below 60kDa are excreted rapidly; for larger ones this process is prolonged and can be achieved by attaching polyethylene glycol (PEG) or serum albumin. PEG is highly water-soluble with high mobility, non-toxic, and triggers antibody reactions. Pegylation at the C-terminus improves peptidase stability, maintains concentration in the blood longer, and prevents rapid excretion.[20] [19]

As we have seen, despite the various benefits of peptides, there are some challenges to overcome in the development of peptide therapeutics, and numerous methods have already been developed to increase metabolic stability. Although there is a growing interest in therapeutic peptides, there are few data on the inhibition of the CYP 450 enzyme by peptides. Nevertheless, there are some reasons to investigate this inhibitory effect, as cyclosporine has already been shown to be a CYP inhibitor. Cyclosporine[23] is a lipophilic, cyclic polypeptide of 11 amino acids and is used as an immunosuppressant in organ transplantation. The calcineurin inhibitor has low bioavailability and is metabolized mainly by CYP3A4. In numerous studies, however, cyclosporine has been observed to inhibit multiple CYP isozymes, resulting in interactions with other important drugs that are also degraded by these enzymes, such as atorvastatin, lovastatin, or simvastatin. [24] [25]

1.4. Computer-aided drug design

In drug development, it is particularly important that new therapeutics have a high safety profile and good pharmacological properties. A large number of drugs fail in the first phase of development due to poor ADMET properties and, in the worst case, can lead to severe drug-drug interactions. In particular, CYP P450s are mainly responsible for these DDIs when multiple drugs are administered simultaneously because it is inhibited or induced by small molecules. For this reason, it is particularly important to have knowledge of metabolism at an early stage of drug development. [26] [4] In vivo tests or experimental analyses are an essential part of any investigation, but they also have some disadvantages. For example, in vitro, and in vivo models which are usually influenced by factors such as age, gender, stress, or drugs, are very time-consuming and expensive. Another problem with experimental approaches is the limited data on the structure-activity relationship for inhibition and induction and the high requirement for technical resources. [26] [27] [28]

Accordingly, numerous in silico methods have already been developed. These theoretical approaches can be used at a an initial point of pharmaceutical research, are less costly, and are capable for predicting the metabolic properties of small molecules as well as inhibition and induction data. However, there are some complications with these models as well. Computational approaches are often based on experimental data that may be incorrect, incomplete, of insufficient quality, or influenced by experimental variability. Literature searches are also very time-consuming and often face the challenge that data is not freely available and is kept secret by pharmaceutical companies. [28] [29]

In general, a distinction can be made between ligand-based and structure-based methods. Ligand-based approaches utilize chemical and biological information on the ligand. They provide structure-activity relationship (QSAR) information, but only consider the substrate and not the interactions of the protein-ligand complex. This method is used for molecular fingerprinting to approximate steric hindrance and orientation problems but has relatively low interpretability on its own.

Structure-based approaches, on the other hand, rely on known 3D crystal structures of proteins or receptor-ligand complexes resolved by X-ray crystallography using energies or energy-like scoring functions.[14] With this method, it is possible to take a closer look into the active site of a protein to determine protein-ligand interactions, generate binding poses or build a pharmacophore model. Docking, molecule dynamic (MD) simulations, or QSAR

models are primarily used because understanding the structural properties of metabolic proteins is a prerequisite for investigating protein-ligand interactions. [26] [27]
In particular, docking is a widely used structure-based method to determine interactions between the protein-ligand pairs. The site of metabolism or inhibition can be predicted by calculating the most likely pose of a ligand in the active side or generate different docking poses. Particularly in the analysis of CYP interactions with different small molecules, docking is a popular method as it provides fast, very accurate results of various docking poses. The flexibility and hydrophobic side of CYP enzymes, on the other hand, complicate the docking process, and algorithms operate differently, so the scoring function is not always reliable in the evaluation of binding affinity. [30] [28]

In summary, there is no clear preference for whether experimental, computational, ligand-based, or structure-based methods should be used. As we have seen each method has its advantages and disadvantages and on its own many restrictions, be it time-consuming, computationally expensive, provides limited information, or requires significant resources. The combination of both, theoretical and experimental, provides the most accurate results and robust predictive models, making an important contribution in the early stages of drug development to understand metabolism and avoid serious drug interactions.

2. Aims

Interest in the development of peptide therapeutics has continued to grow in recent years. NovoNordisk[31] is a global leader in diabetes research and was founded in Denmark in 1923. The company is committed to fighting serious chronic diseases such as hemophilia, obesity, growth disorders, but especially type I and type II diabetes. In diabetes, peptide preparations naturally play a very special role, for example by promoting insulin release from the pancreas. Although NovoNordisk has been working with peptides for years and has detailed knowledge of their metabolism, the question remains whether it is possible to interfere with and inhibit CYP metabolism.
The aim of this work was to determine, whether peptides or smaller fragments of them can inhibit CYP enzymes and thus trigger life-threatening drug interactions. This was conducted using the following three approaches: 1) creating a PCA to visualize if physicochemical properties effect CYP P450 inhibition by peptides 2) Analysis of all available crystal

structures of inhibitors from PDB for their characteristics and interactions with the catalytic center 3) Docking and subsequent ROC curve generation to determine the possible involvement of peptides in ligand binding.

3. Materials and methods

3.1. Inhibition data sets

CYP inhibition datasets were provided by Wojtek Plonka [29] and contain all information on molecules inhibiting CYP enzymes 1A2, 3A4, 2C9, 2C19 and 2D6 from PubChem[32] and ChEMBL[33] databases. Structures were downloaded as SMILES or CSV files and analyzed with the RDKit version 2018.09.150 followed by a cleaning and filtering procedure[29]
PubChem is a public archive and, since 2004, has been part of the Molecular Libraries Roadmap Initiatives of the US National Institutes of Health (NIH).
The Database contains about 157 million descriptions of chemical substances, 60 million chemical structures, and 1 million biological assays with 10.000 protein target sequences, used particularly for the areas of cheminformatics, chemical biology, medicinal chemistry, or drug discovery. To process this large amount of data, PubChem is linked to 3 databases:

- *Substance* (https://www.ncbi.nlm. nih.gov/pcsubstance) with the identifier SID (SubstanceID) extracted chemical structures provided by more than 350 contributors, including university labs, government agencies, pharmaceutical companies, chemical vendors, publishers and chemical biology resources
- *Compound* (https://www.ncbi.nlm.nih.gov/pccompound) with the identifier CID (CompoundID) where the extracted structures are stored
- BioAssay (https://www.ncbi.nlm.nih.gov/pcassay) with identifier AID (AssayID), in which descriptions of biological assays are stored [32]

ChEMBL (https://www.ebi.ac.uk/chembl) was first released in 2011 and contains (as of 2017) over 1,6 million differently composed structures with 14 million activity values ranging from single to protein complexes, tissues, and whole organism, as well as neglected disease screening data, crop protection data or drug metabolism and bioactivity data from patents in many ways.[34] [33] The goal of ChEMBL was to facilitate an easy access to standardized data for users, as data collection is a very time-consuming process. In addition, there is the problem that data is most of the time presented only in pictures and abbreviations

or differs from journal to journal. Data is collected from a large number of chemical journals, converted into a machine-readable form, and subjected to a standardization process.[34] [35] In contrast to PubChem, which focuses mainly on the binding affinity between the ligand and its target, ChEMBL captures all the properties necessary for the structure-activity relationship that makes a successful drug like metabolism, absorption, toxicity or information about application forms or dosage. [36] [35]

In conclusion, these databases are becoming more and more interesting and organizations such as Novartis-GNF[37], the Drugs for Neglected Diseases Initiative (DNDi)[38], the Medicines for Malaria Venture (MMV)[39], the Food and Drug Administration (FDA)[16] or universities are using these platforms for data sharing, which is a great help in the development of new drugs.[33]

3.2. Peptide data set

The peptide data comes from PepTherDia[15] (Peptide Therapeutics and Diagnostics: (http://peptherdia.herokuapp.com/) an online accessible database, developed by PhD students at Liverpool John Moores University[40] and an associate professor at the School of Engineering and Physical Sciences at Heriot-Watt University[41] in Edinburgh. The web server stores around 105 approved peptide drugs with their physicochemical properties, indications, plasma protein binding, approval date, and more. PepTherDia is manually managed and regularly updated to support early phase peptide drug development. Drugs were selected according to defined inclusion and exclusion criteria such as a fixed upper and lower length limit or a molecular weight < 5000g/mol, obtained from DrugBank[42], National Centre for Advancing Translational Sciences web page[43], Drugs. com[44], FDA[16] and EMA[45] web pages, or various pharmaceutical company websites. [15]

The approved drugs used to create the PCA were taken from DrugBank[42].

3.3. Principal component analysis

The PCA was performed using DataWarrior[46], a free software for visualizing chemical data, including all physicochemical properties previously calculated with KNIME. PCA or principal component analysis is a method of compressing data sets with high information density into a simplified one for easier visualization and presentation. The two axes represent

the explained variances based on the physicochemical properties, i.e. how similar each structure is to the other. DataWarrior normalizes and centers the values, and the individual structures are displayed according to their placement in chemical space.

3.4. KNIME

KNIME[47] was used for linking and processing data, calculations, and the creation of curves. KNIME, also known as "Konstanz information miner", is an open-source software, founded in 2004 by the University of Konstanz in Germany, written in java and based on Eclipse. [47] The web interface allows users to process data intuitively, as the workflow can be interrupted at any point, changed rapidly, and intermediate results can be saved permanently, making it easy for users with less programming experience to analyze the data. There are numerous different nodes that are dragged and dropped and are subject to a machine learning or mining algorithm, for data source selection, modeling, visualization, or data preprocessing. [48 49]
In this case, KNIME version 4.3.2, its Schrödinger extension for KNIME, and the RDKIT toolkit, were used to handle the molecules. Usually, the workflow begins with a data reading node that reads in a file and is represented in a table-based format. The extracted data is processed over multiple nodes in the pipeline to add columns, filter rows, or calculate properties before being visualized as curves or images.

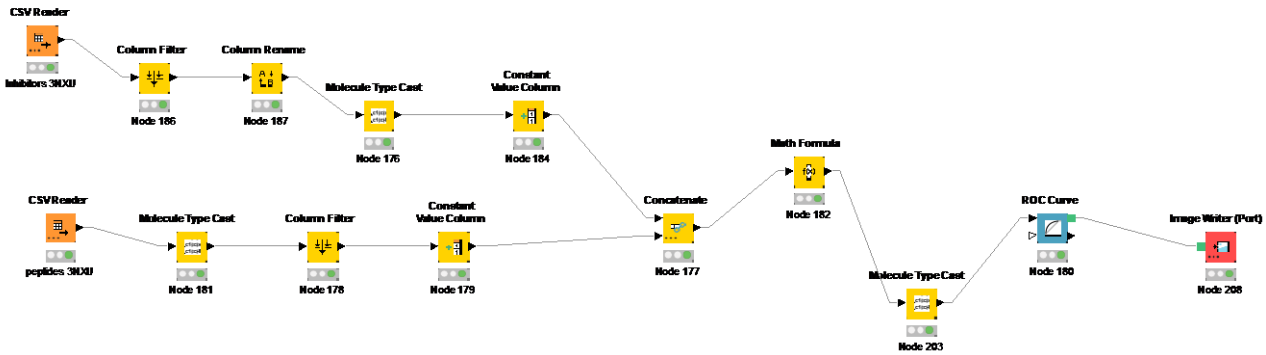


Figure 1: KNIME Workflow for ROC curve generation

3.5. Protein Data Bank

The Protein Data Bank[50] is the world's largest public archive of macromolecules. It was founded in 1971 by Brookhaven National Laboratory and grew from 7 molecules to over 40,000 structures. Since 1998 the database has been managed by the Research Collaboratory for Structural Bioinformatics (RCSB) and is designed to provide users access to information about macromolecules such as 3D coordinates, polymer sequences, ligand interaction determined by X-ray crystallography, NMR and electron microscopy techniques. Each entry is assigned a random PDB ID, a four-letter code, e.g. 3NXU, where 3 is an integer and NXU is alphanumeric. Every code is unique and intended to retrieve the structure from the database.[51] [50]

Before entries are published, they get into the internal core database for data processing, check for errors and follow the rules of the MMCIF dictionary[52], a tool that contains terms for crystallographic experiments and macromolecular structures.

Validation is then performed to verify that the quality of the atomic models matches the experimental data, including protein modifications, metabolic processes, stereochemistry, sequence comparison, and more. [53] [50]

Since the PDB data is now submitted globally, the worldwide PDB (wwPDB)[51] has been established to ensure consistent format and content. Structural biology, especially protein expression, purification, or structure determination, is a very fast-growing field with constantly improving technologies and data management systems and is becoming more and more important. Therefore, the PDB is also constantly updating its data and improving the data quality. Many tools like electron density maps, structure visualizations, a download area for data export for different files or individual filtering of search results facilitate the work with protein structures on desktops as well as on a laptop or mobile device. [51] [53]

3.6. Docking

3.6.1. Calculation of peptide fragments for docking

For docking all peptides from the peptide DrugBank were cleaved into fragments according to the following rules:

- the amide bonds were only cut if they were outside the ring to maintain the ring structure
- to remove too large or too small fragments only those with a molecular weight between 300 and 500 Da were kept

3.6.2. Preparation of protein structures for Docking

All CYP inhibitors retrieved from the ChEMBL and PubChem database and the peptide fragments were prepared with the Protein Preparation Wizard that is part of the Schrödinger Maestro.[54] The preparation includes the following steps:

1. preprocessing of the protein in which the structure is interpreted by assigning bond orders, use CCD database[55], adding hydrogens, creating zero-order bonds to metals, creating disulfide bonds and generating het states using pH 7.0
2. optimization by H-bond assignments and sample water orientations to calculate the most stable structure. For both a pH of 7 was chosen
3. water molecules beyond HET atoms, i.e. all those that are outside the protein and thus cannot form protein-ligand interactions, were removed.
4. restrained minimization with the force field: OPLS3

Problems such as different atom types, missing atoms and overlapping atoms or alternate positions can occur. For the selected crystal structures from the Protein Data Bank, some alternative positions were found, but they were all outside the binding pocket and thus did not pose a problem, as the binding pocket is rigid enough to compensate for this.

3.6.3. Generation of receptor grids for Docking

Maestro's Receptor Grid Generation tool was used to define the protein binding pocket. Position and size were determined by the co-crystallized ligand. Rotatable groups located

within the binding pocket and facing the ligand were selected to remain flexible and to assume different positions during docking. All other settings remain unchanged.

3.6.4. Preparation of the ligands for Docking

The final step before docking is the preparation of peptide fragments and inhibitors with the LigPrep tool of Schrödinger Maestro. Following settings were selected:

1. Generation of possible ionization states at pH 7 and deviations with +/- 0
2. Desalt is excluded
3. Creation of tautomers is not necessary, because all structures are already drawn correctly
4. Retain specified chiralities. For those where it is not defined whether the configuration is R or S, both are generated

All other settings remain unchanged including a maximum ligand size of 500 atoms and the inclusion of the OPLS4 force field.

3.6.5. Ligand docking with glide[56]

Subsequently, both the prepared peptide fragments and the CYP inhibitors were docked to the previously defined binding pocket. Since large data sets were processed here, only one pose per ligand was set as output. All other settings were already preset and were not changed.

Glide[56] (grid-based ligand docking with energetics) is a docking algorithm which, compared to the other most common docking programs such as GOLD[57], FlexX[58] and DOCK[59], not only provides faster results but also has a higher robustness in predicting binding mode. When docking to rigid receptor structures, scientists often face the problem that the ligands are too small for the protein cavity. Therefore, Glide van der Waals reduces radii to increase the space in the binding pocket and facilitate docking to the receptors. Using hierarchical filters, this approach allows for a comprehensive search and evaluation in the receptor's positional, orientation, and conformational space to find possible ligand sites in the active site. Multiple ligand conformations are generated, and those that aren't suitable for binding to the receptor are removed, limiting the search space significantly. The conformations that pass this prescreening are minimized by optimizing the torsional energy on an OPLS-AA potential grid. The best candidates with the lowest energy are then subjected to the so-called Monte

Carlo procedure, in which torsional minima are investigated. Finally, the final evaluation is done by a model energy function combining empirical and force field terms.

4. Results

4.1. Analysis of the relevance of physicochemical properties in CYP inhibition

Physicochemical properties have been shown to play an important role in the inhibition of CYP enzymes. Most CYPs are lipophilic by nature and therefore have a preference for lipophilic molecules, which is why lipophilicity, molecular weight, rotatable bonds, hydrogen donors, and acceptors were determined to be particularly relevant properties for inhibition. For both CYP inhibitors from Plonka's datasets and all approved peptide drug substances from the PepTherDia database, the relevant properties were calculated using a KNIME workflow.[29] [15]

After calculating the mean value, it can be seen in Table 1 that peptides and inhibitors differ significantly in their properties. Peptide drugs are much more hydrophilic, larger, and also have more hydrogen donors or acceptors, as well as rotatable bonds in contrast to the inhibitors. The next step included non-inhibitors, all approved drugs from the drug bank, and co-crystallized inhibitors from the PDB in addition to inhibitors in the analysis.

Table 1: Comparison of peptides and inhibitors in their average physicochemical properties

	logP	MW	rotable bonds	HBD	HBA	AmideBonds	HeteroAtoms	AromaticRings
inhibitors	3,75	383,45	4,96	1,14	4,72	0,61	6,68	2,82
peptides	-4,88	1701,76	44,75	22,47	24,14	14,42	42,64	3,07

4.1.1. Principal component analysis

To visualize coverage of the chemical space, we created a PCA for known inhibitors, non-inhibitors, peptide drugs, co-crystallized inhibitors from the PDB, and all approved drugs from DrugBank, based on the previously calculated properties. The principal component analysis is a method for visually representing high-dimensional data in a low-dimensional space to support interpretation. [60]

The principal component analysis (PCA) scatter plots in Figure 2 show that inhibitors, non-inhibitors and approved drugs are very similar in properties, while peptides differ significantly from inhibitors and non-inhibitors. In some areas, they cover the chemical space of the known inhibitors and non-inhibitors well, but especially in molecular weight, it is shown that some peptide drugs are significantly larger and therefore distribute differently in the chemical space.

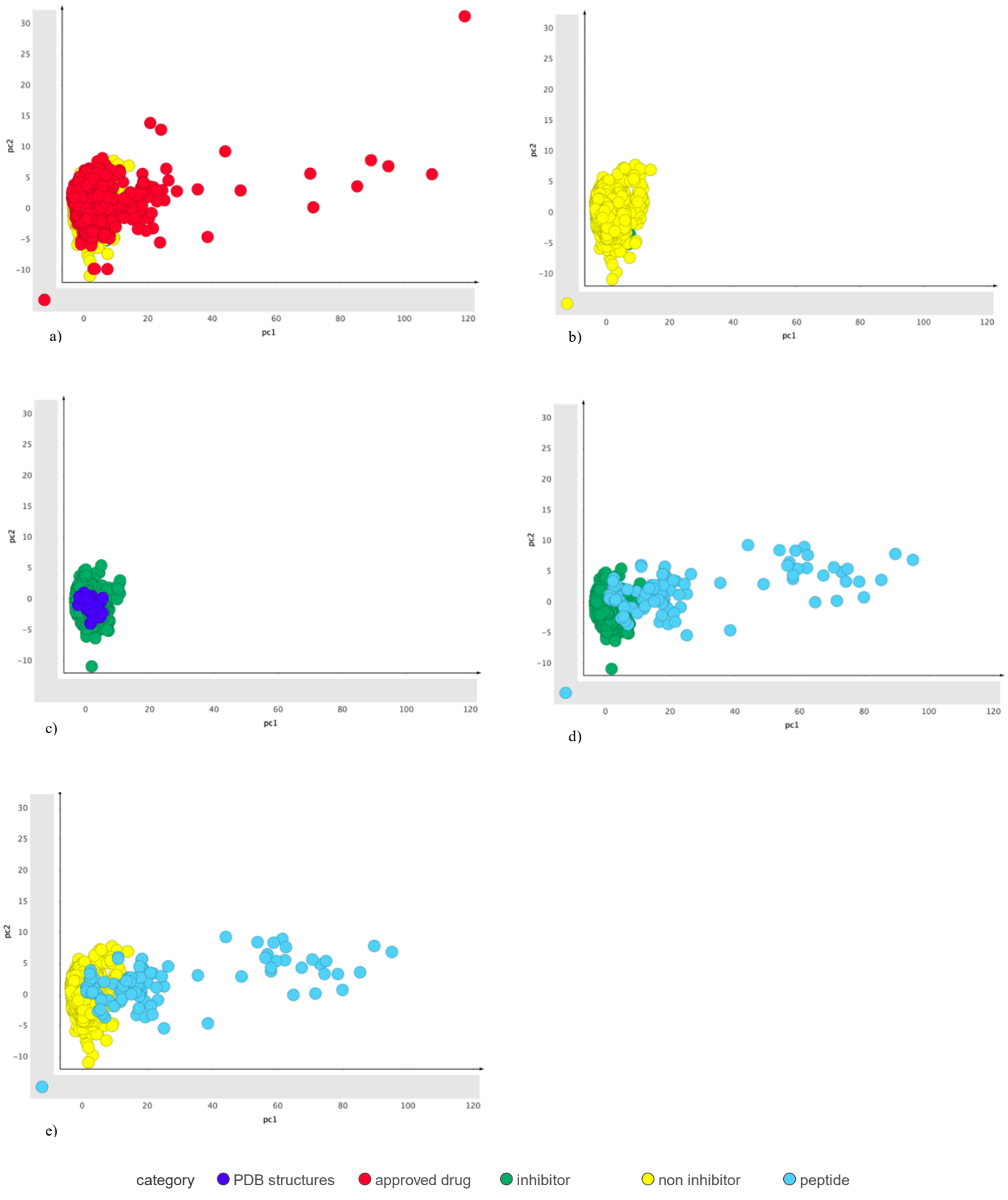


Figure 2: Principal component analysis (PCA): scatter plot comparing the chemical space of (a) inhibitors vs. non-inhibitors vs. approved drugs (b) inhibitors vs. non-inhibitors c) inhibitor vs. PDB structures (d) inhibitor vs. peptides (e) non-inhibitors vs. peptides

4.1.2. Loadings plot

In addition to the PCA, we created a loadings plot to better visualize the distribution in the chemical space. The percentage of explained variances is 42.21% for PC1 and 17.36% for PC2, which shows that the coverage of all structures in the chemical space is relatively low and they differ significantly in their properties. In particular, the approved drugs are not covered in almost any area compared to the other structures.

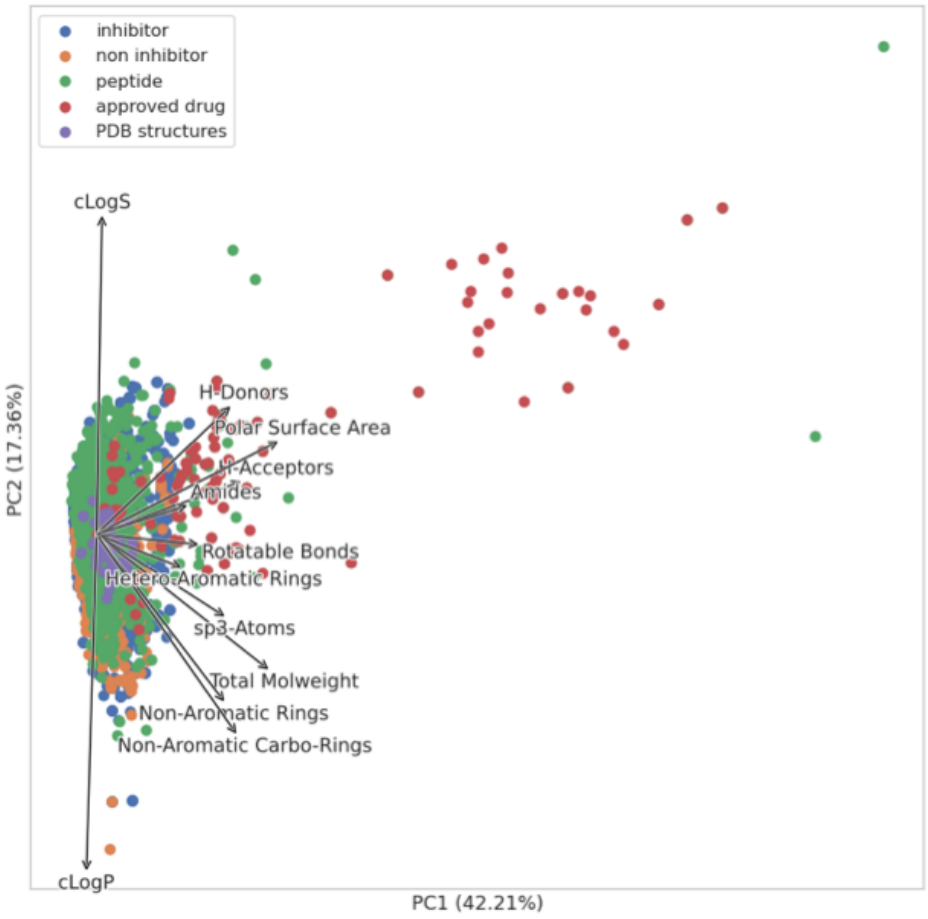


Figure 3: Loadings plot based on the created PCA

4.2. Binding patterns of known inhibitors to CYP enzymes

4.2.1. Cytochrome P450 active site

As already discovered, CYP enzymes catalyze numerous endogenous and exogenous substrates during phase I metabolism, with CYP1A2, 3A4, 2D6, 2C9, and 2C19 being the most important as they are responsible for over 80% of drug metabolism. Since inhibition of these enzymes is a problem and can lead to serious side effects, it is important to understand how inhibitors bind to CYPs. For this purpose, all available PDB structures acting as inhibitors were analyzed. Schrödinger Maestro was used to visualize which heterocycles or other fragments bind to the catalytic heme iron. The information of the crystal structures is derived from the Protein Data Bank (PDB).[51] Another very helpful tool was the web service ProteinsPlus [61] created by the Center of Bioinformatics - Universität Hamburg. The server provides many tools for the early phase of structure-based molecular modeling.[62]
For our purpose, we utilized the SIENA tool to search for alternative conformations and used the resulting pdb.zip file for analysis.

Table 2: All crystal structures of CYP inhibitors available in the PDB

Cytochrome	Ligand name or SMILES	available PDB structure
1A2		
	α-naphtoflavone	2HI4
3A4		
	erythromycine	2J0D
	ketoconazole	2V0M
	ritonavir	3NXU
	desthiazolylmethyloxycarbonyl ritonavir	3TJS
	bromoergocryptine	3UA1
desoxyritonavir analoga	CC(C)(C)OC(=O)NCCCCCC(=O)NCC1=CN=CC=C1	4D6Z
desoxyritonavir analoga	CC(C)(C)OC(=O)NC(CC1=CNC2=CC=CC=C21)CSCCC(=O)NCC3=CN=CC=C3	4D7D
desoxyritonavir analoga	CC(C)(C)OC(=O)NC(CC1=CC=CC=C1)CSCCC(=O)NCC2=CN=CC=C2	4D78
	oxazole–substituted desoxyritonavir	4I4G
	pyridine–substituted desoxyritonavir	4I4H
desoxyritonavir analoga	CC(C)C1=NC(=CS1)CN(C)C(=O)NC(C(C)C)C(=O)NCCCCNC(=O)OCC2=CN=CS2	4K9T
desoxyritonavir analoga	CC(C)C1=NC(=CS1)CN(C)C(=O)NC(C(C)C)C(=O)NC(C)CCC(C)NC(=O)OCC2=CN=CS2	4K9U
desoxyritonavir analoga	CCC(CCC(CC)NC(=O)OCC1=CN=CS1)NC(=O)C(CO)NC(=O)N(C)CC2=CSC(=N2)C(C)C	4K9V
desoxyritonavir analoga	CCCC(CCC(CC1=CC=CC=C1)NC(=O)C(C(C)C)NC(=O)N(C)CC2=CSC(=N2)C(C)C)NC(=O)OCC3=CN=CS3	4K9W
desoxyritonavir analoga	CCC(C)C(=O)NC(CCC(CC1=CC=CC=C1)NC(=O)OCC2=CN=CS2)CC3=CC=CC=C3	4K9X
desoxyritonavir analoga	CC1=C(C=CC(=C1)C2(C3=NCC(CN3C(=N2)N)(F)F)C4=CC(=C(C=C4)F)C5=CN=CC=C5)OC	4NY4
	cysteine depleted CYP3A4 bound to ritonavir	5VCE
	cysteine depleted CYP3A4 bound to bromoergocryptine	5VCG
ritonavir like inhibitor	CC(C)NC(CSCCNC(=O)OC(C)(C)C)C(=O)NCC1=CN=CC=C1	6BCZ
ritonavir like inhibitor	CC(C)(C)OC(=O)NCCSCC(C(=O)NCC1=CN=CC=C1)NC2CCCC2	6BD5
ritonavir like inhibitor	CC(C)(C)OC(=O)NCCSCC(C(=O)NCC1=CN=CC=C1)NC2=CC=CC=C2	6BD6
ritonavir like inhibitor	CC(C)NC(CSCC(CC1=CC=CC=C1)NC(=O)OC(C)(C)C)C(=O)NCC2=CN=CC=C2	6BD7
ritonavir like inhibitor	CC(C)(C)OC(=C)NC(CC1=CC=CC=C1)CSCC(C(=O)NCC2=CN=CC=C2)NC3CCCC3	6BD8
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC=C1)CSCC(C(=O)NCC2=CN=CC=C2)NC3=CC=CC=C3	6BDH
ritonavir like inhibitor	CC(C)NC(CSCC(CC1=CNC2=CC=CC=C21)NC(=O)OC(C)(C)C)C(=O)NCC3=CN=CC=C3	6BDI
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CNC2=CC=CC=C21)CSCC(C(=O)NCC3=CN=CC=C3)NC4CCCC4	6BDK
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CNC2=CC=CC=C21)CSCC(C(=O)NCC3=CN=CC=C3)NC4=CC=CC=C4	6BDM
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC=C1)CSC(CC2=CC=CC=C2)C(=O)NCC3=CN=CC=C3	6DA2
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC=C1)CSC(CC2=CC=CC=C2)C(=O)NCC3=CN=CC=C3	6DA3
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC=C1)CNC(CC2=CC=CC=C2)C(=O)NCC3=CN=CC=C3	6DA8
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC=C1)CSC(CC2=CC=CC=C2)C(=O)NCCC3=CN=CC=C3	6DAB
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC=C1)CSC(CC2=CC=CC=C2)C(=O)NCCC3=CN=CC=C3	6DAC
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC=C1)CSC(CC2=CC=CC=C2)C(=O)NCCC3=CN=CC=C3	6DAG
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC=C1)CSC(CC2=CC=CC=C2)C(=O)NCCC3=CN=CC=C3	6DAJ
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC=C1)CNC(CC2=CC=CC=C2)C(=O)NCCC3=CN=CC=C3	6DAL
	metyrapone	6MA6
	fluonazole	6MA7
	phenylmethansulfonylfluorid	6MA8
	mibefradil	6OO9
	azamulin	6OOA
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CNC2=CC=CC=C21)CSCC(C(=O)NCC3=CN=CC=C3)NC4=CC=CC=C4	6UNE
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CNC2=CC=CC=C21)CSC(CC3=CC=CC=C3)C(=O)NCCC4=CN=CC=C4	6UNI
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC2=CC=CC=C21)CSC(CC3=CC=CC=C3)C(=O)NCCC4=CN=CC=C4	6UNJ
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC2=CC=CC=C21)CSC(CC3=CC=CC=C3)C(=O)NCCC4=CN=CC=C4	6UNK
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC2=CC=CC=C21)CSC(CC3=CC=CC4=CC=CC=C43)C(=O)NCC5=CN=CC=C5	6UNL
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC=C1)CSC(CC2=CC=CC=C2)C(=O)NCCCC3=CN=CC=C3	7KVH
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC=C1)CSC(CC2=CC=CC=C2)C(=O)NCCCC3=CN=CC=C3	7KVI
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC=C1)CSC(CC2=CC=CC=C2)C(=O)NCCCC3=[C-]N=CC=C3	7KVJ
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC=C1)CSC(CC2=CC=CC=C2)C(=O)NCCCC3=CN=CC=C3	7KVK
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC=C1)CSCC(CC2=CC=CC=C2)C(=O)NCCCC3=CN=CC=C3	7KVM
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CNC2=CC=CC=C21)CSC(CC3=CC=CC=C3)C(=O)NCCCC4=CN=CC=C4	7KVN
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC2=CC=CC=C21)CSC(CC3=CC=CC=C3)C(=O)NCCCC4=CN=CC=C4	7KVP
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC2=CC=CC=C21)CSC(CC3=CC=CC=C3)C(=O)NCCCC4=CN=CC=C4	7KVQ
ritonavir like inhibitor	CC(C)(C)OC(=O)NC(CC1=CC=CC2=CC=CC=C21)CSC(CC3=CC=CC=C3)C(=O)NCCCC4=CN=CC=C4	7KVS
2C9		
	CC(C(=O)NC1=C(C=C(C=C1)S(=O)(=O)C2=CC(=CC=C2)Br)Cl)(C(F)(F)F)O	4NZ2
trifluoromethanesulfonamide	C1CC(CCC1NS(=O)(=O)C(F)(F)F)OC2=CC(=C(C=C2)C#N)Cl	5A5I
trifluoromethanesulfonamide	COC1=CC(=CC(=C1OC2=CC(=C(C=C2)C#N)Cl)OC)NS(=O)(=O)C(F)(F)F	5A5J
TCA analog	CCOC(=O)NC(=O)C1=C(SC=C1)NC(=O)C2=NC3=C(S2)C=CN=C3	5W0C
2C19		
benzbromarone analog	CC1=CC(=CC(=C1O)C)C(=O)C2=C(OC3=CC=CC=C32)C	4GQS
2D6		
	prinomastat	3QM4
	ajmalicine	4WNT
	quinidine	4WNU
	quinine	4WNV
BACE1 inhibitor	C1C2CSC(=NC2(COC1C3=CNN=C3)C4=C(C=C(C=C4)F)F)N	4XRZ

4.2.2. Cytochrome 1A2

CYP 1A2 is the most abundant enzyme of CYP P450 enzyme family 1, expressed in extrahepatic tissues like the skin, gastrointestinal tract, or epithelia of the lungs. It plays a major role in the metabolic clearance of many endogenous substances such as caffeine or melatonin but also drugs, including flutamide lidocaine or olanzapine. The ability to metabolize large basic planar compounds in particular the N-oxidation of heterocyclic aryl amines but also the oxidation of polycyclic aromatic hydrocarbons (PAHs) separated CYP1A2 from P450s of other families. The structure contains less than 40% amino acid identity in comparison with the other families and differs mostly in the length and structures of loop regions connecting secondary structure elements. [63 64]

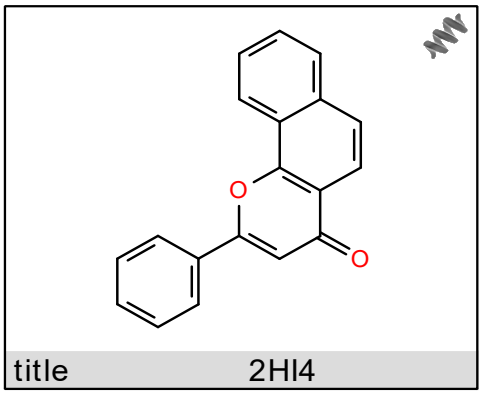


Figure 4: α- Napthoflavone

CYP1A2 with its potent competitive inhibitor α-Naphthoflavone was determined and published in the Protein Data Bank (PDB) under the code 2HI4.
α-Naphthoflavone (Figure 4) is a flat ligand that binds close to the iron atom within a narrow enclosed active site resulting in the displacement of water from the resting Fe^{III} state.
The benzo[h]chromen is located flat against Helix I and the 2- Phenyl substituent toward the catalytic heme iron with a distance of 4,7Å. It can be concluded that α-Naphthoflavone interacts in a single preferred orientation and many structural features contribute to the inhibitory effect.
On the one hand, this planar compound fits perfectly into the proximal binding site, which is shaped like a slit, leading to van der Waals interactions of nonpolar side chains. On the other hand, the hydrophobic effect and ***π***- ***π***- stacking play an important role. According to studies, CYP1A2 is quite capable of metabolizing ANF to ANF -5,6-diol and ANF-5,6-oxide

but due to the large distance of 4.7 Å to C4 carbon, the oxidation rate is too low for the energetically favorable direct attack.
These factors make ANF a potent competitive inhibitor of Cytochrome 1A2. [8 64 63]

4.2.3. Cytochrome 3A4

CYP3A4 is the most important enzyme for drug metabolism, which is highly expressed in the liver, and oxidizes more than 50% of used drugs. Whether and how quickly they are degraded influences their bioavailability and therapeutic effects. CYP3A4 metabolizes large compounds as does CYP 1A2, but there are some structural differences. In particular, the ability to occupy the active site with two or more substrates suggests very high flexibility and a large active site near the catalytic center with a hydrophobic core of phenylalanine residues above. However, the short F and G helix is also characteristic of 3A4 from the other CYP families. In addition, the enzyme can take different shapes and interact with the binding site in many different ways, allowing donors, acceptors, anions, cations, aromatics, and hydrophobic interactions. [65 8]
We can see CYP3A4 is a very flexible enzyme and offers a variety of different recognition features that are important for the metabolism of many drugs, but also play a major role in inhibition. As mentioned above, physicochemical properties play a major role in the inhibition of this enzyme. In addition to a large molecular weight or high lipophilicity, most inhibitors also have one or more aromatic rings as well as hydrogen bond acceptors. A large number of inhibitors have already been identified and deposited in the PDB. In the following section, some of these inhibitors will be examined in more detail.

4.2.3.1. Ritonavir

Ritonavir (ligand of PDB 3NXU) is an HIV protease inhibitor and the most potent CYP3A4 inhibitor known. By irreversibly binding to the heme and active site amino acid residues, it transforms into a reactive intermediate after oxidation and effectively inactivates CYP3A4.[66]
According to the reports, there is no agreement on whether ritonavir functions as a mechanism-based inhibitor or as a competitive inhibitor. [67]
It acts as a type II heme ligand, irreversibly displaces water from the active site with its thiazole group and binds directly to the heme. One phenyl ring lies parallel to the heme and is responsible for the van-der Waals contacts, whereas the other phenyl ring lies in the hydrophobic pocket. This binding results in a reduced redox potential of the CYP450

reductase and is independent of the protein redox state and type I ligands, which can easily be displaced from the active site. The inhibitor is very similar to ketoconazole, but, ketoconazole binds more weakly and engages in fewer hydrophobic interactions than ritonavir, which doesn't fit as well in the active site pocket.[66] Ritonavir is an example where this inhibitory effect can also have a positive influence. Drugs that are also metabolized by 3A4, remain in the body longer because conversion to their water-soluble and more easily excreted metabolites is delayed. This mechanism is used, for example, in the treatment of HIV or hepatitis C.[68]

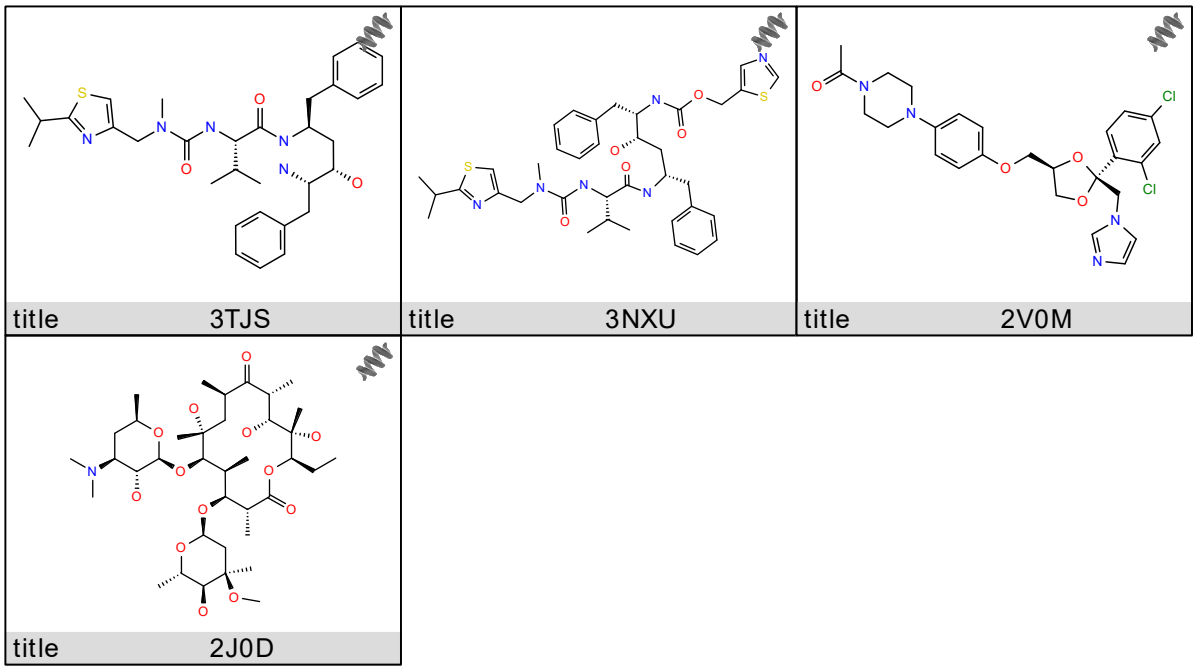


Figure 5: Structures of the potent CYP3A4 inhibitors: desthiazolylmethyloxycarbonyl ritonavir (3TJS), ritonavir (3NXU), ketoconazole (2V0M) and erythromycin (2J0D)

4.2.3.2. Ketoconazole

Ketoconazole (PDB 2V0M) is a well-known type II inhibitor and binds directly at the heme iron at 2,06 Å with its imidazole N. Following its binding, conformational changes occur in structures defining the active site, but also in those located further away. Most of them take place in the region of F and G helices and intervention loops such as a broken hydrophobic cluster, a shift of the loop, or the changed position of Arg212, which is located without ligand in the active site and the ketoconazole complex on the surface of the protein. The keto group

of ketoconazole is located in a polar pocket which leads to an additional stabilization of the inhibitor complex by ***π***- ***π***-- stacking.[69]

4.2.3.3. Metyrapone, Fluconazole

Both compounds are small type II inhibitors with medium strength directly to the heme iron, whereas in the metyrapone (6MA6) the pyridine group binds to the active site at a distance of 2,31 Å.[70] [71] Fluconazole (6MA7) is a slightly weaker ligand than metyrapone with a Fe-N distance of 2,20 Å to the triazole nitrogen and although it is larger and chemically more complex, no conformational change occurs. However, the functional groups cannot arrange themselves as flexibly. Nevertheless, the two inhibitors are very similar in their binding pattern, and in both cases water- and protein-mediated interactions lead to the stabilization of the complex. [70]

4.2.3.4. Pyridine – substituted Desoxyritonavir

Pyridine – substituted Desoxyritonavir (PDB 4I4H) is a ritonavir analog where the thiazole has been replaced by a pyridine. The Ligand is a type II inhibitor and binds with the nitrogen atom directly to the heme iron. The modified desoxyritonavir is similarly placed in the binding pocket as ritonavir, with small differences. Helix I is displaced by the phenyl-1 which is nearest to the heme and enclosed in a hydrophobic pocket. Unlike the corresponding group of ritonavir, the phenyl -2 is not located in the cavity above the heme but parallel to the pyridine ring and the guanidinium group of Arg105 which contributes to the stabilization of the inhibitor-bound complex. Thus, it can be seen that the Fe-N (2,14 Å) bond is a little shorter than with ritonavir (2,29 Å). Furthermore, Ser119 is a key residue in the interaction with ritonavir-like inhibitors by mediating H- binding. Since ritonavir is unable to align in the active site, this stabilizing binding to Ser119 is absent. This and all of the previously mentioned effects make pyridine-substituted desxoyritonavir a more potent inhibitor than ritonavir.[72]

4.2.3.5. Ritonavir analoga

Since ritonavir remains the most potent inhibitory drug, numerous ritonavir analogs have been prepared to map the active site. They have been variously modified at the R1 and R2 phenyl residues by replacing pyridines, indoles, benzimidazole, oxazoles, or a cysteine-based or thioether backbone.

In general, certain properties of ritonavir analogs must be present for them to be potent inhibitors of CYP3A4:

- a nitrogen-containing group with the capacity to interact with heme, like pyridine[73]
- a flexible backbone[73]
- hydrophobic and aromatic side residues (R1, R2) that can relate with the I-helix because the greater the volume, hydrophobicity and aromaticity of R1 the stronger the inhibition. Whereas the R1 residue mainly determines the binding affinity, the R2 residue is responsible for the inhibitory effect through interactions with heme[73]
- tight binding with Ser119 through an H donor or acceptor[74]
- the distance between the functional groups, which influences the binding affinity and inhibitory potency[68]
- strong binding to the heme iron[67]

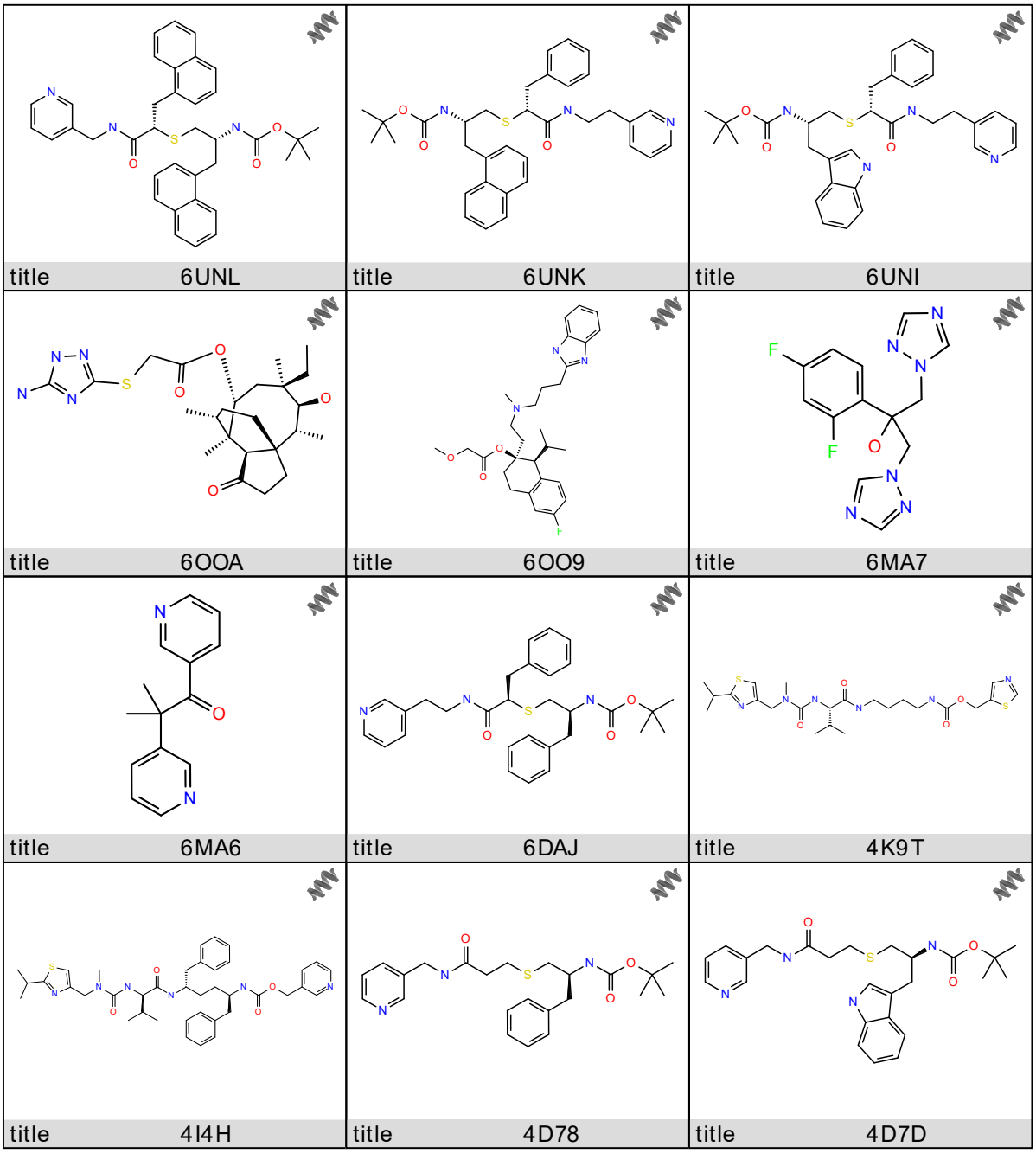


Figure 6: Different ritonavir analogs, deposited in the PDB: all of them contain a strong basic nitrogen atom, several hydrophobic and aromatic residues, tight binding to Ser119, a flexible backbone and the ability to form a strong bond with the heme iron

Pyridine at R1 shows the greatest inhibitory effect of all functional groups. Mainly because of the favorable connections for heme binding and only when it is part of a large ligand. When pyridines are present only alone or as a short-chain derivative, binding and inhibition are very weak.[75]

Indole, on the other hand, is less suitable because it does not allow H-bonding with Ser119 [67] Similarly, bulky side chains decrease matching and interactions with the active site, and

compounds with a hydrophobic naphthalene ring appear to be stronger inhibitors than those containing phenyl. Of all the ritonavir-like molecules synthesized, the 6UNK structure was the most potent inhibitor. It is structurally less complex and showed even stronger inhibition than ritonavir. [76] All analogs are type II ligands with strong heme binding to a basic nitrogen, with backbone length and composition having the greatest influence. CYP3A4 binds more tightly to elongated compounds and prefers molecules in R/S configuration at the R2 and S-configuration in R1.[77] [68]

4.2.3.6. Mibefradil

Mibefradil (6OOA) is a benzimidazole substituted tetralin that was used as a long-acting calcium channel blocker to treat hypertension but has since been withdrawn from the market due to its high inhibitory activity. As a type I ligand, it destroys heme iron by forming an alkyl radical that forms a slowly dissociable complex that prevents substrate molecules from approaching the active site.[78]

4.2.3.7. Azamulin

This substrate (6OOA), which binds closely to heme with its pleuromutilin group, failed already in phase I clinical studies, due to its poor bioavailability. Through several van der Waals interactions and two hydrogen bonds, azamulin was identified as a selective competitive mechanism-based type II inhibitor with high hydrophobicity and low dissociability. [78]

As shown by the two drugs mibefradil and azamulin, both of which are substrates, some compounds can inhibit CYP3A4 by forming a reactive metabolite and are referred to as suicide substrates. [78]

4.2.4. Cytochrome 2D6

CYP2D6 is the second most important enzyme next to CYP3A4 for the metabolism of more than 20% of drugs. However, unlike CYP3A4, which has a very flexible active site and can recognize a wide variety of substrates, CYP2D6 is mainly restricted to those containing a basic N and a planar aromatic ring. A special feature is its ability to interact with the amine nitrogen charge and the anionic chains Glu216 and Asp301, which are in the active site.[79] [80] CYP 2D6 is particularly susceptible to polymorphisms caused by genetic mutation and is observed mainly in the Caucasian population. Affected individuals are then no longer able to completely degrade certain drugs such as debrisoquine, resulting in toxic concentrations. [81] The structure of cytochrome 2D6 is similar to that of 2C9 and shows a typical folding pattern known from other CYP P450 families. As in CYP2C9, the heme iron in 2D6 is connected to the side chains by hydrogen bonds. Because most substrates and inhibitors are lipophilic and contain an aromatic ring, this **π**- **π**-interactions occur with residues Phe120 or Phe483.[81] The surface of the protein is polar and occupied by nonpolar residues. [82] The open cavity is shaped like a right foot or boot with the heel above the heme and the calf protruding from the cavity below the helix F. As observed in other CYPs 2 families, there is also a twist between helix F and helix F' in CYP2D6 and a very short G helix. [82]

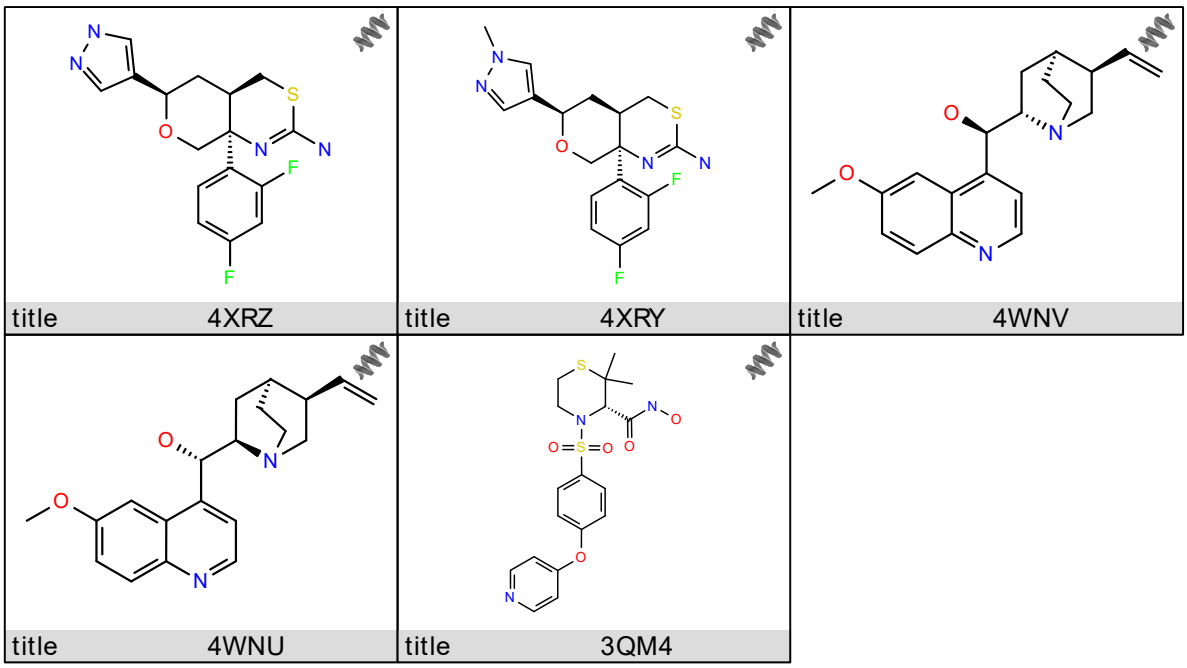


Figure 7: Overview of strong CYP2D6 inhibitors

4.2.4.1. Prinomastat

Prinomastat (3QM4), a type II ligand matrix metalloprotease inhibitor displaces water and attaches directly to the heme iron at a distance of 2,21 Å with its pyridyl nitrogen. It is a neutral molecule in which both the sulfonamide and pyridine groups are not protonated and the hydroxamic acid is also not ionized. After binding to the protein, the structure of CYP changes. This includes a change from an open to a closed cavity and a shift in the F- and G-helix as well as the placement of the large Prinomastat in the binding pocket.[82]

4.2.4.2. Quinidine and Quinine

Quinidine (4WNU) is a potent competitive Inhibitor of CYP2D6. [83] What is unique about this compound is that it exhibits numerous properties that are very characteristic of substrates. These include a type I binding pattern, basic nitrogen, a flat hydrophobic region, a negative electron potential, and hydrogen bonding. [83 84]
In the active site, quinidine fits very well into the binding pocket, but the nitrogen is too far away from the heme for catalytic conversion and displacement of water from the sixth coordination site.[83] The quinolone ring lies below the helix B', G, and I, and a hydrogen bond is formed between the hydroxyl group of quinidine and the negatively charged Glu216, which plays a very important role in substrate recognition of 2D6 with basic nitrogen's.[83] The interactions between the aromatic ring and the protein´s Phe120 or the Phe483 of the protein, which are rather unfavorable for the substrates, also contribute to quinidine acting as an inhibitor.[83]

Quinine (4WNV), the stereoisomer of quinidine, is a much weaker inhibitor.[84]
The bicyclic ring is distal to the active site in quinidine, proximal in quinine and is unable to form hydrogen bonds to the polar residues.[79] This shows that stereochemistry, different location in the binding pocket, and polar interactions have a major impact on binding and inhibition, making quinidine two orders of magnitude more potent CYP2D6 inhibitor. [84 79]

4.2.5. Cytochrome 2C9

Another important enzyme class of CYP450s with a high polymorphic potential is CYP2C9. It is a 2-domain protein with a large cavity and binds mainly to hydrophobic small acidic molecules via basic residues in the active site.[85] As already observed in CYP 3A4, evidence suggests that CYP2C9 is also able to bind several ligands.[85] Warfarin (1OG5) for example is a very well-studied substrate that is hydroxylated to its inactive metabolite and is located at about 10A from the active site. When ligands bind in this binding site, they are too far away for conversion, but through a two-step conformational change, resulting from electron transfer, CYP2C9 can bring the ligand closer.[85] It is possible that warfarin is a multiple binding site, activating CYP2C9 by an allosteric mechanism and making room for other substrates. However, inhibitors may also occupy this binding site and competitively inhibit the enzyme because they cannot get closer to the heme.[85]

Arg108 and Asn204 have been identified as key residues that contribute significantly to the inhibitory effect. Hydrogen bonding between the sulfonamide group of progesterone antagonists (5A5I and 5A5J) and the polar residues increase the negative electrostatic potential and stabilize the complex. [86]

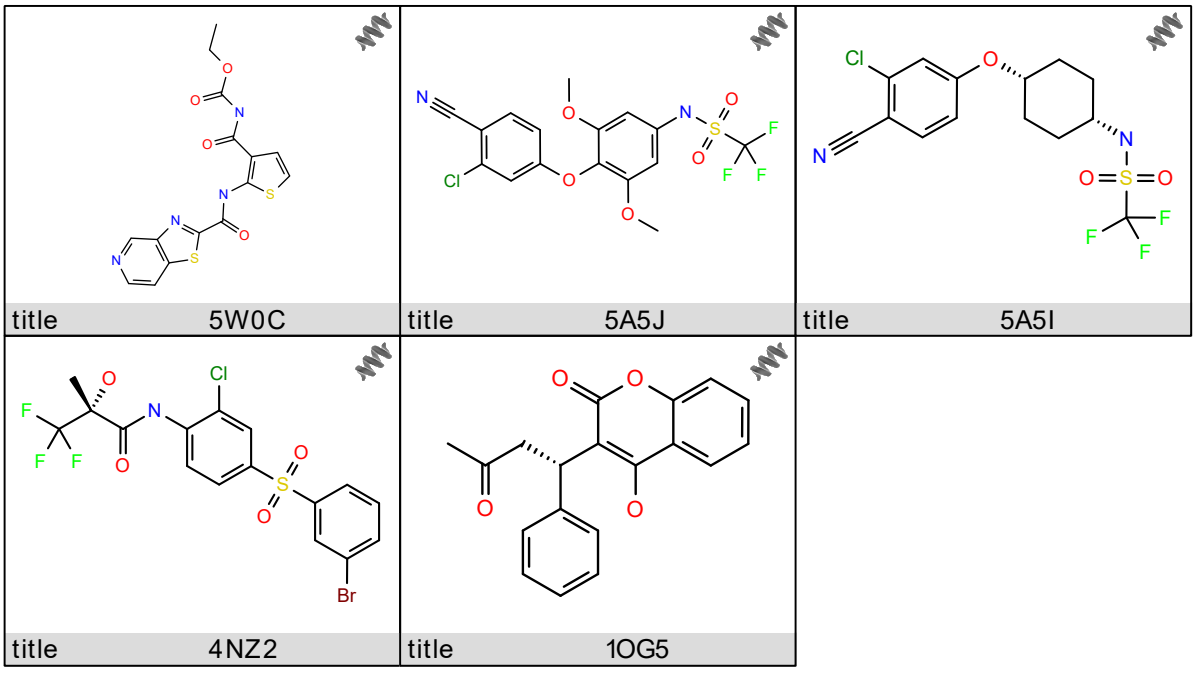


Figure 8: TCA analog (5W0C), progesterone antagonists (5A5I and 5A5J), a potent CYP2C9 inhibitor (4NZ2) that is parallel to the active site and make indirect interaction with the heme iron via two water molecules [87] and the substrate warfarin (1OG5)

The structure of 5WOC shows how diverse the active site of CYP2C9 actually is in comparison to the assumption. It is a TCA analog, an inhibitor of DprE1. DprE1 (decaprenylphosphoryl-b-d-ribofuranose 2-oxidase) is an enzyme in the cell wall of mycobacterium tuberculosis and a target of tuberculosis therapy. TCA007 (5WOC) inhibits cell wall synthesis and has also been shown to strongly inhibit CYP2C9, but not the other isoenzymes. The thiophene ring binds reversibly to Arg108 via hydrophobic interactions and a hydrogen bond, leading to the inhibitory effect.[87]

4.2.6. Cytochrome 2C19

The last group of major enzymes involved in the metabolism of numerous drugs is CYP2C19. 4GQS is the only structure with CYP2C19 that has been crystallized and deposited in the PDB. It's a less acidic benzbromarone analog, where two ethyl groups of the benzofuran ring have been replaced by a methyl group.[88] Unlike other members of enzyme family 2, 2C19 prefers less flexible compounds with low acidity that are neutral at physiological pH as well as phenolic and at least one hydrophobic substitute.[89] However, the structure of the substrate-binding cavity and active site is very similar to 2C8 and 2C9. The preference for neutral over anionic inhibitors is explained by the fact that the key residue for anionic inhibitors, Arg108 does not undergo a conformational change at the active site as in 2C9 and thus does not interact with the inhibitor since it is located on the outside of the B-C loop.[89] The inhibitor lies above the heme at a distance of about 5 Å. There is no clear preference about the site of metabolism but the methyl group of the benzofuran ring is closest to the heme in this binding position.[89] Hydroxylation of the phenol, hydrophobic and aromatic interactions appear to play an important role in inhibition.[89]

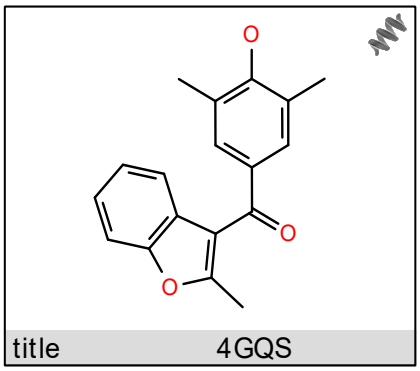


Figure 9: Inhibitor structure of CYP2C19

In summary, most CYP inhibitors contain a heterocyclic nitrogen, particularly pyridine, pyrimidine, imidazole, oxazole, pyrazine, quinoline, isoquinoline, and thiazole, which bind strongly to the heme iron. In some cases, structures containing thiophene, quinoline-cycles, or benzofuran may also exert inhibitory activity. As we have seen, CYP enzymes have very high flexibility with sometimes large active sites and the ability to bind multiple ligands. This feature, but also the fact that inhibitors can bind either close to the heme iron or further away, makes the prediction of inhibition very difficult.

Table 3: Summary of all important CYP isoenzymes and inhibitor properties with their key interactions

	enzyme properties	**inhibitor properties**	**important interactions**
CYP1A2	□ narrow binding site □ planar □ a line of multiple aromatics AS side chains □ especially antipsychotics and antibiotics	□ numerous aromatic moieties: heterocycles, hydrocarbons □ lipophilic □ basic □ planar, flat □ halogens □ at least one hydrophobic feature	□ key amino acid for ligand interactions: phe226 □ pi-pi stacking □ interactions of secondary amines and aromatic residues with heme iron □ hydrophobic effect
CYP3A4	□ large and open binding pocket □ high flexibility □ active site can be occupied by more ligands □ hydrophobic core of phenylalanine residues □ large molecular weight □ high liphophilic	□ hydrogen bond acceptors □ at least one aromatic ring □ various aromatic moieties: heterocycles, amines □ large ligands □ strong nitrogen to interact with heme	□ pi pi stacking interactions with phenylalanine □ key residue: Ser119 by mediating H-bonds
CYP2D6	□ small volume □ highly polymorphic □ acidic active site with phenylalanine □ flexible active site	□ basic nitrogen □ various aromatic moieties: hetereocycles, amines □ lipophilic flat, planar □ small ligands □ hydrogen bond acceptors	□ interactions of carboxylic residues with amine groups □ key residues: acidic Asp301, Glu216 and aromatic Phe120, Phe483
CYP2C9	□ large cavity □ especially non-steroidal anti-inflammatory drugs (NSAIDs) and warfarin □ highly polymorphic	□ aromatic moieties: heterocylces, nitrogens, amines □ lipophilic □ hydrogen bond acceptors □ small to medium size molecules	□ key residues: Arg108 and Asn204 via hydrophobic interactions
CYP2C19	□ less flexible □ similar to CYP2C9	□ at least 2 hydrogen bond acceptors or donor regions □ aromatic moieties: heterocycles, nitrogens □ small ligands □ less liphophilic □ less flexible	□ hydrophobic and aromatic interactions

4.3. Docking of peptide fragments and known inhibitors

Docking is a structure-based computer-aided method for calculating the best position of a ligand in the binding pocket and their interactions based on their 3D structural information. The docking score is used to evaluate binding affinity. [90]

For docking simulation, the CYP inhibition datasets and peptide fragments were docked to each protein structure available in the PDB using Glide from Schrödinger Maestro.[54] The docking scores obtained were used to generate the ROC curves with KNIME.[48] In addition, the corresponding AUC values based on the ROC curves were calculated. The ROC curves show the rate of true positives on the x-axis as a function of false positives on the y-axis. The thick black line represents the inhibitors, and the blue line represents the calculated peptide fragments. With this study, we wanted to find out whether it is likely for peptides to be involved in ligand binding and whether they can bind to the pocket as effectively as inhibitors.

In addition, the AUC values were summarized in Table 4. The area under the receiver operating characteristic curve (AUC) is a measure of the accuracy of my test. In our case, the values indicate how likely it is that the docked peptide fragments can also bind into the binding pocket and thus act as inhibitors. Values between 0 and 1 are calculated, where 0 means that the peptides show no similarities to the inhibitors and 1 means that they match the inhibitors.

In the following section, the individual ROC curves for each CYP isoenzyme are presented:

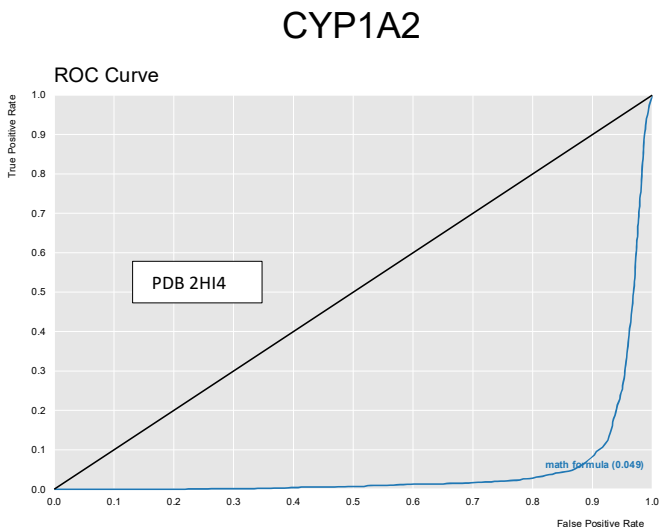
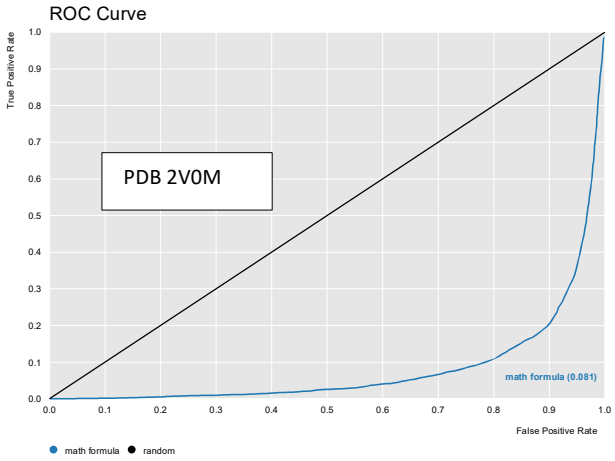
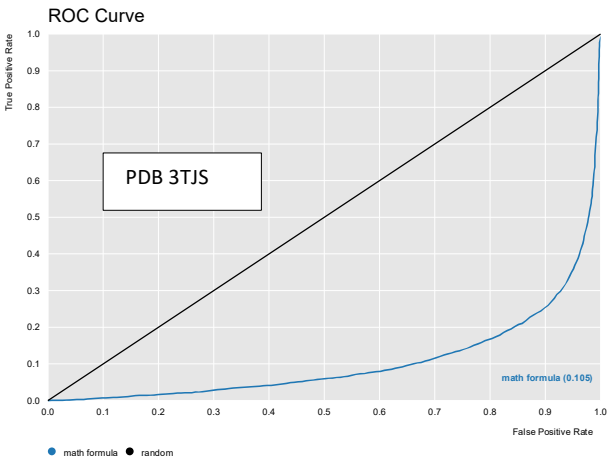
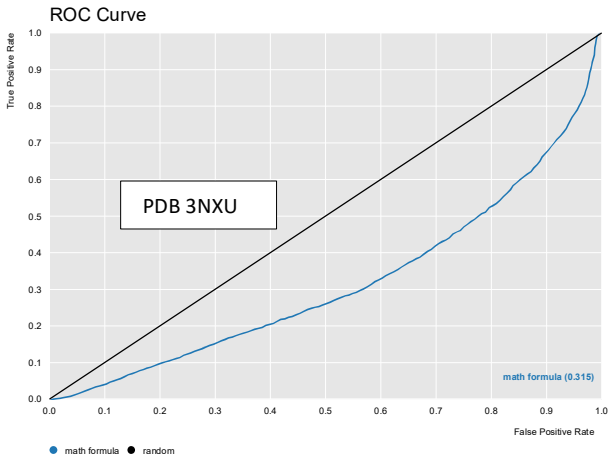


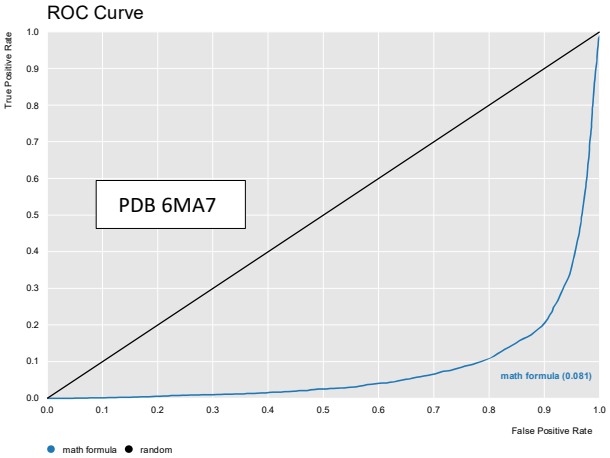
Figure 10: ROC curve of peptides and known inhibitors docked into the binding pocket of the 2HI4 ligand

CYP3A4









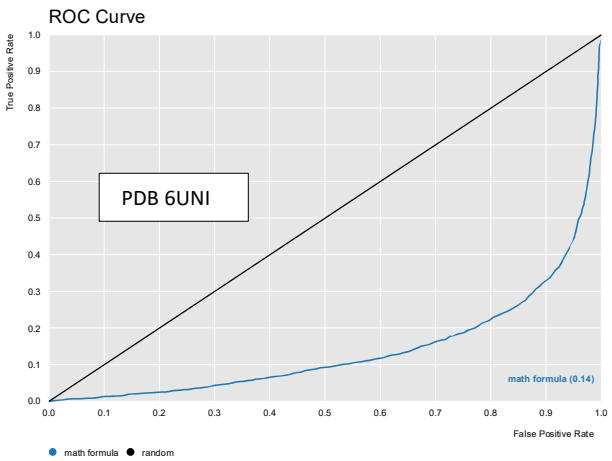


Figure 11: ROC curves of peptides and known inhibitors docked into the binding pockets of the 2V0M, 3TJS, 3NXU, 6MA7 and 6UNI ligand

CYP2D6

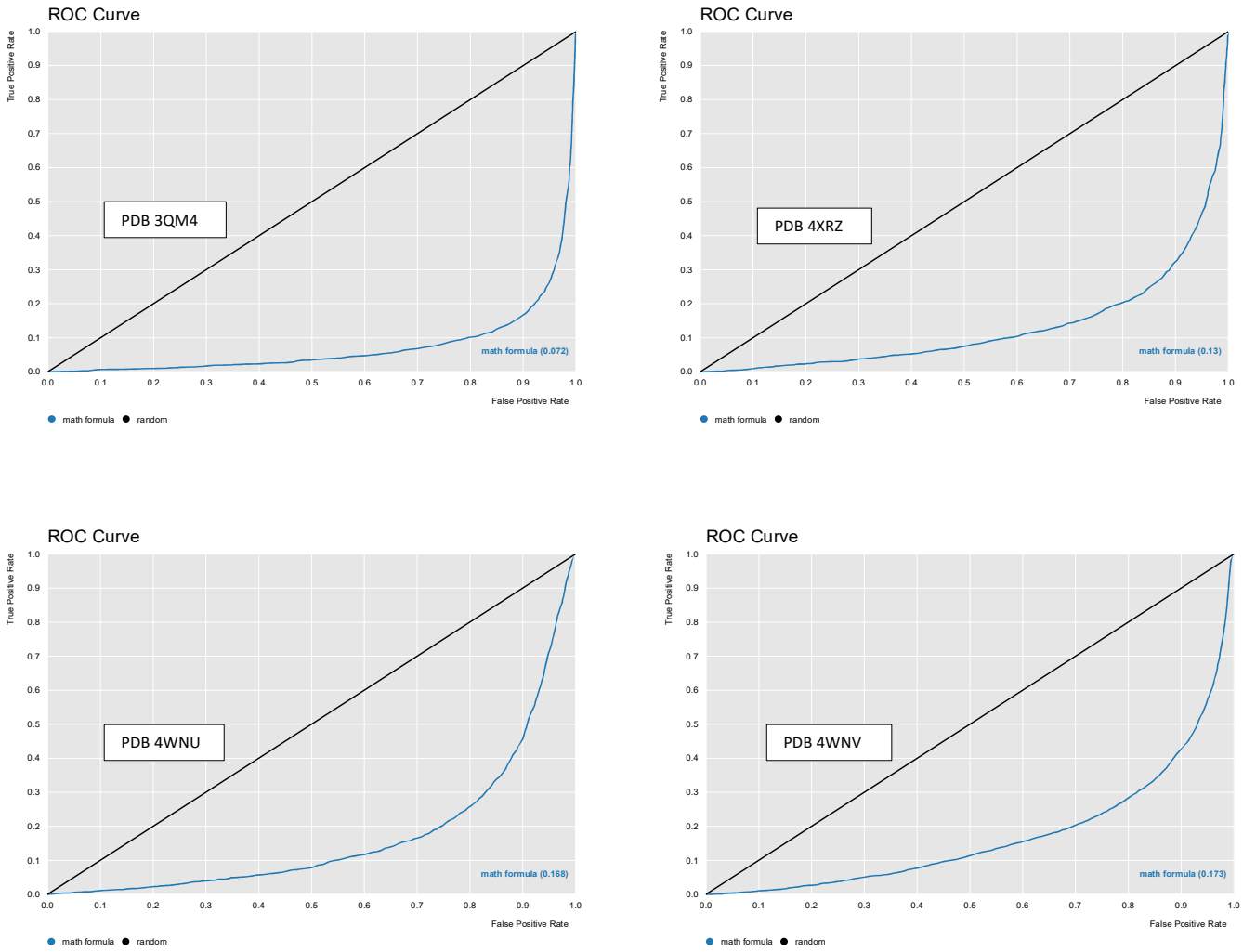


Figure 12: ROC curves of peptides and known inhibitors docked into the binding pockets of the 3QM4, 4XRZ, 4WNU and 4WNV ligands

CYP2C19

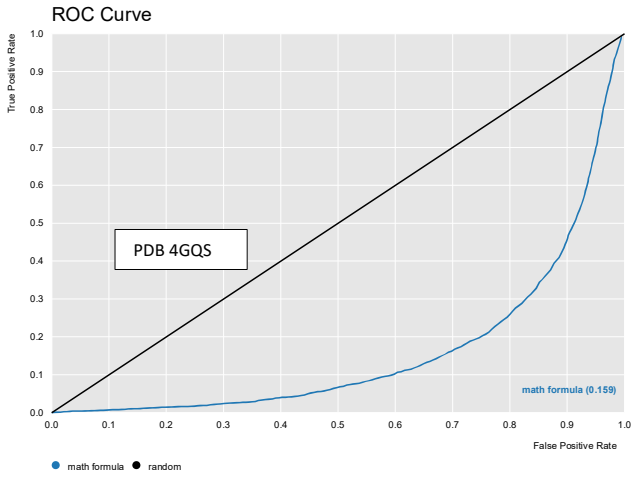


Figure 13: ROC curve of peptides and known inhibitors docked into the binding pocket of the 4GQS ligand

We can see that docking assigned very low ranks to the vast majority of peptides in the hit list. Cytochrome 1A2, 2D6, 2C19, and 3A4 show similar ROC curves which means that docking predicts most of the peptides are less likely to bind to the protein than the known inhibitors. AUC values from 0.049 to 0.173 show that the probability that these fragments can act in the same way as strong CYP inhibitors is very low.

Only the ligand 3NXU shows a slight approximation with a value of 0.315.

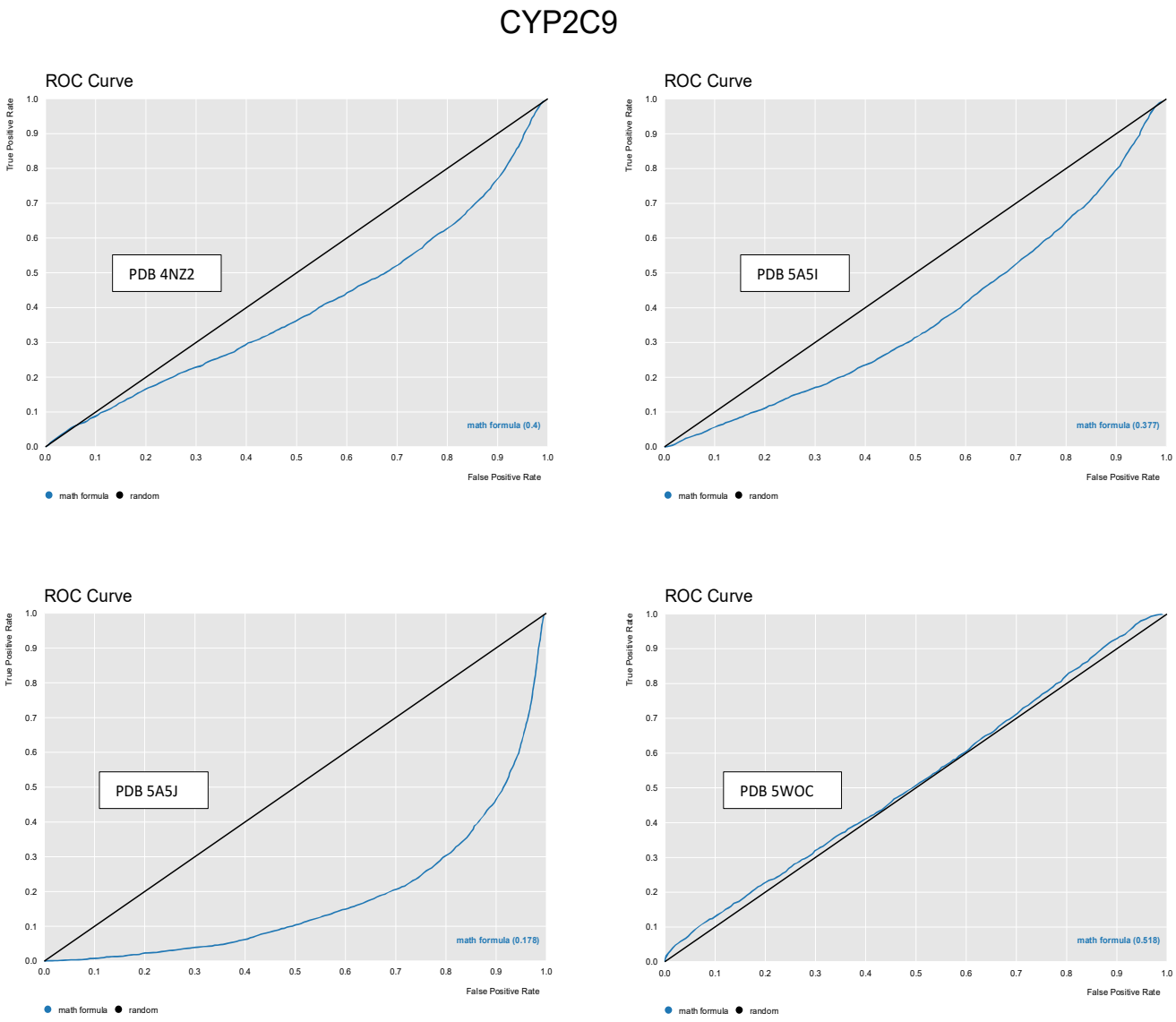


Figure 14: ROC curves of peptides and known inhibitors docked into the binding pockets of the 4NZ2, 5A5I, 5A5J and 5WOC ligands

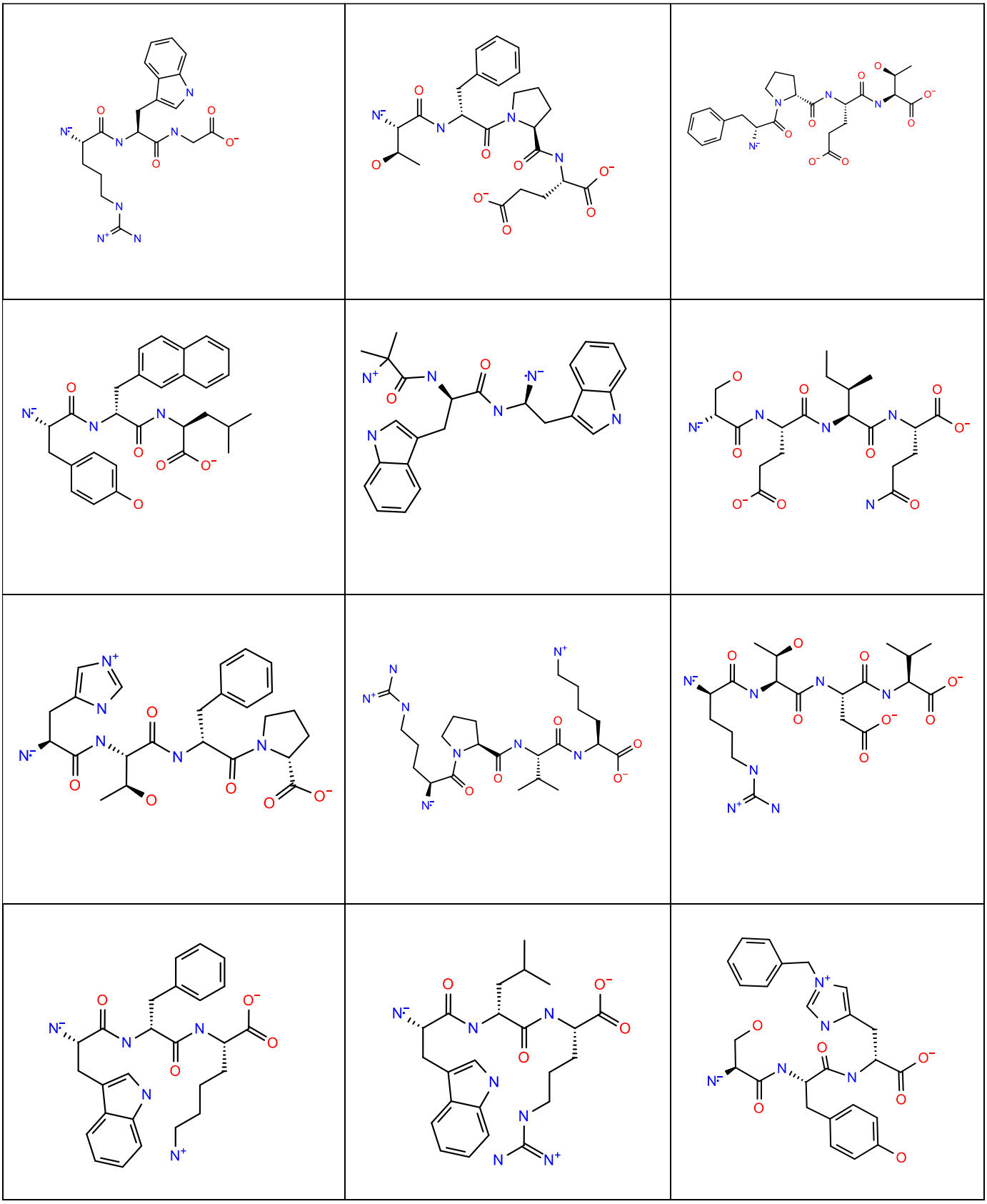
In complete contrast to the others, the results for CYP2C9 are unexpected. The curves indicate clear approximations here, and the peptides almost cover the inhibitors in the hit list of 5WOC. With an AUC of 0.518, the fragments have a high probability of binding to the protein similar to known inhibitors. Here it is very interesting to take a closer look at the peptide fragments that seem to have such related properties as the inhibitors.

Table 4:AUC values based on the received docking scores

Cytochrome	AUC
CYP1A2	
2HI4	0.049
CYP3A4	
2V0M	0.081
3TJS	0.105
3NXU	0.315
6MA7	0.081
6UNI	0.140
CYP2C9	
4NZ2	0.400
5A5I	0.377
5A5J	0.178
5W0C	0.518
CYP2C19	
4GQS	0.159
CYP2D6	
3QM4	0.072
4XRZ	0.130
4WNU	0.168
4WNV	0.173

4.3.1. Commonly occurring peptide fragments

Next, we consider the peptide fragments that have the highest docking scores and are the most abundant. In Figure 15we see that most of these fragments are roughly the same size. As we saw above with the common inhibitors, many basic nitrogen atoms and functional groups such as guanidine, primary and secondary amines seem to be involved. Most importantly, basic heterocycles such as pyrrolidine, indole, and imidazole are also very present.
In summary, the docking results show that, by and large, peptides appear to have little involvement in inhibition. Most peptide fragments show significant deviations from known inhibitors and rank very low in the hit list according to their docking scores. Cytochrome 2C9, on the other hand, stands out since the peptides converge strongly and, in some cases, overlap. A closer look at these fragments reveals relatively large structures and properties similar to those of known inhibitors, highlighting the importance of basic nitrogen compounds and aromatic rings in inhibition.



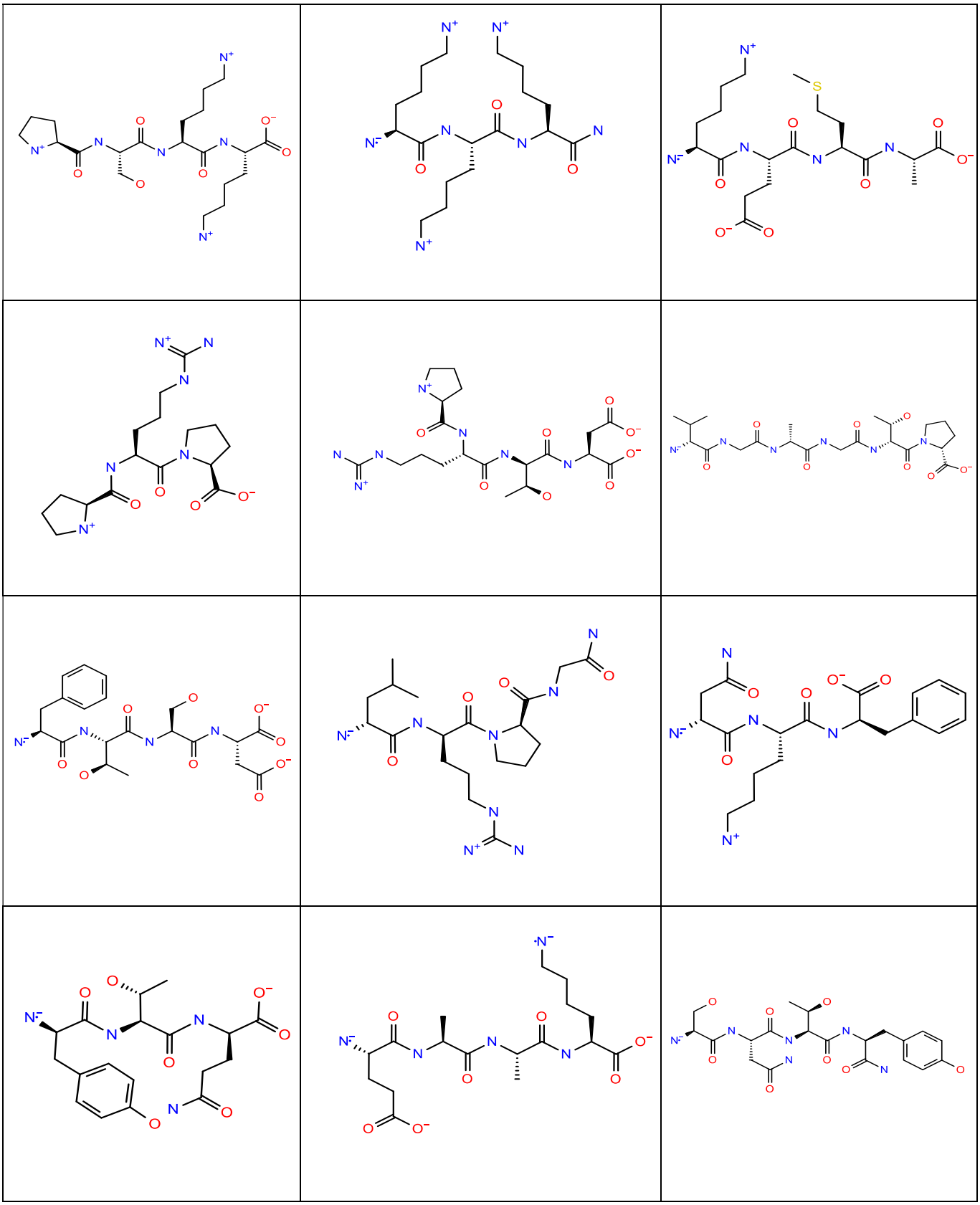


Figure 15:Peptide fragments which are most frequently observed in protein binding and among the top ranked docking scores: all of them are very large, contain various basic nitrogen and aromatic moieties

5. Discussion

Cytochrome P450 enzymes are among the most important metabolic enzymes responsible for numerous catalytic reactions in the body. The conversion of lipophilic molecules into hydrophilic compounds makes further metabolism and excretion easier, which is why inhibition should always be avoided. As people get older, their risk of multimorbidity grows, more and more medications are taken and the danger for CYP-related drug interactions increases due to these polymedications. For this reason, it is particularly important to screen potential candidates for possible CYP interactions in the early stages of drug development. Peptide therapeutics are of particular importance in diabetes treatment and have several advantages due to their low toxicity and good efficacy. However, because of their low metabolic stability and rapid degradation by proteases, scientists are faced with major challenges. Nevertheless, the interest in therapeutic peptide research has increased strongly in recent years and highlights the importance of understanding peptide metabolism and excluding possible CYP inhibition, resulting in serious adverse effects.

We used three different approaches to try to answer this open question of possible CYP inhibition by peptides. The PCA showed that the peptides overlap in their physicochemical properties in some areas with known inhibitors or non-inhibitors, but significant differences in molecular weight or logP value are evident. If we take a closer look at the crystal structure of known inhibitors, we can identify some common features. The presence of a basic nitrogen, aromatic components, lipophilic ligands, and hydrogen bonds as acceptors and donors are features that can be found in most inhibitor molecules. Hydrophobic and aromatic interactions, as well as π- π-stacking or interactions with key residues such as Ser119, Asp301, or Glu216, additionally contribute to a strong inhibitory effect. Because most peptides have a relatively large molecular weight, smaller fragments of all already approved peptide drugs were calculated and docking simulations were performed. For comparison, the inhibitor data set was again used and docked to five different proteins from the PDB. The generated ROC curves presented no major surprises, where most peptides are ranked very late in the hit list. Only in the case of CYP2C9 is the majority of ligands listed highly, suggesting they might be able to bind to the protein in the same way as inhibitors.

In conclusion, the studies in this work show that most peptides, when present as large molecules, are unlikely to interfere with CYP metabolism and exert an inhibitory effect. The situation could be different if they are degraded to smaller metabolites in the body. There is evidence that smaller nitrogen-containing fragments with aromatic rings have features that

are also found in known inhibitors and could therefore act similarly. Nevertheless, this work did not demonstrate that peptides are capable of inhibiting cytochrome P450 enzymes, and further studies are needed.

6. References

(1) *Organic Chemistry Concepts and Applications for Medicinal Chemistry*; Elsevier, 2014. https://doi.org/10.1016/C2013-0-18544-3.

(2) Jancova, P.; Anzenbacher, P.; Anzenbacherova, E. PHASE II DRUG METABOLIZING ENZYMES. *Biomed. Pap.* **2010**, *154* (2), 103–116. https://doi.org/10.5507/bp.2010.017.

(3) Hasler, J. A.; Estabrook, R.; Murray, M.; Pikuleva, I.; Waterman, M.; Capdevila, J.; Holla, V.; Helvig, C.; Falck, J. R.; Farrell, G.; Kaminsky, L. S.; Spivack, S. D.; Boitier, E.; Beaune, P. Human Cytochromes P450. *Mol. Aspects Med.* **1999**, *20* (1–2), 1–137. https://doi.org/10.1016/S0098-2997(99)00005-9.

(4) Beck, T. C.; Beck, K. R.; Morningstar, J.; Benjamin, M. M.; Norris, R. A. Descriptors of Cytochrome Inhibitors and Useful Machine Learning Based Methods for the Design of Safer Drugs. *Pharmaceuticals* **2021**, *14* (5), 472. https://doi.org/10.3390/ph14050472.

(5) Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I. Structure and Chemistry of Cytochrome P450. *Chem. Rev.* **2005**, *105* (6), 2253–2278. https://doi.org/10.1021/cr0307143.

(6) Leach, A. G.; Kidley, N. J. Quantitatively Interpreted Enhanced Inhibition of Cytochrome P450s by Heteroaromatic Rings Containing Nitrogen. *J. Chem. Inf. Model.* **2011**, *51* (5), 1048–1063. https://doi.org/10.1021/ci2000506.

(7) Anzenbacher, P.; Anzenbacherová, E. Cytochromes P450 and Metabolism of Xenobiotics: *Cell. Mol. Life Sci.* **2001**, *58* (5), 737–747. https://doi.org/10.1007/PL00000897.

(8) *Drug Metabolism Prediction*, 1st ed.; Kirchmair, J., Ed.; Methods and Principles in Medicinal Chemistry; Wiley, 2014. https://doi.org/10.1002/9783527673261.

(9) Furge, L. L.; Guengerich, F. P. Cytochrome P450 Enzymes in Drug Metabolism and Chemical Toxicology: An Introduction. *Biochem. Mol. Biol. Educ.* **2006**, *34* (2), 66–74. https://doi.org/10.1002/bmb.2006.49403402066.

(10) Mikov, M.; Đanić, M.; Pavlović, N.; Stanimirov, B.; Goločorbin-Kon, S.; Stankov, K.; Al-Salami, H. The Role of Drug Metabolites in the Inhibition of Cytochrome P450 Enzymes. *Eur. J. Drug Metab. Pharmacokinet.* **2017**, *42* (6), 881–890. https://doi.org/10.1007/s13318-017-0417-y.

(11) Sekiguchi, N.; Higashida, A.; Kato, M.; Nabuchi, Y.; Mitsui, T.; Takanashi, K.; Aso, Y.; Ishigai, M. Prediction of Drug-Drug Interactions Based on Time-Dependent

Inhibition from High Throughput Screening of Cytochrome P450 3A4 Inhibition. *Drug Metab. Pharmacokinet.* **2009**, *24* (6), 500–510. https://doi.org/10.2133/dmpk.24.500.

(12) Lin, J. H.; Lu, A. Y. H. Inhibition and Induction of Cytochrome P450 and the Clinical Implications: *Clin. Pharmacokinet.* **1998**, *35* (5), 361–390. https://doi.org/10.2165/00003088-199835050-00003.

(13) Tompkins, L. M.; Wallace, A. D. Mechanisms of Cytochrome P450 Induction. *J. Biochem. Mol. Toxicol.* **2007**, *21* (4), 176–181. https://doi.org/10.1002/jbt.20180.

(14) Bahar, I.; Jernigan, R. L.; Dill, K. A.; Dill, K. A. *Protein Actions: Principles and Modeling*; Garland Science, Taylor & Francis Group: New York London, 2017.

(15) D'Aloisio, V.; Dognini, P.; Hutcheon, G. A.; Coxon, C. R. PepTherDia: Database and Structural Composition Analysis of Approved Peptide Therapeutics and Diagnostics. *Drug Discov. Today* **2021**, *26* (6), 1409–1419. https://doi.org/10.1016/j.drudis.2021.02.019.

(16) Commissioner, O. of the. U.S. Food and Drug Administration https://www.fda.gov/home (accessed 2022 -03 -18).

(17) Research, C. for D. E. and. Impact Story: Developing the Tools to Evaluate Complex Drug Products: Peptides. *FDA* **2020**.

(18) International Union of Pure and Applied Chemistry http://www.old.iupac.org/publications/books/author/mcnaught.html (accessed 2022 -03 -29).

(19) Werle, M.; Bernkop-Schnürch, A. Strategies to Improve Plasma Half Life Time of Peptide and Protein Drugs. *Amino Acids* **2006**, *30* (4), 351–367. https://doi.org/10.1007/s00726-005-0289-3.

(20) Yao, J.-F.; Yang, H.; Zhao, Y.-Z.; Xue, M. Metabolism of Peptide Drugs and Strategies to Improve Their Metabolic Stability. *Curr. Drug Metab.* **2018**, *19* (11), 892–901. https://doi.org/10.2174/1389200219666180628171531.

(21) Uhlig, T.; Kyprianou, T.; Martinelli, F. G.; Oppici, C. A.; Heiligers, D.; Hills, D.; Calvo, X. R.; Verhaert, P. The Emergence of Peptides in the Pharmaceutical Business: From Exploration to Exploitation. *EuPA Open Proteomics* **2014**, *4*, 58–69. https://doi.org/10.1016/j.euprot.2014.05.003.

(22) Jacobsen, L. V.; Flint, A.; Olsen, A. K.; Ingwersen, S. H. Liraglutide in Type 2 Diabetes Mellitus: Clinical Pharmacokinetics and Pharmacodynamics. *Clin. Pharmacokinet.* **2016**, *55* (6), 657–672. https://doi.org/10.1007/s40262-015-0343-6.

(23) Hebert, M. Contributions of Hepatic and Intestinal Metabolism and P-Glycoprotein to

Cyclosporine and Tacrolimus Oral Drug Delivery. *Adv. Drug Deliv. Rev.* **1997**, *27* (2–3), 201–214. https://doi.org/10.1016/S0169-409X(97)00043-4.

(24) Niwa, T.; Yamamoto, S.; Saito, M.; Shiraga, T.; Takagi, A. Effect of Cyclosporine and Tacrolimus on Cytochrome P450 Activities in Human Liver Microsomes. *YAKUGAKU ZASSHI* **2007**, *127* (1), 209–216. https://doi.org/10.1248/yakushi.127.209.

(25) Gertz, M.; Cartwright, C. M.; Hobbs, M. J.; Kenworthy, K. E.; Rowland, M.; Houston, J. B.; Galetin, A. Cyclosporine Inhibition of Hepatic and Intestinal CYP3A4, Uptake and Efflux Transporters: Application of PBPK Modeling in the Assessment of Drug-Drug Interaction Potential. *Pharm. Res.* **2013**, *30* (3), 761–780. https://doi.org/10.1007/s11095-012-0918-y.

(26) Kato, H. Computational Prediction of Cytochrome P450 Inhibition and Induction. *Drug Metab. Pharmacokinet.* **2020**, *35* (1), 30–44. https://doi.org/10.1016/j.dmpk.2019.11.006.

(27) Tyzack, J. D.; Kirchmair, J. Computational Methods and Tools to Predict Cytochrome P450 Metabolism for Drug Discovery. *Chem. Biol. Drug Des.* **2019**, *93* (4), 377–386. https://doi.org/10.1111/cbdd.13445.

(28) Kirchmair, J.; Göller, A. H.; Lang, D.; Kunze, J.; Testa, B.; Wilson, I. D.; Glen, R. C.; Schneider, G. Predicting Drug Metabolism: Experiment and/or Computation? *Nat. Rev. Drug Discov.* **2015**, *14* (6), 387–404. https://doi.org/10.1038/nrd4581.

(29) Plonka, W.; Stork, C.; Šícho, M.; Kirchmair, J. CYPlebrity: Machine Learning Models for the Prediction of Inhibitors of Cytochrome P450 Enzymes. *Bioorg. Med. Chem.* **2021**, *46*, 116388. https://doi.org/10.1016/j.bmc.2021.116388.

(30) Holmer, M.; de Bruyn Kops, C.; Stork, C.; Kirchmair, J. CYPstrate: A Set of Machine Learning Models for the Accurate Classification of Cytochrome P450 Enzyme Substrates and Non-Substrates. *Molecules* **2021**, *26* (15), 4678. https://doi.org/10.3390/molecules26154678.

(31) Who we are https://www.novonordisk.com/about/who-we-are.html (accessed 2022 -05 -02).

(32) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. PubChem Substance and Compound Databases. *Nucleic Acids Res.* **2016**, *44* (D1), D1202–D1213. https://doi.org/10.1093/nar/gkv951.

(33) Papadatos, G.; Overington, J. P. The ChEMBL Database: A Taster for Medicinal Chemists. *Future Med. Chem.* **2014**, *6* (4), 361–364. https://doi.org/10.4155/fmc.14.8.

(34) Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L. J.; Cibrián-Uhalte, E.; Davies, M.; Dedman, N.; Karlsson, A.; Magariños, M. P.; Overington, J. P.; Papadatos, G.; Smit, I.; Leach, A. R. The ChEMBL Database in 2017. *Nucleic Acids Res.* **2017**, *45* (D1), D945–D954. https://doi.org/10.1093/nar/gkw1074.

(35) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40* (D1), D1100–D1107. https://doi.org/10.1093/nar/gkr777.

(36) Davies, M.; Nowotka, M.; Papadatos, G.; Dedman, N.; Gaulton, A.; Atkinson, F.; Bellis, L.; Overington, J. P. ChEMBL Web Services: Streamlining Access to Drug Discovery Data and Utilities. *Nucleic Acids Res.* **2015**, *43* (W1), W612–W620. https://doi.org/10.1093/nar/gkv352.

(37) Novartis Institutes for BioMedical Research https://www.novartis.com/research-development/novartis-institutes-biomedical-research (accessed 2022 -03 -18).

(38) About | DNDi https://dndi.org/about/ (accessed 2022 -03 -18).

(39) About us | Medicines for Malaria Venture https://www.mmv.org/about-us?gclid=Cj0KCQjw29CRBhCUARIsAOboZbJX5DoeEuh9pA1NdqqMaH0_0xqMUEMgdZlzAeEXe9JJjXxECVmvJfcaAm_4EALw_wcB (accessed 2022 -03 -18).

(40) Homepage | Liverpool John Moores University http://www.ljmu.ac.uk/ (accessed 2022 -03 -18).

(41) United Kingdom https://www.hw.ac.uk/index.htm (accessed 2022 -03 -18).

(42) Wishart, D. S.; Feunang, Y. D.; Guo, A. C.; Lo, E. J.; Marcu, A.; Grant, J. R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; Assempour, N.; Iynkkaran, I.; Liu, Y.; Maciejewski, A.; Gale, N.; Wilson, A.; Chin, L.; Cummings, R.; Le, D.; Pon, A.; Knox, C.; Wilson, M. DrugBank 5.0: A Major Update to the DrugBank Database for 2018. *Nucleic Acids Res.* **2018**, *46* (D1), D1074–D1082. https://doi.org/10.1093/nar/gkx1037.

(43) National Center for Advancing Translational Sciences https://ncats.nih.gov/ (accessed 2022 -03 -18).

(44) Drugs.com | Prescription Drug Information, Interactions & Side Effects https://www.drugs.com/ (accessed 2022 -03 -18).

(45) EMA. European Medicines Agency https://www.ema.europa.eu/en (accessed 2022 -03 -18).

(46) Sander, T.; Freyss, J.; von Korff, M.; Rufener, C. DataWarrior: An Open-Source

Program For Chemistry Aware Data Visualization And Analysis. *J. Chem. Inf. Model.* **2015**, *55* (2), 460–473. https://doi.org/10.1021/ci500588j.

(47) Open Source Story https://www.knime.com/knime-open-source-story (accessed 2022 -03 -04).

(48) KNIME - the Konstanz Information Miner: Version 2.0 and Beyond. *11* (1), 6.

(49) Sieb, C.; Meinl, T.; Berthold, M. R. Parallel and Distributed Data Pipelining with KNIME. 9.

(50) Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J. D.; Zardecki, C. The Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **2002**, *58* (6), 899–907. https://doi.org/10.1107/S0907444902003451.

(51) Berman, H.; Henrick, K.; Nakamura, H.; Markley, J. L. The Worldwide Protein Data Bank (WwPDB): Ensuring a Single, Uniform Archive of PDB Data. *Nucleic Acids Res.* **2007**, *35* (Database), D301–D303. https://doi.org/10.1093/nar/gkl971.

(52) PDBx/mmCIF Dictionary Resources https://mmcif.wwpdb.org/ (accessed 2022 -03 -18).

(53) The RCSB Protein Data Bank: Integrative View of Protein, Gene and 3D Structural Information. *Nucleic Acids Res.* **2016**. https://doi.org/10.1093/nar/gkw1000.

(54) Maestro | Schrödinger https://www.schrodinger.com/products/maestro (accessed 2022 -03 -18).

(55) Home - The Cambridge Crystallographic Data Centre (CCDC) https://www.ccdc.cam.ac.uk/ (accessed 2022 -03 -18).

(56) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739–1749. https://doi.org/10.1021/jm0306430.

(57) Jones, G. et al. Development and Validation of a Genetic Algorithm for Flexible Docking. 22.

(58) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261* (3), 470–489. https://doi.org/10.1006/jmbi.1996.0477.

(59) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search Strategies for Automated Molecular Docking of Flexible Molecule Databases. 18.

(60) Chen, Y.; Kirchmair, J. Cheminformatics in Natural Product-based Drug Discovery. *Mol. Inform.* **2020**, *39* (12), 2000171. https://doi.org/10.1002/minf.202000171.

(61) Zentrum für Bioinformatik: Universität Hamburg - Proteins Plus Server https://proteins.plus/ (accessed 2022 -03 -18).

(62) Fährrolfes et al. - 2017 - ProteinsPlus a Web Portal for Structure Analysis .Pdf.

(63) Sansen, S.; Yano, J. K.; Reynald, R. L.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. Adaptations for the Oxidation of Polycyclic Aromatic Hydrocarbons Exhibited by the Structure of Human P450 1A2. *J. Biol. Chem.* **2007**, *282* (19), 14348–14355. https://doi.org/10.1074/jbc.M611692200.

(64) Walsh, A. A.; Szklarz, G. D.; Scott, E. E. Human Cytochrome P450 1A1 Structure and Utility in Understanding Drug and Xenobiotic Metabolism. *J. Biol. Chem.* **2013**, *288* (18), 12932–12943. https://doi.org/10.1074/jbc.M113.452953.

(65) Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. The Structure of Human Microsomal Cytochrome P450 3A4 Determined by X-Ray Crystallography to 2.05-Å Resolution. *J. Biol. Chem.* **2004**, *279* (37), 38091–38094. https://doi.org/10.1074/jbc.C400293200.

(66) Sevrioukova, I. F.; Poulos, T. L. Structure and Mechanism of the Complex between Cytochrome P4503A4 and Ritonavir. *Proc. Natl. Acad. Sci.* **2010**, *107* (43), 18422–18427. https://doi.org/10.1073/pnas.1010693107.

(67) Kaur, P.; Chamberlin, A. R.; Poulos, T. L.; Sevrioukova, I. F. Structure-Based Inhibitor Design for Evaluation of a CYP3A4 Pharmacophore Model. *J. Med. Chem.* **2016**, *59* (9), 4210–4220. https://doi.org/10.1021/acs.jmedchem.5b01146.

(68) Samuels, E. R.; Sevrioukova, I. F. Rational Design of CYP3A4 Inhibitors: A One-Atom Linker Elongation in Ritonavir-Like Compounds Leads to a Marked Improvement in the Binding Strength. *Int. J. Mol. Sci.* **2021**, *22* (2), 852. https://doi.org/10.3390/ijms22020852.

(69) Ekroos, M.; Sjogren, T. Structural Basis for Ligand Promiscuity in Cytochrome P450 3A4. *Proc. Natl. Acad. Sci.* **2006**, *103* (37), 13682–13687. https://doi.org/10.1073/pnas.0603236103.

(70) Sevrioukova, I. Interaction of Human Drug-Metabolizing CYP3A4 with Small Inhibitory Molecules. *Biochemistry* **2019**, *58* (7), 930–939. https://doi.org/10.1021/acs.biochem.8b01221.

(71) Pamela A. Williams; Cosme, J.; Vinković, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H. Crystal Structures of Human Cytochrome P450 3A4

Bound to Metyrapone and Progesterone. *Science* **2004**, *305* (5684), 683–686. https://doi.org/10.1126/science.1099736.

(72) Sevrioukova, I. F.; Poulos, T. L. Pyridine-Substituted Desoxyritonavir Is a More Potent Inhibitor of Cytochrome P450 3A4 than Ritonavir. *J. Med. Chem.* **2013**, *56* (9), 3733–3741. https://doi.org/10.1021/jm400288z.

(73) Samuels Und Sevrioukova - 2018 - Inhibition of Human CYP3A4 by Rationally Designed .Pdf.

(74) Samuels, E. R.; Sevrioukova, I. Inhibition of Human CYP3A4 by Rationally Designed Ritonavir-Like Compounds: Impact and Interplay of the Side Group Functionalities. *Mol. Pharm.* **2018**, *15* (1), 279–288. https://doi.org/10.1021/acs.molpharmaceut.7b00957.

(75) Kaur et al. - 2016 - Structure-Based Inhibitor Design for Evaluation of.Pdf.

(76) Samuels Und Sevrioukova - 2020 - An Increase in Side-Group Hydrophobicity Largely i.Pdf.

(77) Samuels, E. R.; Sevrioukova, I. Structure–Activity Relationships of Rationally Designed Ritonavir Analogues: Impact of Side-Group Stereochemistry, Headgroup Spacing, and Backbone Composition on the Interaction with CYP3A4. *Biochemistry* **2019**, *58* (15), 2077–2087. https://doi.org/10.1021/acs.biochem.9b00156.

(78) Sevrioukova, I. F. Structural Insights into the Interaction of Cytochrome P450 3A4 with Suicide Substrates: Mibefradil, Azamulin and 6′,7′-Dihydroxybergamottin. *Int. J. Mol. Sci.* **2019**, *20* (17), 4245. https://doi.org/10.3390/ijms20174245.

(79) Wang, A.; Stout, C. D.; Zhang, Q.; Johnson, E. F. Contributions of Ionic Interactions and Protein Dynamics to Cytochrome P450 2D6 (CYP2D6) Substrate and Inhibitor Binding. *J. Biol. Chem.* **2015**, *290* (8), 5092–5104. https://doi.org/10.1074/jbc.M114.627661.

(80) Brodney, M. A.; Beck, E. M.; Butler, C. R.; Barreiro, G.; Johnson, E. F.; Riddell, D.; Parris, K.; Nolan, C. E.; Fan, Y.; Atchison, K.; Gonzales, C.; Robshaw, A. E.; Doran, S. D.; Bundesmann, M. W.; Buzon, L.; Dutra, J.; Henegar, K.; LaChapelle, E.; Hou, X.; Rogers, B. N.; Pandit, J.; Lira, R.; Martinez-Alsina, L.; Mikochik, P.; Murray, J. C.; Ogilvie, K.; Price, L.; Sakya, S. M.; Yu, A.; Zhang, Y.; O'Neill, B. T. Utilizing Structures of CYP2D6 and BACE1 Complexes To Reduce Risk of Drug–Drug Interactions with a Novel Series of Centrally Efficacious BACE1 Inhibitors. *J. Med. Chem.* **2015**, *58* (7), 3223–3252. https://doi.org/10.1021/acs.jmedchem.5b00191.

(81) Rowland, P.; Blaney, F. E.; Smyth, M. G.; Jones, J. J.; Leydon, V. R.; Oxbrow, A. K.;

Lewis, C. J.; Tennant, M. G.; Modi, S.; Eggleston, D. S.; Chenery, R. J.; Bridges, A. M. Crystal Structure of Human Cytochrome P450 2D6. *J. Biol. Chem.* **2006**, *281* (11), 7614–7622. https://doi.org/10.1074/jbc.M511232200.

(82) Wang, A.; Savas, U.; Hsu, M.-H.; Stout, C. D.; Johnson, E. F. Crystal Structure of Human Cytochrome P450 2D6 with Prinomastat Bound. *J. Biol. Chem.* **2012**, *287* (14), 10834–10843. https://doi.org/10.1074/jbc.M111.307918.

(83) McLaughlin, L. A.; Paine, M. J. I.; Kemp, C. A.; Maréchal, J.-D.; Flanagan, J. U.; Ward, C. J.; Sutcliffe, M. J.; Roberts, G. C. K.; Wolf, C. R. Why Is Quinidine an Inhibitor of Cytochrome P450 2D6? *J. Biol. Chem.* **2005**, *280* (46), 38617–38624. https://doi.org/10.1074/jbc.M505974200.

(84) Hutzler, J. M.; Walker, G. S.; Wienkers, L. C. Inhibition of Cytochrome P450 2D6: Structure−Activity Studies Using a Series of Quinidine and Quinine Analogues. *Chem. Res. Toxicol.* **2003**, *16* (4), 450–459. https://doi.org/10.1021/tx025674x.

(85) Williams, P. A.; Cosme, J.; Ward, A.; Angove, H. C.; Matak Vinković, D.; Jhoti, H. Crystal Structure of Human Cytochrome P450 2C9 with Bound Warfarin. *Nature* **2003**, *424* (6947), 464–468. https://doi.org/10.1038/nature01862.

(86) Skerratt, S. E.; de Groot, M. J.; Phillips, C. Discovery of a Novel Binding Pocket for CYP 2C9 Inhibitors: Crystallography, Pharmacophore Modelling and Inhibitor SAR. *MedChemComm* **2016**, *7* (5), 813–819. https://doi.org/10.1039/C6MD00011H.

(87) Brändén, G.; Sjögren, T.; Schnecke, V.; Xue, Y. Structure-Based Ligand Design to Overcome CYP Inhibition in Drug Discovery Projects. *Drug Discov. Today* **2014**, *19* (7), 905–911. https://doi.org/10.1016/j.drudis.2014.03.012.

(88) Locuson, C. W.; Suzuki, H.; Rettie, A. E.; Jones, J. P. Charge and Substituent Effects on Affinity and Metabolism of Benzbromarone-Based CYP2C19 Inhibitors. *J. Med. Chem.* **2004**, *47* (27), 6768–6776. https://doi.org/10.1021/jm049605m.

(89) Reynald, R. L.; Sansen, S.; Stout, C. D.; Johnson, E. F. Structural Characterization of Human Cytochrome P450 2C19. *J. Biol. Chem.* **2012**, *287* (53), 44581–44591. https://doi.org/10.1074/jbc.M112.424895.

(90) Pinzi, L.; Rastelli, G. Molecular Docking: Shifting Paradigms in Drug Discovery. *Int. J. Mol. Sci.* **2019**, *20* (18), 4331. https://doi.org/10.3390/ijms20184331.