HOMOLOGY MODELLING AND CHARACTERIZATION OF THREE-DIMENSIONAL (3D) PROTEIN STRUCTURES OF SELECTED VARIANTS OF CYP3A4

Christianah Taye Omotoso

(STUDENT NUMBER: 3984138)

UNIVERSITY *of the*
WESTERN CAPE

A thesis submitted in fulfilment of the requirements for the degree of Master of Pharmacy, the School of Pharmacy, University of the Western Cape, South Africa.

Supervisor: Dr. Samuel Egieyeh

Co-supervisor: Prof. Alan Christoffels

August 2022

http://etd.uwc.ac.za/

HOMOLOGY MODELLING AND CHARACTERIZATION OF THREE-DIMENSIONAL (3D) PROTEIN STRUCTURES OF SELECTED VARIANTS OF CYP3A4

Christianah Taye Omotoso

Keywords

Drug metabolising enzymes

Cytochrome P450 3A4 (CYP3A4)

Single nucleotide polymorphism

Enzymatic activity

Three-dimensional protein structures

Bioinformatics tools

Homology modelling

Physicochemical properties

Molecular dynamics

Pearson correlation coefficients

Pharmacogenomics

Precision medicine

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

Declaration

I, Christianah Taye Omotoso, declare that this written submission represents my work and that it has not been submitted before for any degree or examination at any other higher education institution. I have cited and referenced the sources where others' ideas or words have been included. I also declare that I have adhered to all academic honesty and integrity ethics and have not misrepresented or fabricated any idea/data/source in my submission.

…… ……

Signed Date

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

Abstract

The efficacy and toxicity of several drugs and prodrugs are influenced by the inter-individual genetic variations in drug-metabolising enzymes. Cytochrome P450 enzymes constitute the major metabolising enzymes in humans. Cytochrome 3A4 (CYP3A4) is the highest abundantly expressed human cytochrome P450 enzyme metabolising about 40% of marketed drugs. Several studies have shown the significant effects of CYP3A4 single nucleotide variability on its enzymatic activity and the pharmacokinetic parameters of metabolized drugs. However, there is a paucity of information on the molecular characteristics (including three-dimensional protein structures, physicochemical properties, and molecular dynamics characteristics) of single nucleotide variants of CYP3A4 enzymes that may be correlated with the reported variability of their enzymatic activities. This study investigated the possible effects of the reported single nucleotide polymorphism on the structural, physicochemical, and molecular dynamic properties of selected variants of CYP3A4. The research also aimed to establish a potential correlation between the previously reported enzymatic activities (as measured by intrinsic clearance of lidocaine) of selected variants of CYP3A4 and molecular characteristics of the variants.

The literature reviewed revealed four CYP3A4 variants that showed significant differences in the clearance of lidocaine relative to the wild type, this include the CYP3A4*2 (S222P) and CYP3A4*24 (Q200H), CYP3A4*11 (T363M) and CYP3A4*23 (R162W) . The differences in the three-dimensional modelled protein surfaces were investigated. Bioinformatics tools such as MOE program, ChimeraX, Expasy protparam, and CharmGUI were used for the prediction of the physicochemical properties of the modelled protein structures of the selected variants. The free energy binding of the modelled protein structures of the CYP3A4 with lidocaine was determined with molecular docking, and the molecular dynamic characteristics of the selected variants with and without the docked lidocaine were investigated. Pearson's correlation coefficients were used to determine the

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

correlation between the reported intrinsic clearance of lidocaine and the predicted molecular protein properties of selected CYP3A4 variants.

Structural differences were observed in the active site region and noticeable differences were also discovered in the molecular lipophilicity potential, coulombic electrostatic potential, and some of the variants' physicochemical and general protein properties relative to the wild type. Notably, the Gibbs free energy change predicted a decrease in the structural stability of the selected variants. The wild type showed the highest solvation energy when compared to all the selected variants. The molecular dynamics study showed a decrease in structural dynamic stability of the selected variants' unbounded protein structures, and increased flexibility was observed in the core active regions of variants CYP3A4*2 and CYP3A4*23. In all, more than sixty-five percent of the total predicted properties (structural, physicochemical, and dynamic properties) showed significant differences in the selected variants relative to the wild type. The correlation study revealed that significant correlation exists between fifty-two percent of the predicted properties with notable differences, and the reported variation in the clearance of the selected variants.

UNIVERSITY of the
WESTERN CAPE

Overall, the data generated from this study provided insight into the effect of single nucleotide polymorphism on the structural, physicochemical properties, and dynamic characteristics of the modelled protein structures of selected variants of CYP3A4. The results from this study provide plausible explanations for the observed differences in the functional activity (metabolism) of the selected variants of CYP3A4 enzymes. The observed correlation between the molecular characteristics and the reported functional differences may help to elucidate the basis for the observed variation in the reported functional activity (metabolism) of the different variants of the CYP3A4 enzyme. This is applicable in drug design and discovery, to reduce the toxicity of drugs associated with genetic variation. This will also facilitate an increase in drug efficacy with appropriate implementation of the knowledge of the effects of single nucleotide polymorphism.

http://etd.uwc.ac.za/

Dedication

This work is dedicated to God, my shield, my glory, and the lifter up of my head.

http://etd.uwc.ac.za/

Acknowledgements

Sincere gratitude and special thanks:

1. To God, who has made me a product of His mercy and has showered His love and kindness to me in diverse ways during this project.
2. To my unquantifiable supervisor, Dr. Samuel Egieyeh. Your support, encouragement, mentorship, constructive corrections, and useful recommendations have left a positive and indelible landmark in my life.
3. To my co-supervisor, Prof. Alan Christoffels for your helpful suggestions, guidance, support, and assistance.
4. To the South African Research Chairs Initiative of the Department of Science and Innovation for their financial support.
5. To the South African National Bioinformatics Institute (SANBI) for the opportunity to participate in their organised training "Introduction to Bioinformatics Training 2020", the training built a helpful foundation for my research work.
6. To the Schrödinger Inc team for granting me access to their program and a free license to their software.
7. To a senior colleague, Mr. Oselusi Samson for your readiness to assist and constant encouragement, thank you for releasing your laptop to me at moments when I needed an extra one.
8. To Dr. Samuel Adegboye, Mrs. Bolu Oladunjoye, and Kian Shafei for your input and contributions at various stages of this research.
9. To all Computational Pharmacology and Cheminformatics Research Group (CPCRG) members, the teamwork with you was a great "push" in this thesis.
10. To my twin sister and precious siblings for your moral support, care, and frequent reminders to put in my best effort into all that my hands have discovered to do.
11. To my parents for your sacrificial and selfless love that helped build the good fount for the now and the future, I aspire to.

http://etd.uwc.ac.za/

12. To my best friend, Abiola Gboyega Kehinde; for staying true and through, your commitment to my progress is one I will forever be grateful for, your constant support towards me being the best God intends has always proved and reassured me of your genuine love.

http://etd.uwc.ac.za/

List of presentations

1. Samuel O. Egieyeh, **Christianah T. Omotoso**, Alan Christoffels, Samuel O. Adegboye (2020). Molecular and mechanistic insight into the varied rate of metabolism by alleles of the CYP3A4 enzyme. *Cold spring harbour laboratories-biological data science virtual conference.* 4th-6th November 2020. Available at: http://dx.doi.org/10.13140/rg.2.2.26089.72804

2. **Christianah T. Omotoso**, Samuel O. Egieyeh, Alan Christoffels (2021). Homology Modelling and Structural Comparison of Three-dimensional Protein Structures of Selected Variants of CYP3A4. *17th International Society of Computational Biology Student Council Symposium.* 23rd - 24th July 2021.
3. **Christianah T. Omotoso**, Samuel O. Egieyeh, Alan Christoffels, Samuel O. Adegboye. (2020). Generation of Mutated Cytochrome P450 Enzymes Sequences with Bioinformatics Techniques. *South Africa Society for Bioinformatics Student Council Online Symposium.* 4th - 6th August 2020.

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

Table of contents

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

List of Figures

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

List of tables

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

List of abbreviations

3D: Three dimensional

BLAST: Basic local alignment search tool (BLAST)

CYP: Cytochrome P450 enzymes

CYP3A4: Cytochrome P450 (3A4) enzymes

DNA: Deoxyribonucleic acid

EMs: Extensive metabolizers

IMs: Intermediate metabolizers

MD: Molecular dynamics

PMs: Poor metabolizers

Rg: Radius of gyration

RMSD: Root mean square deviation

RMSF: Root mean square fluctuation

RSA: Relative solvent accessible area

SASA: Solvent accessible surface area

SNP: Single nucleotide polymorphism

SRS: Substrate recognition sites

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

UMs: Ultra-rapid metabolizers

DMEs: Drug metabolising enzymes

TPMT: Thiopurine S-methyltransferase

dbSNP: Single nucleotide polymorphism database (dbSNP)

XP: Extra precision mode

BE: Binding free energy

PDB: Protein data bank

HMM: Hidden Markov models

QMEAN: Qualitative model energy analysis

SVM: Support vector machine.

MLP: Molecular lipophilicity potential

ELP: Coulombic electrostatic potential

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

Chapter 1

Introduction

1.1 Background of the study

Reports have shown that the efficacy and toxicity of several drugs and prodrugs are influenced by the inter-individual genetic variations in drug-metabolising enzymes (Ahmed *et al.,* 2016; Neamțu, 2020). Pharmacogenomics, which is an essential component in precision medicine, is the study of the response of individuals to medication with their genomics information (Ahmed *et al.,* 2016; Arafah et al.,2021). The advent of pharmacogenomics allows the assessment of genetic variants responsible for an individual's specific drug response (Jørgensen, 2019, Radouani *et al.,* 2020). The alteration in the expression and function of proteins have been linked to variations in the deoxyribonucleic acid (DNA) sequences. (Teama, 2018; Vihinen, 2021), which may also explain the variation in individual drug responses.

Most drugs in use today that undergo hepatic clearance are found to be metabolised by cytochrome P450 (CYP) enzymes (Tornio and Backman, 2018). It has been reported that CYP enzymes undergo gene polymorphism as well as gene duplications which often result in variations in the metabolism of medication (Tverdohleb *et al.,* 2016). In the liver, the CYP3A family is the most common subfamily of CYP isoforms. There are at least four isoforms: 3A3, 3A4, 3A5, and 3A7, the most important of which is 3A4 (Guttman *et al.,* 2019). The variability of CYP3A4 activity in the population has been reported to be extremely high (>100-fold) (Klein and Zanger, 2013; Saiz-Rodríguez *et al.,* 2020).

Several studies have proved the significant effects of CYP3A4 gene variability on its enzymatic activity, and consequently on most pharmacokinetic parameters (Fang *et al.,* 2017; Zhou *et al.,* 2019; Chen *et al.,* 2020). However, there has been no report on the effect of gene variation on the three-dimensional structures, physicochemical and dynamic characteristics of the CYP3A4 protein. This research seeks to explore

http://etd.uwc.ac.za/

the potential link between reported functional differences in CYP3A4 variants and their structural or physicochemical/molecular and dynamic characteristics.

Four CYP3A4 variants that showed significant differences in the clearance of lidocaine relative to the wild type were selected in the report by Fang *et al.*, (2017). These four variants with significant differences in their enzymatic activities were selected to investigate the effect of the variation on the structural properties, molecular dynamic characteristics, physicochemical and general protein properties. Hence, a need to generate three-dimensional protein models of the selected CYP3A4 variants. Homology modelling, also known as comparative protein structure modelling, is a method for obtaining three-dimensional models of a protein from its amino-acid sequence (Ruppé *et al.*, 2019; Adebiyi and Olugbara, 2021). It uses an alignment with a homologous protein of a known structure that serves as the template. Studies have shown that in all circumstances where template structures can be found, homology modelling is the approach of choice for creating accurate three-dimensional in silico models of a protein (Waterhouse *et al.*, 2018; Chikhale *et al.*, 2020). In addition, according to Esfandi and Atabati (2021), to guarantee the structural properties of the 3D protein models' compatibility with the physicochemical rules, the final 3D protein models created via homology modelling must be checked relative to experimentally solved protein structures. Hence, several 3D protein model analysis packages have been developed to validate the quality of the modelled protein structures. A good 3D protein homology model is more likely to give better results in the subsequent protein structural analysis including the effect of single nucleotide polymorphism on the structural, dynamics characteristics, and physicochemical properties of selected protein variants.

According to recent reports, one of the important aspects of the characterization of a protein is the analysis of its physicochemical parameters (Kaur et al., 2020; Munjal, Shukla, and Singh, 2021b). Furthermore, Georgiou (2018) reported that the corresponding characteristics of the amino acids in a protein may determine its physicochemical properties. Therefore, it is important to investigate the effect of

UNIVERSITY of the WESTERN CAPE
http://etd.uwc.ac.za/

the substitution of an amino acid on the physicochemical property of the mutated protein structures.

It is a widely held view that the use of in silico bioinformatics tools and techniques like molecular dynamics to analyse modelled protein structures leads to a better understanding of the molecular mechanism, physicochemical characteristics, protein interaction, and other structural conformation of the protein (Oyugi *et al.*, 2018; Patra *et al.*, 2020). Rezaei *et al.* (2020) also reported that an examination of protein dynamics using molecular dynamics simulations can aid in understanding the effects of mutations on protein structure. This allows the investigation of the effect of a single amino acid change on the characteristics of modelled protein structures of selected CYP3A4 variants.

1.2 Research problem statement

The variations in the implicated enzymes (CYP3A4) led to the observed phenotype of poor metabolizers (PMs), intermediate metabolizers (IMs), extensive metabolizers (EMs), and ultra-rapid metabolizers (UMs). While several studies have proved the significant effects of CYP3A4 variability on its enzymatic activity, especially on most pharmacokinetic parameters, there is a paucity of information on the effect of some of these single nucleotide variations on the molecular, physicochemical, structural, and dynamic properties of the implicated enzymes (CYP3A4 variants). There is also a need to establish the correlation between the molecular characteristics of the variant CYP3A4 enzymes and the observed phenotype of varied rates and extent of the enzymatic functionality (Bins *et al.*, 2019; X. Liu *et al.*, 2019).

1.3 Significance of the study

The data and the knowledge that was generated from this study provide insight into the effect of single nucleotide polymorphism on the structural, physicochemical properties, and dynamic characteristics of modelled and validated protein structures of selected variants of CYP3A4. The significance of this study is hinged on the

http://etd.uwc.ac.za/

potential to be able to predict the intrinsic clearance of a drug by a variant of CYP3A4 enzyme from the structural properties of the variant enzyme.

1.4 Research aim

This study aimed to investigate the effects of single nucleotide polymorphism (SNP) on the structural, physicochemical, and general protein properties of the modelled three-dimensional (3D) protein structures of selected CYP3A4 variants and to correlate such properties to the reported intrinsic clearance of lidocaine.

1.5 Research questions

In this research, four CYP3A4 allelic variants were selected from a set of twenty-two allelic variants reported from a study that assessed the functional enzymatic effect of the variants on the metabolism of lidocaine (Fang *et al.*, 2017). The four selected variants were those with reported significant differences in the metabolism (clearance) of lidocaine relative to the wild type as reported by Fang *et al.* (2017).

Based on this reported study, the following research questions were asked:

1. What is the effect of reported single nucleotide polymorphism (SNP) on the three-dimensional protein structures of the selected variants?
2. What is the effect of reported single nucleotide polymorphism on the physicochemical and general protein characteristics of selected variants of CYP3A4 relative to the wild type?
3. What are the observable differences in the molecular dynamic properties of the three-dimensional protein structures of the wild type and the selected variants of CYP3A4?
4. What is the extent of correlation between the molecular structure, dynamic properties, physicochemical characteristics of selected variants of CYP3A4, and the observed differences in the functional activity (relative clearance) reported by Fang *et al.* (2017)?

UNIVERSITY of the WESTERN CAPE
http://etd.uwc.ac.za/

1.6 Research objectives

The following are the research objectives for this study:

1. Homology modelling and quality evaluation of the protein structures of the CYP3A4 selected variants.
2. Visualisation, structural and molecular comparisons to investigate the effect of single nucleotide polymorphism on modelled and validated protein structures of the selected CYP3A4 variants.
3. Determination of the differences in the physicochemical properties, ligand and solvent interaction, and other general properties of "protein" of the selected CYP3A4 variants, relative to the wild type.
4. Investigate the dynamic properties of the modelled protein structure of the selected variants relative to the wild type of CYP3A4 by molecular dynamic simulations.
5. Determination of the potential correlation between the studied structural, or physicochemical effects of SNPs and the reported effects of SNPs on the relative clearance of selected variants.

UNIVERSITY of the
WESTERN CAPE

1.7 Thesis framework

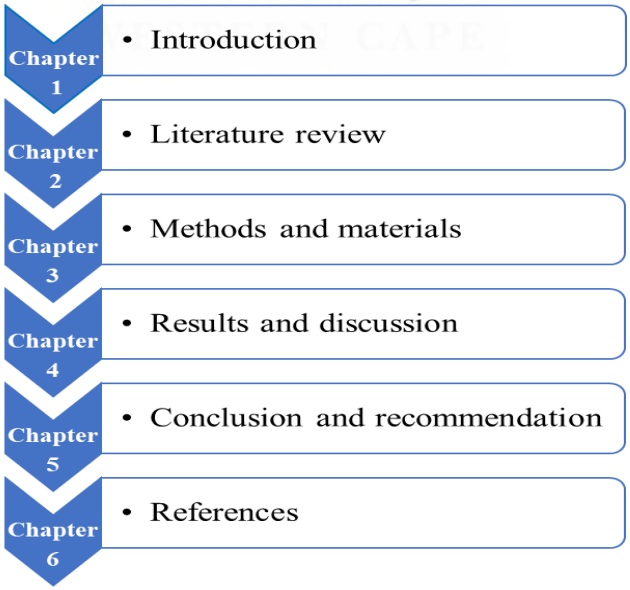


Figure 1.1: Schematic diagram of the thesis framework

http://etd.uwc.ac.za/

Chapter 2

Literature review

This chapter presents an overview of the polymorphism of cytochrome P450 (CYP) enzymes with a major focus on the single nucleotide polymorphism of CYP3A4 and its clinical implications in humans. The review includes a comprehensive report of the protein structure of CYP enzymes and the uniqueness of the CYP3A4 structure including its active site, ligand binding conformation, structural flexibility, topology, and volume. Summary of cytochrome P450, catalytic cycle, catalytic termination, substrate recognition sites, and protein structures' physicochemical properties was also reported in this chapter. Furthermore, the genetic polymorphism in CYPs, with a focus on the reported functional variants of CYP3A4 and their clinical implications are discussed. Finally, a summary of the reported techniques involving homology modelling and validation of the three-dimensional (3D) protein structure as well as docking and molecular dynamics simulation of the 3D structures of CYP are presented in this chapter.

2.1 Overview of polymorphism in cytochrome P450 (with focus on CYP3A4 variants) and their clinical implications

2.1.1 Background of cytochromes P450 enzymes

In 1958, cytochrome P450 (CYP) enzymes were identified and formally named first by Klingenberg while executing research on steroid hormone metabolism. Klingenberg discovered that a spectrum with maximum absorbance at 450 nm resulted from bubbling carbon monoxide that was obtained from the preparation of rat liver microsomes. It was reported that the maximum absorption gave P450s their name “Pigment-450nm (Klingenberg, 1958; McLean and Munro, 2018). Subsequently, cytochrome P450 (CYP) enzymes were discovered to be heme-containing monooxygenases that catalyse a wide variety of both endogenous and xenobiotic compounds (Tornio and Backman, 2018). Approximately 57 CYP genes and 8 pseudogenes have been recorded in humans and currently, about 18 different families and 44 subfamilies are known in humans (Guengerich, Waterman, and

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Egli, 2016; Esteves, Rueff, and Kranendonk, 2021). Previous reports have established that CYP pathways with similar gene sequences are assigned a family number such as CYP1 and CYP3, while a letter is used to indicate the subfamily (e.g., CYP3A and CYP2D). The subfamily is then distinguished with a number that shows the isoform such as CYP3A4 and CYP2D6 (McDonnell and Dang, 2013; Magwanga *et al.*, 2019). Reports have shown that in mammals, CYPs are found in all tissues predominantly in the liver and small intestine, with a widely held view that they are bound in the microsomal fraction of the liver (Pelkonen et al., 2008; Esteves, Rueff, and Kranendonk, 2021). CYPs are said to function mainly as drug-metabolising enzymes (Guengerich, 2018; Guttman, Nudel, and Kerem, 2019).

2.1.2 Cytochrome P450 enzymes as drug-metabolising enzymes

Many scholars hold the view that drug metabolism in the body is a complex biotransformation process in which numerous metabolising enzymes structurally modify chemical compounds into other molecules (metabolites) (Zhang and Tang, 2018; Hilmi Orhan 2021; Soltani *et al.*, 2021). According to Polic (2018), drug metabolism occurs in two phases, namely, phase 1 and phase 2. According to Lakshmanan (2019), the insertion of OH, -SH, or -NH2 functional groups, in phase I reactions converts the parent medication into more polar metabolites, which influences the activation or deactivation of the parent drug. It has been reported that in the phase 2 metabolising process, the drug is conjugated with endogenous charged molecules such as glutathione, glucuronide, and glycine. Since the conjugated metabolite is bigger and more water-soluble, it can be secreted into the bile or urine (Chen, 2020). Recent reports have also shown that the CYPs constitute the major metabolising enzymes in humans, and they are responsible for the metabolism of the most recognized clinically used drugs (Tornio and Backman, 2018; Hakkola *et al.*, 2020). The CYPs enzyme family forms the most essential enzymatic system in phase I drug metabolism in humans (Hakkola *et al.*, 2020). Cytochromes P450 perform a diverse range of oxidative reactions (Guengerich, 2018; Johnson, Su, and Zhang, 2021). CYP enzymes use one atom of oxygen to produce water while the other is incorporated into an organic substrate resulting in

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

the activation of molecular oxygen. This process is also applied in several catalytic reactions undertaken by CYPs, especially hydroxylation, epoxidation, deamination, and dealkylation (Guengerich, 2018; Johnson, Su, and Zhang, 2021).

2.1.2.1 CYP3A4: the most abundant CYPs drug-metabolising enzyme

Several studies have reported that the enzymes in the families 1-3 are responsible for about 80% of activity involved in the hepatic metabolism of drugs while the other families have minor endogenous functions (Ulrich M. Zanger and Schwab, 2013; Rendic and Guengerich, 2015). The CYP3A family is known to be the most common subfamily of CYP enzymes, and includes the CYP3A3, CYP3A4, CYP3A5, and CYP3A7 (S.-F. Zhou 2008; Dostalek *et al.*, 2011). CYP3A4 as an isoform of cytochrome P450 enzymes is a complex heme-containing enzyme that shows non-Michaelis-Menten kinetics and exhibits homotropic and heterotropic cooperativity with various substrates (Sevrioukova and Poulos, 2013; Rendic and Guengerich, 2015). The CYP3A4 gene is reported to be located on chromosome 7q at the q21-q22 locus (Lolodi *et al.*, 2017). It has been shown that CYP3A4 has an abundance of about 40% in the liver and in the intestine, which signifies the highest isoform in the CYPs family (Kumondai *et al.*, 2021). Recent studies have also reported that CYP3A4 is the highest abundantly expressed human CYP, metabolising between 30% and 50% of marketed drugs (Basheer and Kerem, 2015; Ghassabian *et al.*, 2019). Hence, the broad substrate selectivity of CYP3A4 enables it to play an important role in metabolism (Lolodi *et al.*, 2017; Kumondai et al., 2021). Therefore, CYP3A4 is a major metabolizing enzyme with a high potential for more research studies.

2.1.2.2 CYP3A4 broad substrate selectivity

CYP3A4 binds to a very diverse group of molecules with higher molecular weight when compared with the substrates of other CYPs enzymes (Goto, Yamazoe, and Tohkin, 2020). CYP3A4 has shared many substrates and inhibitors with CYP3A5 except that they vary in efficiency, catalysis, and the extent of susceptibility to the inhibitors (Lolodi *et al.*, 2017; Denisov *et al.*, 2021). CYP3A4 is also more active

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

to inhibit fluconazole, nicardipine, erythromycin, diltiazem, and ketoconazole when compared with CYP3A5 (Samuels and Sevrioukova, 2021). Erythromycin, tamoxifen, benzodiazepines, tacrolimus, opioids, and antidepressants are examples of drugs metabolised by CYP3A4 with its preference for mainly lipophilic and large drug compounds in most available drugs categories (Lolodi *et al.*, 2017). CYP3A4 also functions in the metabolism of numerous endogenous steroids such as testosterone, cortisol, progesterone, and bile acids (Niwa *et al.,* 2015). CYP34A4 endogenous biomarkers include 4B-hydroxylation of cholesterol, and 6b hydroxylation of cortisol (Gjestad *et al.,* 2019; Penzak and Rojas-Fernandez, 2019). However, the structural study of the interaction that exists between the CYP3A4 protein and the reported drugs was not identified.

2.1.3 Structure of CYPs enzymes

Various studies have reported that CYP, as a heme protein has one heme prosthetic group, located in the active site (Johnson and Stout, 2013; Midlik et al., 2021). It has been reported that CYP contains about 400-500 amino acid residues and the 3D structural representatives of CYP have similar characteristic folds and common elements present in the structure of CYP (Barr *et al.,* 2020). There are 12 helices and loops with a designated number (A-L) present in the general structure of CYP, and there are also a few B-sheets in the structure (Fig 2.1) (Johnson and Stout, 2013). Reports have shown that CYPs heme moiety is located between the I and L helix (Midlik *et al.*, 2021: Johnson and Stout, 2013). Several lines of evidence also proved that the enzyme's catalytic core is the heme prosthetic group, where a reactive hypervalent oxo-iron protoporphyrin IX radical cation intermediate is generated before the iron-bound oxygen atom is inserted into a substrate bond (Kaur *et al.*, 2016; Guengerich, 2018). Many recent studies established that the B-C and F-G helices play a role in the enzyme specificity and its accessibility to the enzyme substrate (Šrejber *et al.*, 2018; Dong *et al.*, 2021).

http://etd.uwc.ac.za/

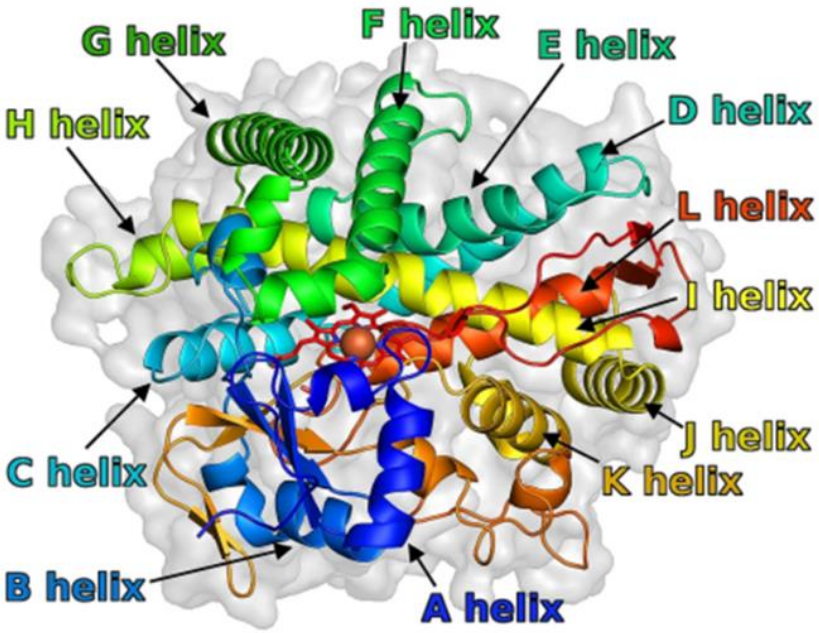


Figure 2.1: General features of CYPs enzyme structures.
The different helixes (A-L) and the centred positioned prosthetic heme group (brown centred circle) (Adapted from Midlik *et al.*, 2021)

2.1.3.1 The uniqueness of CYP3A4 structure in the CYP family

According to Giantin *et al.* (2019), CYP3A4, as it relates to other CYPs enzymes found in mammals, is a membrane-bound protein. The first structures of the drug-metabolising human enzyme CYP3A4 were independently reported by Williams (2004) and Yano et al. (2004) without a bounded ligand. Both structures have a very close structural agreement with a root mean square deviation of 0.6 A. The general tertiary structure of previously observed P450 structures is retained by CYP3A4, which is made up of several Beta -sheets, as well as Alpha helices and, are denoted with letters A-L (Williams, 2004; Yano *et al.*, 2004; P. C. Nair, McKinnon, and Miners, 2019).

It was reported that CYP450 3A4 has a homodimer structure with similar sub-units, including an N-terminal domain made up mainly of β sheet and a larger C-terminal domain rich in α helices (Williams, 2004; Lokwani *et al.*, 2020). Williams (2004) reported that the larger C-terminal domain contains the active site.

http://etd.uwc.ac.za/

Recent studies have shown that CYP3A4 has shorter F and G helices which necessitates the extension by unstructured linkers to the short F' and G' helices (Figure 2.2) (Hsu and Johnson, 2019; Lokwani *et al.*, 2020). Also, an interaction between the phenylalanine from the B-C region, the helix F', and the three phenylalanine residues from the linker regions have been reported to form a unique phenylalanine cluster located on the ceiling of the active-site cavity (Davydov *et al.*, 2012). The homodimer structure of CYP3A4 is reported to include an N-terminal domain made up mainly of β sheet and a larger C-terminal domain rich in α helices (Williams, 2004; Lokwani *et al.*, 2020). According to Williams (2004), the larger C-terminal domain contains the active site.

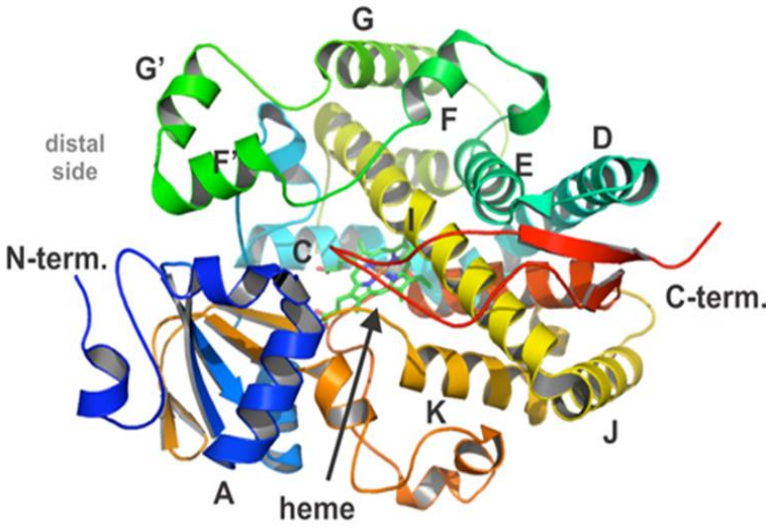


Figure 2.2: The structure of CYP3A4.

The secondary structure elements, β-helices (in stripes), α- helices (curls), The linkages (F' and G'), and the central position heme (Adapted from Otyepka, Berka, and Anzenbacher, 2012)

Figure 2.2 shows the structure of CYP3A4 that indicates the secondary structure elements, β-helices in stripes while α- helices are in curls and labelled with capital letters. The heme is positioned in the hidden active site close to the I helix. Above the heme is the CYP distal side, while the proximal side is below the heme plane, on the opposite side of the distal side (Otyepka, Berka, and Anzenbacher, 2012).

http://etd.uwc.ac.za/

Recent reports have shown that CY3A4 deviates structurally from the other eukaryotic CYP450 enzymes in three principal ways; CP3A4 has a hydrophobic region in which the 36-50 residues in the N-terminal regions help in its interaction with the microsomal membrane (Munjal, Shukla, and Singh, 2021). In addition, the F and G helices constitute the active site upper region in many of the CYP450 enzymes, however, this has been truncated in CYP3A4 (Hsu and Johnson, 2019). Finally, CYP3A4 has a unique cluster of 6 phenylalanine (Phe) residues namely: Phe-213, Phe-215, Phe-219, Phe-220, Phe-241, and Phe-304. The Phe cluster is localised at the upper region of the enzyme's peripheral binding region in the active site, the primary residues in the Phe cluster are reported to take part in the oxidation pathway of the allosteric regulation (Davydov *et al.*, 2012; Munjal, Shukla, and Singh, 2021). It has been reported that no other characterised structure of CYPs has such a phenylalanine residue cluster (Davydov *et al.*, 2012). Therefore, the Phenylalanine cluster is a major component that contributes to CYP3A4's 'uniqueness.

Researchers have reported that there are 256 residues or 55% as part of α helices in the secondary structure of CYP3A4 while only 35 residues or 7% are part of the β sheets (Ekroos and Sjogren, 2006; Hsu and Johnson, 2019; Estrada *et al.*, 2021). Wright, Chenge, and Chen (2019) further stated that the numerous substrates' exit channels, and the high flexibility of the enzyme are due to the high level of alpha-helices contained in the active site.

UNIVERSITY of the
WESTERN CAPE

2.1.3.2 The topology of CYP3A4 active site

Several studies have shown that mono oxidation of various substrates takes place at the CYP3A4 internal active site (Sevrioukova and Poulos, 2015; Guengerich, Waterman, and Egli, 2016). Sevrioukova and Poulos (2017) reported that the heme prosthetic group defines the lower bound of the CYP3A4's active site while the Phenylalanine cluster linking the F and G helices with the F' and G' helices constitute the upper bound. Furthermore, it has been suggested that the binding of the first ligand occurs at the peripheral binding site, and the Phenylalanine cluster initiates the reaction pathway (Ekroos and Sjogren, 2006; Davydov *et al.*, 2012).

http://etd.uwc.ac.za/

Wright, Chenge, and Chen (2019) further stated that the initial binding at the peripheral binding site causes conformational changes to the CYP3A4 structure; and added that this affects the water accessibility and the hydration state of the heme prosthetic group. In addition, Šrejber *et al.* (2018) reported that the F-G and B-C loops of P450 that make up the substrate-access channel could be a part of the membrane binding and orient the substrate-access channel with the membrane surface.

Poulos *et al.* (2015) reported that CYP3A4 has a larger active site cavity as compared to other CYP450 enzymes in families 1 and 2, which account for its ability to bind up to two or more molecules. Godamudunage, Grech, and Scott (2018) reported that CYP3A4 can metabolise various substrates without inhibiting one another competitively. The largeness of the CYP3A4 active site is consistent with its positive cooperation of oxidation or/and binding with some other substrates. The binding of an additional substrate molecule might result in the promotion of metabolism when the first substrate is stabilised in a productive orientation (Godamudunage, Grech, and Scott, 2018). Previous reports from both the homotropic and the heterotropic cooperativity show that the active site of CYP3A4 might sufficiently accommodate up to three molecules simultaneously (Müller, 2014; Müller *et al.*, 2015). The uniqueness of CYP3A4 is further shown in the report of its structure when it is bound to a ligand.

UNIVERSITY of the
WESTERN CAPE

2.1.3.3 The structure of CYP3A4 when bound to the ligand

Yano *et al.* (2004) reported from the study of the crystal structure of CYP3A4 when bound to progesterone as a substrate that the role in the first substrate recognition is consistent with the position of the progesterone binding site located in the F-G region of the CYP3A4 structure. Yano *et al.* (2004) suggested that the substrate access channel could stretch from the peripheral binding site to the heme group with the proper conformational movement of the residues close to the Phenylalanine-cluster. Urban *et al.* (2018) further reported that the substrate access channel gives a route for the movement of the compound to the active site from the first recognition site.

http://etd.uwc.ac.za/

The structure of CYP3A4 with metyrapone which is an inhibitor was also described by Williams (2004). In addition, Williams (2004) reported that there was no change in the protein conformation with the binding of metyrapone in contrast to CYP2C5 which shows changes in its active site topology when different ligands bind to it. Sevrioukova and Poulos (2015) further stated that there might be an expansion of the active cavity of CYP3A4 when substrates or inhibitors bind to it and there could be a contraction in some cases. Furthermore, Yano *et al.* (2004) reported that the expansions are possible with the upper portion of the binding pocket positioned opposite to the prosthetic heme group; and that this flexibility is an indication of the absence of a secondary structure between the F and G regions, which is present in other CYP. Sevrioukova and Poulos (2015) also reported that most of the differences observed in the amino acid sequence between CYP3A4 and CYP3A5 reside in the helixes of the F and G region, which makes up the roof or upper part of the active site cavity as well as in the N-terminal area of the cavity of the structure of CYP3A4. Reports on the structural flexibility of CYP3A4 have also showcased its uniqueness in the CYPs family (Sevrioukova and Poulos, 2015; Urban *et al.*, 2018).

2.1.3.4 The structural flexibility of CYP3A4

UNIVERSITY of the
WESTERN CAPE

The ability of CYP3A4 to recognize and bind to more compounds has been attributed to its well-defined flexibility (Lokwani *et al.*, 2020). Ohkura *et al.* (2009) reported that the size of the volume available for the ligand to bind at the CY3A4 binding site is about 520Å, which is very large. Additionally, Lokwani *et al.* (2020) reported that CY3A4 displays multiple conformations and can bind to a different and huge number of ligands such as its reported multiple binding with ritonavir and ketoconazole by Sevrioukova and Poulos (2017).

The conformation ability of CYP3A4 to the different compounds is attributed to the flexibility of its secondary structure; this flexibility is found to happen in the F-F' region of the CYP3A4 secondary structure. It was reported that there was an extended active site at the ligand bounded structure with an outward expansion also at the F-F' region due to the attributed flexibility of the secondary structure of

http://etd.uwc.ac.za/

CYP3A4 in the amino acid residues located in the roof of its binding pocket (Hsu, Savas, and Johnson, 2018). This change was discovered when the ligand-bound structure of CYP3A4 with two molecules of ketoconazole was compared to the ligand-free structure of CYP3A4 (not bound to any ligand nor coordinated to any water molecules) (Xu and Chen, 2020). None of the reports identified any effect of single nucleotide polymorphism on the structural flexibility of CYP3A4

2.1.3.5 Notable differences in CYP3A4 and CYP3A5 structures

Guttman, Nudel, and Kerem (2019) reported that CYP3A5 is an isoform of the CYP3A family. The difference in the amino acid residues lining the binding pocket of CYP3A4 and CYP3A5 secondary structures recently proved that despite the overall homologous secondary structures in the two, there is a unique shape that had been conferred on each structure in which the active site of CY3A4 is shorter and wider as compared to CYP3A5 which is taller and less wide (Hsu, Savas, and Johnson, 2018).

In studying the differences in the X-ray crystal structure of the Human Mono-Oxygenase Cytochrome CYP3A4 and CYP3A5, Hsu *et al.* (2018) discovered that the amino acid variation in CYP3A4 affects the substrate-binding cavity. This was shown in residue Ile-369 as carried out in the study. It was reported that there is a hydrophobic contact formed with the isopropyl thiazole moiety of ritonavir (IP2) and the thiazole groups with the Ile-369 of the CYP3A4 structure which is not the same in CYPA5, with a different residue (a smaller valine),at that position in CYP3A4 (Hsu, Savas, and Johnson, 2018).However, there was no report to justify if the biochemical characteristics of the different residues could explain the observed differences.

2.1.3.6 The conformational change in the active sites of CYP3A4

Lokwani *et al.* (2020) reported that CYP3A4 has multiple binding sites, and further suggested that these multiple binding sites have been attributed to the homotropic and heterotropic cooperativity displayed by the CYP3A4 enzyme, which influences the metabolic process of CYP3A4 substrates and brings a new level of complexity to the regulation of the CYP3A4. In addition, Kondža *et al.* (2021) reported that the

http://etd.uwc.ac.za/

ability of active sites of CYP3A4 enzymes to accommodate more than one substrate can act to either increase or decrease the metabolic rate. The authors explained that there is an increase in the metabolic rate if the presence of one substrate favourably holds the other in a good position for catalysis, and a decrease occurs when the other substrate is competitively inhibited, thereby preventing the other substrate to be well-positioned for conformational change in the presence of different substrates (Kondža *et al.*, 2021).

Furthermore, Williams (2004) reported that there is an unpredicted peripheral binding portion positioned above the phenylalanine cluster which sometimes is involved in the commencing and recognition of the substrate or that of the allosteric effectors. Li *et al.* (2021) further stated that a noticeable difference occurs in the size of the active site of CYP3A4 due to the conformational change after the binding of ligand to its protein structure. As reported by Yano *et al.* (2004), the active site volume of the CYP3A4 structure has an estimation of 950Å without a bounded ligand. In addition, reports have shown that when CYP3A4 is bounded with ketoconazole and erythromycin, there is an increase from the original 950Å to 1650Å and 2000Å respectively (Ekroos and Sjogren, 2006; Benkaidali *et al.*, 2017). Although the stated volume of the active site without the bounded ligand differs from the recent report by Lokwani *et al.*, (2020) with a reported volume of about 520Å. However, both reports are indicative of a large active site volume. There was however no mention of the effect of variation on the conformational change in the active sites of the implicating enzyme from both reports.

UNIVERSITY of the
WESTERN CAPE

2.1.4 Substrate recognition sites of CYP

The substrate recognition sites (SRS) have been reported to be an important part of CYP (Lokwani *et al.*, 2020). The SRS are regions in the CYP protein where substrates can recognize and bind, and six of them have been recognized with their locations in the CYP enzyme's structure (P. C. Nair, McKinnon, and Miners, 2019). They are, SRS1, SRS2, SRS3, SRS4, SRS5, and SRS6 (Figure 2.3).

http://etd.uwc.ac.za/

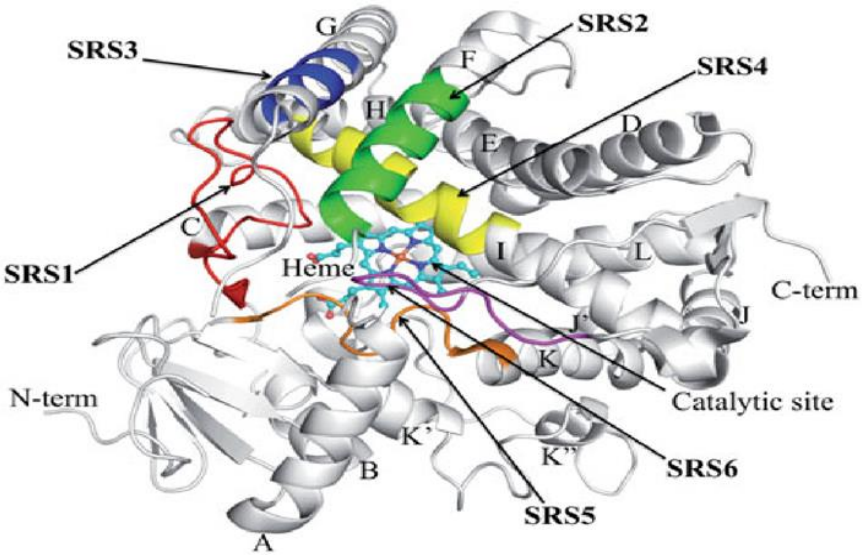


Figure 2.3: Identified substrates recognition sites (SRS) in the X-ray crystal structure of CYP enzymes (CYP29-PDB 1R9O).

The arrows are pointed to; SRS1 (red), SRS2 (green) and SRS3 (blue), SRS4 (yellow), SRS5 (orange), and SRS6 (magenta), while the rest of the secondary structural protein elements are white. The heme is represented with balls and sticks (Adapted from Nair *et al.*, 2016).

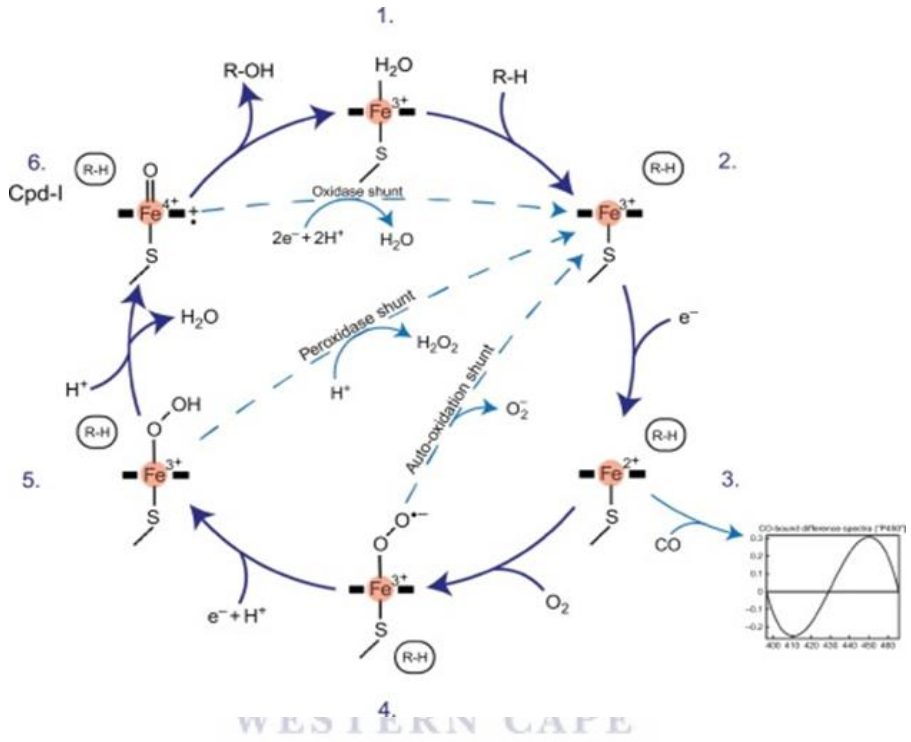
SRS 1 is found between the B and C helixes. SRS 2 and 3 are found between the F and G helices and the upper region of the active site is also located there. SRS 4 is located at the centre portion of the I helix. SRS5 is found at the N- terminus and the start of B sheet 4 (Lokwani et al., 2020). Despite the well-established substrate recognition sites, the studies did not highlight the effect of variation associated with the substrate's recognition sites due to variation.

2.1.5 The catalytic cycle of the cytochrome P450

Several reports have shown that cytochrome P450 enzymes are very active oxidants and can hasten the rates of reaction of substrates through the oxidation of the compound being targeted (Cook *et al.*, 2016; Sellés Vidal *et al.*, 2018). Other reported reactions catalysed by CYP enzymes are the dealkylations and hydroxylation (Girvan *et al.*, 2011; Guengerich, 2018).

http://etd.uwc.ac.za/

Figure 2.4 shows the graphical illustration of the cytochrome P450 catalytic process with a detailed analysis of each step (Cook *et al.*, 2016). The oxidation occurs with the insertion of one of the atoms in its oxygen molecule into the substrate and a subsequent reduction of the other oxygen atom to give a water molecule. This is applied with the use of the two electrons that are provided by NAD(P)H through protein reductase (Cook *et al.*, 2016).



WESTERN CAPE

Figure 2.4: Graphical illustration of Cytochrome P450 catalytic process.

The arrows show the progression of the various steps in the process. (Adapted from Cook *et al.*, 2016)

Hence, CYPs are called monooxygenase enzymes because one oxygen atom is transferred from the initial oxygen molecule to the oxidised substrate as shown in Figure 2.4 (Cook *et al.*, 2016a; Sellés Vidal *et al.*, 2018). These responses are reported to cause a conformational change as not only does the shape of the active site change, but it confers a change at the reductase area as well, especially with adequate catalysts which converge the flavins closer together to aid the movement of electrons down the chain (Girvan *et al.*, 2011). The transformation of iron from

http://etd.uwc.ac.za/

a low spin to a high spin state was suggested to result in this (Girvan *et al.*, 2011). Previous reports have shown that the energy state concerning the iron when it binds to the substrate defines the spin state, with at least nothing less than one unpaired electron in the outer orbital that reflects the higher identified energy state, hence the transition from a lower to a higher spin state (Girvan *et al.*, 2011; Guengerich, 2018).

The energy is reported to be produced when iron attaches to the active site of the substrate and it is known as the spin state, with at least one unpaired electron in the external orbital which makes it have elevated energy and a transition from a low to an elevated spin (Cook *et al.*, 2016). This energy shift was reported to be responsible for the widely observed "Difference spectra " found when calculating the complex of the enzyme-substrate spectrophotometrically (Cook *et al.*, 2016). The spectrum of variation is reported to be the two shifts in the absorbance of the complex; namely, a rise is observed with an absorbance at 390 nm combined with a reduction of absorbance 420 nm absorbance (Cook *et al.*, 2016). Otherwise, Degregorio *et al.* (2017) reported that if carbon monoxide creates a decreased P450, the absorption moves to 450 nm.

Barr *et al.* (2020) showed that conformational alteration occurs when the substrate binds, and it triggers the donation of primary electrons with NAD(P)H present. An electron moves from NAD(P)H to the first molecule of flavin, then FMN, and finally the heme area, where the iron in the porphyrin ring is accepted (Barr *et al.*, 2020). This moves Iron from the initial ferric (Fe3+) state to the ferrous (Fe2+) state. Once there is an initial reduction, then molecular oxygen can combine covalently with the heme iron at the distal axial location (Cook *et al.*, 2016a). The iron-oxygen bond was stated to be dissociated to give superoxide radical at this point (Cook *et al.*, 2016a). The iron-oxygen bond was reported to be very reactive and capable of disrupting the catalytic cycle, and then the second stage reduction occurs which results in the formation of a peroxide group that is negatively charged, in which the iron-bound oxygen gets to be more negatively charged than it was in the previous state before it donated electrons (Cook *et al.*, 2016a). The peroxo group

http://etd.uwc.ac.za/

was reported to be greatly nucleophilic, and this makes the duration of the intermediate stage very short (Girvan *et al.,* 2011). Girvan *et al.* (2011) suggested that the complex is quickly protonated twice because of the hydrogen bonding of the surrounding water and the amino acid side chains. This results in the splitting of the dioxygen which attaches to the given hydrogen resulting in the release of the water molecule and the formation of the intermediate cytochrome P450 compound I (Girvan *et al.,* 2011). This intermediate was reported to have a Fe4+ group due to a donated electron pair from the nucleophilic peroxide group and has a double-bonded oxygen atom (Girvan *et al.*, 2011). Cytochrome P450 compound I (Cpd-I) is generally recognized as the final intermediate stage in the P450 catalytic cycle and is thought to be the key oxidising agent that produces the hydroxylated substance (Rittle and Green, 2010). The bound substrate was stated to be oxidised by the Cpd-1 and thereafter released as a hydroxylated substance. Water is then reported to bind to the active site and returns the enzyme to its original resting state (Rittle and Green, 2010). The studied and reported catalytic cycle of CYP provides a broader knowledge and understanding of the conformational changes of each step in the cycle. However, the process of catalytic termination was not involved in the cycle.

UNIVERSITY of the
WESTERN CAPE

2.1.6 The process of catalytic termination

The process of catalytic cessation was thereafter reported to occur in three different major ways after the oxygen has bound to the heme iron leading to dissociation (Polic, 2018). First is through an auto-oxidation shunt process (Figure 2.4), which involves the degradation of the intermediate compound; peroxo-ferrous, and the formation of anion superoxide radical. Furthermore, the addition of protons to the hydroxyl ferric intermediate results in the dissociation and production of hydrogen peroxide hence this process occurs through the peroxidase shunt process. Finally, the termination occurs with the oxidase shunt pathway with the release of a water molecule after the deprotonation of Cytochrome P450 compound I and double reduction (Polic, 2018; Zhao *et al.*, 2021).

http://etd.uwc.ac.za/

Several studies have shown that there are several identified amino acids which are located around the heme region that play an important function in the passage of electrons required for the cleavage of the dioxygen. One of such revealed the significance of leucine amino acid located at the "cysteine pocket" of the CYP450, the mutation of the leucine with proline results in an increase in the uncoupling rates of external substrates and a subsequent increase in the rate of enzymatic activity (Pedroso, Zampieri, and Donato, 2015).

2.1.7 Genetic polymorphism and reported phenotypes in CYP

For many years, several landmarks of pharmacogenetic research have taken place and reported with other drug-metabolising enzymes such as N-acetyltransferase-2 and pseudocholinesterase. However, there were none in Cytochrome P450 enzymes until 1975 when the CYP2D6 incident occurred at St. Mary's Hospital Medical School in London (Ahmed *et al.*, 2016; Tornio and Backman, 2018; and Mitchell, 2020). This led to the discovery of two phenotypes: the "poor" and the "extensive" metabolizers in genetic polymorphism that occur in drug oxidation. Further discoveries of numerous CYP2D6 alleles and different distinct variants occurred that can be mostly associated with four metabolism phenotypes: poor, intermediate, extensive, and ultra-rapid (Lam, 2019; Smith and Mitchell, 2020).

UNIVERSITY of the
WESTERN CAPE

Del Re *et al.* (2016) reported that the metabolism of certain drugs varies depending on the phenotype encoded by the genes. Klein and Zanger (2013) and Jørgensen (2019) also thought that the ability of an individual to metabolise drugs is due to the paired individual allele inherited from the parents and this allele could either be the wild type or the variant. They further stated that the wild type is referred to as the "normal" and is found mostly in the general population while the variant, also known as the defective allele, may be responsible for reduced or no activity (Klein and Zanger, 2013; Jørgensen, 2019).

Similarly, it was reported that the intermediate metabolizers are characterised by decreased enzymatic activity. Enzyme activity may become higher than normal when there is gene amplification or duplication in more than two gene copies of the

http://etd.uwc.ac.za/

wild-type alleles. These individuals are referred to as ultra metabolizers and they occur in some cases (Figure 2.5) (Taylor *et al.*, 2020; Arafah *et al.*, 2021).

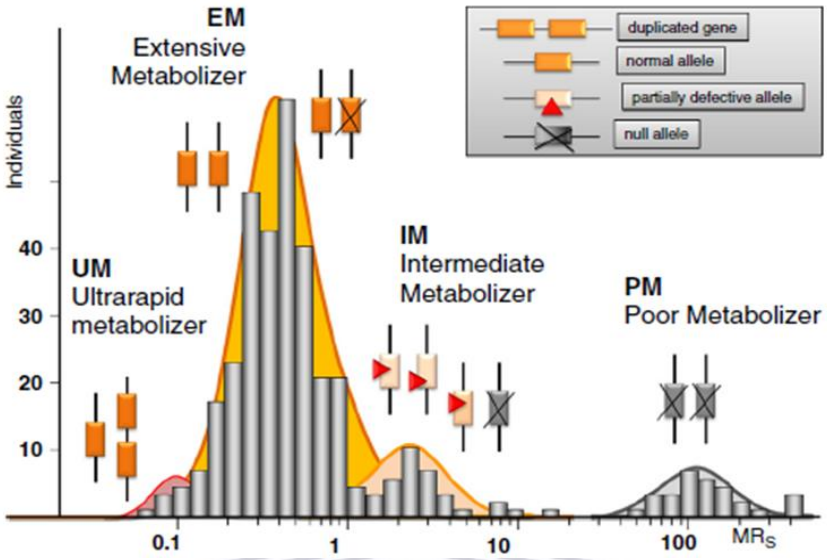


Figure 2.5: A graphical representation of phenotype and genotype of Sparteine oxidation in a German population of three hundred and eight individuals. The different phenotypes (ultrarapid, extensive, intermediate, poor metabolizers) encoded by the different alleles (duplicated, normal, partially defective, null) are shown (Adapted from Zanger and Schwab, 2013).

UNIVERSITY of the
WESTERN CAPE

2.1.8 Single nucleotide polymorphism in CYP3A4

There is a rather large inter-individual variability in CYP3A4 activity, which has been linked to genetic polymorphism (Denisov *et al.*, 2021; Kumondai *et al.*, 2021). Many recent studies have postulated that single nucleotide polymorphisms (SNPs) are one of the major molecular mechanisms that affect the uniform expression and the activity of CYP3A4 (Berno *et al.*, 2014; Mulder *et al.*, 2021). With the reported large inter-individual variability in CYP3A4 activity, a further study into the structural effects could also provide valuable knowledge of the effect of SNPs on the uniform expression as well as the structural conformation.

http://etd.uwc.ac.za/

2.1.9 Functional variants of CYP3A4

The first identified variant of CYP3A4 was assigned in the Single Nucleotide Polymorphism database (dbSNP) as rs2740574 (CYP3A4*1B) with an A>G substitution at the 5' promoter potion. It was found in 66% of African Americans, and 4% of the European population. (Stillemans *et al.*, 2018). Wang *et al.*, (2011) identified the CYP3A4*22 (rs35599367 C>T) variant, which has been reported to occur mainly in 3% of African Americans and 5% of Europeans. The 664T>C variant (rs55785340, CYP3A4*2) was reported to have a frequency of 0.03 in Caucasians. CYP3A4*4 were identified in three subjects among 102 Chinese selected individuals (Y. Zhou *et al.*, 2017).

The detailed annotation for the available SNPs is well documented in the Human Cytochrome P450 (CYP) Allele Nomenclature Database (www.cypalleles.ki.se) and (dbSNP). More than 20 SNPs have been reported to take place within the coding region of CYP3A4 (Kumondai *et al.*, 2021). Variants of CYP3A4 have been reported with clinical implications on its enzymatic activity.

2.1.10 Recent reports on the clinical implications of single nucleotide polymorphism in the enzymatic activities of CYP3A4 variants

UNIVERSITY of the
WESTERN CAPE

It was reported that CYP3A4*22 has a profound impact on the area under the curve (AUC) ratio in volunteers treated with atorvastatin (Mulder *et al.*, 2021). With the attainment of maximum lipid control, as shown in the volunteers as well as in clinical studies, there was a 1.7 to 5-fold reduction in the statin dose of T-allele carriers when compared with the non-T carriers (Mulder *et al.*, 2021). There was also an indication of the relationship of the CYP3A4*22 variant with an observed lipid-lowering response that occurs with the use of simvastatin in a clinical study (Ragia *et al.*, 2015). Reduced CYP3A4 activity was observed in renal transplant recipients who were carriers of the T-allele with about 33% reduction in the mean daily dose that is required for the blood concentration of Tacrolimus as compared to those of the wild type (Abdel-Kahaar *et al.*, 2019).

http://etd.uwc.ac.za/

Findings revealed that rs67666821 (CYP3A4*20) which is a variant that does not include heme and is therefore not functional (Saiz-Rodríguez *et al.*, 2020). In addition, Klein *et al.* (2012) established that this variant was responsible for the reduction in the protein levels when the study was done with 150 liver samples, however, it was not proven with multiple sample testing.

CYP3A4*1B was linked to lower dose-adjusted trough concentrations (C0/D) and higher tacrolimus and cyclosporine dosage needs. It was also linked to a decreased likelihood of dose reduction or switching therapy while taking simvastatin (Ben-Fredj *et al.*, 2020). The summary of recent reports on clinical implications of single nucleotide polymorphism in the enzymatic activities of CYP3A4 variants is presented in Table 2.1.

The clinical implications of CYP3A4 variants are increasing with available drugs of different classes. However, none of the report identified the structural implications of the identified variations. To enable the study of the structural properties and dynamic characteristics of the reported variants of CYPs enzymes, especially the CYP3A4 enzymes, it is expedient to model the three-dimensional structures of the implicating variants.

UNIVERSITY of the
WESTERN CAPE

Table 2.1: The summary of recent reports on clinical implications of single nucleotide polymorphism in the enzymatic activities of CYP3A4 variants.

Identified variants with significantly reduced activity	Identified variants with significantly increased activity	Identified variants with lost Enzymatic activity	Identified variants without any effect activity	Affected drug	References
CYP3A4*11 CYP3A4*12 CYP3A4*13 CYP3A4*20 CYP3A4*23	CYP3A4*15 CYP3A4*29	CYP3A4*6 CYP3A4*26 , CYP3A4*30	CYP3A4*28 CYP3A4*34	Quinine	Zhou et al. (2019)

http://etd.uwc.ac.za/

CYP3A4*8 CYP3A4*17 CYP3A4*21					
CYP3A4*2,CYP 3A4*5, CYP3A4*9, CYP3A4*16, and CYP3A4*24	CYP3A4*29,CYP 3A4*31, CYP3A4*32, CYP3A4*34	CYP3A4 *17 CYP3A4 *30		Lidocaine	Fang et al. (2017)
CYP3A4*9, CYP3A4*14, CYP3A4*16, CYP3A4*19, CYP3A4*23, CYP3A4*24, CYP3A4*28, CYP3A4*32	CYP3A4*11CYP3 A4*2, CYP3A4*3, CYP3A4*11, CYP3A4*29, CYP3A4*33	CYP3A4 *6, CYP3A4 *7, CYP3A$ *8, CYP3A4 * 12, CYP3A4 *13, CYP3A4 * 17, CYP3A4 * 18, CYP3A4 * 20	n CYP3A4 *4, CYP3A4 *5, CYP3A4 * 10, CYP3A4 * 15, CYP3A4 * 31, and CYP3A4 *34 r	Acalabrut inib	Han et al. (2021) a
CYP3A4*6, CYP3A4*17, CYP3A4*20, CYP3A4*30				Loperami de	Lin et al., 2019
(CYP3A4*2, CYP3A4*3, CYP3A4*4, CYP3A4*5, CYP3A4*.7, CYP3A4*8, CYP3A4*9, CYP3A4*10, CYP3A4*11, CYP3A4*12, CYP3A4*13, CYP3A4*16, CYP3A4*17, CYP3A4*18, CYP3A4*20, CYP3A4*23, CYP3A4*29, 27 CYP3A4*31, CYP3A4*32, CYP3A4*33 and CYP3A4*34)	CYP3A4*14 and CYP3A4*15 w		CYP3A4 *19 and CYP3A4 *24. CYP3A4 *28.	Brexpipra zole	(Chen et al., 2020

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

CYP3A4*2, CYP3A4*7, CYP3A4*16, CYP3A4*18, CYP3A4*19, CYP3A4*23, CYP3A4*24, CYP3A4*28, CYP3A4*29, CYP3A4*31, CYP3A4*34	CYP3A4*6, CYP3A4*17, CYP3A4*20, CYP3A4*21, CYP3A4*26, CYP3A4*30)			Oxycodone	Cai et al. (2021)

2.2 Overview of the developments in the prediction and validation of three-dimensional protein structures

2.2.1 The advances in homology modelling

Most recent reports have shown that the knowledge of biological systems and the operation of protein complexes can be known with a detailed description of their protein interaction and an overall three-dimensional structure (Townshend et al., 2019; Bryant, Pozzati, and Elofsson, 2022). Jha *et al.* (2022) further stated that the modulation of protein networks and complexes is also made possible with the knowledge of the three-dimensional structure. Reports have shown that the three-dimensional structure of proteins informs the understanding of protein function on a molecular level (Fuller, Burgoyne, and Jackson, 2009; Nim *et al.*, 2016).

According to Jha *et al.* (2022) and Denisov *et al.* (2021), it has been established that the major insights into the active site regions, the portions for the binding of the inhibitors and substrates, and the recognition of individual residues involved in the binding of ligand are obtained via the understanding of the three-dimensional (3D) structures of the CYP enzymes. In addition, Selvam *et al.*, (2017) reported that the knowledge of the structure of CYP enzymes may be maximised in findings relating to the structure-function relationships, altered enzymatic activity associated with genetic polymorphism, or investigations that relate to drug-drug interactions with emphasis on its molecular basis. It was also reported that the efficiency of the CYP

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

enzymes can also be facilitated with adequate knowledge of the structures of CYP (Selvam *et al.*, 2017; Haddad, Adam, and Heger, 2020).

Different methods have been developed such as co-evolution methods, which have been reported to utilise the correlated amino acid mutations in deep multiple sequence alignments (MSA) and can be used for the identification of the protein interaction with the application of the sequence details alone (de Oliveira and Deane, 2017). Electron-Microscopy which was developed recently had been utilised in the computational prediction of macromolecular assemblies (Marsh and Teichmann, 2015; Soni and Madhusudhan, 2017).

However, it has been shown that high quality models are only obtained when models are made from available multi-chain template information (Peterson *et al.*, 2018). Waterhouse *et al.* (2018) reported that interacting interfaces are always conserved in homology complexes as observed in increasing available experimentally modelled protein structures. It was further stated that the templates of the known protein-protein interactions are available in homology modelling (Waterhouse *et al.*, 2018).

According to Akapo *et al.* (2021), homology modelling was shown to be a useful tool in the close study of the interaction between substrates and Cytochrome P450 enzymes when crystal structures are not present. These observed reasons as discussed, have provided the principle for the comparative or homology modelling of protein complexes.

Homology Modelling, also known as Comparative Modelling, has been defined as the prediction of a protein's three-dimensional structure based on its amino acid sequence (Nikolaev et al., 2018; Barcelos *et al.*, 2020;). The literature has highlighted that the modelling technique can be broken down into the following steps: First, a suitable template(s) relevant to the target sequence is chosen from the Protein Data Bank (PDB) (the sequence identity must be greater or equal to 30%). Second, using the Basic Local Alignment Search Tool (BLAST), an alignment of the target sequence to the template(s) is generated. Third, the three-dimensional

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

model's coordinates are constructed using the alignment and the template structures (Haddad, Adam, and Heger, 2020).

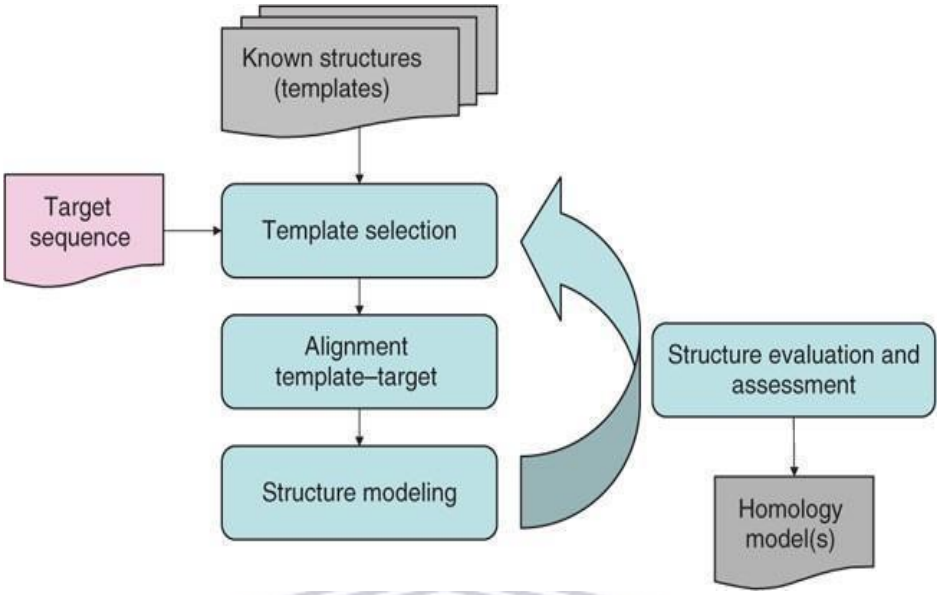


Figure 2.6: The major steps involved in homology modelling of protein structures (Adapted from Bordoli *et al.*, 2009)

Reports have shown that alternative alignment approaches, such as profile-profile alignments, Hidden Markov models (HMMs), and position-specific iterated BLAST (psi-BLAST), are employed to eliminate shifts and gaps in the case of low homology below 35 percent sequence identity (Gupta *et al.*, 2020). Recently, it was reported that the application of AlphaFold based on Machine learning is a useful tool to design new protein folds (Jumper *et al.*, 2021; Wei, Guo-Wei 2021). Vatansever *et al.* (2021) further stated that the AlphaFold invention has made enormous gains in predicting structures by demonstrating the power of Machine Learning in identifying patterns in fundamental sequences that predict three-dimensional folds with high accuracy. Jumper *et al.* (2021) reported that the core of AlphaFold is a neural network that is based on many structures in the Protein Data Bank to estimate the distributions of distances between the C atoms of pairs of residues in a protein and build an artificial force field to direct folding. Gao *et al.* (2020) in their study established that AlphaFold also uses sequence databases and

http://etd.uwc.ac.za/

multiple sequence alignments extensively. However, researchers are doing further study to the Alphafold, as not all the three-dimensional protein structures have been stated to be established by the Alphafold.

Although several crystal structures of CYP3A4 are available in various repositories; more than 78 crystal structures are available on PBD with different resolutions which could all serve as templates in homology modelling, however the structures of the CYP3A4 variants models have been understudied. Homology modelling could therefore be used effectively in the structural prediction of CYP3A4 variants. However, the quality of the predicted three-dimensional protein structures must be checked before it is used for further analysis

2.2.2 The quality estimation of modelled protein structures

Several quality assessment tools are available and have been used to estimate the quality of modelled protein structures, this includes Verify3D, QMEAN, and the PROCHECK (Reddy, 2020; Adebiyi and Olugbara, 2021; Akapo *et al.*, 2021).

Table 2.2 further gives a summary of the reported available quality assessment tools.

Table 2.2: The summary of available quality assessment tools

Quality Analysis Package	Method Employed	Links	References
Verify 3D	It determines an atomic model's (3D) compatibility using its amino acid sequence (1D) by using a structural class depending on its location and environment in comparison with results from good structures	https://servicesn.mbi.ucla.edu/Verify3D/	(Tan *et al.*, 2020)
ERRAT	It evaluates the statistics in interactions that are non-bonded that exist in distinct atom types and derives plots to show the value of the error function against the location of the 9-residue	https://servicesn.mbi.ucla.edu/ERRAT/	(Panda *et al.*, 2021)

http://etd.uwc.ac.za/

	sliding window obtained by the comparison in statistics from structures that are well defined.		
PROCHECK,	Analyse the residue-by-residue geometry as well as the overall structure's geometry to determine the stereochemical quality of the protein structure	https://servicesn.mbi.ucla.edu/PROCHECK/	(Park *et al.*, 2020)
ProSA	Offers a number (Z-Score) that indicates the model's overall quality; when the score falls within the range for typical protein structures estimated within the same length, the built model is most likely error-free.	https://prosa.services.came.sbg.ac.at/prosa.php	(Okella *et al.*, 2020)
PROVE	It estimates the atom volume of macromolecules with an algorithm that regards atoms as hard spheres and determines the model's statistical Z-score deviation from highly resolved structures found in the PDB	https://servicesn.mbi.ucla.edu/PROVE/	(Sahay, Piprodhe, and Pise, 2020)
WHATCHECK	Created from a subset of WHATIF protein verification tools that thoroughly examines various stereochemical properties of the model's residues.	https://servicesn.mbi.ucla.edu/WHATCHECK/	(Yenenler, Gerlevik, and Sezerman, 2020)
QMEAN	It allows the server to rank the models by providing access to two scoring algorithms for estimating the quality of structured protein (mean score, Z-score). It also identifies the regions of the protein structure that are possibly unreliable.	https://swissmodel.expasy.org/qmean/	(Studer *et al.*, 2020)

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

2.2.4 The relevance of molecular docking in CYP modelled protein structures

Several reports have established that molecular docking exists as an in-silico approach for the flexible generation and sampling of possible ligand binding poses against the ligand-binding region of receptor protein structures (Salmaso and Moro, 2018; Rahman *et al.*, 2019). The presence of a three-dimensional generated high-resolution protein structure obtained by computational methods such as NMR, homology modelling, and X-ray crystallography is, therefore, a must to obtain the possible poses between the ligand and the binding site of the protein (Butt *et al.*, 2020).

The docking capacity to imitate ligand-receptor recognition at the atomic level might provide useful insight into complex and empirically challenging processes, like enzyme reaction mechanics or ligand-receptor interaction (Salmaso and Moro, 2018; Munjal, Shukla, and Singh, 2021b). A wealth of research has described how molecular docking interaction with CYP3A4 structures has been used to predict potential interacting compounds with the CYP3A4 enzyme (Harahap et al.,2022; Patil *et al.*2022). There is little or no information on the molecular docking interactions of ligands with the different protein structures of variants of CYP3A4.

UNIVERSITY of the
WESTERN CAPE

2.2.5 The significance of molecular dynamics in CYP3A4

Reports have established that protein is both dynamic and densely folded (DuBay, Bowman, and Geissler, 2015). It has also been established that the important cellular function of a protein can be determined most importantly by its dynamic properties and not only by the notable relative rigid structures (Yang *et al.*, 2014; Tokunaga *et al.*, 2020). The catalytic activity of enzymes including cytochrome P450 is being altered by mutation situated away from the active sites of the enzymes (Osuna, 2021). Studies have shown that mutations within the cytochrome P450 respective active sites exhibit dynamically correlated motion (Tyukhtenko *et al.*, 2018).

http://etd.uwc.ac.za/

The dynamic correlation exists between residues, and this can be detected through molecular dynamics simulations (Barbera *et al.*, 2018). Molecular dynamics provides individual particle function to time, and the difficulty in the experimental access of protein global molecular motions has been solved with molecular dynamics techniques (Vidal-Limon, Aguilar-Toalá, and Liceaga 2022). The simulation accuracy has been greatly increased with the improvement in force fields and advanced presentation of the solvent (Stroet, 2018).

Molecular dynamics simulations work based on Newton's equations of motion to generate subsequent conformations of the system to time with known potential energy, conformational coordinates, and a set of starting velocities (De Vivo et al., 2016; Nair, McKinnon, and Miners, 2016). Molecular dynamics simulation enables the determination of the force affecting individual atoms to determine the new sets of conformations by projecting the system forward with time (Gkeka *et al.*, 2020). The bimolecular force fields undergo the process of parametrization to fit well in the calculations of Quantum mechanical and experimental data of spectroscopy (Oostenbrink *et al.*, 2004). The definition of all bonded and non-bonded terms is done during parameterization, and the way each of the force fields is parameterized differs with similar outputs. Major programs used in MD include CHARM (Brooks) GROMACS (Abraham *et al.*, 2015), and NAMD (Phillips *et al.*, 2005).

UNIVERSITY of the
WESTERN CAPE

Other experimental methods can be used to understand the dynamic nature of CYP structures, however, there are certain known limitations attached. Nuclear magnetic resonance (NMR) does provide knowledge of the protein dynamic but the protein size of approximately 20kDa is a limitation to this approach (Yee *et al.*, 2014) which does not apply to molecular dynamics. Since CYP3A4 and other CYPs enzymes have a larger size, molecular dynamics have shown that the observed channels that connect the heme active domain to the protein surface as observed in the X-ray structure underwent conformational changes (Johnson and Stout, 2013; Yee *et al.*, 2014). Hence, molecular dynamics simulation has been generally accepted by researchers for major studies in CYP protein structures.

http://etd.uwc.ac.za/

The high flexibility of the active sites of CYP enzymes has been proved in studies, and this flexibility is associated with the conformational changes when the ligand binds with the subsequent ability of the enzyme's active site to structurally interact with various compounds (Ekroos and Sjogren, 2006; Zhao *et al.*, 2006). Hence, this major reason necessitates and predicts the importance of the study of the CYP enzymes in line with their dynamic behaviour effects on ligand binding, to have a better understanding of its CYP enzyme structure-function (Xiao et al. 2020; Ducharme *et al.* 2021). With Molecular dynamics simulation, the flexibility of the CYP structures can be modelled which could also elucidate the influence of flexibility on ligand binding (Salo-Ahen *et al.*, 2021).

With the capture of the CYP structures dynamics, several numbers of the chemical, physical and biological properties of the CYP protein enzymes can be elucidated such as its molecular interactions, binding of ligand to the receptor, folding of the protein, and so on.

The 214-220 amino acid region of CYP3A4 makes up the F' helix which is a part of the region F-G previously recorded to be hidden in the CYP3A4 membrane (Zhao *et al.*, 2006; Denisov, Shih, and Sligar, 2012). Molecular dynamics simulation (MD) has revealed that there is a huge movement occurring when the F, F', G', and the G helices interact with the lipid bilayer (Denisov, Shih, and Sligar, 2012). Studies have proved the use of MD to make findings on the major features present in the structure of CYP3A4 (Šrejber et al., 2018; Sabiu and Idowu 2022). An MD study performed by Hendrychová *et al.* (2011) showed that all CYP enzymes have an active site with notable flexibility with a difference of more than 50% in the binding site volume with time. MD has also proved the elasticity of CYP3A4 and its ability to recognize and accommodate large numbers of different substrates with some of the substrates having a higher molecular mass when compared with other compounds (Hendrychova *et al.*, 2012). The application of molecular dynamic simulations in CYP3A4 analysis is continually increasing.

http://etd.uwc.ac.za/

2.3 Conclusion

The reported interindividual variability in drug response has been linked to the polymorphisms of the drug-metabolising enzymes, especially the cytochrome P450 (CYP). The most abundant isoform of CYP, CYP3A4, has shown greater importance and uniqueness in its metabolic activity. Its polymorphic nature has been reported with the discovery of different CYP3A4 variants. From the literature reviewed, several investigations brought an understanding of the clinical implications of the effects of polymorphism, mostly in the enzymatic activity in some of the discovered variants.

There is an advent of knowledge on the structure of CYPs enzyme, with a focus on reports on the uniqueness of CYP3A4 structural components. Understanding the flexibility and conformational changes of the CYP3A4 complex with ligands has undoubtedly advanced the knowledge of the major features involved in the enzymatic activity. The major components of the CYP3A4 structure prove their importance in the metabolic process. However, the available information from the literature focuses only on the effects of polymorphism on enzymatic activity. There has been a paucity of information on the effects of polymorphism on the CYP3A4 protein structure. Little or no information exists to show the effect of polymorphism on the dynamic properties of the implicating enzymes, especially on major reported structural components involved in the metabolic process.

Several reports have shown the relevance of homology modelling in the prediction of a protein's three-dimensional structure based on its amino acid sequence. This is further supported by information on the increasing available quality assessment tools to estimate the quality of the modelled protein structures. Hence, an assured possibility in the prediction and quality validation of the three-dimensional protein structures of available CYP3A4 variants. Further analysis of the structural and dynamic effects of polymorphism on the modelled structures, therefore, becomes a possibility.

Applications of molecular dynamics bring light to the dynamic properties and flexibility of CYP variants modelled protein structures. The literature reviewed

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

shows the advantage of molecular dynamics simulation in CYP. This information buttresses its usefulness in the study of the effect of polymorphism on the dynamic properties of the available CYP3A4 variants protein structures.

http://etd.uwc.ac.za/

Chapter 3

Research materials and methods

This chapter describes the research study methodology. It starts with a synopsis of the materials used for the research study (Figure 3:1). Thereafter, a detailed description of the research methods utilized was discussed.

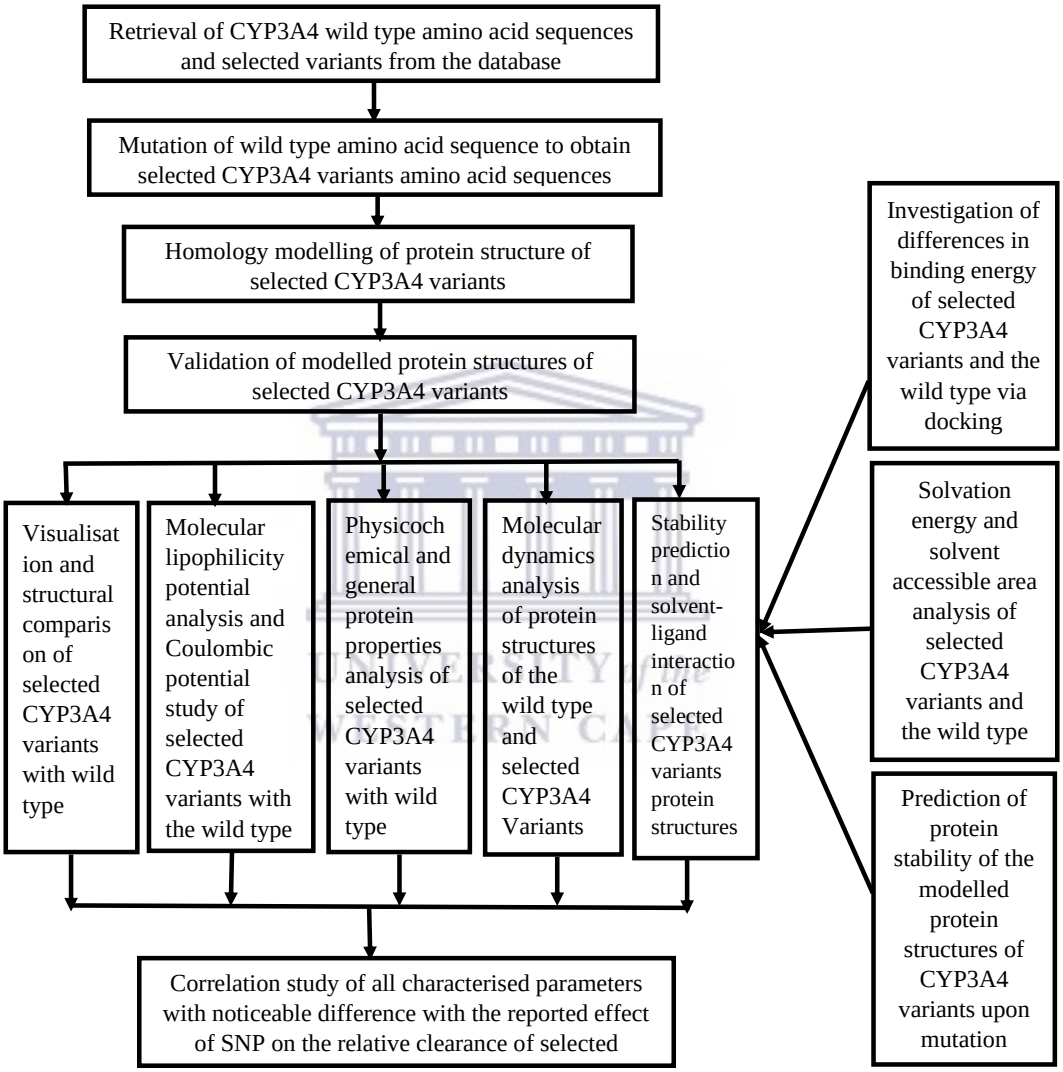


Figure 3.1: A flowchart showing the synopsis of the method used in the current research.

3.1 Research materials

All computational analysis was carried out on windows 10 operating system with processor 3.1GHz Quad-core intel 17, memory 16GB 2133MHz, and windows 10

http://etd.uwc.ac.za/

operating system, with 8GB memory, Intel Core i5-7200U processor, and 64-bit. The biological databases employed are the National Centre for Biotechnological Information (NCBI)(https://www.ncbi.nlm.nih.gov/), Protein Data Bank (PDB) (https://www.rcsb.org/),SNPedia(https://www.snpedia.com/),SwissModel (https://swissmodel.expasy.org/) and the Human Cytochrome P450 Allele Nomenclature (CYP-allele) (http://www.cypalleles.ki.se/). Software: MOE program (2019.01), OSIRIS DataWarrior (version 5.5.0), UCSF Chimera (version 1.14), UCSF ChimeraX (version1.3), Git for Windows (version2.33.0), Graphpad prism version 9.0 were all utilised in this study.

3.2 Retrieval of CYP3A4 wild-type amino acid sequence, and mutation of the wildtype sequence with Bio.SeqIO module on Python package (biopython)

The methods below outlined the steps taken to obtain (from the NCBI database) and mutate the amino acid sequence of the wild type. It includes retrieval of the CYP3A4 wild type sequence, collation of information on amino acid sequences of selected variants on databases, and the method used in the mutation of the amino acid sequences of the wild type required for the modelling of the selected variants' protein structures.

UNIVERSITY of the
WESTERN CAPE

3.2.1 CYP3A4 wild-type amino acid sequences retrieval

The CYP3A4 wild-type amino acid sequence (NP_059488.2) was retrieved from the (National Centre for Biotechnological Information) NCBI database (Sayers *et al.*, 2019) in the Fasta format (accessed in December 2020).

3.2.2 Collation of information of amino acid sequences of the selected CYP3A4 variants on databases

Four CYP3A4 allelic variants were selected based on the report of SNPs on the functional assessment of CYP3A4 allelic variants in lidocaine metabolism (Fang *et al.*, 2017). Twenty-two allelic variants were subjected to the in vivo study; however, the four chosen variants were those that indicated a significant difference in the clearance of lidocaine relative to the wild type as reported by Fang *et al.* (2017).

http://etd.uwc.ac.za/

CYP3A4*2 and CYP3A4*24 have values significantly lower (27.93 and 30.29 respectively) than the relative clearance value of the wildtype (100), while CYP3A4*11 and CYP3A4*23 have values that are significantly higher (213.61 and 206.96 respectively) than the relative clearance value of the wild type (100).

The position of mutation and information of the substituted amino acid of the selected variants of CYP3A4 for this study were retrieved from SNPedia (a wiki resource that collects information on single nucleotide variants), and on the Human Cytochrome P450 Allele Nomenclature (CYP-allele) website) database (Gaedigk *et al.*, 2018).

Table 3.1: The selected variants with an indication of the position of mutation on the amino acid sequences of CYP3A4 selected variants

Variant name	Mutation on Amino Acid Sequence	References
CYP3A4*1	-	(Gonzalez *et al.*, 1988; Córdova *et al.*, 2017)
CYP3A4*2	S222P	(Sata, 2000; Tang *et al.*, 2020)
CYP3A4*11	T363M	(Murayama *et al.*, 2002; Han *et al.*, 2021)
CYP3A4*23	R162W	(Drögemöller *et al.*, 2013; Chen *et al.*, 2020)
CYP3A4*24	Q200H	(Fang *et al.*, 2017; Tang *et al.*, 2020)

The position of the mutation in the amino acid sequence and the substituted amino acid of the CYP3A4 selected variants (Table 3.1) were available in the database sources. The CYP3A4*2 variant has **S-P**, which implies that serine was replaced by proline. In the CYP3A4*11 variant, threonine was replaced by methionine **(T-**

http://etd.uwc.ac.za/

M). CYP3A4*23* **R-W** denotes that arginine was replaced by tryptophan CYP3A4*24, and Q-**H** denotes that glutamine was replaced by histidine (Table 3-1). All selected variants have amino acid changes located in the coding or exonic regions (www.pharmvar.org/gene/CYP3A4).

3.2.3 Mutation of the wildtype amino acid sequence using Bio.SeqIO module on Python package (bio-python)

The steps taken to obtain the amino acid sequences of the selected variants were outlined in this section (Figure 3.2). With the known position of the amino acid substitution as obtained from the SNPedia and on the Human Cytochrome P450 Allele Nomenclature (CYP-allele) website), the amino acid sequence of the wild type of sequence was mutated in Python with its underlying bio python package to obtain the selected variants sequences.

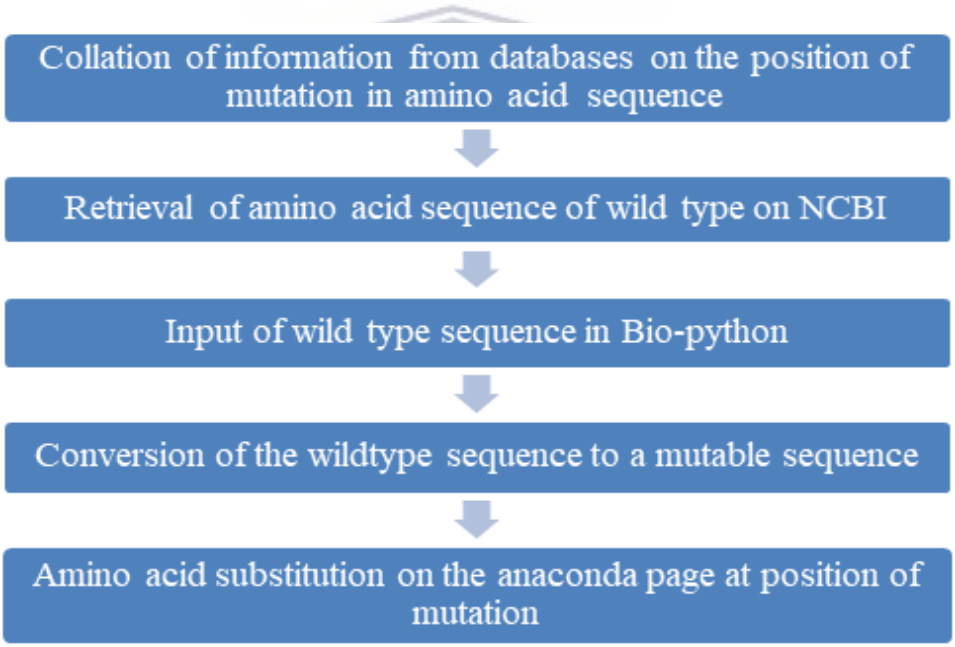


Figure 3.2: The process of generating amino acid sequences of selected CYP3A4 variants on the Python package

The bio python package was installed into the Jupyter notebook (Cock *et al.*, 2009; Leong, 2020). The Fasta format of the wild-type sequence was converted to a mutable sequence (initially in a string form), and the sequence became immutable

http://etd.uwc.ac.za/

as a Bio-Seq object type. The substitution of the amino acid was implemented at each indicated position of mutation of each variant from the SNPedia and on the Human Cytochrome P450 Allele Nomenclature (CYP-allele) website). The amino acid sequences of the selected CYP3A4 variants were generated and downloaded as a mutated sequence.

3.3 Homology modelling of the selected CYP3A4 variants

Homology modelling was employed to model the three-dimensional protein structures of the selected CYP3A4 variants. Homology modelling is the comparative modelling that generates a three-dimensional protein model from a targeted sequence. It uses a protein structure that has deduced experimental information and is evolutionarily related to the targeted sequence as template (Jacob K. *et al.,* 2017; Chikhale *et al.,* 2020). The SWISS-Model, an automated protein homology server, was used in the comparative modelling (Waterhouse *et al.,* 2018). The SwissModel was the pioneered automated modelling platform and has been undergoing continuous development since its release about 25years ago. Since the template of the wildtype structure was available on SwissModel, it was chosen as the platform for the homology modelling.

Crystal structure of CYP3A4 ligated to pyridine-substituted desoxyritonavir (4i4h.1A Cytochrome P450 3A4) was selected as the template from the protein data bank (PDB). It has a good resolution of 2.90 Å, a Rfree-0.263, a clash score of 17,and Ramachandran outliers of 0.4%. Each of the variant's amino acid sequences was separately inputted in Fasta format as the query. The protein structures of the four variants were built on SWISS-Model, with the crystal structure of CYP3A4 (4i4h.1A Cytochrome P450 3A4) as the template. The SWISS-Model used the computational structural biological method on its "openStructure" to model the protein structures of the selected variants, as described in previous studies (Biasini *et al.,* 2013; Bienert *et al.,* 2017).

http://etd.uwc.ac.za/

3.4 Evaluation and quality check of the modelled protein structures of the selected CYP3A4 variants.

The quality and accuracy of the predicted CYP3A4 selected variant protein models were evaluated for further structural analysis and characterization using the QMEAN (Qualitative Model Energy Analysis) (Pramanik *et al.,* 2017), the Verify 3D (Choudhary, Malik, and Tomar, 2020), and the ERRAT (Rahman *et al.,* 2019). These are tools on the SAVES server (The SAVES v6.0) (Akapo *et al.,* 2021). The Ramachandran plot was also plotted on PROCHECK (also a tool on the SAVES server) to evaluate the structural stereochemical property of the modelled protein structures of the CYP3A4 selected variants. The PROCHECK examines the overall model geometry and generates the Ramachandran plot after a modelled protein file has been inputted (Selvam *et al.,* 2017).

3.5 Structural alignment and molecular surface visualization of the CYP3A4 wild-type and the selected variants' models

The visualisation of each of the modelled protein structures was done on UCSF Chimera version: 1.14 (Pettersen *et al.,* 2021). The alignment of each of the modelled variants was done with the modelled structure of the wild type as the reference. Each of the modelled protein structures was inputted and superimposed on the reference protein structure. The Needleman-Wunsch alignment algorithm with matrix BLOSUM-62 on matchmaker UCSF Chimera (version: 1.14) at 2.0Angstroms was used for the structural comparison.

The molecular surface characteristics of the wildtype and the modelled variants proteins were estimated with UCSF ChimeraX version: 1.3 (2021-12-08) (Goddard *et al.,* 2018; Pettersen *et al.,* 2021). The molecular lipophilicity potential (MLP) for the protein structures of the wild type and the variants were estimated.

The Coulombic electrostatic potential (ESP) was also calculated on the UCSF ChimeraX version:1.3. Coulomb's law is used to compute the Coulombic electrostatic potential given the atomic partial charges and atomic coordinates:

$\varphi = \Sigma [qi / (\varepsilon di)]$.

http://etd.uwc.ac.za/

φ is the potential (which fluctuates in space), q is the atomic partial charges, d is the distances between the atoms, and ε is the dielectric, which represents media screening. At 298 K, the resultant potential is measured in kcal/(mole). The estimated ESP was used as colour representation on the molecular surfaces of the protein structures with pre-set colours ranging from red for the negative potential to white for positive potential and blue for neutral potential.

3.6 Investigation of the differences in the physicochemical and general protein properties of the selected CYP3A4 variants relative to the wild type

To investigate the notable differences in the protein properties of the CYP3A4 wildtype and the selected variants, the physicochemical properties, and general protein properties were; predicted. The analysis of all the physicochemical properties was carried out on the Expasy protparam tool (Suhaibun, Elengoe, and Poddar, 2020) and Molecular Operating Environment (MOE) 2019 (accessed on 2nd January 2022). The protein structures of the selected variant structures of CYP3A4 were loaded on MOE. The MOE protein property descriptor was used to calculate the molecular protein properties, and the descriptors were based on the entire protein shape (Gupta, Baudry, and Menon, 2022; Sripriya Akondi *et al.*, 2022). The property calculator was set at NaCl (salt), viscosity (0.89), dielectric (78), probe radius (1.8), hydrophobic cut-off (0.09), hydrophobic minimum area (50), charge cutoff (40), charged min area (40). The simulation environment was set at the default levels of pH 7.4, the temperature at 300 Kelvin, and a salt concentration of 0.1M. More general protein properties for the wildtype and the variants were also obtained from the patch surface analyser on Maestro 2019 (Sankar *et al.*, 2018).

3.7 Prediction of the protein stability of the modelled protein structures of variants.

The changes in the stability of the protein structures of the selected variants' sequences were predicted with I-Mutant 2.0. I-Mutant 2.0 is a support vector

http://etd.uwc.ac.za/

machine (SVM) tool that gives an automatic prediction of protein stability with a single point mutation. It indicates the effect of mutation on the protein structure as a destabilised or stabilised surface (López-López, Naveja, and Medina-Franco, 2019). The relative solvent accessible area and free energy change value (DDG) were also obtained from the support vector machine (SVM) tool. The relative solvent accessible area (RSA) value was determined from the protein structure by dividing the accessible surface area value of the altered amino acid region by the free residue surface (López-López, Naveja, and Medina-Franco, 2019). The PDB code of the wild type of protein structure was inputted with an indication of chain A and the position of the mutation. The new amino acid residue at the position of mutation was also inputted at a temperature of 25-celsius degrees and a pH value of 7. The direction (the DDG sign) of the protein stability changes and the DDG associated values were predicted and interpreted.

3.8 Determination of solvation energy of protein structures of variants

The PBEQ solver on the CHARMM-GUI (Brooks *et al.,* 2009), was used to generate the solvation energy of the protein structures of the wild type and the selected CYP3A4 variants. The PBEQ solver uses the Poisson-Boltzmann (PB) equations to generate the solvation energy with the PBEQ module in CHARMM (Jo *et al.,* 2008; Brooks *et al.,* 2009). Each of the protein structures (both the wild type and the variants) was uploaded in PDB format to obtain an output of the solvation energies.

3.9 Determination of binding score of the wild type and the selected CYP3A4 variants with lidocaine ligand

Molecular docking was done using the protein structures of the CYP3A4 wild type and the variants as the receptor. Lidocaine was used as the ligand, according to the study reported by Fang *et al.* (2017). The docking calculations were performed using the methods described in previous studies (Bahuguna *et al.,* 2020 Fadaka *et al.,* 2020).

http://etd.uwc.ac.za/

The modelled structures were prepared using the “Protein Preparation Wizard" of the Schrodinger suite (Schrödinger Release 2020–3: Maestro, Schrödinger, LLC, New York, NY, 2020. version 12.2). At this stage, a series of processes were carried out, which included the addition of hydrogen atoms, fixing of bond orders, removal of water molecules, optimization of the protonation states, and hydrogen bond network. The structures underwent energy minimisation using the OPLS-2005 force field, and all other parameters were kept at default (Kumar *et al.*, 2018).

For the ligand preparation step, the two-dimensional structure of lidocaine in 2-Dimensional format was downloaded from PubChem (https://pubchem.ncbi.nlm.nih.gov/) (Kim *et al.*, 2016), and prepared using the “LigPrep” module of Maestro. At this stage, the physiological and ionisation states were set at 7.0 pH using ± 2 with the Epik ionisation program. The OPLS-2005 force field was selected for minimization, keeping all other parameters as default. The binding site was specified to generate the receptor grid at the site of the co-crystallized ligand, using the ‘‘Grid generation’’ panel. Thereafter, the prepared ligand was docked into each of the prepared proteins using the ‘‘Ligand docking’’ panel with extra precision mode (XP) of the Glide module on the Schrödinger suite. The docked poses were analysed and the best pose of the selected CYP3A4 variants’ complex was selected based on their binding energy (BE).

UNIVERSITY of the
WESTERN CAPE

The Molecular Mechanics-Generalised Born Surface Area (MM-GBSA) approach as reported according to previous studies (Mali and Chaudhari, 2018; Rajagopal, Arumugasamy, and Byran 2019; Choudhary et al., 2020), was done with the Prime module of Schrödinger suite 2016-2, the OPLS3 power field and VSGB dissolvable model were used to analyse the calculations.

3.10 Molecular dynamics analysis

The dynamic properties of the modelled protein structure of the selected variants relative to the wild type of CYP3A4 were investigated with molecular dynamics simulation. Molecular dynamic (MD) simulation was done with Schrodinger software on Maestro molecular modelling platform (Schrödinger Release 2020–4: Maestro, Schrödinger, LLC, New York, NY, 2020. version 12.2.012). The MD

http://etd.uwc.ac.za/

simulations were carried out on a workstation having an Ubuntu platform, with Intel ® Xenon (R) Gold 6130 CPU @ 2.10 GHz × 64 processors, Quadro P620/PCIe/SSE2 graphics card, and 134.8 GB RAM. Molecular dynamics studies were carried out on the docked complexes of the wild type and the variants of CYP3A4 according to the method described in previous studies (Choudhary *et al.*, 2020; Fadaka *et al.,* 2020). This was performed using the Desmond module of Schrödinger with the OPLS3e force field (Harder *et al.,* 2016). The system was prepared using "System Builder" in Maestro, where the best-docked pose of the selected protein-ligand complex was bounded with a predefined TIP3P water model in the orthorhombic box. The volume of the box was minimised, and the overall charge of the system was neutralised by adding Na and Cl ions. The simulations were made under the Isothermal-Isobaric (NPT) ensemble. The temperature and atmospheric pressure of the system were kept constant at 310 Kelvin and 1.013 bar respectively, and the job was set to run at 100 ns. The system was minimized in a phase of 1000 steps with a restraint applied on solute atoms, and only water and salt ions were allowed to move. Minimization was followed by heating of the system from 0 to 298 K. The simulation interaction diagram tool of Maestro was used to analyse and sketch the plots and figures. The final trajectories was used for the calculation of the root mean square deviation (RMSD),root mean square fluctuations(RMSF), and radius of gyration (RG), secondary structure elements(SSE), and solvent accessible surface area (SASA).

UNIVERSITY of the
WESTERN CAPE

3.11 The correlation study between the studied structural and physicochemical effects of SNPs on selected variants and the reported effects of SNP on the relative clearance of the selected variants

To investigate the link between the studied structural and physiological differences in the variants and the reported functional differences in the pharmacokinetic parameter (intrinsic clearance) as reported by Fang *et al.* (2017), a correlation study

http://etd.uwc.ac.za/

was done on Datawarrior (version 5.5.0). Datawarrior is a free, versatile data analysis and visualisation program (López-López, Naveja, and Medina-Franco, 2019). The correlation plot was plotted, and Pearson's correlation coefficients were obtained from the plotted graph. The values obtained were interpreted and reported accordingly.

http://etd.uwc.ac.za/

Chapter 4

Results and discussion

This chapter presents and discusses the findings/results of the effects of single nucleotide polymorphisms (SNP) on the structural, physicochemical, and general protein properties of the modelled three-dimensional (3D) protein structures of selected CYP3A4 variants. The comparative study of the reported effect of SNPs on the relative clearance of lidocaine as reported by Fang *et al.* (2017) and the studied properties, to identify any significant correlation were also presented in this chapter.

The chapter starts with the outcome of the selection and generation of amino acid sequences of the selected variants used for the modelling of the variants' protein structures. A detailed report of the result of the CYP3A4 variants protein modeling and validation was reported and interpreted. The result from the investigation of the effect of SNPs on the structural and molecular properties of the selected variants was presented and discussed. The data obtained from the evaluation of the physicochemical and general protein properties of the selected variants were reported in this chapter. This was followed by the estimated effect of SNPs on the dynamic characteristics of the protein structures of the CYP3A4 selected variants.

UNIVERSITY of the
WESTERN CAPE

4.1 Selection and collation of sequences of the selected variants and the wild type of CYP3A4

The four selected CYP3A4 variants, namely, CYP3A4*2, CYP3A4*24, CYP3A4*11 AND CYP3A4*23, and their substituted amino acids, the location of substitution, and the reported clearance were summarized in Table 4.1

S222P indicates Serine (S) was replaced by Proline (P) at position 222, T363M indicates that Threonine (T) was replaced by Methionine (M) at position 363, R162W indicates that Arginine (R) was replaced by Tryptophan (W) at position 162, Q200H indicates that Glutamine (Q) was replaced by Histidine (H) at position 200.

http://etd.uwc.ac.za/

Table 4.1: The selected CYP3A4 variants with the location of substitution, the substituted amino acids, and the reported relative clearance.

Selected Mutants	Mutation location	Class of amino acid being substituted	Class of new amino acid present after substitution	Reported Relative clearance (%)
CYP3A4*1	Wild type	N/A	N/A	100
CYP3A4*2	S222P	S-(Hydrophobic) (non-polar)	P (Hydrophilic) (Polar)	27.93
CYP3A4*11	T363M	T-(Hydrophilic) (neutral)	M-(Hydrophobic)	213.61
CYP3A4*23	R162W	R- (Hydrophilic) (Basic)	W-(Hydrophobic)	206.96
CYP3A4*24	Q200H	Q- (Hydrophilic) (Acidic)	H- (Hydrophilic) (Basic)	30.29

UNIVERSITY of the
WESTERN CAPE

At the position of mutation, the newly substituted amino acid was shown with proline as the newly substituted amino acid at position 222 for variant 2 (see example in Fig.4.1). This process was also employed to obtain other selected variants' sequences that were studied. The substitution of each variant was confirmed with pairwise and multiple sequence alignment of all variants as indicated in Figure 4.2 to Figure 4.5.

The pairwise alignment of the wildtype and selected CYP3A4 variants' amino acid sequences indicated the missing residues at the positions of the mutation (S222P, T162M, R222W, Q200H for CYP3A4*2, CYP3A4*11, CYP3A4*23, CYP3A4*24 respectively) in the selected CYP3A4 variants (Fig 4.2 to 4.5). The study of the class of substituted amino acids for each of the selected variants indicated a change from a hydrophobic amino acid to a hydrophilic amino acid (CYP3A4*11, CYP3A4*23) or a change from a hydrophilic amino acid to a hydrophobic amino

http://etd.uwc.ac.za/

acid (CYP3A4*2) except for variant CYP3A4*24 (Table 4.1). In CYP3A4*24, the substituted amino acids were in the same class (though the substituted amino acid is more acidic than the originally positioned amino acid). It can be estimated that the substitution of amino acids with different characteristics (hydrophilicity and hydrophobicity) at an exon position might interfere with the protein properties and subsequently, the enzymatic activity of the implicating variants. This is consistent with previous reports by Drögemöller et al. (2013) and Cai *et al.* (2021) . The reports stated that an amino acid substitution of a hydrophilic amino acid to a hydrophobic amino acid could affect the function of the enzyme variants. Therefore, varied classes of the substituted amino acid and the new amino acid at positions of mutation might account for the observed differences in the rate of metabolism in the selected variants of CYP3A4 with lidocaine as reported by Fang *et al.* (2017).

Additionally, an increase of 113.6 and 106.96 was observed in the relative clearance of variants CYP3A4*11 and CYP3A4*23 (213.61 and 206.96 respectively). They both have hydrophobic newly substituted amino acids. A change from a hydrophilic amino acid to a hydrophobic amino acid might affect the rate of metabolism of the implicating enzymes. This is consistent with the finding by Koulgi *et al.* (2022) who predicted that increased interaction with hydrophobic residues causes an increase in the rate of metabolism of the ligands.

UNIVERSITY of the
WESTERN CAPE

```
from Bio.Seq import Seq
from Bio.Alphabet import IUPAC
import os

import Bio.SeqIO
isoform1=r'C:\Users\Omotoso Ogooluwa\Desktop\isoform1.fa.fasta'
seq_rec=Bio.SeqIO.read(isoform1,'fasta')
mutable_seq= seq_rec.seq.tomutable()



mutable_seq= seq_rec.seq.tomutable()
mutable_seq[222] = 'P'
new_seq_rec = mutable_seq.toseq()
```

Figure 4.1: An extract of the process to obtain the mutated sequence of the CYP3A4*2 variant on the python Jupyter page.

Note: the substitution at position 222 to proline(P) is indicated in the third last line.

http://etd.uwc.ac.za/

Figure 4.2 A CLUSTAL generated multiple alignment of CYP3A4*1-S222P, CYP3A4*11-T162M CYP3A4*23-R222W and the wild-type sequence (4i4th.1A Cytochrome P450 3A4).

An asterisk (*) denotes identical amino acids. The red arrow depicts an amino acid substitution

http://etd.uwc.ac.za/

Figure 4.3: A CLUSTAL generated pairwise alignment of CYP3A4*2-S222P and the wild-type sequence (4i4th.1A Cytochrome P450 3A4).

An asterisk (*) denotes identical amino acids. The red arrow depicts an amino acid substitution.

Figure 4.4: A CLUSTAL generated pairwise alignment of CYP3A4*11-T162M and the wild-type sequence (4i4th.1A Cytochrome P450 3A4).

An asterisk (*) denotes identical amino acids. The red arrow depicts an amino acid substitution.

http://etd.uwc.ac.za/

Figure 4.5: A CLUSTAL generated pairwise alignment of CYP3A4*23-R222W and the wild-type sequence (4i4th.1A Cytochrome P450 3A4).

An asterisk (*) denotes identical amino acids. The red arrow depicts an amino acid substitution

Figure 4.6: A CLUSTAL generated pairwise alignment of CYP3A4*24-Q200H and the wild-type sequence (4i4th.1A Cytochrome P450 3A4).

An asterisk (*) denotes identical amino acids. The red arrow depicts an amino acid substitution.

http://etd.uwc.ac.za/

4.2 Homology modelling of the selected variants

Homology modelling was employed to generate the three-dimensional protein structures of the selected variants of CYP3A4. An initially established template, a crystal structure of CYP3A4 ligated to pyridine-substituted desoxyritonavir (4i4h.1A Cytochrome P450 3A4) was used as the template sequence, together with the query sequences (variants sequences) to model the three-dimensional protein structure of the selected variants. The protein structures of the four variants are shown in Figure 4.6.

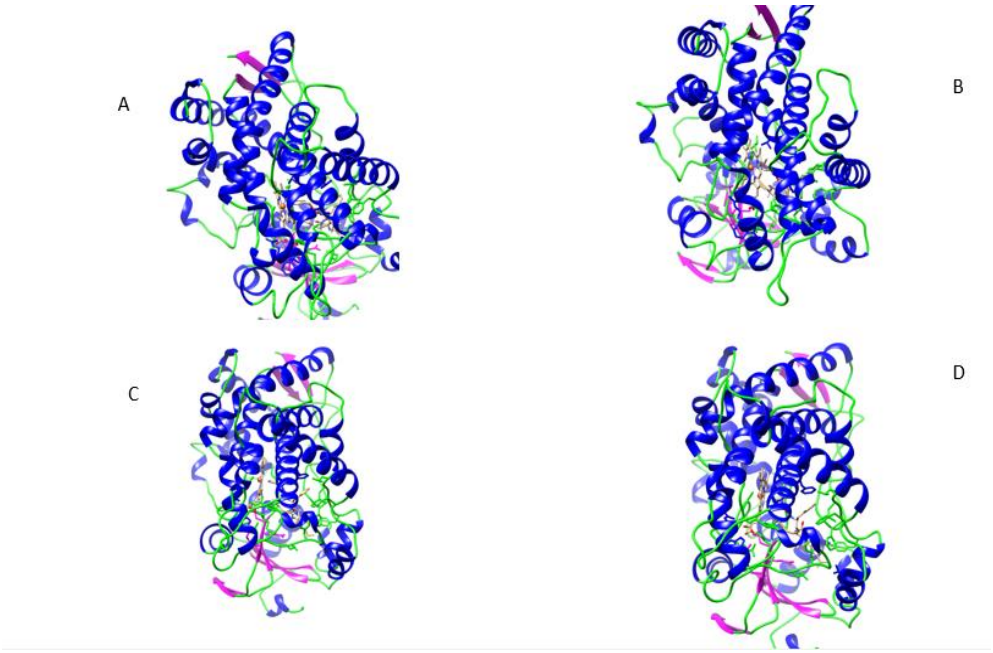


Figure 4.7: Three-dimensional protein structures of the selected CYP3A4 variants. Secondary structures are indicated in different colours: helixes in blue, coils in green, strands in purple and the centre positioned heme in golden brown (A-CYP3A4*2, B-CYP3A4*11, C-CYP3A4*23, D-CYP3A4*24)

4.3 Estimation of the quality of the modelled protein structures

The estimation of the reliability and quality of a modelled protein structure is a vital step in protein structure characterization and prediction (Shin *et al.*, 2017; Haddad, Adam, and Heger, 2020). To predict the quality of the modelled protein structures,

http://etd.uwc.ac.za/

the QMEAN (Qualitative Model Energy Analysis) (Pramanik *et al.,* 2017), the Verify 3D (Choudhary, Malik, and Tomar, 2020), and the ERRAT (Rahman *et al.*, 2019) were employed as highlighted in the method (Chapter 3, sub-section 3.4).

4.3.1 Quality evaluation of modelled protein structures on QMEAN

The GMQE was used to assess the quality of the target (query) to template alignment in modelled protein structures. It combined target-template alignment and template structure properties (Waterhouse *et al.,* 2018; Crook *et al.,* 2021). QMEANDisCO generated a plot of local distance difference comparison values for each residue between the range of 0 and 1. The reliability of the modelled structures was shown from the obtained values (Table 4.2).

The GMQE and QMEANDisCo >0.6 were predictive of a highly reliable model (David *et al.,* 2020). The GMQE and QMEANDisCO values obtained were high (>0.8), and the acquired data demonstrated that the modelled protein structures of CYP3A4 variants had a high level of confidence and were credible for further research investigations.

UNIVERSITY of the
WESTERN CAPE

Table 4.2: Estimation of the quality of modelled protein structure on QMEAN with GMQE and QMEANDisCO obtained values (>0.8).

Protein Structure	QMEANDisCo Global	Global Model Quality Estimate (GMQE)
CYP3A4*1	0.86±0.05	0.85
CYP3A4*2	0.86±0.05	0.85
CYP3A4*11	0.85±0.05	0.85
CYP3A4*23	0.86±0.05	0.85
CYP3A4*24	0.86±0.05	0.85

http://etd.uwc.ac.za/

4.3.2 Quality evaluation of modelled protein structures with the Ramachandran plot

The prediction of the structural stereochemical properties was obtained on the Ramachandran plot. The PROCHECK evaluates the geometry of each residue to determine the stereochemical quality of a predicted model (Pramanik *et al.*, 2017).

The Ramachandran plot generated by the PROCHECK tool in SAVES was used to generate the Ramachandran plots and it showed that 90.7% of the residues were in most of the favoured regions of the plot, 0.0% in generously allowed regions and 0.5% in disallowed regions for all the modelled protein structures of selected variants (Fig.4.7 to 4.10).

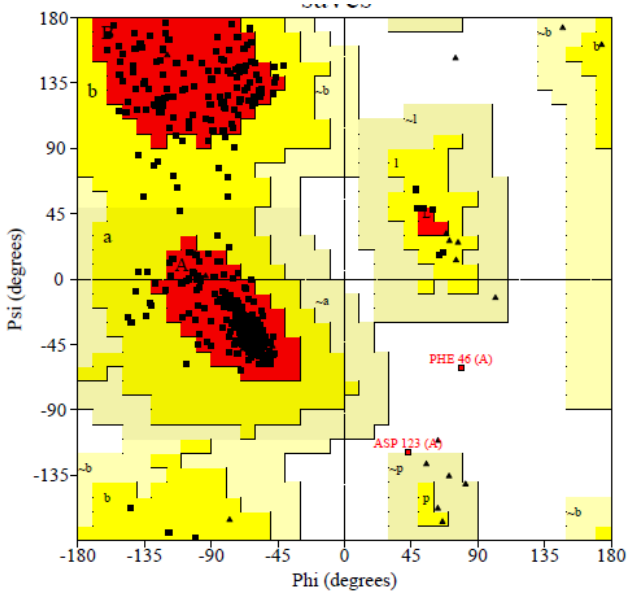


Figure 4.8: The Ramachandran plot of the modelled CYP3A4*1 structure generated on PROCHECK.

Residues are shown by little black squares. The red section denotes the most favoured regions. Yellow denotes allowed areas. Disallowed areas were indicated by white space with no colour. It indicates that 90.7% of the residues were in most of the favoured regions of the plot, 8.8% in the additional allowed regions, 0.0% in generously allowed regions and 0.5% in disallowed regions

http://etd.uwc.ac.za/

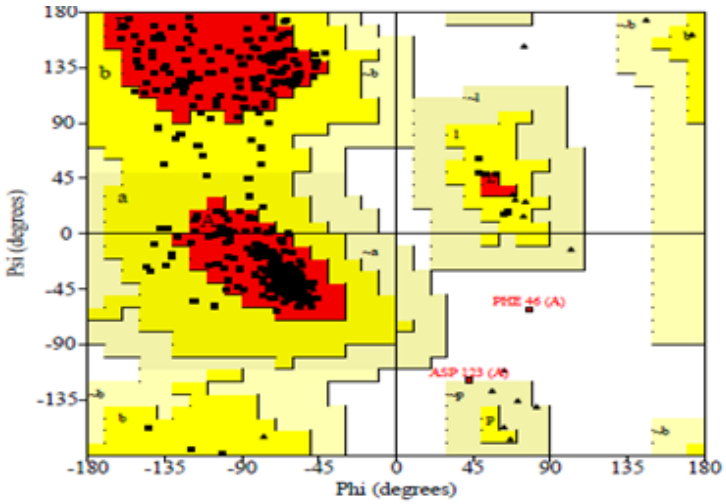


Figure 4.9: The Ramachandran plot of the modelled CYP3A4*2 structure generated on PROCHECK.

Residues are shown by little black squares. The red section denotes the most favoured regions. Yellow denotes allowed areas. Disallowed areas were indicated by white space with no colour. It indicates that 90.7% of the residues were in most of the favoured regions of the plot, 8.8% in additional allowed regions, 0.0% in generously allowed regions and 0.5% in disallowed regions

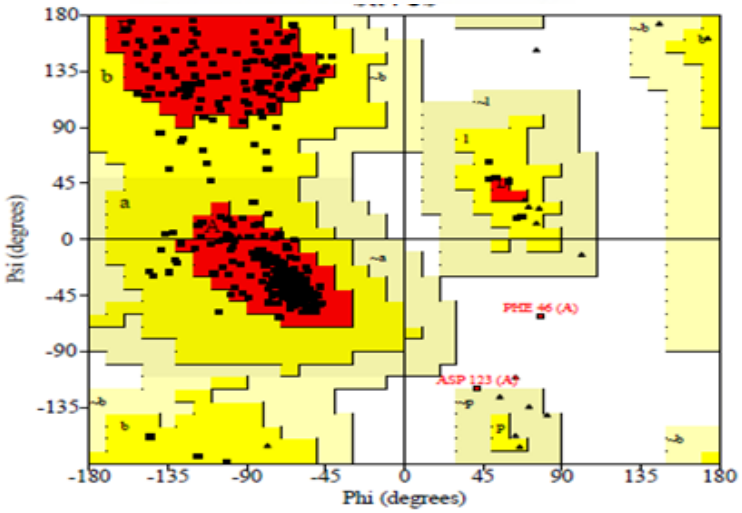


Figure 4.10: The Ramachandran plot of the modelled CYP3A4*11 structure generated on PROCHECK

http://etd.uwc.ac.za/

Residues are shown by little black squares. The red section denotes the most favoured regions. Yellow denotes allowed areas. Disallowed areas were indicated by white space with no colour. It indicates that 90.7% of the residues were in most of the favoured regions of the plot, 8.8% in additional allowed regions, 0.0% in generously allowed regions and 0.5% in disallowed regions

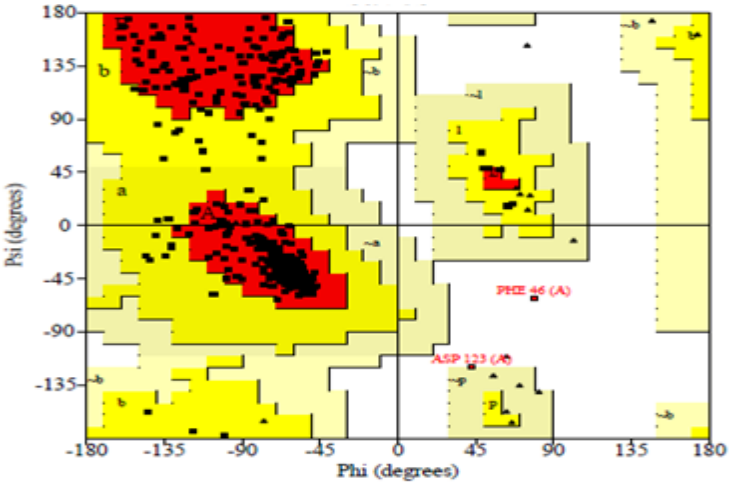


Figure 4.11: The Ramachandran plot of the modelled CYP3A4*23 structure generated on PROCHECK.

Residues are shown by little black squares. The red section denotes the most favoured regions. Yellow denotes allowed areas. Disallowed areas were indicated by white space with no colour. It indicates that 90.7% of the residues were in most of the favoured regions of the plot,8.8% in additional allowed regions, 0.0% in generously allowed regions and 0.5% in disallowed regions

http://etd.uwc.ac.za/

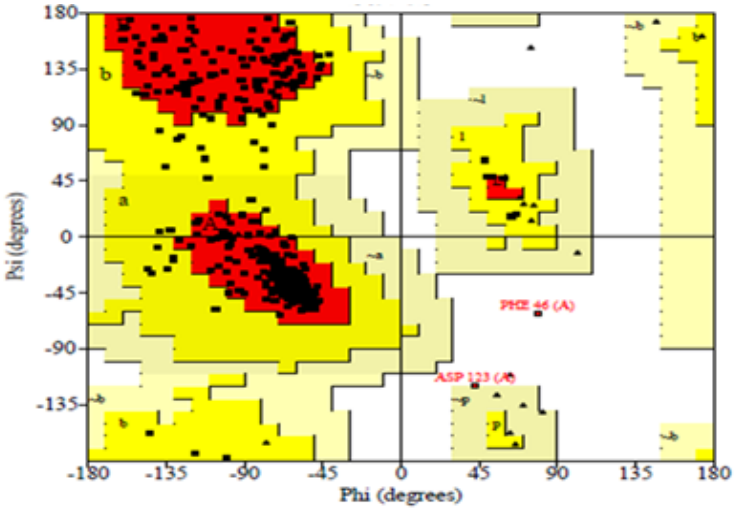


Figure 4.12: The Ramachandran plot of the modelled CYP3A4*24 structure generated on PROCHECK.

Residues are shown by little black squares. The red section denotes the most favoured regions. Yellow denotes allowed areas. Disallowed areas were indicated by white space with no colour. It indicates that 90.7% of the residues were in most of the favoured regions of the plot, 8.8% in additional allowed regions,0.0% in generously allowed regions and 0.5% in disallowed regions

4.3.3 Quality evaluation of modelled protein structures on ERRAT plot

The ERRAT uses a database of highly refined protein structures to produce a graph of the position of the nine-residue sliding window against the error function. The Errat graph was based on the improved structure database's nonbonded interaction statistics between distinct atom kinds. (Kumar *et al.*, 2019). The validation report from the ERRAT quality plot derived from the SAVE server (Table 4.3) shows that all the modelled structures inputted had good resolution with an overall quality factor >95%.

http://etd.uwc.ac.za/

Table 4.3: The overall quality factor of modelled protein structure on ERRAT

Modelled Protein Structure	ERRAT Overall Quality Factor
Wild type	95.55556
CYP3A4*2	95.12
CYP3A4*11	+95.79
CYP3A4*23	95.56
CYP3A4*24	95.56

The output graphs of the modelled protein structure (Figure 4.11 to 4.14) indicated that the red or yellow colour regions are very minimal while the ash-coloured regions are predominant.

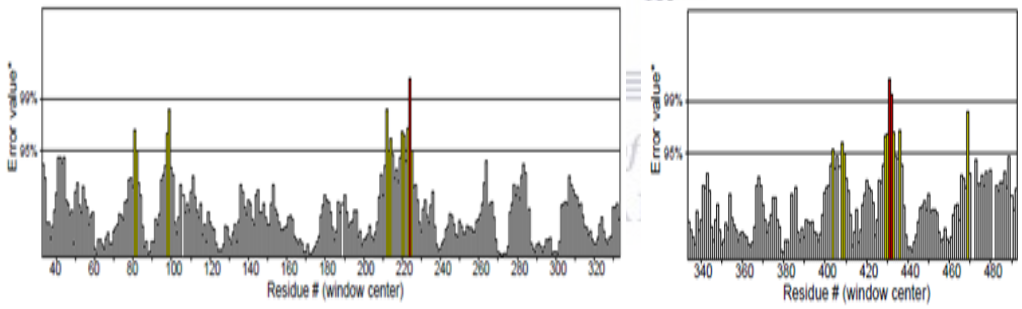


Figure 4.13: Errat model for CYP3A4*2

The red- and yellow-coloured sections in the graph represent the faulty parts of the structure, while the ash-coloured portions show the allowed regions

http://etd.uwc.ac.za/

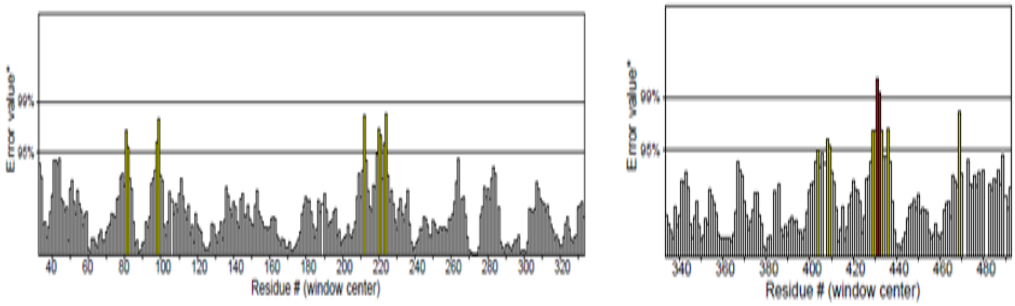


Figure 4.14: Errat model for CYP3A4*11.

The red- and yellow-coloured sections in the graph represent the faulty parts of the structure, while the ash-coloured portions show the allowed regions

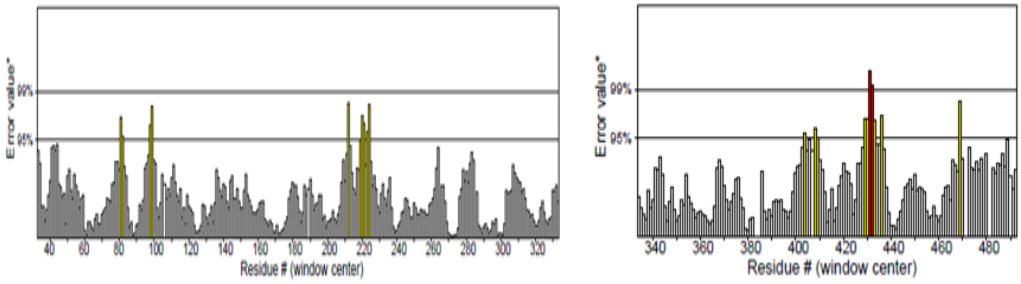


Figure 4.15: Errat model for CYP3A4*23.

The red- and yellow-coloured sections in the graph represent the faulty parts of the structure, while the ash-coloured portions show the allowed regions

UNIVERSITY of the
WESTERN CAPE

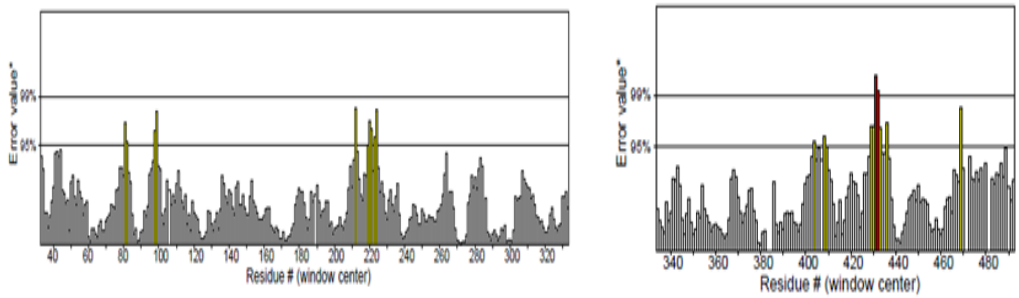


Figure 4.16: Errat model for CYP3A4*24

The red- and yellow-coloured sections in the graph represent the faulty parts of the structure, while the ash-coloured portions show the allowed regions

http://etd.uwc.ac.za/

4.3.4 Quality evaluation of modelled protein structures on Verify 3D

Verify 3D makes use of the structural database of proteins to determine if a 3D structure will be compatible with a 1D (one-dimensional) amino acid sequence based on structure assignments. Examples of such structure assignments include the loops, the alpha-helix, and the sheets (Khan *et al.*, 2021). In the Verify 3D, at least 80% of the amino acids must have scored >= 0.2 in the 3D/1D profile for quality and reliable modelled protein structures (Oyugi *et al.,* 2018; Khalid *et al.,* 2020).

All modelled protein structures have 96.58% of the residues with a 3D-1D score >=0.2 which is the same as the wild type of protein structure, and therefore met the requirement (Table 4.4). In summary, all the modelled protein structures passed all quality check steps indicating high reliability and were therefore suitable for further characterization and analysis of the modelled protein structures of the selected CYP3A4 variants.

Table 4.4: The estimation of the quality of modelled protein structure on Verify 3D

Modelled Protein Structure	Verify 3D
Wild-type	96.58% of the residues have averaged 3D-1D score >= 0.2 Pass. At least 80% of the amino acids have scored >= 0.2 in the 3D/1D profile.
CYP3A4*2	96.58% of the residues have averaged 3D-1D score >= 0.2 Pass. At least 80% of the amino acids have scored >= 0.2 in the 3D/1D profile.
CYP3A4*11	96.58% of the residues have averaged 3D-1D score >= 0.2 Pass. At least 80% of the amino acids have scored >= 0.2 in the 3D/1D profile.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

CYP3A4*23	96.58% of the residues have averaged 3D-1D score >= 0.2 Pass. At least 80% of the amino acids have scored >= 0.2 in the 3D/1D profile.
CYP3A4*24	96.58% of the residues have averaged 3D-1D score >= 0.2 Pass. At least 80% of the amino acids have scored >= 0.2 in the 3D/1D profile.

4.4 Structural alignment and molecular surface visualization of the CYP3A4 wild-type and the selected variants' models

4.4.1 Structural alignment of the CYP3A4 selected variants and the wild type.

From the visualised modelled protein structures of the wild-type and the variants, it was observed that the position of the heme ligand of all the variants' structures did not show any significant difference from the wild type. None of the mutations likely affected the positioning of the heme moiety. It has been reported that the enzyme's catalytic core is the heme prosthetic group (Kaur *et al.,* 2016; Guengerich, 2018). This could explain why there was no reported loss of activity in the reported effects of mutation for the selected variants in the previously reported study (Fang *et al.,* 2017). A projection of a reduction or increase in activity has been the difference reported (Fang *et al.,* 2017). The single nucleotide polymorphism in the selected variants might not have affected the cysteine-heme property of CYP3A4. Therefore, the heme prosthetic group property might not be a determinant of the reported varied rate of metabolism in the selected variants.

In CYP3A4*2(S222P), the Ser222 is in the loop, in between the F and G helices, and the substitution with proline occurred at the F-G helices region (Figure 4.15). Previous reports have shown that the F-G helixes played a role in enzyme specificity and accessibility to the enzyme substrate (Šrejber *et al.,* 2018; Dong *et al.,* 2021). The substitution at this position (S222P) with proline might interfere with the enzyme specificity which might account for the variation in the enzymatic

UNIVERSITY of the WESTERN CAPE
http://etd.uwc.ac.za/

activity (reduced intrinsic clearance relative to the wildtype) of CYP3A4*2 with lidocaine in the previous report (Fang *et al.*, 2017). In addition, in previous reports, Proline has been reported to break helixes (Miyazaki *et al.*, 2008), The effect of the substituted proline as reported by Miyazaki *et al.*, (2008) is consistent with the prediction of the effect of SNPs on the F and G helices, where the substitution occurred.

In CYP3A4*11 (T363M), threonine was replaced by methionine **(T-M)**. It has been reported that Threonine 363 at the SRS-5 region was a major part of hydrogen interaction in CYP3A4 interaction with ligands (Murayama *et al.*, 2002; Zhou *et al.*, 2019). Reports have shown that the hydrogen bond is important in CYP3A4 interaction, and an alteration could interfere with its enzymatic function (Eiselt *et al.*, 2001; El-Sayed *et al.*, 2016). Hence, the substitution at T363M (Figure 4.16) could lead to alteration in hydrogen interaction and was estimated to be responsible for the varied rate of enzymatic activity (relative to the wildtype) in the metabolism of CYP3A4*11 with lidocaine as reported by Fang *et al.* (2017).

In CYP3A4*23 (R162W) arginine was replaced by tryptophan, the arginine at position162 has a very close distance to the active site (Figure 4.17). The substitution at the position occurred between amino acids of different classes: Arginine(R) is a hydrophilic amino acid while Tryptophan(W) is a hydrophobic amino acid. Drögemöller *et al.* (2013) in their findings predicted that an amino acid substitution of a hydrophilic amino acid for a hydrophobic amino acid could affect the catalytic activity of the enzyme variants. In addition, Koulgi *et al.* (2022) predicted in the study of the conformation of CYP3A4 structures that increased interactions with hydrophobic residues may cause an increase in the rate of metabolism of the ligands. Therefore, the substitution of hydrophilic residue (Arginine) with hydrophobic residue (Tryptophan) might account for the reported difference in the rate of metabolism of CYP3A4*23 with lidocaine (higher relative clearance of 206.96 as against the wild type of 100) (Fang *et al.*, 2017).

In variant CYP3A4*24 (Q200H), glutamine was substituted with histidine at position 200. Both amino acids belong to the same class(hydrophilic) and no

http://etd.uwc.ac.za/

noticeable difference was observed in the superimposed structure of the CYP3A4*24 variant three-dimensional structure with the wildtype. It was estimated that other protein properties might account for the reported varied rate of metabolism by Fang *et al.* (2017). Variants 2 and 23 showed a mismatch at the helix (Figure 4.18 and Figure 4.21). This noticeable structural difference could account for the difference in the reported relative clearance, relative to the wild type.

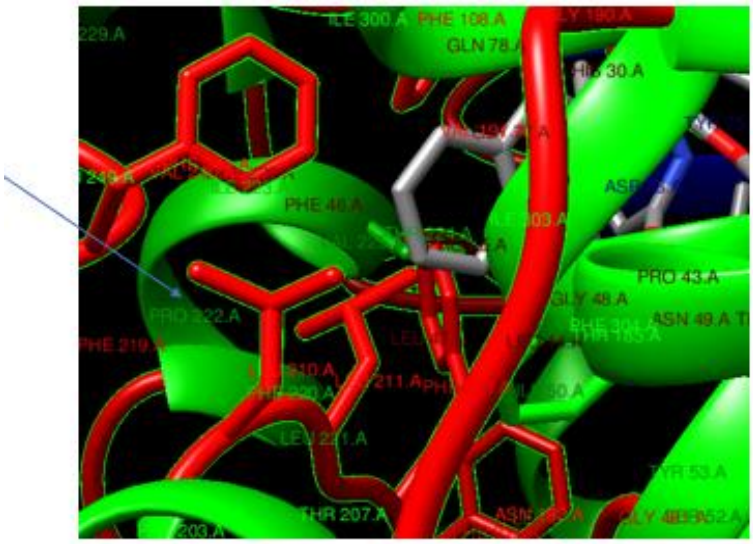


Figure 4.17: The molecular surface structure of CYP3A4*2(coils in red, helix in green, strands in blue, heme moiety in ash) .

UNIVERSITY of the
WESTERN CAPE

The headed arrow shows the location of the new amino acid Proline at position 222 (PRO222).

http://etd.uwc.ac.za/

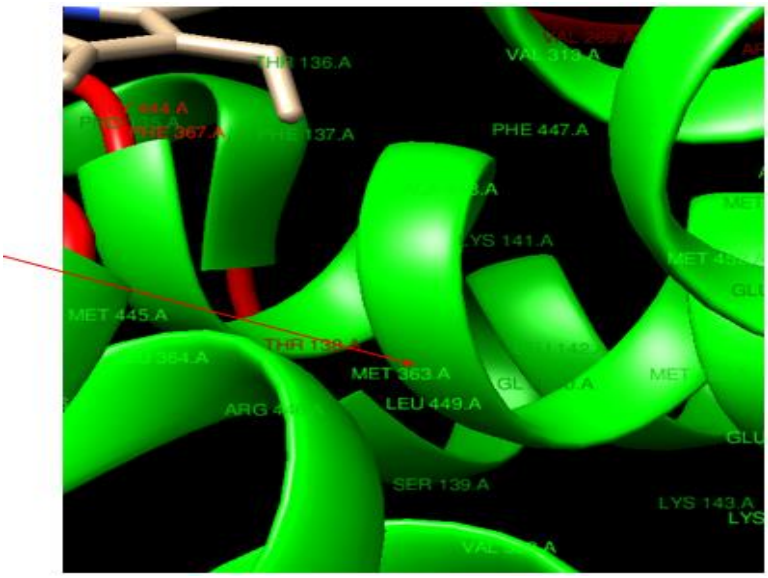


Figure 4.18: The molecular surface structure of CYP3A4*11(coils in red, helix in green, strands in blue, heme moiety in ash).

The headed arrow shows the location of the new amino acid Methionine at position 363 (MET363).

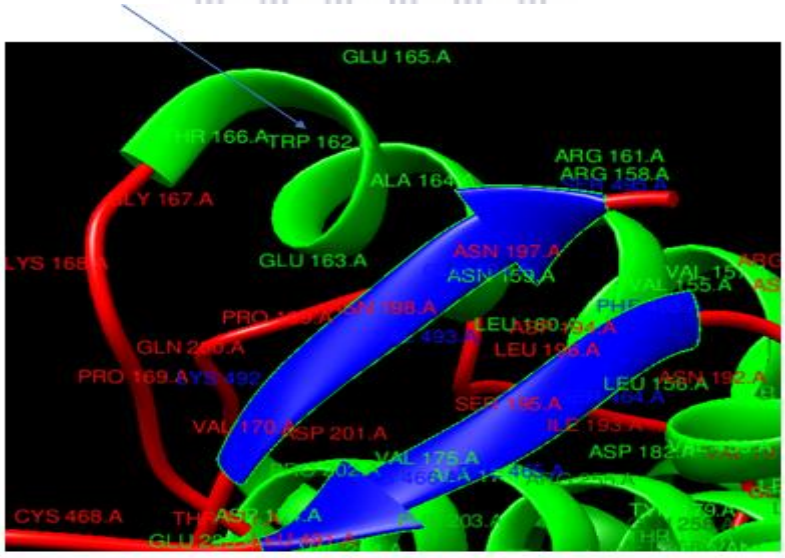


Figure 4.19: The molecular surface structure of CYP3A4*23(coils in red, helix in green, strands in blue, heme moiety in ash).

The headed arrow shows the location of the new amino acid Tryptophan at position 162 (TRP162).

http://etd.uwc.ac.za/

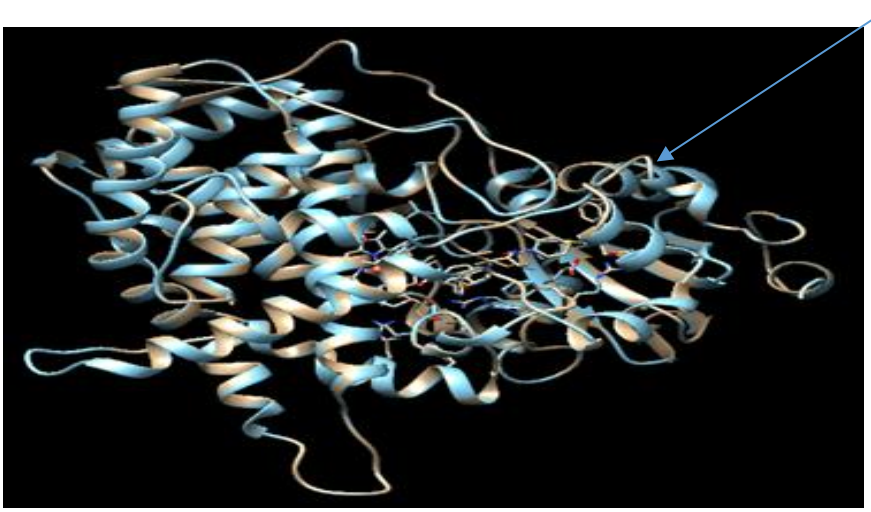


Figure 4.20: The structural comparison of CYP3A4*1(blue) and CYP3A4*2(ash).

The superimposed three-dimensional protein structural models, with the headed arrow showing the position of mix match at the helix

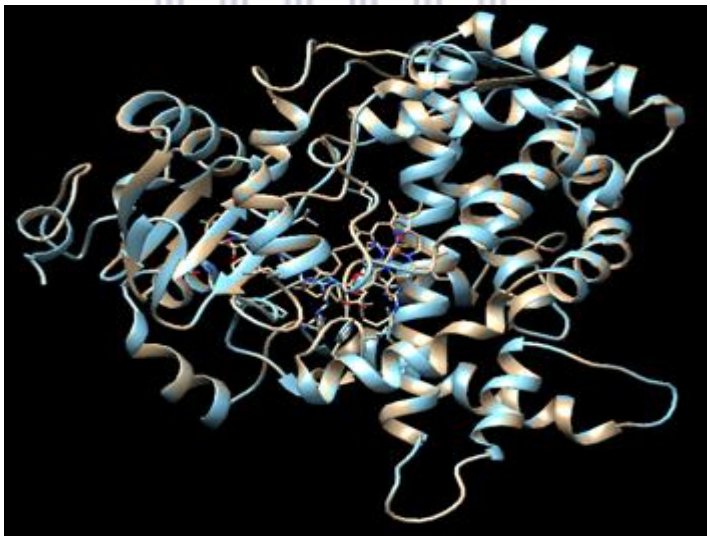


Figure 4.21: The structural comparison of CYP3A4*1(blue) and CYP3A4*11(ash) .

The superimposed three-dimensional protein structural models with no noticeable mix match

http://etd.uwc.ac.za/

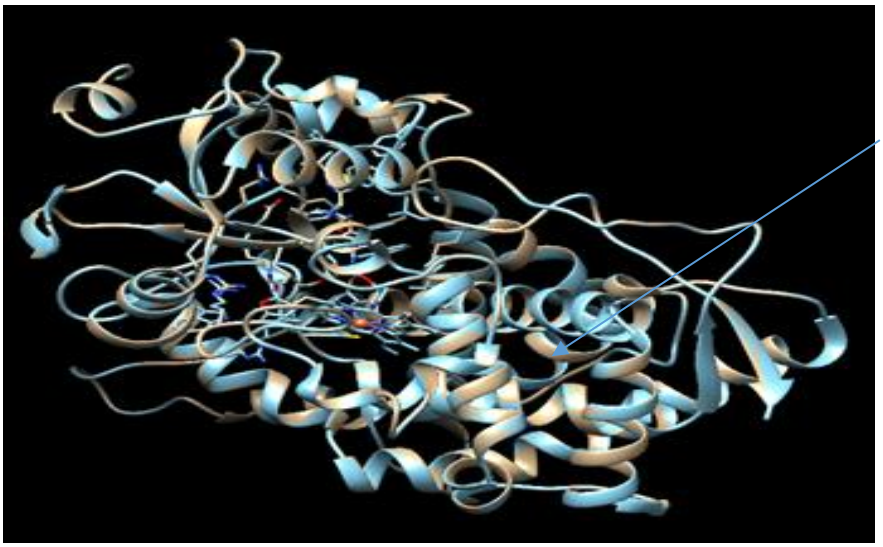


Figure 4.22: The structural comparison of CYP3A4*1(blue) and CYP3A4*23 (ash) .

The superimposed three-dimensional protein structural models and the centralised heme moiety in a ball(brown) with the headed arrow showing the position of the mix match at the helix

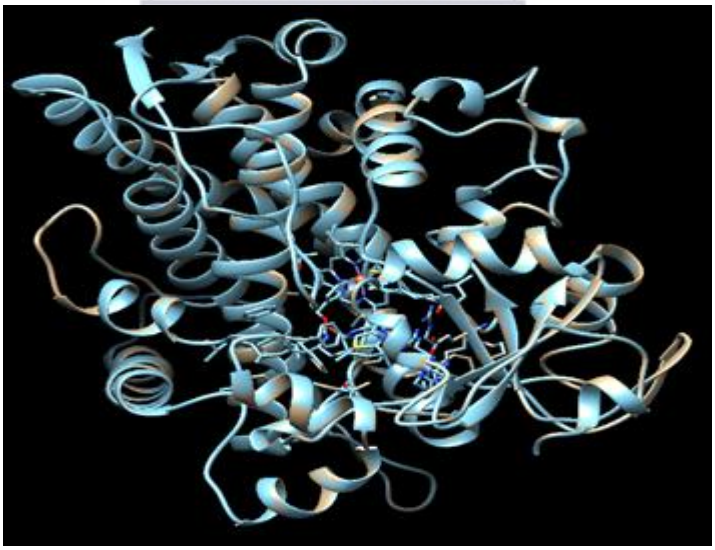


Figure 4.23: The structural comparison of CYP3A4*1 (blue) and CYP3A4*24 (ash) .

The superimposed three-dimensional protein structural models, with no noticeable mix match

http://etd.uwc.ac.za/

The observed RMSD has minimal deviation (Table 4.5) in the superimposed structure of the wild type and selected variants protein structures on the matchmaker. It was noticeable that variants 2 and 24, though negligible, had the highest RMSD values relative to the other selected variants (CYP3A4*11 and CYP3A4*23) (Table 4.5).

Table 4.5: RSMD values for structural comparison of CYP3A4 wild type and the selected variant models

Protein Model	RMSD (Root Mean Square Deviation) of superimposed protein structures
CYP3A*2	0.029
CYP3A*11	0.015
CYP3A4*23	0.003
CYP3A4*24	0.02

4.4.2 The molecular surface visualization of the CYP3A4 wild-type and the selected variants' models

The molecular lipophilicity potential of the modelled protein structures of the wild type and the variants was shown in Table 4.6.

UNIVERSITY of the
WESTERN CAPE

Table 4.6: The molecular lipophilicity potential values of CYP3A4 wild-type and selected variant protein models

Modelled Protein Structures	Minimum	Mean	Maximum
Wild type	-28.39	-3.448	23.93
Model 2	-28.46	-3.426	23.83
Model 11	-28.53	-3.447	24.09
Model 23	-27.44	-3.372	23.95
Model 24	-28.39	-3.444	23.93

http://etd.uwc.ac.za/

The colour range on the molecular surface ranges from dark cyan (most hydrophilic) to white (less hydrophilic) to dark golden (most lipophilic).The differences in the hydrophobicity of the various regions of the variants relative to the wild type were visualised. There was a noticeable variation in the molecular surface colour in variants CYP3A4*11, CYP3A4*23, and CYP3A4*24 when compared with the wild type (Figure 4.22 -4.25). The wild type has a darker cyan (pointed region Figure 4.22), while a white colour was observed in CYP3A4*11 and CYP3A24 (Figure 4.23 and Figure 4.25). A golden colour region was observed in CYP3A4*23 (Figure 4.24). This implies that the wild type (darker cyan blue) was more hydrophilic than the selected variants (white and golden colour). The observation also gives a prediction that CYP3A4*23 was the most hydrophobic (golden colour).

It was observed from the mean values of the molecular lipophilicity potential (Table 4.6) that the selected CYP3A4 have varied values, relative to the wild type. The wild type has the least value (-3.448) as compared to the selected variants (-3.426, -3.447, -3.372, -3.444) for CYP3A4*2, CYP3A4*11, CYP3A4*23, CYP3A4*24 respectively). This further supports the observation of the molecular surface and gives an estimation that the wild type was likely more hydrophilic than the selected variants. The mean value of the molecular lipophilicity potential also suggests that CYP3A4*23 is the most lipophilic of the selected variants with the highest mean value of (-3.372).

Lipophilicity is a fundamental physicochemical property that affects the metabolism and toxicity of a drug, and it is evident that lipophilicity is important to P450 substrate selectivity and clearance (Lewis, Jacobs, and Dickins, 2004; Stephens, Lucena, and Andrade, 2018). It has been reported that drug clearance increases as lipophilicity rises (Broccatelli, Aliagas, and Zheng, 2018). This report agrees with the finding in CP3A4 variants. For example, the high relative clearance (206.96) observed in CYP3A4*23 has the highest mean MLP (-3.372). With the altered lipophilicity (due to SNPs) in the selected variants, it is estimated that the

http://etd.uwc.ac.za/

molecular lipophilicity potential might account for the reported variation in the rate of metabolism in the selected variants. The molecular lipophilic potential result was subjected to a correlation study (section 4.10) to further investigate the correlation between the physicochemical parameter (molecular lipophilicity potential) and the reported difference in clearance of the selected CYP3A4 relative to the wildtype.

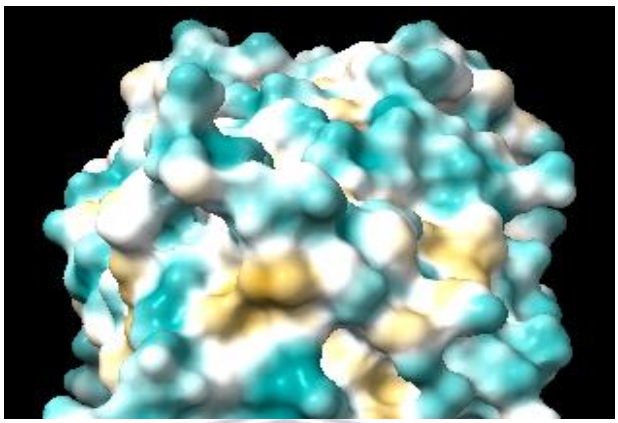


Figure 4.24: The molecular hydrophobic surface colour of wild-type structure.

Dark cyan (most hydrophilic) to white (less hydrophilic) to dark golden (most lipophilic.

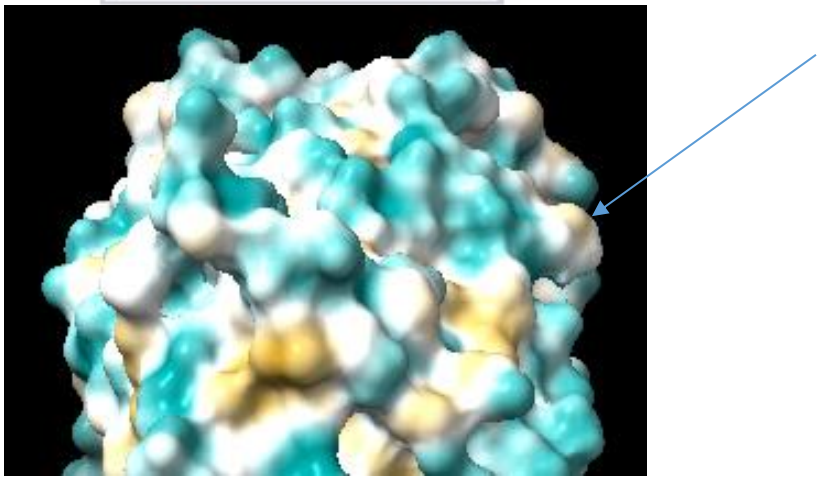


Figure 4.25: The molecular hydrophobic surface colour of CYP3A4*11.

Dark cyan (most hydrophilic) to white (less hydrophilic) to dark golden (most lipophilic.

http://etd.uwc.ac.za/

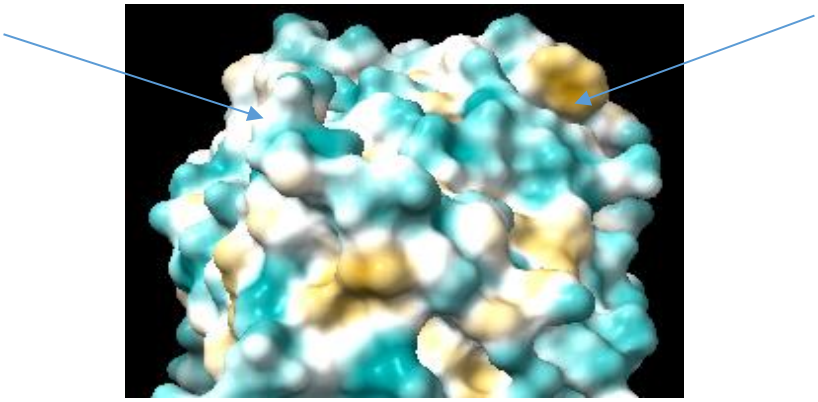


Figure 4.26: The molecular hydrophobic surface colour of CYP3A4*23

Dark cyan (most hydrophilic) to white (less hydrophilic) to dark golden (most lipophilic

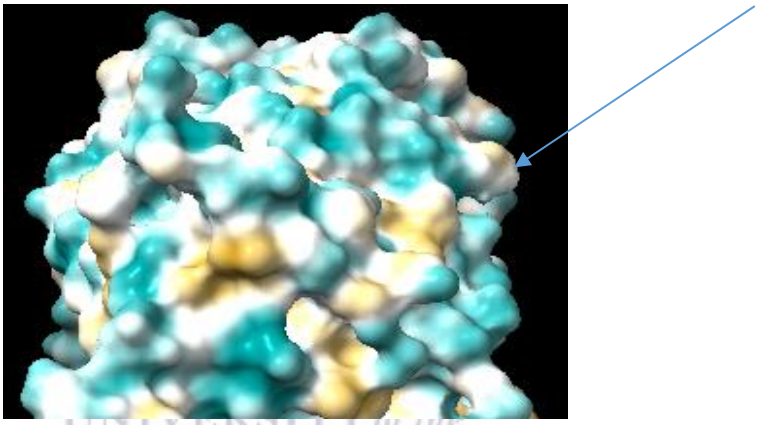


Figure 4.27: The molecular hydrophobic surface colour of CYP3A4*24.

Dark cyan (most hydrophilic) to white (less hydrophilic) to dark golden (most lipophilic

The differences in the electrostatic potential of the various regions of the variants relative to the wildtype were visualised. No colour difference was observed in the wild type and selected variants' colour surface range. The Coulombic electrostatic potential (ESP) of the wild type and the selected variants were calculated. The obtained result is shown in Table 4.7. It was observed that the mean values of the Coulombic electrostatic potential (ELP) of the selected variants differ from the ELP value of the wild type. For example, a very significant difference was observed in CYP3A4*23 (0.18) relative to the wildtype (0.38). This gives a prediction that SNPs might affect the ELP of a variant's protein structure. It was, however,

http://etd.uwc.ac.za/

estimated that the Coulombic electrostatic potential might be a determining property in the reported difference in the rate of metabolism (clearance) of the selected CYP3A4 variants. The Coulombic electrostatic potential result was subjected to a correlation study (section 4.10) to establish the potential link between the Coulombic electrostatic potential and the reported difference in clearance of the selected CYP3A4 relative to the wildtype.

Table 4.7: The coulombic electrostatic potential values of CYP3A4 wild-type and the selected variants protein models

Modelled Protein Structures	Minimum	Mean	Maximum
Wild type	-11.21	0.38	15.70
Model 2	-11.32	0.39	15.61
Model 11	-13.88	0.37	15.68
Model 23	-11.24	0.18	15.58
Model 24	-15.41	0.37	15.69

4.5 The differences in the physicochemical and the general protein properties of the selected CYP3A4 mutants, relative to the wild type.

Previous reports have shown that one of the important aspects of the characterization of a protein is its physicochemical parameters analysis (Kaur *et al.*, 2020; Munjal, Shukla, and Singh, 2021b). The physicochemical properties of each of the selected mutated sequences of CYP3A4 and the wild-type sequence investigated in this study were outlined in Table A.0.1 (Appendix A).

Each of the CYP3A4 studied sequences contains 503 amino acids. Varied molecular weight was obtained for each variant, none of it was the same as that obtained for the wild type due to the different residues substituted at the point of mutation. It can be suggested that the varied values of the molecular weight could be a determining property in the reported difference in lidocaine clearance of the

http://etd.uwc.ac.za/

selected variant. The isoelectric point (pI) is an important physicochemical property that indicates the point at which a solution gives an absolute charge of zero due to an equal amount of positive and negative charges. It gives an estimation of the surface charge of proteins at different pH conditions (Kozlowski, 2021). All variants studied have a pI of 8.27 except for CYP3A4*23 with a pI of 8.00. All values were an indication of a mildly alkaline protein which was the same as the wild type. The sequence and structure-based pI did not show significant differences; hence, they might not be a determining factor for the reported difference in enzymatic activity. The extinction coefficient depicts the interaction between the protein-protein and protein-ligand (Munjal, Shukla, and Singh, 2021). The quantitative finding of the extinction coefficient has the same value for all the variants; hence, it might not be a determinant of the observed varied enzymatic activity of the wild type and variants of CYP3A4 (Table A.0.1). The instability index for all selected mutants and variants was above 40. Verma, Singh, and Gaur (2016) reported that values higher than 40 are unstable. This study predicts a higher instability in CYP3A4*2 and CYP3A4*24 relative to the wild type. Interestingly, both variants also have higher RMSD values as reported in the superimposition of the variants' protein structures. Therefore, this suggests that the instability index might be a factor in the reported difference in the clearance of the selected variant relative to the wild type. The property (instability index) shows a relationship with the reported relative clearance values. It was observed that variants with lower clearance relative to the wildtype have higher instability index (Table A.0.1).

The Aliphatic index (which is the relative volume of protein that the aliphatic side chains occupied) (Kaur *et al.*, 2020) was the same in all studied variants and the wild type (Table A.0.1). The aliphatic index is therefore not a determinant of the reported varied enzymatic activity in the selected CYP3A4 variants. The Gravy (Grand Average of Hydropathicity) is the sum of the relative hydrophobicity and hydrophilicity values of all the amino acids when divided by the total number of residues in each sequence for all variants (Nagai *et al.*, 2016). According to Nagai *et al.* (2016), low values indicate that the interaction was more between water and the protein. The obtained value for each variant differed from the wild type and all

http://etd.uwc.ac.za/

obtained values were lower (relative to the wildtype) except CYP3A4*2 (Table A.0.1). This could be a predictive property of the observed difference in the enzymatic activity (clearance) of the selected CYP3A variants. The Debye screening length is a protein property that depends on the salt concentration and the temperature of the settings (Piccinini *et al.*, 2018). The values obtained in both the variants and the wildtype were the same, (Table A.0.1) (Section 3.6). The Debye screening length may not account for the reported varied rate of clearance of CYP3A4 selected variants. The frictional coefficient (an individual molecule's frictional drag coefficient quantified by the size and viscosity of the protein) (Erickson, 2009; Su *et al.*, 2020), diffusion coefficient (diffusion coefficient of Stokes-Einstein, quantity is affected by temperature, protein size, and viscosity) (Poolman *et al.*, 2021; Bellotto *et al.*, 2022), the radius of gyration (the root mean square distance of the molecule atoms from the protein's centre of mass scaling laws) (Ahmed, Crehuet, and Lindorff-Larsen, 2019; Pražnikar, 2021), and the hydrodynamic radius (the apparent size of the dynamic hydrated/solvated particle, calculated based on the macromolecule's diffusion characteristics) (Ahmed, Crehuet, and Lindorff-Larsen, 2020) were not indicative of the reported varied rate of enzymatic activity in the selected variants. The values obtained from these studied protein properties do not indicate a significant difference between the wild type and the selected variants (Table A.0.1).The sedimentation constant is the dimension of time in seconds, given in Svedbergs and it is used to characterise sedimentation processes such as centrifugation. The duration of one Svedberg is 10-13 seconds, and the value is based on a constant buoyancy in water (Sétáló Jr, 2013). The sedimentation constant did not present a significant difference in the CYP3A4 selected variants (Table A.0.1). It was predicted not to be a determinant property in the reported variation in the enzymatic property of the CYP3A4 variants. Protein eccentricity is a shape descriptor that is determined by the ratio of the two main components (the perfect sphere and the linearized protein); the perfect sphere has a value of 1.0, while a linearized protein has a value of 0(Feng, Wang, and Wang, 2017). The data obtained for this property for the wild type was the same for the selected CYP3A4 variants and therefore the protein eccentricity is likely not

http://etd.uwc.ac.za/

to be a determinant property of the varied rate of metabolism in the reported CYP3A4 variants. Accessible surface area (a protein's water-accessible surface area) (Konstantinidis *et al.,* 2021), hydrophobic surface area (hydrophobic atoms of a protein with a water-accessible surface area.) and the hydrophilic surface area (hydrophilic atoms of a protein with the water-accessible surface area) have varied values for the wildtype and the selected CYP3A4 (Table A.0.1). These properties might therefore be predictors of the reported difference in enzymatic activities in selected variants of CYP3A4.The protein volume in the protein volume calculation was made with the addition of its solvation shell (Dahal and Schmit, 2018), and there was a noticeable significant difference in the property (protein volume) for the wild type and the selected variants of CYP3A4 (Table A.0.1). There is an estimation of the protein volume being a determining property of the observed variation in the rate of metabolism of the selected variants. Protein mobility was calculated according to Kim *et al.* (2006), and the value represents the protein migration velocity during an electrophoretic separation for a particular electric potential (J.-B. Kim *et al.,* 2006; Webster *et al.,* 2019). No significant difference was observed in the protein mobility values (Table A.0.1); hence it might not be a determinant property of the reported difference in relative clearance of the selected variants.The protein helix ratio is the percentage of residues assigned to a helical secondary structure (Lu *et al.,* 2019). All the selected variants except for variant CYP3A4*2 (though not a significant difference) have the same value as the wild type (Table A.0.1), the protein helix is not likely to be a determining protein property of the varied enzymatic activity of the selected variants.

Protein Net charge (the formal charge on the protein at an ascertained pH) (Lu *etal.,* 2019) might not be an indicative property of the observed varied rate of activity of the selected variants because the values of the variants were the same as the wildtype (Table A.0.1). Sum negative surface area, sum positive surface area, sum donor surface area, and sum acceptor surface area, all have values with significant differences between the wild type and the selected variants (Table A.0.1). It was estimated that they might be determinant properties of the difference observed in the relative clearance of the selected CYP3A4 variants. The sum of aggregation-

http://etd.uwc.ac.za/

prone regions (sum of AggScore) in proteins are short sequence stretches that aggregate in a protein and they can form a cross-steric hindrance (Buck, Kumar, and Singh, 2013; Sankar et al., 2018). Studies have shown that the sum of Aggscore often has more capacity to stabilise proteins than equal length segments from the same protein. The sum of Aggscore could be a determinant of the difference observed in the relative clearance of the selected variants of CYP3A4, its values significantly differ in the selected variants relative to the wild type (Table A.0.1) (Buck, Kumar, and Singh, 2013; Sankar *et al.*, 2018). Protein Dipole Moment is based on the inconsistent distribution of positive and negative charges on the various atoms in a protein. The charges allocated by the forcefield involved were used in the computations (Tartaglia *et al.*, 2004; Das *et al.*, 2019). The protein dipole moment might account for the reported variation in the rate of metabolism as the values obtained for the selected CYP3A4 variants significantly differ from the wild type (Table A.0.1). Hydrophobicity Moment is a weighted moment computation in which amino acid residues are given quantities based on their intrinsic hydrophobicity (Kraml *et al.*, 2019; Stone *et al.*, 2019). There was a noticeable significant difference in the property (protein hydrophobicity moment) for the wild type and the selected variants of CYP3A4 (Table A.0.1). There is an estimation of the hydrophobicity moment being a determining property of the observed variation in the rate of metabolism of the selected variants.

Zeta potential at debye length has an aggregation tendency that is indicated by a value of zero. Typical values vary from positive to negative values (Piccinini *et al.*, 2018; Foteini *et al.*, 2019). The obtained value for each variant differs from the wild type and all obtained values were different from the wild type (Table A.0.1). This could be a predictive property of the observed difference in the enzymatic activity (clearance) of the selected CYP3A4 variants. Zeta Dipole Moment is the quantity of the dipole moment of the zeta potential. It is mainly measured by direct examination of the angular velocity that exists in rotational electrophoresis. It was also defined as the measurement of charge homogeneity at the surface of a solvated protein that can be utilised as an additional indicator of particle stability, particularly when the zeta potential is approximately 0 mV. Zeta Quadrupole

http://etd.uwc.ac.za/

Moment is the quantity of quadrupole moments of the zeta potential (Long and Labute, 2010). No significant differences were observed in the Zeta dipole moment and the zeta quadrupole moment values (Table A.0.1); hence it might not be a determinant property of the reported difference in relative clearance of the selected variant.

In summary, a total of thirty-five physicochemical and general protein properties were studied out of which seventeen parameters indicated significant differences relative to the wildtype (Table A.0.1). It was estimated that single nucleotide polymorphism might affect the studied seventeen parameters with noticeable differences relative to the wild type. The physicochemical and general protein properties of the selected CYP3A4 variants that differ significantly from that of the wild type were selected. The selected physicochemical and general protein properties with notable differences would be subjected to correlation analysis in section 4.10. The correlation analysis was to estimate the quantitative correlation between the obtained values of the protein physicochemical properties and the reported relative clearance of the selected variants.

4.6 The prediction of the protein stability of the modelled protein structures of the selected variants upon mutation.

The prediction of stability changes of the modelled protein structures of selected variants was done on I-Mutant 2.0, a tool used for stability prediction upon mutation in protein structures. It indicates whether the mutation of the protein is destabilised or stabilised (Sohaib, Smiline, and Vijayashree, 2019). The Relative Solvent Accessible Area (RSA) value was determined from the protein structure by dividing the accessible surface area value of the altered amino acid region by the free residue surface.

The unfolding Gibbs free energy value of the wildtype was subtracted from the Gibbs free energy value of the mutant protein to get the free energy change value (DDG) (Kcal/mol) used to express all values. A negative symbol (-) denotes a reduction in its stability, whereas a positive sign (+) denotes increased stability.

http://etd.uwc.ac.za/

Table 4.8: The stability index, energy change value and relative solvent area result

Protein Models	Position	WT	NEW	Temperature	pH	RSA	Energy Change Value (DDG) (Kcal/mol)	Stability
CYP3A4 *2	222	S	P	25	7.0	41.0	-1.27	Decrease
CYP3A4 *11	363	T	M	25	7.0	0.7	-0.07	Decrease
CYP3A4 *23	162	R	W	25	7.0	59.4	-0.24	Decrease
CYP3A4 *24	200	Q	H	25	7.0	93.5	-0.98	Decrease

WT: Amino acid in the wild-type protein

New: New substituted amino acid after mutation

DDG: DG (New Protein Energy value) - DG (Wild Type protein energy value) in Kcal/mol

DDG<0: Decrease stability

DDG>0: Increase stability

UNIVERSITY of the
WESTERN CAPE

pH: -log[H+]

RSA: Relative solvent accessible area

The energy change values have significant quantities for all the selected variants (Table 4.8). Jayakanthan *et al.* (2010) in the analysis of CYP3A4 interaction with HIV-1 protease drugs reported that the CYP3A4 complex (CYP3A4 and the reacting molecule) with a higher free energy change value has a higher preference for metabolism (metabolic clearance). This report is consistent with findings in the energy change values for the selected CYP3A4 variants with significant values (Table 4.8). The energy change value might therefore account for the varied rate of metabolism of lidocaine with the selected CYP3A4 variants.

http://etd.uwc.ac.za/

The result obtained (Table 4.8) from the DDG, therefore, gives a prediction that the stability of all selected mutated protein structures decreased from that of the wild type. This indicates that the single nucleotide polymorphism might affect the stability of the selected variants' protein structures.

The relative solvent accessible area (RSA) values of the selected variants also indicated noticeable differences relative to the wildtype (Table 4.8). The solvent-accessible surface area of substrates is a determinant of susceptibility to CYP3A4-mediated metabolism (Singh *et al.,* 2003). RSA might be a determinant property for the reported variation in the rate of metabolism of selected CYP3A4 variants. Further investigation was done in this study (section 4.10) to investigate if there is a link between the free energy change value and the functional differences (relative clearance) in the selected CYP3A4 variants. The likeable link would also be investigated for the RSA values.

4.7 The solvation free energy of protein structures of variants

The solvation energy was reported to be one of the important forces responsible for protein folding and interactions in water. It also guides the stability of the protein structure in water (Matubayasi, 2017; Kraml *et al.,* 2019). The investigation was carried out to see the effect of SNP on the solvation free energy of the selected variants of CYP3A4. The result (Figure 4.26) indicates that the solvation energy value for the wild type differs from that of the selected variants. The highest solvation free energy was observed in the wildtype with a noticeable decrease in the solvation free energy in the other variants. Studies have shown that solvation energy affects the structural behaviour of drug molecules, this includes their ability to bind to target receptors (Dasari and Mallik, 2020; Li *et al.,* 2020). Li *et al.* 2020 reported that a drug molecule interacts with a target receptor when both species desolvate and this change in free energy is included in the total binding constant. This implies that a decrease or increase in solvation energy might have an equivalent effect on the free binding energy. Findings have shown that the binding affinity (energy) of a drug to a receptor could affect the clearance of the drug (Eigenmann *et al.,* 2017; Smith, Gagnon, and Waters, 2017; Wang *et al.,* 2020;).

http://etd.uwc.ac.za/

This indicates that the amino acid substitution in the variants might influence the solvation free energy and subsequently the binding energy.

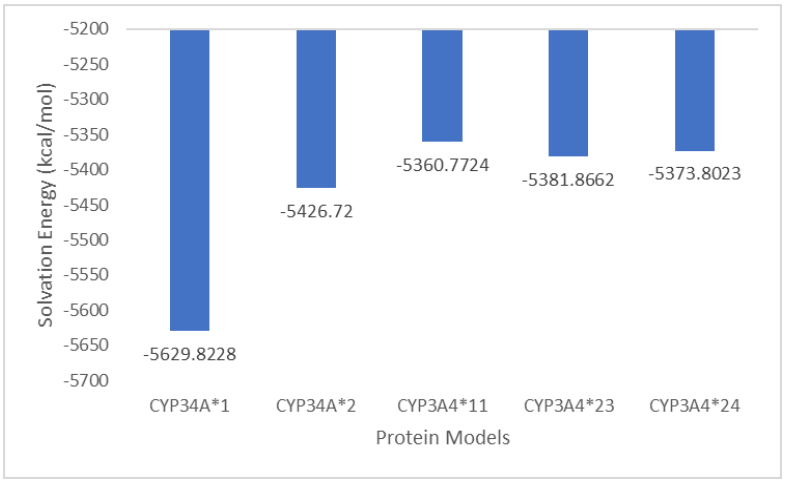


Figure 4.28: Graphical representation of the free energy of solvation of the protein structures of the variants and the wildtype.

There is an indication of higher free energy of solvation in CYP3A4*1(wildtype) relative to the selected variants.

4.8 The free binding energy of the CYP3A4 wild type and the selected CYP3A4 variants

To investigate the effect of single nucleotide polymorphism (SNP) in the free binding energy of CYP3A4 variants-lidocaine complexes, the free binding energy of the wildtype/selected variants-lidocaine complexes was obtained from the docking procedure and the Molecular Mechanics-Generalised Born Surface Area (MM-GBSA) method (Mali and Chaudhari, 2018; Rajagopal, Arumugasamy, and Byran, 2019; Choudhary *et al.*, 2020). It was observed that the calculated free binding energy from the MM/GBSA of the selected variants differs from the wild type (Table 4.9). This was also observed in the initial docking scores obtained for the variants and the wild type. However, the values of the free binding free energy obtained from the MM/GBSA calculation differ from the initial docking score. This may be due to the solvation energy parameters incorporated in the MM/GBSA calculations (Choudhary *et al.*, 2020). It may also be associated with additional

http://etd.uwc.ac.za/

energy components used in the calculation of the binding energy in MM GBSA (E. Wang *et al.,* 2021).

The varied binding energy in both methods (docking procedure and MMGBSA) for the wildtype and the selected variants may imply that the single nucleotide polymorphism may affect the free binding energy of the selected CYP3A4 variants-lidocaine complexes. Previous studies have shown that the binding affinity (energy) of a drug to a receptor could modulate the clearance of the implicating drug (Eigenmann *et al.*, 2017; Smith, Gagnon, and Waters, 2017; Wang *et al.*, 2020). Therefore, the free binding energy might be a determining parameter in the reported variation in the relative clearance of the selected CYP3A4 variants.

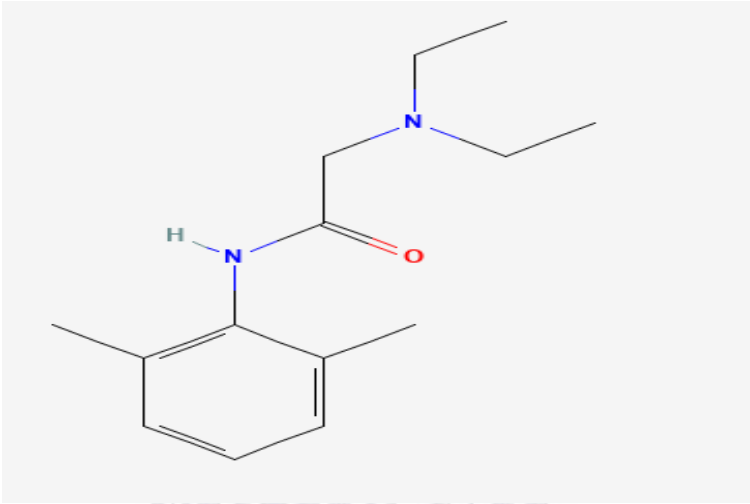
The result of the protein-ligand interaction indicated varied positions and numbers of hydrophobic bonds in the selected CYP3A4 variants, relative to the wild type (Fig 4.28-4.32). Although all the selected CYP3A4 variants had the same number of hydrogen bonds as the wild type (one hydrogen bond each), it was observed that the positions of the hydrogen bond interaction for variants CYP3A4*23 and CYP3A4*24 differ from the wildtype (Table 4.9). It was estimated that the single nucleotide polymorphism is likely to influence the protein-ligand interaction and subsequently, the binding energy of the implicating enzymes.

UNIVERSITY of the
WESTERN CAPE

The results of the binding free energy, hydrogen bonding, lipophilic energy, and the pi-pi packing correction are shown in Table 4.10. The calculated energy components (hydrogen, lipophilic, van der Waals, and pi-pi energy) of the selected CYP3A4 variants differ significantly from the wild type. Yao *et al.* (2018) reported that hydrogen bonding with other hydrophobic energies can significantly increase the binding affinity and selectivity of a receptor. It has also been reported that the binding affinity (energy) of a drug to a receptor could modulate the clearance of the implicating drug (Eigenmann *et al.,* 2017; Smith, Gagnon, and Waters, 2017; Wang *et al.,* 2020;). In addition, previous findings have established that protein-ligand interaction energies can be used to estimate binding potency trends (Kohut *et al.*, 2018; Thapa *et al.*, 2018; Thapa and Raghavachari, 2019;). Therefore, the

http://etd.uwc.ac.za/

differences in the free binding energy and the calculated component energies could be determinant properties of the reported variation in the metabolic rate of the selected CYP3A4 variants. The possible relationship between the calculated energy components and the reported variation in the CYP3A4 rate of metabolism would be investigated in section (4.10).



WESTERN CAPE

Figure 4.29: 2D structure of lidocaine ligand. Molecular formula: $C_{14}H_{22}N_2O$, Molecular weight: 234.34g/mol, IUPAC name: 2-(diethylamino)-N-(2,6-dimethylphenyl)acetamide, Canonical SMILES: CCN(CC)CC(=O)NC1=C(C=CC=C1C)C), International Chemical Identifier (InChI): InChI=1S/C14H22N2O/c1-5-16(6-2)10-13(17)15-14-11(3)8-7-9-12(14)4/h7-9H,5-6,10H2,1-4H3,(H,15,17)

http://etd.uwc.ac.za/

Table 4.9: The docking score and bond interactions of CYP3A4 wildtype and the selected variants with lidocaine ligand.

Wild type/Variants	Hydrophobic bond	Hydrogen bonds interacting with amino acid	Docking score(kcal/mol)	Total binding free energy	Hydrogen-bonding correction (H-bond)	Lipophilic energy (Lipo)	Van der Waals energy (vdW)	Pi-pi packing correction
CYP3A4*1	ALA305, PHE304, ILE301, ILE300, ILE369, ILE120, PHE108, LEU210, LEU211, PHE241, MET114	SER119	-5.916	-5.90	-0.39	-23.25	-29.16	-0.81
CYP3A4*2	PHE241, LEU210, LEU211, PHE304, ALA305, LEU482, ILE301, ILE300, MET114, ILE120, PHE108	SER 119	-5.589	-13.57	-0.51	-22.41	-31.08	-1.36
CYP3A4*11	PHE241, MET114, LEU211, ILE21O, ILE300, ILE301, ILE120, PHE304, ALA30, ILE369, ALA370	SER 119	-4.595	-13.36	-0.38	-23.25	-29.12	-2.02
CYP3A4*23	LEU373, ALA370, PHE57, PHE108, TYR53, ILE50, PHE215	GLU 374	-3.374	-35.53	-1.36	-16.04	-32.35	-3.14
CYP3A4*24	LEU462, LEU483, ALA30, PHE304, ILE369, ALA370, MET371, PHE57, LEU482, LEU483, GLN484.	ILE369	-4.667	-33.39	-0.04	-14.90	-31.96	-4.22

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

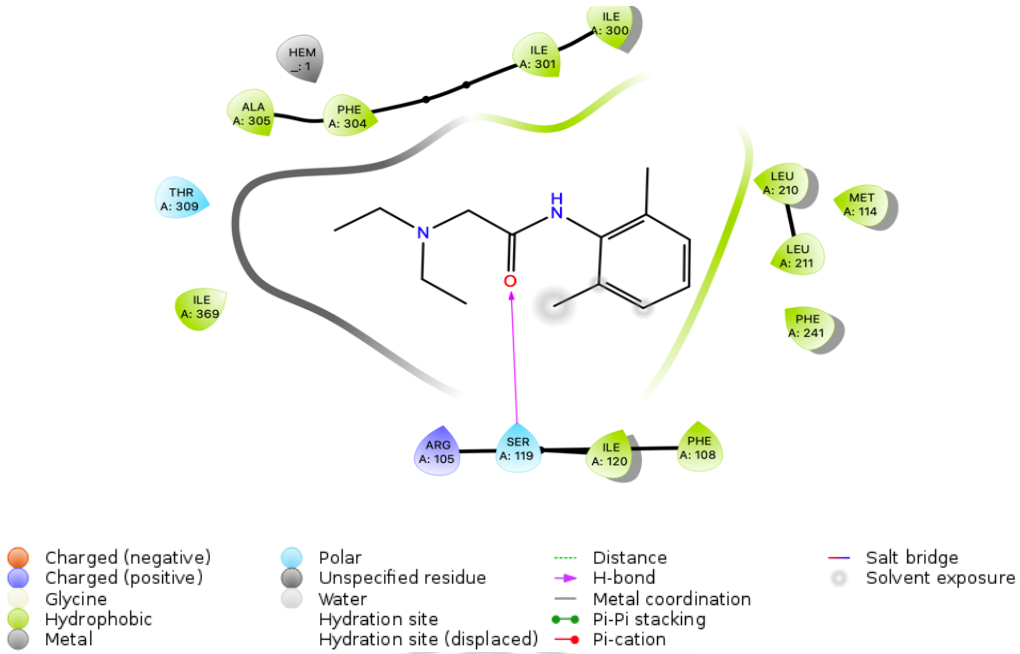


Figure 4.30: Protein-ligand interaction of lidocaine and CYP3A4*1(wildtype).

It shows the 2D structure of the ligand(lidocaine), the hydrophobic bond interacting residues (green), hydrogen bonds (arrow-headed purple line)

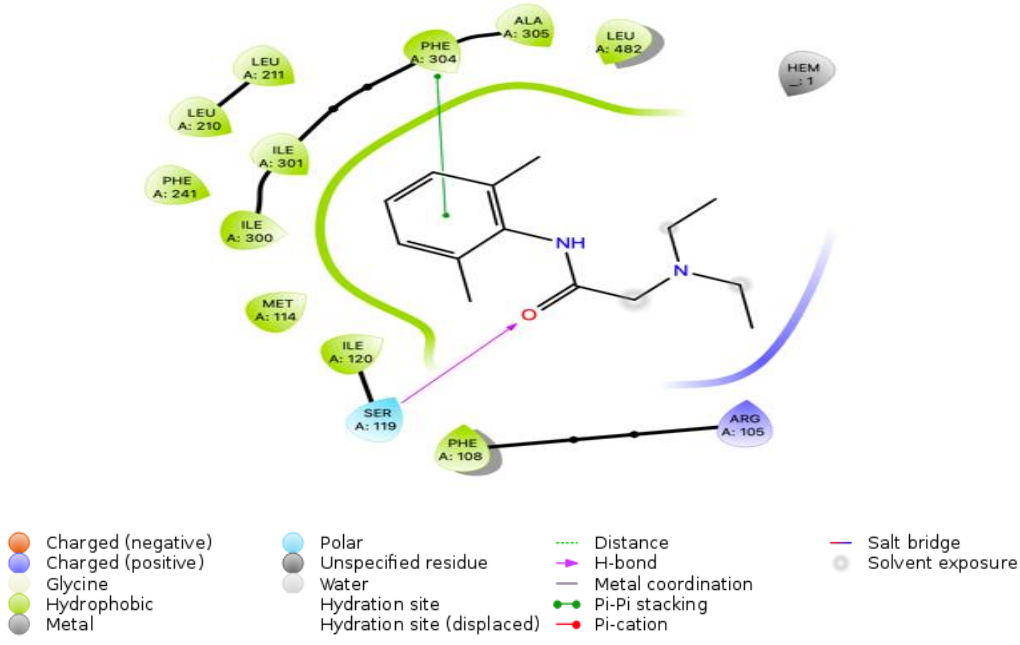


Figure 4.31: Protein-ligand interaction of lidocaine and CYP3A4*2.

http://etd.uwc.ac.za/

It shows the 2D structure of the ligand(lidocaine), the hydrophobic bond interacting residues (green), hydrogen bonds (arrow-headed purple line)

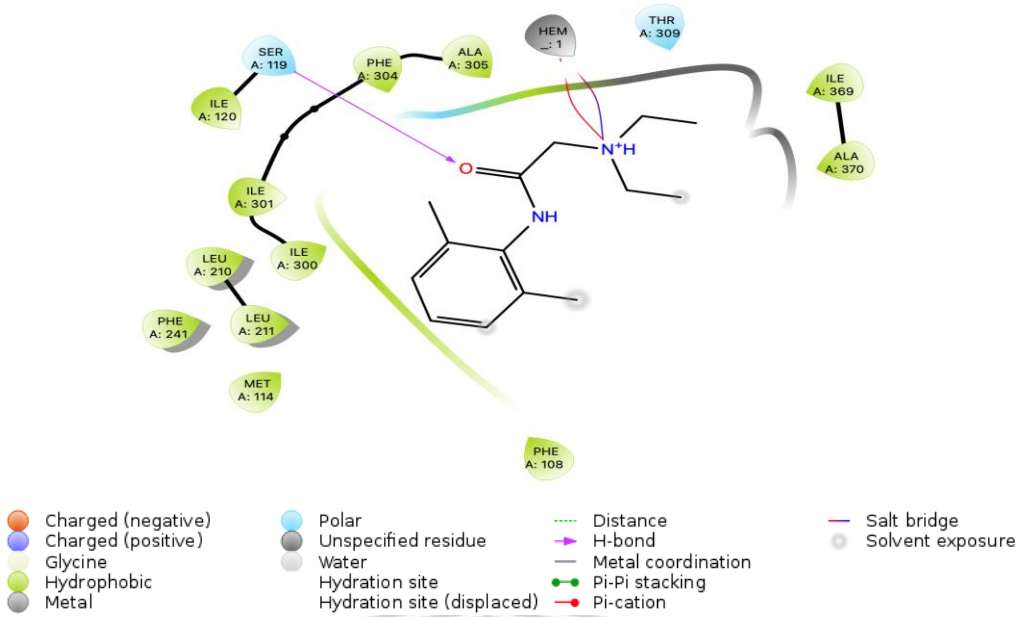


Figure 4.32: Protein-ligand interaction of lidocaine and CYP3A4*11.

It shows the 2D structure of the ligand(lidocaine), the hydrophobic bond interacting residues (green), hydrogen bonds (arrow-headed purple lines)

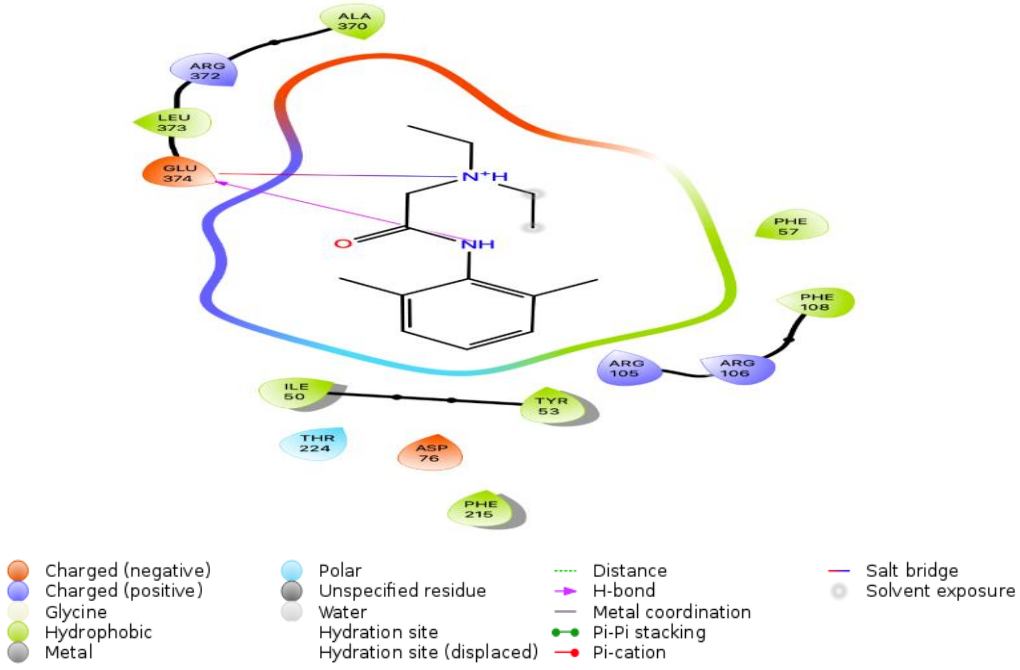


Figure 4.33: Protein-ligand interaction of lidocaine and CYP3A4*23.

http://etd.uwc.ac.za/

It shows the 2D structure of the ligand (lidocaine), the hydrophobic bond interacting residues (green), and hydrogen bonds (arrow-headed purple lines).

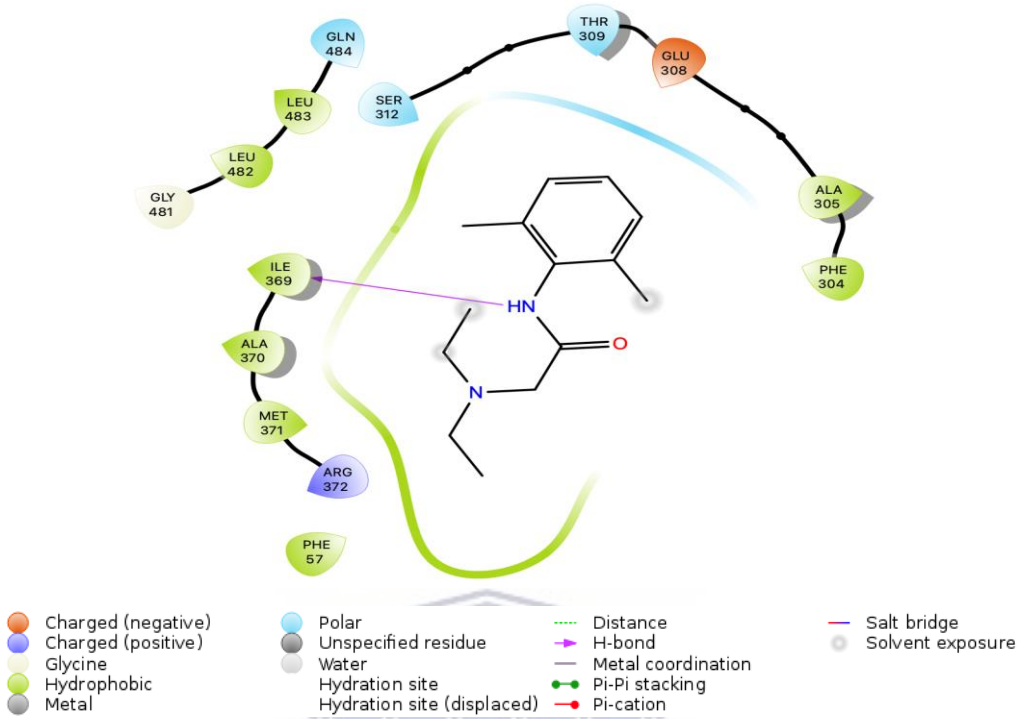


Figure 4.34: Protein-ligand interaction of lidocaine and CYP3A4*24.

It shows the 2D structure of the ligand(lidocaine), the hydrophobic bond interacting residues (green), hydrogen bonds (arrow-headed purple line)

UNIVERSITY of the
WESTERN CAPE

The binding free energies of the CYP3A4 variants/wild-type protein structures and the lidocaine ligand were also calculated with the Molecular Mechanics-Generalised Born Surface Area (MM-GBSA) method (Mali and Chaudhari, 2018; Rajagopal, Arumugasamy, and Byran, 2019; Choudhary *et al.*, 2020). MM-GBSA has been described as an accurate and efficient computational approach to calculating protein-ligand binding free energies (Mali and Chaudhari, 2018; Arumugasamy, and Byran, 2019). The accuracy has been ascribed to its ability to calculate the solvation energy parameters that analyse the solute and solvent interactions (Wang *et al.*, 2021). The results of the binding free energy, hydrogen bonding, lipophilic energy and the pi-pi packing correction are shown in Table 4.10.

http://etd.uwc.ac.za/

Table 4.10: The calculated component interacting energies result in Kcal/mol.

The table contains the component interacting energies (hydrogen bond, lipophilic energy, van der Waals energy and pi-pi packing) of the CYP3A4 wildtype/selected variants-lidocaine complexes.

CYP3A4 enzymes	Total binding free energy	Hydrogen-bonding correction (H-bond)	Lipophilic energy (Lipo)	Van der Waals energy(vdW)	Pi-pi packing correction
Wildtype	-5.90	-0.39	-23.25	-29.16	-0.81
CYP3A4*2	-13.57	-0.51	-22.41	-31.08	-1.36
CYP3A4*11	-13.36	-0.38	-23.25	-29.12	-2.02
CYP3A4*23	-35.53	-1.36	-16.04	-32.35	-3.14
CYP3A4*24	-33.39	-0.04	-14.90	-31.96	-4.22

UNIVERSITY of the
WESTERN CAPE

It was observed that the calculated free binding energy from the MM/GBSA of the selected variants differs from the wild type (Table 4.10). This was also observed in the initial docking scores obtained for the variants and the wild type, however, the values of the free binding free energy obtained from the MM/GBSA calculation differ from the initial docking score. This may be due to the solvation energy parameters incorporated in the MM/GBSA calculations (Choudhary *et al.,* 2020). It may also be associated with additional energy components used in the calculation of the binding energy in MM GBSA (E. Wang *et al.,* 2021).

The calculated energy components (hydrogen, lipophilic, van der Waals and pi-pi energy) of the selected CYP3A4 variants differ significantly from the wild type. Yao *et al.* (2018) reported that hydrogen bonding with other hydrophobic energies can significantly increase the binding affinity and selectivity of a receptor. It has also been reported that the binding affinity (energy) of a drug to a receptor could

http://etd.uwc.ac.za/

modulate the clearance of the implicating drug (Eigenmann *et al.*, 2017; Smith, Gagnon, and Waters, 2017; Wang *et al.*, 2020). In addition, previous findings have established that protein-ligand interaction energies can be used to estimate binding potency trends (Kohut *et al.*, 2018; Thapa *et al.*, 2018; Thapa and Raghavachari, 2019). Therefore, the differences in the free binding energy and the calculated component energies could be determinant properties of the reported variation in the metabolic rate of the selected CYP3A4 variants. The possible relationship between the calculated energy components and the reported variation in the CYP3A4 rate of metabolism would be investigated in section (4.10).

4.9 Molecular dynamics simulations studies of the protein structures of the wild type and the selected variants.

According to previous reports, mutations may produce changes in interaction patterns or conformations with notable effects on the dynamic properties of the mutated protein structures which could affect their enzymatic activity (Tyukhtenko *et al.*, 2018; Rodrigues, Pires, and Ascher, 2018). The effects of single nucleotide polymorphism (SNPs) on the dynamic properties of the mutated protein structures were investigated with molecular dynamics simulation in which the root mean square deviation (RMSD), root mean square fluctuation (RMSF), and the radius of gyration (Rg) of the wild type and the selected variants were studied. The **s**olvent-accessible surface area **(**SASA**)** and the secondary structure elements (SSE) of the variants were also investigated in this section.

4.9.1 The root mean square deviation of the wild type and the selected variants of CYP3A4

The standard measurement of the structural average distance that exists between coordinates of atoms (mostly backbones) of superimposed protein was achieved with root mean square deviation (RMSD). RMSD helps to determine the conformational change in the protein backbone (Salmaso and Moro, 2018; Lazar *et al.*, 2020). The trajectory evaluation of the RMSD of the backbone was done for 100ns, and the average and standard deviation of the RMSD were calculated from the data obtained from the simulation (Table 4.11).

http://etd.uwc.ac.za/

Table 4.11: Average RMSD of the wild type and the selected mutants of CYP3A4

Protein	Average RMSD(Å)
CYP3A4*1	1.57±0.137
CYP3A4*2	1.71±0.147
CYP3A4*11	2.02±0.247
CYP3A4*23	2.19±0.292
CYP3A4*24	2.00±0.162

As shown in the result obtained in Table 4.11 the wild type has the least value of overall mean RMSD of 1.57 with a statistically significant difference with all the selected variants except variant 2. This was an indication that the wild type has a higher stability in comparison with all the selected variants except variant 2. The result obtained shows a notable change in the RMSD of all variants relative to the wild type; there is a higher RMSD value for all variants, an indication of decreased stability as shown in Figure 4.33.

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

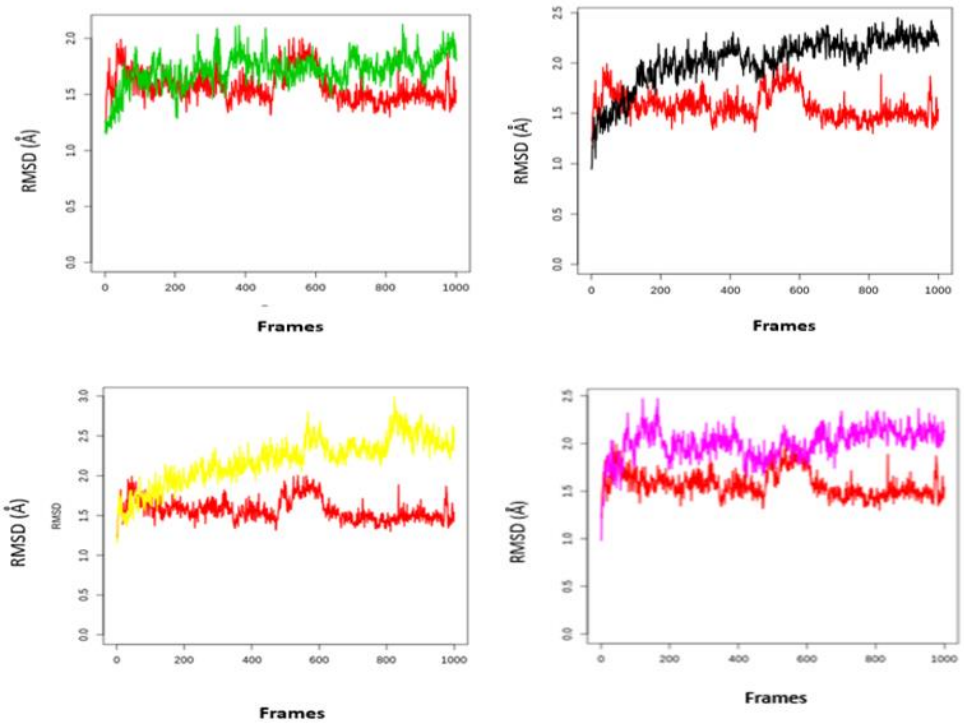


Figure 4.35: The comparative RMSD plot of the wildtype and the selected variants.

The wild type(red), CYP3A4*2 (green), CYP3A4*11(black), CYP3A4*23 (yellow), and CYP3A4*24 (purple), indicating a higher RMSD in the selected CYP3A4 variants

This might imply that the substituted amino acid at the position of mutation significantly influences the stability of the protein structures of all variants except variant 2. This agrees with the result obtained from the initial prediction of stability in the method reported previously in section 4.6.1 of this study. The comparative plots of each selected variant against the wild type (Fig 4.33) indicated that CYP3A4*11 has the highest overall mean RMSD variation value, it was predicted to be the least stable of the selected CYP3A4 variants. However, all selected CYP3A4 variants and wild type reached an equilibrium between 800 to 1000 frames.

The molecular dynamics analysis of the protein-ligand complexes and protein-ligand interactions of the selected variants and wild type were done to study the differences in the CYP3A4 complex structures and their interactions with lidocaine. It was observed from the analysis that both the wildtype and the selected CYP 3A4

http://etd.uwc.ac.za/

variants protein structures (unbounded) have higher stability than the protein-ligand complex (Figure 4.34). However, from the plots (Fig 4:34), it was observed that the difference between the RMSD values for the CYP3A4 protein structures (unbounded) and protein-ligand complex is larger in the selected variants than in the wild type (at the 400 to 1000frames end of simulation equilibration) except in CYP3A4*24. This might imply that the lidocaine ligand was more aligned to the protein receptor in the wildtype than the selected CYP3A4 variants except in CYP3A4*24. This further supports the estimation that the single nucleotide polymorphism might affect the stability of the protein-lidocaine complex of the selected variants. However, a correlation study would be done to investigate any likeable link between the average RMSD of the complexes and the reported differences in the rate of metabolism in section 4.10.

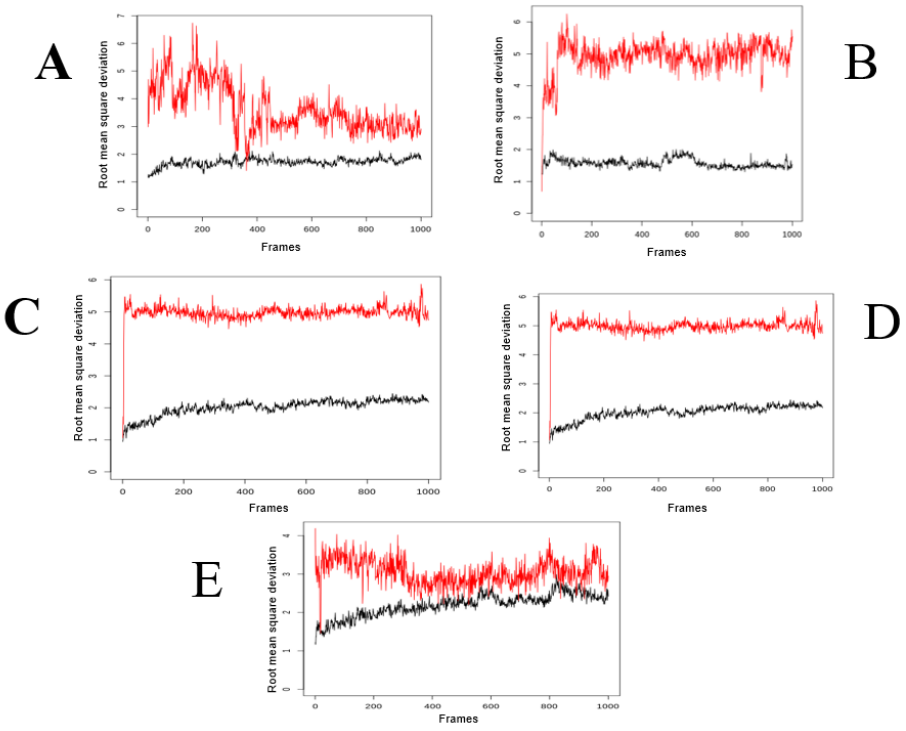


Figure 4.36: The comparative plots of the CYP3A4 protein-lidocaine complex (red) and the free protein structures (black).

The difference between the RMSD values for the CYP3A4 protein structures (unbounded) and protein-ligand complex is larger in the selected variants than in the wildtype (at the 400 to 1000frames end of simulation equilibration) except in

http://etd.uwc.ac.za/

CYP3A4*24. (A-wildtype, B-CYP3A4*2, C-CYP3A4*11, D-CYP3A4*23, E-CYP3A4*24).

4.9.2 The analysis of the root mean square fluctuation of the selected variants and the wild type

The RMSF of the protein backbone was calculated to determine the effect of the mutations on the structural flexibility of the protein structures of the selected CYP3A4 variants when compared to the wild type (Salo-Ahen *et al.*, 2021). The result obtained as shown in Table 4.12 indicated that the wild type has the lowest value for the overall average RMSF, however, the difference between the RMSF value for the variants and the wildtype was not statistically significant, therefore the mutation might not have affected the rigidity of the protein structure's interaction with ligands.

Table 4.12: The average root mean square fluctuation of the wild type and the selected variants of CYP3A4

Protein	Average RMSF(Å)
CYP3A4*1	0.89±0.583
CYP3A4*2	0.94±0.454
CYP3A4*11	0.93±0.539
CYP3A4*23	1.01±0.618
CYP3A4*24	0.95±0.442

From the plots of the RMSF against the residue index (Figure 4.35), each of the variants has more numbers of flexible regions than the wild type. Although there was an observed fluctuation in residue 380 to 400 in the wildtype. Variant 24 has the highest number of regions as seen in the RMSF value and the plot. The regions of fluctuation in two of the variants (CYP3A4*2 and CYP3A4*23) fall between the residue 209-217 and 237-242, which were the core active regions in the CYP3A4 structures (Yano *et al.*, 2004). The mutation at position 222 from serine to proline in CYP3A4*2 falls within the reported core active regions. This implies that the

http://etd.uwc.ac.za/

mutation could affect the flexibility of the active regions of the mutant's protein structures of CYP3A4*2 and CYP3A4*23.

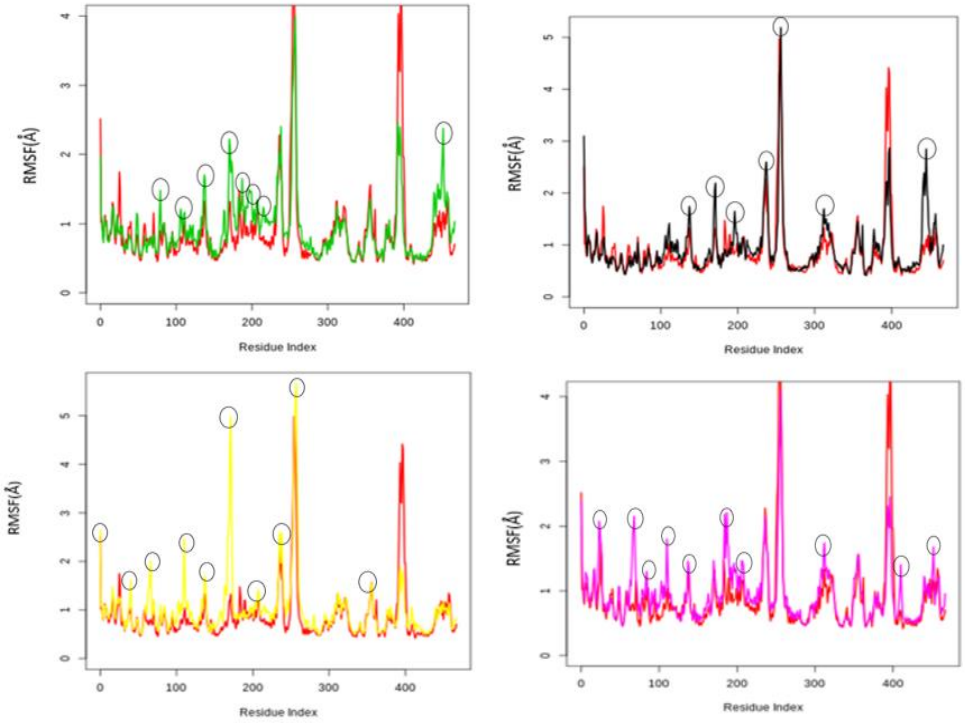


Figure 4.37: The comparative RMSF plot of the wildtype and the selected variants.

The wild type(red) and CYP3A4*2 (green), CYP3A4*11(black), CYP3A4*23 (yellow), and CYP3A4*24 (purple), indicate higher flexibility in the selected CYP3A4 variants

There were observed fluctuations between residue 215 to 220 in CYP3A4*24(Fig. 4.35) which was the position of the Phenylalanine cluster in the CYP3A4 that was responsible for the oxidative allosteric regulation pathway (Davydov *et al.*, 2012). Although the reported substitution in CYP3A4*24 was said to occur at position 200, however its proximity to the phenylalanine cluster region might be responsible for the observed fluctuations. This increased flexibility may affect the allosteric regulation of the implicated protein structure. All variants appear to be less rigid than the wild type, with increased regions of fluctuation when compared with the wild type as shown in Figures 4.35.

http://etd.uwc.ac.za/

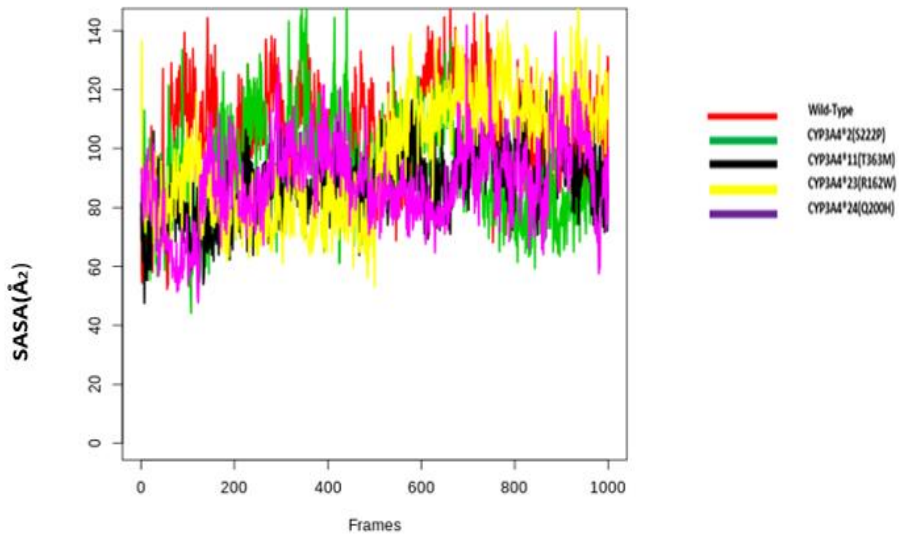
4.9.3 Analysis of the solvent-accessible surface area (SASA) of the selected variants and the wild type

The solvent-accessible surface area (SASA) of a protein is the region of the protein structures that can be in contact with water molecules, that is the hydrophilic regions (Durham *et al.*, 2009; Ausaf *et al.*, 2014). The change in solvent accessible surface area of the wild type and variants was calculated to study the main hydrophilic domain. As shown in Table 4.13, the result for the average SASA for both the variants and the wild-type, there was a significant decrease in the average SASA for all selected variants (though with varying degrees of decrease) when compared with the wild-type.

Table 4.13: The average solvent accessible surface area of the wild type (highest value) and the selected mutants of CYP3A4

Protein	Average SASA (Å^2)
CYP3A4*1	105.06±16.069
CYP3A4*2	92.37±16.377
CYP3A4*11	87.94±12.164
CYP3A4*23	96.67±19.863
CYP3A4*24	88.43±14.175

UNIVERSITY of the
WESTERN CAPE



http://etd.uwc.ac.za/

Figure 4.38 The comparative SASA plot of the wild type (red) and the selected variants.

Indicating a higher SASA in the wild type and a lower SASA in the other variants, CYP3A4*2 (green), CYP3A4*11 (black), CYP3A4*23 (yellow), CYP3A4*24 (purple)

4.9.4 The analysis of the radius of gyration (Rg) of the selected variants and the wild type.

The radius of gyration (Rg) is a measure of the extent of compactness of a protein structure (Sharn, Singh, and Singh, 2021). To investigate the effect of mutation on the backbone of the protein structure compactness of selected variants, the average Rg for all variants was calculated and compared with the Rg of the wild type. There was a minimal increment in the values obtained for variants CYP3A4*2 and CYP3A4*23 while a minimal decrease was observed in CYP3A4*11 and CYP3A4*24 (Table 4.14). It was predicted that the mutation of selected variants may not significantly affect the compactness of interacting protein structures.

Table 4.14: The radius of gyration of the selected CYP3A4 variants and the wild type

Protein	Radius of Gyration(Å)
CYP3A4*1	3.30
CYP3A4*2	3.35
CYP3A4*11	3.11
CYP3A4*23	3.32
CYP3A4*24	3.20

UNIVERSITY of the WESTERN CAPE

4.9.5 The analysis of protein secondary structure of the selected variants and the wild type

The protein secondary structure elements (SSE) which include the alpha helices and the beta strands were monitored in the simulation process. The total summary of the SSE distribution in the protein structure for the selected variants and wild type trajectory frames throughout the simulation process is shown in Table 4.15.

http://etd.uwc.ac.za/

Table 4.15: The total summary of the SSE distribution in the protein structure of the selected variants and the wild type. CYP3A4 wild type (CYP3A4*1) has the highest percentage helix.

Protein	Percentage Helix	Percentage Strand	Percentage Total Secondary Structure Element
CYP3A4*1	41.92	6.43	48.35
CYP3A4*2	41.38	7.51	48.88
CYP3A4*11	41.64	7.04	48.68
CYP3A4*23	40.89	6.82	47.71
CYP3A4*24	41.34	7.73	49.07

The result shows that the wildtype has the highest percentage of the helix, while the mutated protein structures have a reduced percentage of the helix. It has been reported structurally that the high percentage of alpha-helices present in the catalytic regions leads to significant enzyme flexibility and more substrate egress channels (Fishelovitch *et al.*, 2009). This suggests that the mutation affects the SSE, specifically the percentage helix, hence, having an equivalent effect on the metabolic activity of the protein structures of variants.

UNIVERSITY of the
WESTERN CAPE

4.9.6 The analysis of Hydrogen bond interactions between the ligand(lidocaine) and the selected CYP3A4 variants and the wild type

The protein-ligand interactions analysis indicated that CYP3A4*11 and CYP3A4*23 have a higher hydrogen bond interaction relative to the wildtype (Figure 4.37). More hydrophobic interactions were also observed in the wildtype and CYP3A4*2. The observed differences in hydrogen bonding interactions in the selected CYP3A4 variants relative to the wild type might account for the reported variations in the relative clearance. This was consistent with previous reports as hydrogen bonds in CYP3A4 interactions have been reported to play an important function in ligand binding (Mukhtar, Sajid Kiani, and Jabeen 2017; Yao *et al.*,2018)

http://etd.uwc.ac.za/

also reported that hydrogen bonding with other hydrophobic energies can significantly increase the binding affinity and selectivity to a receptor.

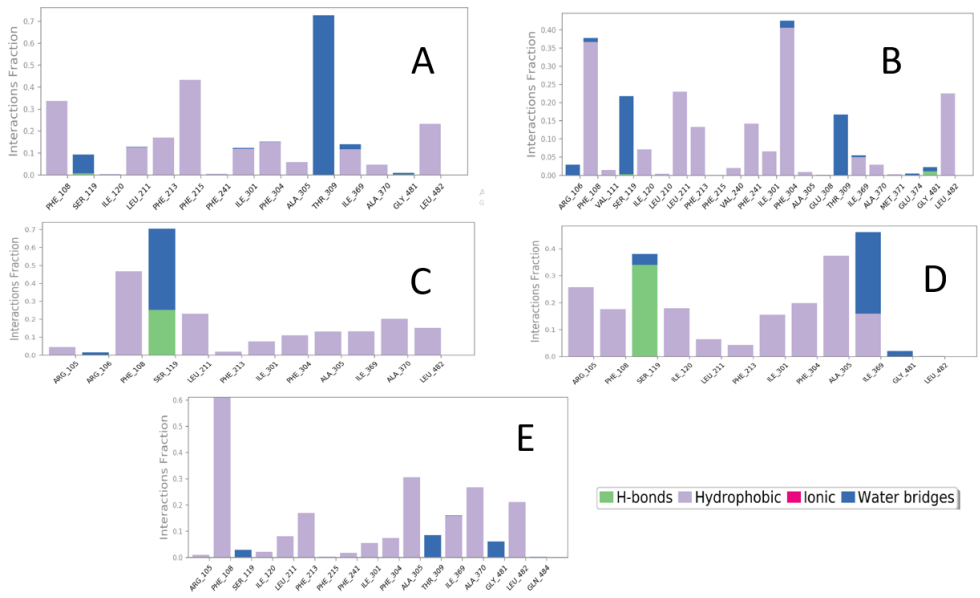


Figure 4.39: The comparative plots of the CYP3A4 protein-ligand interaction. (A - wildtype, B - CYP3A4*2, C - CYP3A4*11, D - CYP3A4*23, E - CYP3A4*24) with a higher hydrogen bond interaction in CYP3A4*11 and CYP3A4*23 have a higher hydrogen bond interaction relative to the wildtype

UNIVERSITY of the
WESTERN CAPE

4.10 The correlation plot result and discussion

The complete result obtained from the characterised parameters with noticeable differences in the selected variants was summarised in *Table 4.16*. A correlation graph was plotted for each parameter as shown in **Appendix B** (Figure B.0.1 to B.0.34) to investigate if there was a quantifiable correlation between the reported functional difference (relative clearance) and the differences observed in the data obtained from the characterization investigation. It has been reported in previous studies that the Pearson correlation coefficient gives an indication of the strength of correlation as reported in previous studies (Stefaniu, Rasul, and Hussain, 2020; Oselusi, Egieyeh, and Christoffels, 2021).

http://etd.uwc.ac.za/

Table 4.16: The summary data of characterised properties of the wild type and the selected variants with a significant difference.

Characterised properties of wildtype and selected variants with a noticeable difference	Wildtype	Variant 2	Variant 11	Variant 23	Variant 24
Reported relative clearance*	100	27.93	213.61	206.96	30.29
Molecular Weight (Da)	57343.18	57353.22	57373.27	57373.20	57352.19
Gravy	0.040	0.042	0.035	0.033	0.039
RMSD(Å)	N/A	0.015	0.033	0.020	0.029
Energy change value Kcal/mol)	N/A	-0.07	0.24	0.98	-1.27
Solvation energy (Kcal/mol)	-5629.82	-5426.77	-5360.77	-5381.87	-5373.80
Docking score (Kcal/mol)	-5.916	-5.589	-4.4595	-3.374	-4.4667
Relative solvent accessible area ($Å^2$)	N/A	41.0	0.7	0.24	0.98
Average RMSD (complex structures) (Å)	4.93	3.66	4.97	3.044	1.41
Molecular lipophilicity potential minimum	-28.39	-28.46	-28.53	-27.44	-28.39

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

Molecular lipophilicity potential maximum	23.93	23.83	24.09	23.95	23.93
Molecular lipophilicity potential mean	-3.45	-3.43	-3.45	-3.37	-3.44
Coulombic electrostatic potential minimum kcal/(mol·e)	–11.21	-11.22	-13.88	-11.24	-15.41
Coulombic electrostatic potential mean kcal/(mol·e)	0.38	0.39	0.37	0.18	0.37
Coulombic electrostatic potential maximum kcal/(mol·e)	15.70	15.61	15.68	15.58	15.69
Average RMSF(Å) for the protein backbone	0.89	0.94	0.93	1.01	1.02
Average solvent-accessible surface area ($Å^2$)	105.06	92.37	255.59	258.41	88.43
Radius of Gyration(Å)	3.30	3.35	3.11	3.32	3.20
Percentage Helix (%)	41.92	41.38	41.64	40.89	41.34
Percentage Strand (%)	6.43	7.51	7.04	6.82	7.73
Percentage Total SSE (%)	48.35	48.88	48.68	47.71	49.07

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Accessible surface Area (Water probe) *(Å^2)	20369.7	20407.5	20375.3	20419.1	20353.5
Hydrophobic Surface Area*(Å^2)	11941.1	11973.5	11925.1	12037.4	11944.4
Hydrophilic surface Area*(Å^2)	7448.2	7448.2	74616.6	7433.2	7386.7
Protein volume Å^3	50087.6	50033.6	50083.6	50113.4	50098.0
Protein mobility(cm2/V)	4.3	4.4	4.3	3.5	4.4
Protein Net charge	5.56	5.63	5.57	4.60	5.66
Sum positive surface area*(Å^2)	12080.30	12090.82	12095.13	12254.23	12337.73
Sum negative surface area*(Å^2)	6996.98	6860.53	6858.70	6591.23	6606.73
Sum donor surface area*(Å^2)	3379.18	3296.91	3313.10	3134.33	3146.54
Sum acceptor surface area*(Å^2)	4015.71	4048.43	4062.33	3973.62	4001.24
Protein Charge at Debye Length	1.43	1.45	1.44	1.18	1.46
Sum of Aggregation Score	630.13	621.64	611.90	687.59	687.17

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

Hydrogen-bonding correction (Kcal/mol)	-0.39	-0.51	-0.38	-1.36	-0.04
Lipophilic energy (Kcal/mol)	-23.25	-22.41	-23.25	-16.04	-14.90
Van der Waals energy (Kcal/mol)	-29.16	-31.08	-29.12	-32.35	-31.96
Pi-pi packing correction	-0.81 -	-1.36	-2.02	-3.14	-4.22
Hydrophobicity Moment	2989.8	2070.2	2069.9	2038.0	2089.3
Zeta potential at Debye Length*(mV)	7.48	7.58	7.49	7.62	7.62
Protein Dipole Moment (Debye)	1056.67	1039.69	1056.96	1152.37	1048.99
Debye screening length K-1	1056.67	1056.67	1056.67	1056.67	1056.67

**-reported relative clearance value by Fang et al. (2017).*

Note: Check Figure B.0.1 to B.0.33 for the plotted graphs of relative clearance against the studied parameters (with a noticeable difference) of the selected variants.

Surprisingly, more than sixty-five percent (forty out of the total sixty-one) of the studied and characterised parameters of the CYP3A4 selected variants showed noticeable differences relative to the wild type (Table 4.16). All studied parameters (physicochemical, dynamics, and structural) with noticeable differences between the wildtype and the selected CYP3A4 variants were subjected to a correlation study. The correlation study was done to investigate the relationship between the

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

reported variation in the rate of metabolism (relative clearance) of the selected CYP3A4 variants and the structural, physicochemical, and dynamic properties of the selected variants. Twenty-one of the characterised properties (structural, physicochemical, and dynamics) with noticeable differences indicated a positive and negative correlation (Table 4.17) with the reported variation in the relative clearance of the CYP3A4 selected variants. This was observed in the Pearson's correlation coefficients obtained from the plotted graphs as shown in **Appendix B** (Figure B.0.1 to B.0.33).

Table 4.17: The correlation interpretation of characterised parameters (physicochemical, dynamic, and structural properties) with the reported difference in the rate of metabolism (relative clearance) of the wild type and the selected variants of CYP3A4

Properties	Pearson Correlation factor(r)	Interpretation
Molecular weight (Da)	0.816	Strong positive linear correlation
Gravy	-0.819	Strong negative linear correlation
Root mean square Deviation(Å)	-0.945	Strong negative linear correlation
Free energy change value (DDG). *(Kcal/mol)*	0.818	Strong positive linear correlation
Average RMSD (complex structures) (Å)	0.453	Good positive linear correlation
Solvation Energy *(Kcal/mol)*	0.226	Very weak correlation
Docking Score (Binding energy) *(Kcal/mol)*	0.623	Strong positive linear correlation
RMSF(Å)	0.141	Very weak correlation
The radius of Gyration(Å)	-0.349	Very weak correlation
Percentage Helix *(%)*	-0.175	Very weak correlation
Percentage Strand *(%)*	-0.581	Good negative correlation
Percentage Total Secondary structure elements *(%)*	-0.690	Strong negative linear correlation

http://etd.uwc.ac.za/

Relative solvent accessible area (Å^2)	0.420	Good correlation value
Average solvent-accessible surface area(Å^2)	0.966	Strong positive linear correlation
Radius of Gyration (Å)	-0.349	Very weak correlation
Accessible surface area (Water probe) Å^2	0.311	Weak correlation value
Hydrophobic Surface Area(Å^2)	0.252	Weak correlation value
Hydrophilic surface Area (Å^2)	0.601	Strong positive linear correlation
Protein volume	0.533	Good positive linear correlation
Protein mobility*(cm2/Vs)*	-0.647	Strong negative linear correlation
Protein Net charge	-0.616	Strong negative correlation value
Sum positive surface area (Å^2)	-0.13	Weak correlation value
Sum negative surface area (Å^2)	0.081	Very weak correlation
Sum donor surface area (Å^2)	0.043	Very weak correlation
Sum acceptor surface area (Å^2)	-0.071	Very weak correlation
Protein Charge at Debye Length	-0.603	Strong negative linear correlation
Sum of Aggregation Score	0.049	Very weak correlation
Protein Dipole Moment (*Debye)*	0.65	Strong positive linear correlation
Molecular lipophilicity potential minimum (MLP)	0.48	Good positive linear correlation
Molecular lipophilicity potential mean (MLP)	0.40	Good positive linear correlation
Molecular lipophilicity potential maximum (MLP)	0.78	Strong positive linear correlation
Coulombic electrostatic potential minimum kcal/(mol·e)	0.15	Weak correlation value
Coulombic electrostatic potential mean kcal/(mol·e)	-0.60	Strong negative linear correlation

http://etd.uwc.ac.za/

Coulombic electrostatic potential maximum kcal/(mol·e)	0.15	Weak correlation value
Hydrophobicity Moment	-0.126	Weak correlation value
Hydrogen bond (Kcal/mol)	-0.60	Strong negative linear correlation
Lipophilic energy *(Kcal/mol)*	0.085	Weak correlation value
Van der Waals energy *(Kcal/mol)*	0.218	Weak correlation value
Pi-Pi energy (Kcal/mol)	0.016	Weak correlation value
Zeta potential at Debye Length *(mV)*	-0.28	Weak correlation value

Note: Check Figure B.0.1 to B.0.33 for the plotted graphs with the Pearson correlation values.

From the correlation plots analysis, as shown in **Appendix B** (Figure B.0.1 to B.0.33), it was predicted that twelve of the studied parameters with noticeable variations have a good or strong linear positive correlation (Pearson coefficient ≥4). This implies that the parameters with strong or good positive correlation values are directly correlated to the differences observed in the relative clearance of the selected variants in the metabolism of lidocaine. Nine characterised parameters have a good or strong negative Pearson correlation coefficient (Pearson coefficient ≤-4). This was an indication that they are inversely correlated to the reported differences in the relative clearance of the selected variants. Other characterised parameters showed weak or non-significant correlations with the obtained Pearson correlation coefficient values. All 21 of the total characterised parameters with noticeable differences have a strong correlation (positive and negative correlations) with the reported difference in intrinsic clearance of the variants with lidocaine ligand (Table 4.17).

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Minimum lipophilic potential (MLP) (minimum (0.48), maximum (0.78), and mean (0.40) values) all have a good positive correlation. This gives a prediction that the higher the MLP the greater the rate of metabolism(clearance) and vice versa, this finding was consistent with the report given by Yusra Sajid Kiani and Ishrat Jabeen (2020). It was reported that the metabolic stability and intrinsic clearance of CYP450 substrates were connected to their lipophilicity (log P or log D). It was stated that the higher the log D value of a substrate, the more eagerly it binds to a certain CYP450 enzyme, and the greater it shows more intrinsic clearance (Kiani and Jabeen, 2020).

A strong linear positive correlation (0.642) was observed between the molecular weight and the relative clearance of the selected CYP3A4 variants. Asides from the reports on the effect of varied molecular weight of ligands that react with protein receptors (El-Sayed *et al.,* 2016; Zhang *et al.,* 2021), works of literature haven't given a relevant report on the effect of varied molecular weight (due to mutation) of CY3A4 protein receptors (enzymes). Therefore, this prediction would require further investigation to probe into the effect of the varied molecular weight of CYP3A4 variants on the rate of metabolism and to validate the relationship estimated in this study.

UNIVERSITY of the
WESTERN CAPE

The docking score (free binding energy) and the free energy change value(kcal/mol) have a strong linear positive correlation (0.69 and 0.94), this implies that an increase or decrease in the value of the free binding energy/free energy change value could lead to an equivalent increase or decrease in the enzymatic activity of the selected CYP3A4 enzymes. This finding was consistent with previous reports (Jayakanthan *et al.,* 2010; Saba and Seal 2018). For example, Jayakanthan *et al.* (2010) in the analysis of CYP3A4 interaction with certain drug molecules reported that the CYP3A4 complex (CYP3A4 and the reacting molecule) with higher free binding energy and free energy change value have a higher preference for metabolism (metabolic clearance). The finding was further supported with an in vitro study reported in the study.

http://etd.uwc.ac.za/

The average solvent-accessible surface area (SASA) is a measure of how much solvent molecules are exposed to enzymes. It was estimated in this study that a strong positive linear relationship (0.8) exists between SASA and the reported variation in the rate of metabolism of the selected CYP3A4 variants, relative to the wildtype in lidocaine metabolism. The linear correlation indicates that an increase or decrease in the SASA results in the same enzymatic activity of the selected CYP3A4 variants with lidocaine. This was consistent with the previous study, as reports have shown that the structural integrity of the CYP3A4 structure can be altered with a decrease in SASA. However, an increase in SASA maintains the structural stability (Tang *et al.* 2021; Kehinde *et al.*, 2022;). The report also has it that stability enhances the rate of metabolism (Munjal, Shukla, and Singh 2021), therefore the finding on the direct positive effect of SASA on the rate of metabolism of the selected variants in lidocaine metabolism was consistent with these reports.

The root mean square deviation (RMSD) for the protein-ligand complex has a good positive correlation (0.453) with the reported rate of metabolism of the selected variants of CYP3A4. This implies that a higher RMSD might produce a proportional higher relative clearance of the variants. The relative solvent accessible area did not indicate any significant relationship with the rate of metabolism in the selected variants. This also applies to the radius of gyration. The hydrophobic surface area indicated a weak correlation value (0.252) with the rate of metabolism (clearance) in the selected CYP3A4 variants. A strong positive linear correlation (0.601) was observed between the hydrophilic surface area and the reported rate of metabolism of the selected CYP3A4 variants. This agrees with previous reports by Cai *et al.* (2021) and Drögemöller *et al.* (2013). It was reported that hydrophilicity and hydrophobicity might interfere with the protein properties and subsequently, the enzymatic activity of the implicating variants.

The grand average of hydropathicity (Gravy) has a strong negative correlation of -0.79. This implies that a change in Gravy might cause an opposite effect in the rate of metabolism of the selected CYP3A4. Free solvation energy was estimated to have a very weak or insignificant correlation value of 0.0046.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

The root mean square fluctuation (RMSF) was used to investigate the effect of mutation on the structural flexibility of the protein structures of the selected CYP3A4 variants. The correlation study between the RMSF and reported variation (relative to the wild type) in the rate of metabolism (clearance) of the selected variants was estimated to be very weak or insignificant (0.14).

The percentage of total secondary structure elements (SSE) indicated a strong negative correlation (-0.69) with the varied rate of metabolism in the selected CYP3A4 variants. This implies that an increase or decrease in the percentage helix present in the structure of the CYP3A4 might cause an opposite effect in the rate of metabolism (clearance) in the selected CYP3A4 variants. This same relationship was observed in the percentage strand with a good negative correlation of -0.581. The correlation study of the relationship between the percentage of the helix and the reported difference in the enzymatic activities of selected CYP3A4 variants was estimated to be insignificant. Fishelovitch *et al.* (2009) reported that structurally, the high percentage of alpha-helices present in the catalytic regions leads to significant enzyme flexibility and more substrate egress channels (Fishelovitch *et al.*, 2009), but the report did not well define the insignificant correlation observed in the percentage of the helix in this study.

UNIVERSITY of the
WESTERN CAPE

Protein volume has a positive linear correlation (0.533) while protein mobility has a strong negative correlation value (-0.647) with the rate of metabolism in the selected CYP3A4 variants. An increase or decrease in the protein volume might result in an equivalent outcome in the relative clearance of the selected variants. An increase or decrease in protein mobility could cause an opposite effect on the rate of metabolism. The protein net charge has a negative correlation value (-0.616) in its relationship with the relative clearance of the selected CYP3A4 variants and wild type. It was predicted that a change in the value of the protein net charge would result in an opposite effect on the rate of metabolism. Sum positive surface area, sum negative surface area, sum donor surface area, and sum acceptor surface area all have very insignificant correlation values with the rate of metabolism of the selected CYP3A4 variants relative to the wildtype.

http://etd.uwc.ac.za/

The protein dipole moment has a positive correlation of 0.65 with the rate of metabolism of the selected variants of CYP3A4. An increase in protein dipole moment might have an equivalent effect on the rate of metabolism. Protein charge at Debye length has a negative correlation of -0.603 with the rate of metabolism (clearance) of the selected CYP3A4 variants. This implies that an increase in the protein charge at Debye length results in an opposite effect on the rate of metabolism in the CYP3A4. The sum of the Aggregation score did not show any significant correlation (0.0489) with the rate of metabolism of the selected CYP3A4 variants. ESP minimum and maximum (0.15) have insignificant correlation values with the rate of metabolism of the selected variants and the wild type. While the ESP mean has a negative correlation (-0.60) with the rate of metabolism with the selected CYP3A4 variants. This implies that a change in the ESP might cause an opposite effect on the rate of metabolism.

The hydrogen bond has a strong negative linear correlation (-0.60) with the rate of metabolism of the selected This indicates that an increase in the hydrogen bond may lead to an opposite effect in the rate of metabolism of the selected CYP3A4 variants. This agrees with the study by Yao *et al.* (2018), which reported that hydrogen bonding may significantly increase the binding affinity and selectivity of a receptor. It has also been reported that the binding affinity (energy) of a drug to a receptor could modulate the clearance of the implicating drug (Eigenmann *et al.*, 2017; Smith, Gagnon, and Waters, 2017; Wang *et al.*, 2020).

Lipophilic energy, van der Waals energy and pi-pi energy did not indicate any significant relationship with the rate of metabolism in selected CYP3A4. Previous reports have shown the presence of lipophilic, van der Waals energy and pi-pi energy in CYP3A4 ligand interactions (Fa, Cong, and Wang, 2015; Prajapati *et al.*, 2022), however, the relationship between them and the rate of metabolism has not been reported. The hydrophobicity moment has no significant correlation (-0.126) with the relative clearance of CYP3A4 variants. Zeta potential at Debye Length has no significant correlation value (-0.28) with the rate of metabolism of CYP3A4 variants.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

4.11. Summary of the chapter

The chapter outlined the result obtained from the homology modelling, and quality evaluation of the modelled CYP3A4 selected variants. The result obtained from the effect of single nucleotide polymorphism on the structural, physicochemical, and dynamic properties of the selected CYP3A4 variants were also presented, interpreted, and discussed in this chapter. Interestingly, more than sixty-five percent of the total predicted properties (structural, physicochemical, and dynamic properties) showed significant differences in the selected variants relative to the wild type. The chapter further presents the result of the correlation study between the identified parameters (structural, physicochemical, and dynamic properties) and the reported variation in the clearance of the selected variants with an estimated significant difference. Amazingly, the correlation study revealed that a significant correlation exists between fifty-two percent of the predicted properties with notable differences. The implications of the established link in the correlation study to the reported rate of metabolism(clearance) were discussed.

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

Chapter 5

Conclusion and recommendation

This chapter provides an overview of the primary findings and conclusions derived from the study's aims. It also outlined the study's limitations and makes recommendations for future research.

The variations in CYP3A4 have led to the observed phenotype of poor metabolizers (PMs), intermediate metabolizers (IMs), extensive metabolizers (Ems), and ultra-rapid metabolizers (Ums), which influence the efficacy and toxicity of most drugs and prodrugs (Ahmed *et al.,* 2016; Neamțu, 2020). While several studies have proved the significant effects of CYP3A4 variability on its enzymatic activity, especially on most pharmacokinetic parameters, there is a paucity of information on the effect of some of these single nucleotide variations on the molecular, physicochemical, structural, and dynamic properties of the implicated enzymes (CYP3A4 variants).

This study aimed to investigate the effects of single nucleotide polymorphism (SNP) on the structural, physicochemical, and general protein properties of the modelled 3D protein structures of selected CYP3A4 variants. The possible effects of the SNPs on the dynamic characteristics of selected variants of CYP3A4 via molecular dynamics were evaluated. A comparative study with the reported effect of SNP on the relative clearance of lidocaine as reported by Fang *et al.* (2017) was done to validate any correlation with the structural, physicochemical and dynamics properties. The overall aim was achieved, and the major findings from the research objectives were outlined below.

5.1 The summary of the major findings

1. ***Homology modelling and quality evaluation of the protein structures of the CYP3A4 selected variants***: The sequences of the selected variants were successfully obtained with the biopython package on python. The obtained sequences were used to model the 3-dimensional protein structures of the selected

http://etd.uwc.ac.za/

variants via homology modelling. The reliability and quality were all attested to by the various protein quality assessment procedures they were subjected to.

2. *Visualisation, structural and molecular comparisons to investigate the effect of single nucleotide polymorphism on modelled and validated protein structures of the selected CYP3A4 variants:* The findings on the effect of single nucleotide polymorphism on the structural properties showed that the position of the heme ligand of all the variants' structures did not indicate any significant difference from the wild type. None of the mutations likely affected the positioning of the heme moiety. The visualised superimposed structures of the modelled protein structures of the wild-type and the variants indicated that variants 2 and 23 had a mismatch at the helix. It was estimated that the substitution at the position (S222P) with proline (a helix breaker) in CYP3A4*2 might interfere with the enzyme specificity and could account for the variation in the enzymatic activity (reduced intrinsic clearance relative to the wildtype) of CYP3A4*2 with lidocaine in the previous report (Fang *et al.*,2017). In CYP3A4*11 (T363M), it was estimated that the T363M substitution could lead to alteration in hydrogen interactions (the substituted Threonine 363 at the SRS-5 region functions majorly in hydrogen interaction in CYP3A4), this was estimated to be responsible for the varied rate of enzymatic activity (relative to the wildtype) in the metabolism of CYP3A4*11 with lidocaine as reported by Fang *et al.* (2017). In CYP3A4*23 (R162W), the substituted Arginine occurred at position 162 (very close to the active site). It was estimated that the substitution of hydrophilic residue (Arginine) with hydrophobic residue (Tryptophan) might account for the reported difference in the rate of metabolism of CYP3A4*23 with lidocaine (higher relative clearance of 206.96 as against the wildtype of 100) (Fang *et al.*, 2017).

It was discovered from the molecular lipophilicity potential (MLP)mean values and the variation in the surface colour of the (MLP) analysis, that the MLP might account for the reported variation in the rate of metabolism in the selected CYP3A4 variants, with an estimation that the wild type was more lipophilic than the selected variants. The mean value of the molecular lipophilicity potential also suggested that

UNIVERSITY of the WESTERN CAPE
http://etd.uwc.ac.za/

CYP3A4*23 was the most lipophilic of the selected variants with a high MLP mean value of (-3.372) which could explain the reported high relative clearance (206.96). The Coulombic electrostatic potential (ELP) selected variants varied significantly from the ELP value of the wild type, especially the CYP3A4*23, a prediction that SNPs might affect the ELP of a variant's protein structure. It was also estimated that the Coulombic electrostatic potential might be a determining property in the reported difference in the rate of metabolism (clearance) of the selected CYP3A4 variants.

3. ***Study of the differences in the physicochemical properties, ligand and solvent interaction, and other general properties of "protein" of the selected CYP3A4 variants, relative to the wild type:*** A total of thirty-five physicochemical and general protein properties of the selected CYP3A4 variants were studied out of which seventeen parameters indicated significant difference relative to the wildtype. It was predicted that single nucleotide polymorphism might affect the studied seventeen parameters with noticeable differences relative to the wild type. The study estimated that the single nucleotide polymorphism (SNPs) affected the stability of the selected variants' protein structures, from the prediction of the protein structure stability with the free energy change value. The relative solvent accessible (RSA) values of the selected CYP3A4 variants also indicated noticeable differences relative to the wildtype. A significantly varied negative binding energy was observed in the selected variants relative to the wild type. It was estimated that the single nucleotide polymorphism might affect the protein-ligand interaction and subsequently, the binding energy of the implicating enzymes. The calculated energy components (hydrogen, lipophilic, van der Waals and pi-pi energy) of the selected CYP3A4 variants differ significantly from the wild type. It was predicted that the calculated component energies could be influenced by SNPs.

4. ***Investigate the dynamic properties of the modelled protein structure of the selected variants relative to the wild type of CYP3A4 by molecular dynamic simulations:*** The effects of single nucleotide polymorphism on the selected variants' enzymatic dynamics were investigated with molecular dynamics

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

simulation. The difference was observed in the average root mean square of the protein structures relative to the wild type; this agrees with the result obtained in the stability prediction carried out in the initial characterization study. There was an indication from the study that the variant's protein structures were less rigid with more observable flexible regions. The fluctuation regions in two of the mutants (CYP3A4*11 and CYP3A4*23) were in the core active regions in the CYP3A4 structures, this suggested that the mutation may affect the conformation of the active regions of the mutant's protein structures of CYP3A4*11 and CYP3A4*23. An observed fluctuation in CYP3A4*24 at the position of the phenylalanine cluster in the CYP3A4 responsible for the oxidative allosteric regulation pathway indicated that single nucleotide polymorphism may affect the allosteric regulation of the implicated protein structure. There was a decrease in the average SASA for all selected variants with the wild type, which predicted a decrease in hydrophilic interaction in the variants, this finding agreed with the result obtained from the Gravy result in the physicochemical analysis study. The study predicted that mutation of the selected variants may not significantly affect the compactness of interacting protein structures from the values obtained from the radius of gyration. It was observed that the wild type had the highest percentage of a helix, while the mutated protein structures had a reduced percentage of the helix (It has been reported that the highest percentage of helix was found in the active site of the protein structures of the CYP3A4), this reflected in the result obtained in the binding energy of the interacting protein structures. The wild type with the highest helix percentage has the highest binding energy while CYP3A4*23 with the lowest helix percentage has the least binding energy. It was estimated that the mutation may affect the SSE, specifically the percentage helix. The complex protein-ligand interaction indicated a higher hydrogen bond interaction in CYP3A4*11 and CYP3A4*23.

5. ***Study the potential correlation between the studied structural or physicochemical effects of SNPs and the reported effects of SNPs on the relative clearance of selected variants:*** Forty characterised properties from the total characterised parameters (sixty-one) of the 3D structures of the selected CYP3A4

http://etd.uwc.ac.za/

variants properties indicated noticeable differences and were subjected to a correlation study. From the correlation analysis, it was predicted that twelve of the studied parameters with noticeable variations had a good or strong linear positive correlation, these include the molecular weight, free energy change value, free binding energy, average solvent-accessible surface area, relative solvent accessible area, hydrophilic surface area, protein volume, protein dipole moment, RMSD(protein-ligand complex) molecular lipophilicity potential minimum, molecular lipophilicity potential mean and the molecular lipophilicity potential maximum. Nine parameters had a good or strong negative correlation, these include the Gravy, root mean square deviation, percentage strand total, secondary structure elements, protein mobility, protein net charge, coulombic electrostatic potential, hydrogen bond, protein charge at Debye length, protein dipole moment. Overall, forty of sixty-one of the total characterised parameters (physicochemical, structural, and dynamic properties) indicated noticeable differences because of the effect of SNPs, while twenty-one (21) of the parameters with noticeable differences indicated significant correlation with the varied rate of metabolism (relative clearance) in the selected variants.

5.2 Conclusion and overall goal

UNIVERSITY of the
WESTERN CAPE

The overall aim was achieved, and the data obtained showed the effects of single nucleotide polymorphisms on the structural, physicochemical, and general protein properties of the modelled 3D protein structures of selected CYP3A4 variants. The effects of the SNP on the dynamic characteristics of selected variants of CYP3A4 via molecular dynamics were also predicted. The comparative study with the reported effect of SNP on the relative clearance of lidocaine as reported by Fang *et al.* (2017) estimated the correlation between the structural, physicochemical and dynamics properties and the varied rate of metabolism in the selected variants.

Overall, the knowledge and the data generated from this study provided insight into the effect of single nucleotide polymorphism on the structural, physicochemical properties, and dynamic characteristics of modelled and validated protein structures of selected variants of CYP3A4. This may serve as the determinant of the observed

http://etd.uwc.ac.za/

differences in the functional activity (metabolism) of the selected variants of CYP3A4 enzymes. The established link between the molecular differences and the reported functional differences elucidated the basis of the observed differences in the functional activity (metabolism) of the different variants of CYP3A4 enzyme; this is applicable in drug design and discovery, to reduce the toxicity of drugs associated with genetic variation. This will also facilitate an increase in drug efficacy with appropriate implementation of the knowledge of the effects of single nucleotide polymorphism.

5.3 Limitations of the study and recommendations for future work

There has been an acceleration of sequencing data generated for African populations. The work in this study provides the basis for analysing CYP3A enzymes in African populations to understand the genetic variation and pharmacodynamics of drug metabolism in populations where there is a paucity of data. The research might also be used as a template in the characterization of other reported variants of CYPs isoforms. Further study with other major CYP3A4 substrates is beyond the scope of this research study; it is therefore recommended for future research work.

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

References

Abdel-Kahaar, Emaad, Stefan Winter, Roman Tremmel, Elke Schaeffeler, Christoph J. Olbricht, Eberhard Wieland, Matthias Schwab, Maria Shipkova, and Simon U. Jaeger. 2019. “The Impact of CYP3A4*22 on Tacrolimus Pharmacokinetics and Outcome in Clinical Practice at a Single Kidney Transplant Center.” *Frontiers in Genetics* 10 (September): 871. https://doi.org/10.3389/fgene.2019.00871.

Abraham, Mark James, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. 2015. “GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers.” *SoftwareX* 1–2 (September): 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

Adebiyi, Marion, and Oludayo Olugbara. 2021. “Binding Site Identification of COVID-19 Main Protease 3D Structure by Homology Modeling.” *Indonesian Journal of Electrical Engineering and Computer Science* 21 (3): 1713. http://dx.doi.org/10.11591/ijeecs.v21.i3.pp1713-1721

Ahmed, Mustapha Carab, Ramon Crehuet, and Kresten Lindorff-Larsen. "Computing, analyzing, and comparing the radius of gyration and hydrodynamic radius in conformational ensembles of intrinsically disordered proteins." In Intrinsically Disordered Proteins, pp. 429-445. Humana, New York, NY, 2020. https://doi.org/10.1007/978-1-0716-0524-0_21

Ahmed, Shabbir, Zhan Zhou, Jie Zhou, and Shu-Qing Chen. 2016. “Pharmacogenomics of Drug Metabolizing Enzymes and Transporters: Relevance to Precision Medicine.” *Genomics, Proteomics & Bioinformatics*, SI: Big Data and Precision Medicine, 14 (5): 298–313. https://doi.org/10.1016/j.gpb.2016.03.008.

Akapo, Olufunmilayo Olukemi, Joanna M. Macnar, Justyna D. Kryś, Puleng Rosinah Syed, Khajamohiddin Syed, and Dominik Gront. 2021. “In Silico

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Structural Modeling and Analysis of Interactions of Tremellomycetes Cytochrome P450 Monooxygenases CYP51s with Substrates and Azoles." *International Journal of Molecular Sciences* 22 (15): 7811. https://doi.org/10.3390/ijms22157811.

Arafah, Azher, Shafat Ali, Sabhiya Majid, Samia Rashid, Shabhat Rasool, Hilal Ahmad Wani, Iyman Rasool, and Muneeb U. Rehman. 2021. "Single Nucleotide Polymorphisms and Pharmacogenomics." In *Genetic Polymorphism and Cancer Susceptibility*, edited by Aga Syed Sameer, Mujeeb Zafar Banday, and Saniya Nissar, 23–52. Singapore: Springer. https://doi.org/10.1007/978-981-33-6699-2_2.

Arora, Tanmay, and Asrar A. Malik. 2021. "Chapter 14 - An Introduction to BLAST: Applications for Computer-Aided Drug Design and Development." In *Chemoinformatics and Bioinformatics in the Pharmaceutical Sciences*, edited by Navneet Sharma, Himanshu Ojha, Pawan Kumar Raghav, and Ramesh k. Goyal, 423–53. Academic Press. https://doi.org/10.1016/B978-0-12-821748-1.00015-4.

Asgharzadeh, Sanaz, Behzad Shareghi, and Sadegh Farhadian. 2019. "Experimental and Theoretical Investigations on the Interaction of L-Methionine Molecules with α-Chymotrypsin in the Aqueous Solution Using Various Methods." *International Journal of Biological Macromolecules* 131: 548–56. https://doi.org/10.1016/j.ijbiomac.2019.03.080

Ausaf Ali, Syed, Imtaiyaz Hassan, Asimul Islam, and Faizan Ahmad. 2014. "A Review of Methods Available to Estimate Solvent-Accessible Surface Areas of Soluble Proteins in the Folded and Unfolded States." Current Protein and Peptide Science 15 (5): 456–76. https://pubmed.ncbi.nlm.nih.gov/24678666/

Barbera, Nicolas, Manuela A. A. Ayee, Belinda S. Akpa, and Irena Levitan. 2018. "Molecular Dynamics Simulations of Kir2.2 Interactions with an Ensemble of Cholesterol Molecules." *Biophysical Journal* 115 (7): 1264–80. https://doi.org/10.1016/j.bpj.2018.07.041.

http://etd.uwc.ac.za/

Barcelos, Mariana Pegrucci, Leonardo Bruno Federico, Carlton A. Taft, and Carlos HT de Paula da Silva. 2020. “Prediction of the Three-Dimensional Structure of Phosphate-6-Mannose PMI Present in the Cell Membrane of Xanthomonas Citri Subsp. Citri of Interest for the Citrus Canker Control.” In *Emerging Research in Science and Engineering Based on Advanced Experimental and Computational Strategies*, 259–76. Springer. doi: 10.1007/978-3-030-31403-3_10

Barr, John T., Zhican Wang, Xiaoshan Min, Henry J. Wienkers, Brooke M. Rock, Dan A. Rock, and Larry C. Wienkers. 2020. “Mechanistic Studies of Cytochrome P450 3A4 Time-Dependent Inhibition Using Two Cysteine-Targeting Electrophiles.” *Drug Metabolism and Disposition* 48 (6): 508–14. https://doi.org/10.1124/dmd.119.089813.

Basheer, Loai, and Zohar Kerem. 2015. “Interactions between CYP3A4 and Dietary Polyphenols.” *Oxidative Medicine and Cellular Longevity* 2015: 1–15. https://doi.org/10.1155/2015/854015.

Bellotto, N., J. Agudo-Canalejo, R. Colin, R. Golestanian, G. Malengo, and V. Sourjik. 2022. “Dependence of Diffusion in Escherichia Coli Cytoplasm on Protein Size, Environmental Conditions and Cell Growth.” BioRxiv. doi: https://doi.org/10.1101/2022.02.17.480843

Ben-Fredj, Nadia, Ibtissem Hannachi, Zohra Chadli, Haifa Ben-Romdhane, Naceur A Boughattas, Najah Ben-Fadhel, and Karim Aouam. 2020. “Dosing Algorithm for Tacrolimus in Tunisian Kidney Transplant Patients: Effect of CYP 3A4*1B and CYP3A4*22 Polymorphisms.” *Toxicology and Applied Pharmacology* 407 (November): 115245. https://doi.org/10.1016/j.taap.2020.115245.

Benkaidali, Lydia, François André, Gautier Moroy, Bahoueddine Tangour, François Maurel, and Michel Petitjean. 2017. “The Cytochrome P450 3A4 Has Three Major Conformations: New Clues to Drug Recognition by This Promiscuous Enzyme.” *Molecular Informatics* 36 (10): 1700044. https://doi.org/10.1002/minf.201700044.

http://etd.uwc.ac.za/

Berno, Giulia, Mauro Zaccarelli, Caterina Gori, Massimo Tempestilli, Andrea Antinori, Carlo Federico Perno, Leopoldo Paolo Pucillo, and Roberta D'Arrigo. 2014. "Analysis of Single-Nucleotide Polymorphisms (SNPs) in Human CYP3A4 and CYP3A5 Genes: Potential Implications for the Metabolism of HIV Drugs." *BMC Medical Genetics* 15 (July): 76. https://doi.org/10.1186/1471-2350-15-76.

Biasini, M., T. Schmidt, S. Bienert, V. Mariani, G. Studer, J. Haas, N. Johner, A. D. Schenk, A. Philippsen, and T. Schwede. 2013. "*OpenStructure* : An Integrated Software Framework for Computational Structural Biology." *Acta Crystallographica Section D Biological Crystallography* 69 (5): 701–9. https://doi.org/10.1107/S0907444913007051.

Biasini, Marco, Stefan Bienert, Andrew Waterhouse, Konstantin Arnold, Gabriel Studer, Tobias Schmidt, Florian Kiefer, *et al.* 2014. "SWISS-MODEL: Modelling Protein Tertiary and Quaternary Structure Using Evolutionary Information." *Nucleic Acids Research* 42 (W1): W252–58. https://doi.org/10.1093/nar/gku340.

Bienert, Stefan, Andrew Waterhouse, Tjaart A. P. de Beer, Gerardo Tauriello, Gabriel Studer, Lorenza Bordoli, and Torsten Schwede. 2017. "The SWISS-MODEL Repository—New Features and Functionality." *Nucleic Acids Research* 45 (D1): D313–19. https://doi.org/10.1093/nar/gkw1132.

Blundell, Tom L. 2021. "The First Resolution Revolution in Protein Structure Analysis: X-Ray Diffraction of Polypeptide Conformations and Globular Protein Folds in 1950s and 1960s." *Progress in Biophysics and Molecular Biology*, September,S007961072100105X.https://doi.org/10.1016/j.pbiomolbio.2021.09.002.

Bordoli, Lorenza, Florian Kiefer, Konstantin Arnold, Pascal Benkert, James Battey, and Torsten Schwede. 2009. "Protein Structure Homology Modeling Using SWISS-MODEL Workspace." *Nature Protocols* 4 (1): 1–13. https://doi.org/10.1038/nprot.2008.197.

http://etd.uwc.ac.za/

Broccatelli, Fabio, Ignacio Aliagas, and Hao Zheng. 2018. “Why Decreasing Lipophilicity Alone Is Often Not a Reliable Strategy for Extending IV Half-Life.” ACS Medicinal Chemistry Letters 9 (6): 522–27. https://doi.org/10.1021/acsmedchemlett.8b00047.

Broccatelli, Fabio, Ignacio Aliagas, and Hao Zheng. 2018. “Why Decreasing Lipophilicity Alone Is Often Not a Reliable Strategy for Extending IV Half-Life.” *ACS Medicinal Chemistry Letters* 9 (6): 522–27. https://doi.org/10.1021/acsmedchemlett.8b00047.

Brooks, B.R., C.L. Brooks, A.D. MacKerell, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, *et al.* 2009. “CHARMM: The Biomolecular Simulation Program.” *Journal of Computational Chemistry* 30 (10): 1545–1614. https://doi.org/10.1002/jcc.21287.

Bryant, Patrick, Gabriele Pozzati, and Arne Elofsson. 2022. “Improved Prediction of Protein-Protein Interactions Using AlphaFold2.” *Nature Communications* 13 (1): 1–11. https://doi.org/10.1038/s41467-022-28865-w

Buck, Patrick M., Sandeep Kumar, and Satish K. Singh. 2013. “On the Role of Aggregation Prone Regions in Protein Evolution, Stability, and Enzymatic Catalysis: Insights from Diverse Analyses.” *PLoS Computational Biology* 9 (10): e1003291. https://doi.org/10.1371/journal.pcbi.1003291.

Butt, Sania Safdar, Yasmin Badshah, Maria Shabbir, and Mehak Rafiq. 2020. “Molecular Docking Using Chimera and Autodock Vina Software for Nonbioinformaticians.” *JMIR Bioinformatics and Biotechnology* 1 (1): e14232. https://doi.org/10.2196/14232.

Cai, Yaoyao, Qianmeng Lin, Zhousheng Jin, Fangfang Xia, Yingchao Ye, Yun Xia, Thomas J. Papadimos, *et al.* 2021. “Evaluation of Recombinant CYP3A4 Variants on the Metabolism of Oxycodone In Vitro.” *Chemical Research in Toxicology* 34 (1): 103–9. https://doi.org/10.1021/acs.chemrestox.0c00361.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Chaudhury, Sidhartha, Monica Berrondo, Brian D. Weitzner, Pravin Muthu, Hannah Bergman, and Jeffrey J. Gray. 2011. “Benchmarking and Analysis of Protein Docking Performance in Rosetta v3.2.” Edited by Vladimir N. Uversky. *PLoS ONE* 6 (8): e22477. https://doi.org/10.1371/journal.pone.0022477.

Chen, Bingbing, Xiao-dan Zhang, Jian Wen, Bowen Zhang, Daoxing Chen, Shuanghu Wang, Jian-pin Cai, and Guo-xin Hu. 2020. “Effects of 26 Recombinant CYP3A4 Variants on Brexpiprazole Metabolism.” *Chemical Research in Toxicology* 33 (1): 172–80. https://doi.org/10.1021/acs.chemrestox.9b00186.

Chen, Chang-Hwei. 2020. *Xenobiotic Metabolic Enzymes: Bioactivation and Antioxidant Defense*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-41679-9.

Chikhale, Rupesh V., Vivek K. Gupta, Gaber E. Eldesoky, Saikh M. Wabaidur, Shripad A. Patil, and Md Ataul Islam. 2020. “Identification of Potential Anti-TMPRSS2 Natural Products through Homology Modelling, Virtual Screening and Molecular Dynamics Simulation Studies.” *Journal of Biomolecular Structure and Dynamics* 0 (0): 1–16. https://doi.org/10.1080/07391102.2020.1798813.

Choudhary, M. Iqbal, Muniza Shaikh, Atia- tul-Wahab, and Atta- ur-Rahman. 2020. “In Silico Identification of Potential Inhibitors of Key SARS-CoV-2 3CL Hydrolase (Mpro) via Molecular Docking, MMGBSA Predictive Binding Energy Calculations, and Molecular Dynamics Simulation.” *PLoS ONE* 15 (7): e0235030. https://doi.org/10.1371/journal.pone.0235030.

Cock, Peter J. A., Christopher J. Fields, Naohisa Goto, Michael L. Heuer, and Peter M. Rice. 2010. “The Sanger FASTQ File Format for Sequences with Quality Scores, and the Solexa/Illumina FASTQ Variants.” *Nucleic Acids Research* 38 (6): 1767–71. https://doi.org/10.1093/nar/gkp1137.

Cook, D. J., J. D. Finnigan, K. Cook, G. W. Black, and S. J. Charnock. 2016a. “Chapter Five - Cytochromes P450: History, Classes, Catalytic Mechanism, and

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Industrial Application." In *Advances in Protein Chemistry and Structural Biology*, edited by Christo Z. Christov, 105:105–26. Insights into Enzyme Mechanisms and Functions from Experimental and Computational Methods. Academic Press. https://doi.org/10.1016/bs.apcsb.2016.07.003.

Córdova, Pamela, Ana-María Gonzalez, David R. Nelson, María-Soledad Gutiérrez, Marcelo Baeza, Víctor Cifuentes, and Jennifer Alcaíno. 2017. "Characterization of the Cytochrome P450 Monooxygenase Genes (P450ome) from the Carotenogenic Yeast Xanthophyllomyces Dendrorhous." *BMC Genomics* 18 (1): 540. https://doi.org/10.1186/s12864-017-3942-9.

Crook, Jack M., Ivana Murphy, Daniel P. Carter, Steven T. Pullan, Miles Carroll, Richard Vipond, Andrew A. Cunningham, and Diana Bell. 2021. "Metagenomic Identification of a New Sarbecovirus from Horseshoe Bats in Europe." *Scientific Reports* 11 (1): 14723. https://doi.org/10.1038/s41598-021-94011-z.

Dahal, Yuba Raj, and Jeremy D. Schmit. 2018. "Ion Specificity and Nonmonotonic Protein Solubility from Salt Entropy." *Biophysical Journal* 114 (1): 76–87. https://doi.org/10.1016/j.bpj.2017.10.040

Das, Lopamudra, Anand Kumar, Sarita Nanda, and J. K. Das. 2019. "Improved Protein Coding Region Prediction Using Dipole Moment Based SVD Algorithm." In *2019 5th International Conference on Signal Processing, Computing and Control (ISPCC)*, 90–95. IEEE.doi; https://doi.org/10.1109/ISPCC48220.2019.8988320

Dasari, Sathish, and Bhabani S. Mallik. 2020. "Solubility and Solvation Free Energy of a Cardiovascular Drug, LASSBio-294, in Ionic Liquids: A Computational Study." *Journal of Molecular Liquids* 301: 112449. https://doi.org/10.1016/j.molliq.2020.112449

David, Alessia, Tarun Khanna, Melina Beykou, Gordon Hanna, and Michael J E Sternberg. 2020. "Structure, Function and Variants Analysis of the Androgen-

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Regulated *TMPRSS2*, a Drug Target Candidate for COVID-19 Infection." Preprint. Bioinformatics. https://doi.org/10.1101/2020.05.26.116608.

Davydov, Dmitri R., Jessica A. O. Rumfeldt, Elena V. Sineva, Harshica Fernando, Nadezhda Y. Davydova, and James R. Halpert. 2012. "Peripheral Ligand-Binding Site in Cytochrome P450 3A4 Located with Fluorescence Resonance Energy Transfer (FRET)." Journal of Biological Chemistry 287 (9): 6797–6809. https://doi.org/10.1074/jbc.M111.325654.

De Vivo, Marco, Matteo Masetti, Giovanni Bottegoni, and Andrea Cavalli. 2016. "Role of Molecular Dynamics and Related Methods in Drug Discovery." *Journal of Medicinal Chemistry* 59 (9): 4035–61. https://doi.org/10.1021/acs.jmedchem.5b01684.

Degregorio, Danilo, Serena D'Avino, Silvia Castrignanò, Giovanna Di Nardo, Sheila J. Sadeghi, Gianluca Catucci, and Gianfranco Gilardi. 2017. "Human Cytochrome P450 3A4 as a Biocatalyst: Effects of the Engineered Linker in Modulation of Coupling Efficiency in 3A4-BMR Chimeras." *Frontiers in Pharmacology* 8: 121. https://doi.org/10.3389/fphar.2017.00121.

Del Re, M., V. Citi, S. Crucitta, E. Rofi, F. Belcari, R. H. van Schaik, and R. Danesi. 2016. "Pharmacogenetics of CYP2D6 and Tamoxifen Therapy: Light at the End of the Tunnel?" *Pharmacological Research* 107 (May): 398–406. https://doi.org/10.1016/j.phrs.2016.03.025.

Denisov, I. G., A.Y. Shih, and S.G. Sligar. 2012. "STRUCTURAL DIFFERENCES BETWEEN SOLUBLE AND MEMBRANE BOUND CYTOCHROMES P450." *Journal of Inorganic Biochemistry* 108 (March): 150–58. https://doi.org/10.1016/j.jinorgbio.2011.11.026.

Denisov, Ilia G., Yelena V. Grinkova, Tyler Camp, Mark A. McLean, and Stephen G. Sligar. 2021. "Midazolam as a Probe for Drug-Drug Interactions Mediated by

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

CYP3A4: Homotropic Allosteric Mechanism of Site-Specific Hydroxylation." *Biochemistry* 60 (21): 1670–81. https://doi.org/10.1021/acs.biochem.1c00161.

Dong, Amelia Nathania, Boon Hooi Tan, Yan Pan, and Chin Eng Ong. 2021. "The CYP2R1 Enzyme: Structure, Function, Enzymatic Properties and Genetic Polymorphism." Journal of Pharmacy & Pharmaceutical Sciences 24 (February): 94–112. https://doi.org/10.18433/jpps31305.

Dong, Yun-wei, Ming-ling Liao, Xian-liang Meng, and George N. Somero. 2018. "Structural Flexibility and Protein Adaptation to Temperature: Molecular Dynamics Analysis of Malate Dehydrogenases of Marine Molluscs." *Proceedings of the National Academy of Sciences* 115 (6): 1274–79. https://doi.org/10.1073/pnas.1718910115

Dostalek, Miroslav, Michael H. Court, Bingfang Yan, and Fatemeh Akhlaghi. 2011. "Significantly Reduced Cytochrome P450 3A4 Expression and Activity in Liver from Humans with Diabetes Mellitus." *British Journal of Pharmacology* 163 (5): 937–47. https://doi.org/10.1111/j.1476-5381.2011.01270.x

Drögemöller, Britt Ingrid, Marieth Plummer, Lundi Korkie, Gloudi Agenbag, Anke Dunaiski, Dana Niehaus, Liezl Koen, *et al.* 2013. "Characterization of the Genetic Variation Present in CYP3A4 in Three South African Populations." *Frontiers in Genetics* 4. https://doi.org/10.3389/fgene.2013.00017.

Ducharme, Julie, Irina F. Sevrioukova, Christopher J. Thibodeaux, and Karine Auclair. 2021. "Structural Dynamics of Cytochrome P450 3A4 in the Presence of Substrates and Cytochrome P450 Reductase." *Biochemistry* 60 (28): 2259–71. https://doi.org/10.1021/acs.biochem.1c00178

Durham, Elizabeth, Brent Dorr, Nils Woetzel, René Staritzbichler, and Jens Meiler. 2009. "Solvent Accessible Surface Area Approximations for Rapid and Accurate Protein Structure Prediction." Journal of Molecular Modeling 15 (9): 1093–1108. https://doi.org/10.1007/s00894-009-0454-9

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Egieyeh, Samuel, James Syce, Sarel F. Malan, and Alan Christoffels. 2018. "Predictive Classifier Models Built from Natural Products with Antimalarial Bioactivity Using Machine Learning Approach." PLOS ONE 13 (9): e0204644. https://doi.org/10.1371/journal.pone.0204644.

Eigenmann, Miro J., Ludivine Fronton, Hans Peter Grimm, Michael B. Otteneder, and Ben-Fillippo Krippendorff. 2017. "Quantification of IgG Monoclonal Antibody Clearance in Tissues." In MAbs, 9:1007–15. Taylor & Francis. https://doi.org/10.1080/19420862.2017.1337619

Ekroos, M., and T. Sjogren. 2006. "Structural Basis for Ligand Promiscuity in Cytochrome P450 3A4." *Proceedings of the National Academy of Sciences* 103 (37): 13682–87. https://doi.org/10.1073/pnas.0603236103.

El-Sayed, Ramy, Kunal Bhattacharya, Zonglin Gu, Zaixing Yang, Jeffrey K. Weber, Hu Li, Klaus Leifer, *et al.* 2016. "Single-Walled Carbon Nanotubes Inhibit the Cytochrome P450 Enzyme, CYP3A4." *Scientific Reports* 6 (1): 21316. https://doi.org/10.1038/srep21316.

Erickson, Harold P. 2009. "Size and Shape of Protein Molecules at the Nanometer Level Determined by Sedimentation, Gel Filtration, and Electron Microscopy." *Biological Procedures Online* 11 (1): 32–51. https://doi.org/10.1007/s12575-009-9008-x

Esfandi, Babak, and Morteza Atabati. 2021. "Sequential Dihedral Angles (SDAs): A Method for Evaluating the 3D Structure of Proteins." *The Protein Journal* 40 (1): 1–7. https://doi.org/10.1007/s10930-020-09961-6.

Esteves, Francisco, José Rueff, and Michel Kranendonk. 2021. "The Central Role of Cytochrome P450 in Xenobiotic Metabolism—A Brief Review on a Fascinating Enzyme Family." *Journal of Xenobiotics* 11 (3): 94–114. https://doi.org/10.3390/jox11030007.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Estrada, D. Fernando, Amit Kumar, Christopher S. Campomizzi, and Natalie Jay. 2021. “Crystal Structures of Drug-Metabolizing CYPs.” In *Enzyme Kinetics in Drug Metabolism: Fundamentals and Applications*, edited by Swati Nagar, Upendra A. Argikar, and Donald Tweedie, 171–92. Methods in Molecular Biology. New York, NY: Springer US. https://doi.org/10.1007/978-1-0716-1554-6_7.

Fa, Botao, Shan Cong, and Jingfang Wang. 2015. “Pi-Pi Stacking Mediated Cooperative Mechanism for Human Cytochrome P450 3A4.” *Molecules* 20 (5): 7558–73. https://doi.org/10.3390/molecules20057558

Fadaka, Adewale Oluwaseun, Raphael Taiwo Aruleba, Nicole Remaliah Samantha Sibuyi, Ashwil Klein, Abram Madimabe Madiehe, and Mervin Meyer. n.d. “Inhibitory Potential of Repurposed Drugs against the SARS-CoV-2 Main Protease: A Computational-Aided Approach.” *Journal of Biomolecular Structure & Dynamics*, 1–13. https://doi.org/10.1080/07391102.2020.1847197.

Fang, Ping, Peng-Fei Tang, Ren-Ai Xu, Xiang Zheng, Jian Wen, Su-Su Bao, Jian-Ping Cai, and Guo-Xin Hu. 2017. “Functional Assessment of CYP3A4 Allelic Variants on Lidocaine Metabolism in Vitro.” *Drug Design, Development and Therapy* 11: 3503–10. https://doi.org/10.2147/DDDT.S152366.

Feng, Yanghe, Qi Wang, and Tengjiao Wang. 2017. “Drug Target Protein-Protein Interaction Networks: A Systematic Perspective.” *BioMed Research International* 2017. https://doi.org/10.1155/2017/1289259

Fishelovitch, Dan, Sason Shaik, Haim J. Wolfson, and Ruth Nussinov. 2009. “Theoretical Characterization of Substrate Access/Exit Channels in the Human Cytochrome P450 3A4 Enzyme: Involvement of Phenylalanine Residues in the Gating Mechanism.” The Journal of Physical Chemistry B 113 (39): 13018–25. https://doi.org/10.1021/jp810386z.

Foteini, Papageorgiou, Natassa Pippa, Nikolaos Naziris, and Costas Demetzos. 2019. “Physicochemical Study of the Protein–Liposome Interactions: Influence of

http://etd.uwc.ac.za/

Liposome Composition and Concentration on Protein Binding." *Journal of Liposome Research* 29 (4): 313–21. https://doi.org/10.1080/08982104.2018.1468774

Fuller, Jonathan C., Nicholas J. Burgoyne, and Richard M. Jackson. 2009. "Predicting Druggable Binding Sites at the Protein–Protein Interface." *Drug Discovery Today* 14 (3–4): 155–61. https://doi.org/10.1016/j.drudis.2008.10.009.

Gaedigk, Andrea, Magnus Ingelman-Sundberg, Neil A. Miller, J. Steven Leeder, Michelle Whirl-Carrillo, Teri E. Klein, and the PharmVar Steering Committee. 2018. "The Pharmacogene Variation (PharmVar) Consortium: Incorporation of the Human Cytochrome P450 (*CYP*) Allele Nomenclature Database." *Clinical Pharmacology & Therapeutics* 103 (3): 399–401. https://doi.org/10.1002/cpt.910.

Gao, Wenhao, Sai Pooja Mahajan, Jeremias Sulam, and Jeffrey J. Gray. "Deep learning in protein structural modeling and design." Patterns 1, no. 9 (2020): 100142. https://doi.org/10.1016/j.patter.2020.100142

Georgiou, Christos D. 2018. "Functional Properties of Amino Acid Side Chains as Biomarkers of Extraterrestrial Life." *Astrobiology* 18 (11): 1479–96. https://doi.org/10.1089/ast.2018.1868.

Ghassabian, Sussan, Tina B. Gillani, Tristan Rawling, Severine Crettol, Pramod C. Nair, and Michael Murray. 2019. "Sorafenib N-Oxide Is an Inhibitor of Human Hepatic CYP3A4." *The AAPS Journal* 21 (2): 15. https://doi.org/10.1208/s12248-018-0262-1.

Giantin, Mery, Minna Rahnasto-Rilla, Roberta Tolosi, Lorena Lucatello, Marianna Pauletto, Giorgia Guerra, Francesca Pezzato, *et al.* 2019. "Functional Impact of Cytochrome P450 3A (CYP3A) Missense Variants in Cattle." *Scientific Reports* 9 (1): 19672. https://doi.org/10.1038/s41598-019-56271-8.

Girvan, Hazel M., Adrian J. Dunford, Rajasekhar Neeli, Idorenyin S. Ekanem, Timothy N. Waltham, M. Gordon Joyce, David Leys, *et al.* 2011.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

"Flavocytochrome P450 BM3 Mutant W1046A Is a NADH-Dependent Fatty Acid Hydroxylase: Implications for the Mechanism of Electron Transfer in the P450 BM3 Dimer." *Archives of Biochemistry and Biophysics* 507 (1): 75–85. https://doi.org/10.1016/j.abb.2010.09.014.

Gjestad, Caroline, Kristine Hole, Tore Haslemo, Ulf Diczfalusy, and Espen Molden. 2019. "Effect of Grapefruit Juice Intake on Serum Level of the Endogenous CYP3A4 Metabolite 4β-Hydroxycholesterol—an Interaction Study in Healthy Volunteers." *The AAPS Journal* 21 (4): 58. https://doi.org/10.1208/s12248-019-0330-1.

Gkeka, Paraskevi, Gabriel Stoltz, Amir Barati Farimani, Zineb Belkacemi, Michele Ceriotti, John D. Chodera, Aaron R. Dinner, Andrew L. Ferguson, Jean-Bernard Maillet, and Hervé Minoux. 2020. "Machine Learning Force Fields and Coarse-Grained Variables in Molecular Dynamics: Application to Materials and Biological Systems." *Journal of Chemical Theory and Computation* 16 (8): 4757–75. https://doi.org/10.1021/acs.jctc.0c00355

Godamudunage, Malika P., Anne M. Grech, and Emily E. Scott. 2018. "Comparison of Antifungal Azole Interactions with Adult Cytochrome P450 3A4 versus Neonatal Cytochrome P450 3A7." *Drug Metabolism and Disposition* 46 (9): 1329–37. https://doi.org/10.1124/dmd.118.082032.

Goddard, Thomas D., Conrad C. Huang, Elaine C. Meng, Eric F. Pettersen, Gregory S. Couch, John H. Morris, and Thomas E. Ferrin. 2018. "UCSF ChimeraX: Meeting Modern Challenges in Visualization and Analysis: UCSF ChimeraX Visualization System." *Protein Science* 27 (1): 14–25. https://doi.org/10.1002/pro.3235.

Gonzalez, Frank J., Balthazar J. Schmid, Morio Umeno, O. Wesley Mcbride, James P. Hardwick, Urs A. Meyer, Harry V. Gelboin, and Jeffrey R. Idle. 1988. "Human P450PCN1: Sequence, Chromosome Localization, and Direct Evidence through CDNA Expression That P450PCN1 Is Nifedipine Oxidase." *DNA* 7 (2): 79–86. https://doi.org/10.1089/dna.1988.7.79.

http://etd.uwc.ac.za/

Goto, Takahiro, Yasushi Yamazoe, and Masahiro Tohkin. 2020. “Applications of a Grid-Based CYP3A4 Template System to Understand the Interacting Mechanisms of Large-Size Ligands; Part 4 of CYP3A4 Template Study.” *Drug Metabolism and Pharmacokinetics* 35 (6): 485–96. https://doi.org/10.1016/j.dmpk.2020.06.008.

Gudin, Jeff, and Sri Nalamachu. 2020. “Utility of Lidocaine as a Topical Analgesic and Improvements in Patch Delivery Systems.” *Postgraduate Medicine* 132 (1): 28–36. https://doi.org/10.1080/00325481.2019.1702296

Guengerich, F. Peter, Michael R. Waterman, and Martin Egli. 2016. “Recent Structural Insights into Cytochrome P450 Function.” *Trends in Pharmacological Sciences* 37 (8): 625–40. https://doi.org/10.1016/j.tips.2016.05.006.

Guengerich, F. Peter. 2018. “Mechanisms of Cytochrome P450-Catalyzed Oxidations.” *ACS Catalysis* 8 (12): 10964–76. https://doi.org/10.1021/acscatal.8b03401.

Gupta, Manoj Kumar, Gayatri Gouda, N. Rajesh, Ravindra Donde, S. Sabarinathan, Pallabi Pati, Sushil Kumar Rathore, Ramakrishna Vadde, and Lambodar Behera. "Sequence Alignment." In Bioinformatics in Rice Research, pp. 129-162. Springer, Singapore, 2021. https://doi.org/10.1007/978-981-16-3993-7_7

Gupta, Ritika, Ankita Dey, Anu Vijan, and Bitu Gartia. 2017. “In Silico Structure Modeling and Characterization of Hypothetical Protein YP_004590319.1 Present in Enterobacter Aerogens.” *Journal of Proteomics & Bioinformatics* 10 (June). https://doi.org/10.4172/jpb.1000436.

Gupta, Shivangi, Jerome Baudry, and Vineetha Menon. 2022. “Using Big Data Analytics to ‘Back Engineer’ Protein Conformational Selection Mechanisms.” *Molecules* 27 (8): 2509. https://doi.org/10.3390/molecules27082509.

Guttman, Yelena, Adi Nudel, and Zohar Kerem. 2019. “Polymorphism in Cytochrome P450 3A4 Is Ethnicity Related.” *Frontiers in Genetics* 0. https://doi.org/10.3389/fgene.2019.00224.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Haddad, Yazan, Vojtech Adam, and Zbynek Heger. 2020. “Ten Quick Tips for Homology Modeling of High-Resolution Protein 3D Structures.” Edited by Francis Ouellette. *PLOS Computational Biology* 16 (4): e1007449. https://doi.org/10.1371/journal.pcbi.1007449.

Hakkola, Jukka, Janne Hukkanen, Miia Turpeinen, and Olavi Pelkonen. 2020. “Inhibition and Induction of CYP Enzymes in Humans: An Update.” *Archives of Toxicology* 94 (11): 3671–3722. https://doi.org/10.1007/s00204-020-02936-7.

Han, Mingming, Jianchang Qian, Zhize Ye, Renai Xu, Daoxing Chen, Saili Xie, Jianping Cai, and Guoxin Hu. 2021. “Functional Assessment of the Effects of CYP3A4 Variants on Acalabrutinib Metabolism in Vitro.” *Chemico-Biological Interactions* 345 (August): 109559. https://doi.org/10.1016/j.cbi.2021.109559.

Harder, Edward, Wolfgang Damm, Jon Maple, Chuanjie Wu, Mark Reboul, Jin Yu Xiang, Lingle Wang, *et al.* 2016. “OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins.” *Journal of Chemical Theory and Computation* 12 (1): 281–96. https://doi.org/10.1021/acs.jctc.5b00864.

Harrington, Matthew J., and Peter Fratzl. 2021. “Natural Load-Bearing Protein Materials.” *Progress in Materials Science* 120: 100767. https://doi.org/10.1016/j.pmatsci.2020.100767

Hendrychová, Tereza, Eva Anzenbacherová, Jiří Hudeček, Josef Skopalík, Reinhard Lange, Peter Hildebrandt, Michal Otyepka, and Pavel Anzenbacher. 2011. “Flexibility of Human Cytochrome P450 Enzymes: Molecular Dynamics and Spectroscopy Reveal Important Function-Related Variations.” *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1814 (1): 58–68. https://doi.org/10.1016/j.bbapap.2010.07.017.

Hendrychova, Tereza, Karel Berka, Veronika Navratilova, Pavel Anzenbacher, and Michal Otyepka. 2012. “Dynamics and Hydration of the Active Sites of Mammalian Cytochromes P450 Probed by Molecular Dynamics Simulations.”

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Current Drug Metabolism 13 (2): 177–89. https://doi.org/10.2174/138920012798918408.

Hsu, Mei-Hui, and Eric F. Johnson. 2019. "Active-Site Differences between Substrate-Free and Ritonavir-Bound Cytochrome P450 (CYP) 3A5 Reveal Plasticity Differences between CYP3A5 and CYP3A4." *Journal of Biological Chemistry* 294 (20): 8015–22. https://doi.org/10.1074/jbc.RA119.007928.

Hsu, Mei-Hui, Uzen Savas, and Eric F. Johnson. 2018. "The X-Ray Crystal Structure of the Human Mono-Oxygenase Cytochrome P450 3A5-Ritonavir Complex Reveals Active Site Differences between P450s 3A4 and 3A5." *Molecular Pharmacology* 93 (1): 14–24. https://doi.org/10.1124/mol.117.109744.

Idris, Mukhtar Oluwaseun, Abeeb Abiodun Yekeen, Oluwaseun Suleiman Alakanse, and Olanrewaju Ayodeji Durojaye. 2021. "Computer-Aided Screening for Potential TMPRSS2 Inhibitors: A Combination of Pharmacophore Modeling, Molecular Docking and Molecular Dynamics Simulation Approaches." *Journal of Biomolecular Structure and Dynamics* 39 (15): 5638–56. https://doi.org/10.1080/07391102.2020.1792346.

Jacob K., Sony, Swastika Ganguly, Pravin Kumar, Raju Poddar, and Anoop Kumar. 2017. "Homology Model, Molecular Dynamics Simulation and Novel Pyrazole Analogs Design of *Candida Albicans* CYP450 Lanosterol 14 α-Demethylase, a Target Enzyme for Antifungal Therapy." *Journal of Biomolecular Structure and Dynamics* 35 (7): 1446–63. https://doi.org/10.1080/07391102.2016.1185380.

Jayakanthan, Mannu, Sanniyasi Chandrasekar, Jayaraman Muthukumaran, and Premendu P. Mathur. 2010. "Analysis of CYP3A4-HIV-1 Protease Drugs Interactions by Computational Methods for Highly Active Antiretroviral Therapy in HIV/AIDS." *Journal of Molecular Graphics and Modelling* 28 (5): 455–63. https://doi.org/10.1016/j.jmgm.2009.10.005

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Jha, Rajat Kumar, Rameez Jabeer Khan, Ekampreet Singh, Ankit Kumar, Monika Jain, Jayaraman Muthukumaran, and Amit Kumar Singh. 2022. “An Extensive Computational Study to Identify Potential Inhibitors of Acyl-Homoserine-Lactone Synthase from Acinetobacter Baumannii (Strain AYE).” *Journal of Molecular Graphics and Modelling* 114: 108168. https://doi.org/10.1016/j.jmgm.2022.108168

Jo, Sunhwan, Miklos Vargyas, Judit Vasko-Szedlar, Benoît Roux, and Wonpil Im. 2008. “PBEQ-Solver for Online Visualization of Electrostatic Potential of Biomolecules.” *Nucleic Acids Research* 36 (Web Server issue): W270–75. https://doi.org/10.1093/nar/gkn314.

Johnson, Eric F., and C. David Stout. 2013. “Structural Diversity of Eukaryotic Membrane Cytochrome P450s.” *The Journal of Biological Chemistry* 288 (24): 17082–90. https://doi.org/10.1074/jbc.R113.452805.

Johnson, Kevin M., Dian Su, and Donglu Zhang. 2021. “Characteristics of Major Drug Metabolizing Cytochrome P450 Enzymes.” In *Cytochrome P450: In Vitro Methods and Protocols*, edited by Zhengyin Yan and Gary W. Caldwell, 27–54. Methods in Pharmacology and Toxicology. New York, NY: Springer US. https://doi.org/10.1007/978-1-0716-1542-3_2.

Jørgensen, Jan Trøst. 2019. “Twenty Years with Personalized Medicine: Past, Present, and Future of Individualized Pharmacotherapy.” *The Oncologist* 24 (7): e432–40. https://doi.org/10.1634/theoncologist.2019-0054.

Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool *et al.* "Highly accurate protein structure prediction with AlphaFold." Nature 596, no. 7873 (2021): 583-589. https://doi.org/10.1038/s41586-021-03819-2

Kasai, Kosuke, Yudai Ito, Akihide Nitta, Kentaro Ariyoshi, Toshiya Nakamura, and Tomisato Miura. 2020. “Metal Coordination by L-Amino Acid Oxidase Derived

http://etd.uwc.ac.za/

from Flounder Platichthys Stellatus Is Structurally Essential and Regulates Antibacterial Activity." *Applied Microbiology and Biotechnology* 104 (22): 9645–54. https://doi.org/10.1007/s00253-020-10914-3.

Kaur, Amritpreet, Pratap Kumar Pati, Aparna Maitra Pati, and Avinash Kaur Nagpal. 2020. "Physico-Chemical Characterization and Topological Analysis of Pathogenesis-Related Proteins from Arabidopsis Thaliana and Oryza Sativa Using in-Silico Approaches." Edited by Diego F. Gomez-Casati. PLOS ONE 15 (9): e0239836. https://doi.org/10.1371/journal.pone.0239836.

Kaur, Parminder, A. Richard Chamberlin, Thomas L. Poulos, and Irina F. Sevrioukova. 2016. "Structure-Based Inhibitor Design for Evaluation of a CYP3A4 Pharmacophore Model." *Journal of Medicinal Chemistry* 59 (9): 4210–20. https://doi.org/10.1021/acs.jmedchem.5b01146.

Kehinde, Idowu, Pritika Ramharack, Manimbulu Nlooto, and Michelle Gordon. 2022. "Molecular Dynamic Mechanism (s) of Inhibition of Bioactive Antiviral Phytochemical Compounds Targeting Cytochrome P450 3A4 and P-Glycoprotein." *Journal of Biomolecular Structure and Dynamics* 40 (3): 1037–47. https://doi.org/10.1080/07391102.2020.1821780

Khalid, Ramsha, Muhammad Faraz Anwar, Muhammad Aanish Raees, Sadaf Naeem, Syed Hani Abidi, and Syed Ali. 2020. "An in Silico Approach to Analyze HCV Genotype-Specific Binding-Site Variation and Its Effect on Drug–Protein Interaction." Scientific Reports 10 (1): 20885. https://doi.org/10.1038/s41598-020-77720-9.

Khan, Rameez Jabeer, Rajat Kumar Jha, Gizachew Muluneh Amera, Monika Jain, Ekampreet Singh, Amita Pathak, Rashmi Prabha Singh, Jayaraman Muthukumaran, and Amit Kumar Singh. 2021. "Targeting SARS-CoV-2: A Systematic Drug Repurposing Approach to Identify Promising Inhibitors against 3C-like Proteinase and 2′-O-Ribose Methyltransferase." Journal of Biomolecular Structure and Dynamics 39 (8): 2679–92. https://doi.org/10.1080/07391102.2020.1753577.

http://etd.uwc.ac.za/

Kiani, Yusra Sajid, and Ishrat Jabeen. 2020. “Lipophilic Metabolic Efficiency (LipMetE) and Drug Efficiency Indices to Explore the Metabolic Properties of the Substrates of Selected Cytochrome P450 Isoforms.” *ACS Omega* 5 (1): 179–88. https://doi.org/10.1021/acsomega.9b02344.

Kim, J.-B., J. Sig Choi, Y.-M. Yu, K. Nam, C.-S. Piao, S.-W. Kim, M.-H. Lee, P.-L. Han, J.-s. Park, and J.-K. Lee. 2006. “HMGB1, a Novel Cytokine-Like Mediator Linking Acute Neuronal Death and Delayed Neuroinflammation in the Postischemic Brain.” *Journal of Neuroscience* 26 (24): 6413–21. https://doi.org/10.1523/JNEUROSCI.3815-05.2006.

Kim, Sunghwan, Paul A. Thiessen, Evan E. Bolton, Jie Chen, Gang Fu, Asta Gindulyte, Lianyi Han, *et al.* 2016. “PubChem Substance and Compound Databases.” *Nucleic Acids Research* 44 (Database issue): D1202–13. https://doi.org/10.1093/nar/gkv951.

Klein, Kathrin, and Ulrich M. Zanger. 2013. “Pharmacogenomics of Cytochrome P450 3A4: Recent Progress Toward the ‘Missing Heritability’ Problem.” *Frontiers in Genetics* 4. https://doi.org/10.3389/fgene.2013.00012.

Klein, Kathrin, Maria Thomas, Stefan Winter, Andreas K. Nussler, Mikko Niemi, Matthias Schwab, and Ulrich M. Zanger. 2012. “PPARA: A Novel Genetic Determinant of CYP3A4 In Vitro and In Vivo.” *Clinical Pharmacology and Therapeutics* 91 (6): 1044–52. https://doi.org/10.1038/clpt.2011.336.

Klingenberg, Martin. 1958. “Pigments of Rat Liver Microsomes.” *Archives of Biochemistry and Biophysics* 75 (2): 376–86. https://doi.org/10.1016/0003-9861(58)90436-3.

Kohut, Gergely, Adam Liwo, Szilvia Bősze, Tamás Beke-Somfai, and Sergey A. Samsonov. 2018. “Protein–Ligand Interaction Energy-Based Entropy Calculations: Fundamental Challenges for Flexible Systems.” *The Journal of Physical Chemistry B* 122 (32): 7821–27. https://doi.org/10.1021/acs.jpcb.8b03658

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Kondža, Martin, Mirza Bojić, Ivona Tomić, Željan Maleš, Valentina Rezić, and Ivan Ćavar. 2021. “Characterization of the CYP3A4 Enzyme Inhibition Potential of Selected Flavonoids.” *Molecules* 26 (10): 3018. https://doi.org/10.3390/molecules26103018.

Konstantinidis, Konstantinos, Ioannis Karakasiliotis, Kostas Anagnostopoulos, and Georgios C. Boulougouris. 2021. “On the Estimation of the Molecular Inaccessible Volume and the Molecular Accessible Surface of a Ligand in Protein–Ligand Systems.” *Molecular Systems Design & Engineering* 6 (11): 946–63. https://doi.org/10.1039/D1ME00053E

Koulgi, Shruti, Vinod Jani, Samiron Phukan, Uddhavesh Sonavane, Rajendra Joshi, Rajender Kumar Kamboj, and Venkata Palle. 2022. “A Deep Dive into the Conformational Dynamics of CYP3A4: Understanding the Binding of Homotropic and Non-Homotropic Ligands for Mitigating Drug-Drug Interaction (DDI).” *ChemistrySelect* 7 (17): e202200249. https://doi.org/10.1002/slct.202200249.

Kouza, Maksim, Anirban Banerji, Andrzej Kolinski, Irina Buhimschi, and Andrzej Kloczkowski. 2018. “Role of Resultant Dipole Moment in Mechanical Dissociation of Biological Complexes.” *Molecules* 23 (8): 1995. https://doi.org/10.3390/molecules23081995

Kozlowski, Lukasz Pawel. 2021. “IPC 2.0: Prediction of Isoelectric Point and p K a Dissociation Constants.” Nucleic Acids Research 49 (W1): W285–92. https://doi.org/10.1093/nar/gkab295.

Kraml, Johannes, Anna S. Kamenik, Franz Waibl, Michael Schauperl, and Klaus R. Liedl. 2019. “Solvation Free Energy as a Measure of Hydrophobicity: Application to Serine Protease Binding Interfaces.” *Journal of Chemical Theory and Computation* 15 (11): 5872–82. https://doi.org/10.1021/acs.jctc.9b00742.

Kumar, Manoj, Maharishi Tomar, Jayashree Potkule, Reetu, Sneh Punia, Jyoti Dhakane-Lad, Surinder Singh, *et al.* 2022. “Functional Characterization of Plant-

http://etd.uwc.ac.za/

Based Protein to Determine Its Quality for Food Applications." *Food Hydrocolloids* 123 (February): 106986. https://doi.org/10.1016/j.foodhyd.2021.106986.

Kumar, TV Ajay, Alias Anand S. Athavan, C. Loganathan, K. Saravanan, S. Kabilan, and V. Parthasarathy. 2018. "Design, 3D QSAR Modeling and Docking of TGF-β Type I Inhibitors to Target Cancer." Computational Biology and Chemistry 76: 232–44. https://doi.org/10.1016/j.compbiolchem.2018.07.011

Kumondai, Masaki, Evelyn Marie Gutiérrez Rico, Eiji Hishinuma, Akiko Ueda, Sakae Saito, Daisuke Saigusa, Shu Tadaka, *et al.* 2021. "Functional Characterization of 40 CYP3A4 Variants by Assessing Midazolam 1′-Hydroxylation and Testosterone 6β-Hydroxylation." *Drug Metabolism and Disposition* 49 (3): 212–20. https://doi.org/10.1124/dmd.120.000261.

Kundra, Pankaj, and Stalin Vinayagam. 2020. "Perioperative Intravenous Lidocaine: Crossing Local Boundaries and Reaching Systemic Horizons." *Indian Journal of Anaesthesia* 64 (5): 363. https://doi.org/10.4103%2Fija.IJA_431_20

Lakshmanan, Mageshwaran. 2019. "Drug Metabolism." In *Introduction to Basics of Pharmacology and Toxicology: Volume 1: General and Molecular Pharmacology: Principles of Drug Action*, edited by Gerard Marshall Raj and Ramasamy Raveendran, 99–116. Singapore: Springer. https://doi.org/10.1007/978-981-32-9779-1_7.

Lam, Y. W. Francis. 2019. "Principles of Pharmacogenomics." In *Pharmacogenomics*, 1–53. Elsevier. https://doi.org/10.1016/B978-0-12-812626-4.00001-2.

Lazar, Tamas, Mainak Guharoy, Wim Vranken, Sarah Rauscher, Shoshana J. Wodak, and Peter Tompa. 2020. "Distance-Based Metrics for Comparing Conformational Ensembles of Intrinsically Disordered Proteins." https://doi.org/10.1101/2020.04.06.027979.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Leong, Marieanne C. 2020. *Python 3 Essentials For Absolute Beginners and Curious Cats 1st Edition*. UMK PRESS. http://hdl.handle.net/123456789/302

Lewis, Benjamin C., Peter I. Mackenzie, and John O. Miners. 2011. "Application of Homology Modeling to Generate CYP1A1 Mutants with Enhanced Activation of the Cancer Chemotherapeutic Prodrug Dacarbazine." Molecular Pharmacology 80 (5): 879–88. https://doi.org/10.1124/mol.111.072124.

Lewis, David F.V., Miriam N. Jacobs, and Maurice Dickins. 2004. "Compound Lipophilicity for Substrate Binding to Human P450s in Drug Metabolism." *Drug Discovery Today* 9 (12): 530–37. https://doi.org/10.1016/S1359-6446(04)03115-0.

Lewis, David FV, Brian G. Lake, and Maurice Dickins. "Quantitative structure–activity relationships (QSARs) in CYP3A4 inhibitors: the importance of lipophilic character and hydrogen bonding." Journal of enzyme inhibition and medicinal chemistry 21, no. 2 (2006): 127-132. https://doi.org/10.1080/14756360500532747

Li, Jing, Luyong Zhang, Hang Zhou, Mark Stoneking, and Kun Tang. 2011. "Global Patterns of Genetic Diversity and Signals of Natural Selection for Human ADME Genes." *Human Molecular Genetics* 20 (3): 528–40. https://doi.org/10.1093/hmg/ddq498.

Li, Junhao, Yue Chen, Yun Tang, Weihua Li, and Yaoquan Tu. 2021. "Homotropic Cooperativity of Midazolam Metabolism by Cytochrome P450 3A4: Insight from Computational Studies." *Journal of Chemical Information and Modeling* 61 (5): 2418–26. https://doi.org/10.1021/acs.jcim.1c00266.

Li, Wanxin, Ali Farajtabar, Rong Xing, Yiting Zhu, and Hongkun Zhao. 2020. "Equilibrium Solubility Determination, Solvent Effect and Preferential Solvation of Amoxicillin in Aqueous Co-Solvent Mixtures of N, N-Dimethylformamide, Isopropanol, N-Methyl Pyrrolidone and Ethylene Glycol." *The Journal of Chemical Thermodynamics* 142: 106010. https://doi.org/10.1016/j.jct.2019.106010

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Li, Ying-Hui, Qian-Meng Lin, Ni-Hong Pang, Xiao-Dan Zhang, Huan-Le Huang, Jian-Ping Cai, and Guo-Xin Hu. 2019. “Functional Characterization of 27 CYP3A4 Protein Variants to Metabolize Regorafenib in Vitro.” *Basic & Clinical Pharmacology & Toxicology* 125 (4): 337–44. https://doi.org/10.1111/bcpt.13246

Lin, Qian-Meng, Ying-Hui Li, Qian Liu, Ni-Hong Pang, Ren-Ai Xu, Jian-Ping Cai, and Guo-Xin Hu. 2019. “Functional Characteristics of CYP3A4 Allelic Variants on the Metabolism of Loperamide in Vitro.” *Infection and Drug Resistance* 12 (September): 2809–17. https://doi.org/10.2147/IDR.S215129.

Liu, Tao, Ge Qian, WenTing Wang, and YanGuo Zhang. "Molecular docking to understand the metabolic behavior of GNF-351 by CYP3A4 and its potential drug–drug interaction with ketoconazole." European Journal of Drug Metabolism and Pharmacokinetics 40, no. 2 (2015): 235-238. https://doi.org/10.1007/s13318-014-0201-1

Lokwani, Deepak K., Aniket P. Sarkate, Kshipra S. Karnik, Anna Pratima G. Nikalje, and Julio A. Seijas. "Structure-based site of metabolism (SOM) prediction of ligand for CYP3A4 enzyme: comparison of Glide XP and Induced Fit Docking (IFD)." Molecules 25, no. 7 (2020): 1622. https://doi.org/10.3390/molecules25071622

Lokwani, Deepak K., Aniket P. Sarkate, Kshipra S. Karnik, Anna Pratima G. Nikalje, and Julio A. Seijas. 2020. “Structure-Based Site of Metabolism (SOM) Prediction of Ligand for CYP3A4 Enzyme: Comparison of Glide XP and Induced Fit Docking (IFD).” *Molecules* 25 (7): 1622. https://doi.org/10.3390/molecules25071622.

Lolodi, Ogheneochukome, Yue-Ming Wang, William C. Wright, and Taosheng Chen. 2017. “Differential Regulation of CYP3A4 and CYP3A5 and Its Implication in Drug Discovery.” *Current Drug Metabolism* 18 (12): 1095–1105. https://doi.org/10.2174/1389200218666170531112038.

http://etd.uwc.ac.za/

Long, William F., and Paul Labute. 2010. “Calibrative Approaches to Protein Solubility Modeling of a Mutant Series Using Physicochemical Descriptors.” *Journal of Computer-Aided Molecular Design* 24 (11): 907–16. https://doi.org/10.1007/s10822-010-9383-z

López-López, Edgar, J. Jesús Naveja, and José L. Medina-Franco. 2019. “DataWarrior: An Evaluation of the Open-Source Drug Discovery Tool.” *Expert Opinion on Drug Discovery* 14 (4): 335–41. https://doi.org/10.1080/17460441.2019.1581170.

Lu, Xiang, Peipei Xu, Hong-Ming Ding, You-Sheng Yu, Da Huo, and Yu-Qiang Ma. 2019. “Tailoring the Component of Protein Corona via Simple Chemistry.” *Nature Communications* 10 (1): 1–14. https://doi.org/10.1038/s41467-019-12470-5

Ma, Qiang, and Anthony Y. H. Lu. 2011. “Pharmacogenetics, Pharmacogenomics, and Individualized Medicine.” *Pharmacological Reviews* 63 (2): 437–59. https://doi.org/10.1124/pr.110.003533.

Magwanga, Richard Odongo, Pu Lu, Joy Nyangasi Kirungu, Qi Dong, Xiaoyan Cai, Zhongli Zhou, Xingxing Wang, *et al.* 2019. “Knockdown of Cytochrome P450 Genes Gh_D07G1197 and Gh_A13G2057 on Chromosomes D07 and A13 Reveals Their Putative Role in Enhancing Drought and Salt Stress Tolerance in Gossypium Hirsutum.” *Genes* 10 (3): 226. https://doi.org/10.3390/genes10030226.

Mali, Suraj N., and Hemchandra K. Chaudhari. 2018. “Computational Studies on Imidazo [1, 2-a] Pyridine-3-Carboxamide Analogues as Antimycobacterial Agents: Common Pharmacophore Generation, Atom-Based 3D-QSAR, Molecular Dynamics Simulation, QikProp, Molecular Docking and Prime MMGBSA Approaches.” *Open Pharmaceutical Sciences Journal* 5 (1). http://dx.doi.org/10.2174/1874844901805010012

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Marsh, Joseph A., and Sarah A. Teichmann. 2015. “Structure, Dynamics, Assembly, and Evolution of Protein Complexes.” *Annual Review of Biochemistry* 84 (1): 551–75. https://doi.org/10.1146/annurev-biochem-060614-034142.

Matubayasi, Nobuyuki. 2017. “Free-Energy Analysis of Protein Solvation with All-Atom Molecular Dynamics Simulation Combined with a Theory of Solutions.” Current Opinion in Structural Biology 43 (April): 45–54. https://doi.org/10.1016/j.sbi.2016.10.005.

McDonnell, Anne M, and Cathyyen H. Dang. 2013. “Basic Review of the Cytochrome P450 System.” *Journal of the Advanced Practitioner in Oncology* 4 (4): 263–68. https://doi.org/10.6004%2Fjadpro.2013.4.4.7

McLean, Kirsty J., and Andrew W. Munro. 2018. “Cytochrome P450 (Cyp).” In *Encyclopedia of Signaling Molecules*, edited by Sangdun Choi, 1288–1305. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-67199-4_101615.

Michel-Levy, Javier Marcos. 2020. “Pharmacokinetics and Pharmacodynamics of Local Anesthetics.” In *Topics in Local Anesthetics*, 31. IntechOpen. DOI: 10.5772/intechopen.91700

Midlik, Adam, Veronika Navrátilová, Taraka Ramji Moturu, Jaroslav Koča, Radka Svobodová, and Karel Berka. 2021. “Uncovering of Cytochrome P450 Anatomy by SecStrAnnotator.” *Scientific Reports* 11 (1): 12345. https://doi.org/10.1038/s41598-021-91494-8.

Miyazaki, Mitsue, Katsunori Nakamura, Yukiyoshi Fujita, F. Peter Guengerich, Ryuya Horiuchi, and Koujirou Yamamoto. 2008. “Defective Activity of Recombinant Cytochromes P450 3A4. 2 and 3A4. 16 in Oxidation of Midazolam, Nifedipine, and Testosterone.” *Drug Metabolism and Disposition* 36 (11): 2287–91. https://doi.org/10.1124/dmd.108.021816

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Mukhtar, Sadia, Yusra Sajid Kiani, and Ishrat Jabeen. "Molecular docking simulations and GRID-independent molecular descriptor (GRIND) analysis to probe stereoselective interactions of CYP3A4 inhibitors." Medicinal Chemistry Research 26, no. 10 (2017): 2322-2335. https://doi.org/10.1007/s00044-017-1933-7

Mulder, Tessa A. M., Ruben A. G. van Eerden, Mirjam de With, Laure Elens, Dennis A. Hesselink, Maja Matic, Sander Bins, Ron H. J. Mathijssen, and Ron H. N. van Schaik. 2021. "CYP3A4∗22 Genotyping in Clinical Practice: Ready for Implementation?" *Frontiers in Genetics* 12: 1237. https://doi.org/10.3389/fgene.2021.711943.

Müller, Christian S. 2014. "Functional Investigations of Cytochrome P450 3A4." https://doi.org/10.5167/UZH-103953.

Müller, Christian S., Tim Knehans, Dmitri R. Davydov, Patricia L. Bounds, Ursula von Mandach, James R. Halpert, Amedeo Caflisch, and Willem H. Koppenol. 2015. "Concurrent Cooperativity and Substrate Inhibition in the Epoxidation of Carbamazepine by Cytochrome P450 3A4 Active Site Mutants Inspired by Molecular Dynamics Simulations." *Biochemistry* 54 (3): 711–21. https://doi.org/10.1021/bi5011656.

Munjal, Nupur S., Rohit Shukla, and Tiratha Raj Singh. 2021. "Physicochemical Characterization of Paclitaxel Prodrugs with Cytochrome 3A4 to Correlate Solubility and Bioavailability Implementing Molecular Docking and Simulation Studies." *Journal of Biomolecular Structure and Dynamics* 0 (0): 1–13. https://doi.org/10.1080/07391102.2021.1875881.

Munjal, Nupur S., Rohit Shukla, and Tiratha Raj Singh. 2021a. "Physicochemical Characterization of Paclitaxel Prodrugs with Cytochrome 3A4 to Correlate Solubility and Bioavailability Implementing Molecular Docking and Simulation Studies." Journal of Biomolecular Structure and Dynamics 0 (0): 1–13. https://doi.org/10.1080/07391102.2021.1875881.

http://etd.uwc.ac.za/

Murayama, Norie, Takahiro Nakamura, Mayumi Saeki, Akiko Soyama, Yoshiro Saito, Kimie Sai, Seiichi Ishida, *et al.* 2002. "CYP3A4 Gene Polymorphisms Influence Testosterone 6beta-Hydroxylation." *Drug Metabolism and Pharmacokinetics* 17 (2): 150–56. https://doi.org/10.2133/dmpk.17.150.

Nagai, Mika, Yoshihiro Konno, Masahiro Satsukawa, Shinji Yamashita, and Kouichi Yoshinari. 2016. "Establishment of In Silico Prediction Models for CYP3A4 and CYP2B6 Induction in Human Hepatocytes by Multiple Regression Analysis Using Azole Compounds." Drug Metabolism and Disposition: The Biological Fate of Chemicals 44 (8): 1390–98. https://doi.org/10.1124/dmd.115.068619.

Nair, Pramod, Ross McKinnon, and John Miners. 2016. "Cytochrome P450 Structure–Function: Insights from Molecular Dynamics Simulations." *Drug Metabolism Reviews* 48 (May): 1–19. https://doi.org/10.1080/03602532.2016.1178771.

Neamțu, Monica. 2020. "MOLECULAR BACKGROUNDS OF INDIVIDUALISED PHARMACOTHERAPY." *FARMACIA* 68 (3): 396–405. https://doi.org/10.31925/farmacia.2020.3.4.

Nikolaev, Dmitrii M., Andrey A. Shtyrov, Maxim S. Panov, Adeel Jamal, Oleg B. Chakchir, Vladimir A. Kochemirovsky, Massimo Olivucci, and Mikhail N. Ryazantsev. 2018. "A Comparative Study of Modern Homology Modeling Algorithms for Rhodopsin Structure Prediction." *ACS Omega* 3 (7): 7555–66.

Nim, Satra, Jouhyun Jeon, Carles Corbi-Verge, Moon-Hyeong Seo, Ylva Ivarsson, Jason Moffat, Nadya Tarasova, and Philip M Kim. 2016. "Pooled Screening for Antiproliferative Inhibitors of Protein-Protein Interactions." *Nature Chemical Biology* 12 (4): 275–81. https://doi.org/10.1038/nchembio.2026.

UNIVERSITY of the
WESTERN CAPE
http://etd.uwc.ac.za/

Niwa, Toshiro, Norie Murayama, and Hiroshi Yamazaki. 2011. “Stereoselectivity of Human Cytochrome P450 in Metabolic and Inhibitory Activities.” *Current Drug Metabolism* 12 (6): 549–69. https://doi.org/10.2174/138920011795713724.

Niwa, Toshiro, Norie Murayama, Yurie Imagawa, and Hiroshi Yamazaki. 2015. “Regioselective Hydroxylation of Steroid Hormones by Human Cytochromes P450.” *Drug Metabolism Reviews* 47 (2): 89–110. https://doi.org/10.3109/03602532.2015.1011658.

Ohkura, Kazuto, Yuki Kawaguchi, Yasuko Watanabe, Yasuhiro Masubuchi, Yasuo Shinohara, and Hitoshi Hori. 2009. “Flexible Structure of Cytochrome P450: Promiscuity of Ligand Binding in The CYP3A4 Heme Pocket.” *Anticancer Research* 29 (3): 935–42. http://orcid.org/0000-0002-4045-2760

Okella, Hedmon, John J. Georrge, Sylvester Ochwo, Christian Ndekezi, Kevin Tindo Koffi, Jacqueline Aber, Clement Olusoji Ajayi, *et al.* 2020. “New Putative Antimicrobial Candidates: In Silico Design of Fish-Derived Antibacterial Peptide-Motifs.” *Frontiers in Bioengineering and Biotechnology* 8. https://doi.org/10.3389/fbioe.2020.604041.

Oliveira, Saulo de, and Charlotte Deane. 2017. “Co-Evolution Techniques Are Reshaping the Way We Do Structural Bioinformatics.” *F1000Research* 6 (July): 1224. https://doi.org/10.12688/f1000research.11543.1.

Oostenbrink, Chris, Alessandra Villa, Alan E. Mark, and Wilfred F. van Gunsteren. 2004. “A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6.” *Journal of Computational Chemistry* 25 (13): 1656–76. https://doi.org/10.1002/jcc.20090.

Oselusi, Samson O., Samuel A. Egieyeh, and Alan Christoffels. 2021. “Cheminformatic Profiling and Hit Prioritization of Natural Products with Activities against Methicillin-Resistant Staphylococcus Aureus (MRSA).” Molecules 26 (12): 3674. https://doi.org/10.3390/molecules26123674.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Osuna, Sílvia. 2021. “The Challenge of Predicting Distal Active Site Mutations in Computational Enzyme Design.” *WIREs Computational Molecular Science* 11 (3). https://doi.org/10.1002/wcms.1502.

Otyepka, Michal, Karel Berka, and Pavel Anzenbacher. 2012. “Is There a Relationship Between the Substrate Preferences and Structural Flexibility of Cytochromes P450?” *Current Drug Metabolism* 13 (2): 130–42. https://doi.org/10.2174/138920012798918372.

Oyugi, Martin Omulindi, Johnson Kangethe Kinyua, Esther Nkirote Magiri, Milcah Wagio Kigoni, Evenilton Pessoa Costa, and Naftaly Wang’ombe Githaka. 2018. “In Silico Characterization and Structural Modeling of Dermacentor Andersoni P36 Immunosuppressive Protein.” *Advances in Bioinformatics* 2018 (April): e7963401. https://doi.org/10.1155/2018/7963401.

Paiz, Elisia A., Karen A. Lewis, and Steven T. Whitten. 2021. “Structural and Energetic Characterization of the Denatured State from the Perspectives of Peptides, the Coil Library, and Intrinsically Disordered Proteins.” *Molecules* 26 (3): 634. https://doi.org/10.3390/molecules26030634

Panda, Saroj Kumar, Shalini Saxena, Parth Sarthi Sen Gupta, and Malay Kumar Rana. 2021. “Inhibitors of Plasmepsin X Plasmodium Falciparum: Structure-Based Pharmacophore Generation and Molecular Dynamics Simulation.” *Journal of Molecular Liquids* 340 (October): 116851. https://doi.org/10.1016/j.molliq.2021.116851.

Park, Jin-Young, Yuno Lee, Hee Jae Lee, Yong-Soo Kwon, and Wanjoo Chun. 2020. “In Silico Screening of GABA Aminotransferase Inhibitors from the Constituents of Valeriana Officinalis by Molecular Docking and Molecular Dynamics Simulation Study.” *Journal of Molecular Modeling* 26 (9): 228. https://doi.org/10.1007/s00894-020-04495-1.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Patel, Rushika, Rajesh Chudasama, Rutujaben Solanki, Priya Patel, Krupali Parmar, and Nasreen S. Munshi. 2020. “Structure Prediction and Molecular Docking Studies of Aromatic Hydrocarbon Sensing Proteins TbuT, HbpR and PhnR to Detect Priority Pollutants.” Journal of Environmental Science and Health, Part A 55 (2): 126–41. https://doi.org/10.1080/10934529.2019.1672457.

Patil, Prajakta H., Puralae Channabasavaiah Jagadish, Fajeelath Fatima, Sumit Birangal, Gurupur Gautham Shenoy, Mahadev Rao, Junaid Farooqui, Himanshu Rastogi, Tarun Sharma, and Jakir Pinjari. "Inhibition of Cytochrome P450 Enzyme and Drug-Drug Interaction Potential of Acid Reducing Agents Used in Management of CDK Inhibitors for Breast Cancer Chemotherapy." Current Drug Metabolism 23, no. 2 (2022): 137-149. https://doi.org/10.2174/1389200223666220218090948

Patra, Prasanta, Pratik Ghosh, Bidhan Chandra Patra, and Manojit Bhattacharya. 2020. “Biocomputational Analysis and In Silico Characterization of an Angiogenic Protein (RNase5) in Zebrafish (Danio Rerio).” *International Journal of Peptide Research and Therapeutics* 26 (4): 1687–97. https://doi.org/10.1007/s10989-019-09978-1.

Pedroso, João A.B., Thais T. Zampieri, and Jose Donato. 2015. “Reviewing the Effects of L-Leucine Supplementation in the Regulation of Food Intake, Energy Balance, and Glucose Homeostasis.” *Nutrients* 7 (5): 3914–37. https://doi.org/10.3390/nu7053914.

Penzak, Scott R., and Carlos Rojas-Fernandez. 2019. “4β-Hydroxycholesterol as an Endogenous Biomarker for CYP3A Activity: Literature Review and Critical Evaluation.” *The Journal of Clinical Pharmacology* 59 (5): 611–24. https://doi.org/10.1002/jcph.1391.

Peterson, Lenna X., Woong-Hee Shin, Hyungrae Kim, and Daisuke Kihara. 2018. “Improved Performance in CAPRI Round 37 Using LZerD Docking and Template-Based Modeling with Combined Scoring Functions.” *Proteins: Structure,*

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Function, and Bioinformatics 86 (March): 311–20. https://doi.org/10.1002/prot.25376.

Pettersen, Eric F., Thomas D. Goddard, Conrad C. Huang, Elaine C. Meng, Gregory S. Couch, Tristan I. Croll, John H. Morris, and Thomas E. Ferrin. 2021. "UCSF ChimeraX: Structure Visualization for Researchers, Educators, and Developers." *Protein Science* 30 (1): 70–82. https://doi.org/10.1002/pro.3943.

Phillips, James C., Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kalé, and Klaus Schulten. 2005. "Scalable Molecular Dynamics with NAMD." *Journal of Computational Chemistry* 26 (16): 1781–1802. https://doi.org/10.1002/jcc.20289.

Piccinini, Esteban, Sebastián Alberti, Gabriel S. Longo, Teresa Berninger, Josef Breu, Jakub Dostalek, Omar Azzaroni, and Wolfgang Knoll. 2018. "Pushing the Boundaries of Interfacial Sensitivity in Graphene FET Sensors: Polyelectrolyte Multilayers Strongly Increase the Debye Screening Length." *The Journal of Physical Chemistry C* 122 (18): 10181–88. https://doi.org/10.1021/acs.jpcc.7b11128

Pickard, J. E., J. Fisher, E. Ingham, and J. Egan. 1998. "Investigation into the Effects of Proteins and Lipids on the Frictional Properties of Articular Cartilage." *Biomaterials* 19 (19): 1807–12. https://doi.org/10.1016/S0142-9612(98)00147-1

Polic, Vanja. 2018. *Investigating Cytochrome P450 3A4: Exploring Its Biocatalytic, Allosteric and Kinetic Properties*. McGill University Libraries.

Poolman, Bert, Buu Minh Tran, Haritha Prabha, Aditya Iyer, Conor P. O'Byrne, and Tjakko Abee. 2021. "Measurement of Protein Mobility in Listeria Monocytogenes Reveals a Unique Tolerance to Osmotic Stress and Temperature Dependence of Diffusion." Frontiers in Microbiology 12: 207. http://dx.doi.org/10.3389/fmicb.2021.640149

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Poulos, Thomas L., and Eric F. Johnson. 2015. “Structures of Cytochrome P450 Enzymes.” In *Cytochrome P450*, edited by Paul R. Ortiz de Montellano, 3–32. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-12108-6_1.

Prajapati, Jignesh, Rohit Patel, Priyashi Rao, Meenu Saraf, Rakesh Rawal, and Dweipayan Goswami. 2022. “Perceiving SARS-CoV-2 Mpro and PLpro Dual Inhibitors from Pool of Recognized Antiviral Compounds of Endophytic Microbes: An in Silico Simulation Study.” *Structural Chemistry*, 1–25.

Pramanik, Krishnendu, Tithi Soren, Soumik Mitra, and Tushar Kanti Maiti. 2017. “In Silico Structural and Functional Analysis of Mesorhizobium ACC Deaminase.” *Computational Biology and Chemistry* 68 (June): 12–21. https://doi.org/10.1016/j.compbiolchem.2017.02.005.

Pražnikar, Jure. 2021. “Scaling Laws of Graphs of 3D Protein Structures.” Journal of Bioinformatics and Computational Biology 19 (03): 2050050. https://doi.org/10.1007/s11224-022-01932-0

Radouani, Fouzia, Lyndon Zass, Yosr Hamdi, Jorge da Rocha, Reem Sallam, Sonia Abdelhak, Samah Ahmed, *et al.* 2020. “A Review of Clinical Pharmacogenetics Studies in African Populations.” *Personalized Medicine* 17 (2): 155–70. https://doi.org/10.2217/pme-2019-0110.

Ragia, Georgia, Vana Kolovou, Anna Tavridou, Laure Elens, Alexandros D. Tselepis, Moses Elisaf, Ron H. N. Van Schaik, Genovefa Kolovou, and Vangelis G. Manolopoulos. 2015. “No Effect of CYP3A4 Intron 6 C>T Polymorphism (CYP3A4*22) on Lipid-Lowering Response to Statins in Greek Patients with Primary Hypercholesterolemia.” *Drug Metabolism and Personalized Therapy* 30 (1): 43–48. https://doi.org/10.1515/dmdi-2014-0021.

Rahman, Noor, Ijaz Muhammad, Gul-E-Nayab, Haroon Khan, Michael Aschner, Rosanna Filosa, and Maria Daglia. 2019. “Molecular Docking of Isolated Alkaloids

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

for Possible α-Glucosidase Inhibition." *Biomolecules* 9 (10): 544. https://doi.org/10.3390/biom9100544.

Rajagopal, Kalirajan, Pandiselvi Arumugasamy, and Gowramma Byran. 2019. "In-Silico Drug Design, ADMET Screening, MM-GBSA Binding Free Energy of Some Chalcone Substituted 9-Anilinoacridines as HER2 Inhibitors for Breast Cancer." *International Journal of Computational and Theoretical Chemistry* 7 (1): 6. https://doi.org/10.2174/2589977511666190912154817

Ramachandran, G. N., C. Ramakrishnan, and V. Sasisekharan. 1963. "Stereochemistry of Polypeptide Chain Configurations." *Journal of Molecular Biology* 7 (1): 95–99. https://doi.org/10.1016/S0022-2836(63)80023-6.

Reddy, P. Praveen. 2020. "Homology Modeling of Microbial Nattokinase Enzyme, An Anti-Blood Clotting (Fibrinolytic) Agent Using Computational Tools." *Research Journal of Pharmacy and Technology* 13 (9): 4135. https://doi.org/10.5958/0974-360X.2020.00730.1.

Rendic, Slobodan, and F. Peter Guengerich. 2015. "Survey of Human Oxidoreductases and Cytochrome P450 Enzymes Involved in the Metabolism of Xenobiotic and Natural Chemicals." *Chemical Research in Toxicology* 28 (1): 38–42. https://doi.org/10.1021/tx500444e.

Rezaei, Shokouh, Yahya Sefidbakht, and Vuk Uskoković. 2020. "Comparative Molecular Dynamics Study of the Receptor-Binding Domains in SARS-CoV-2 and SARS-CoV and the Effects of Mutations on the Binding Affinity." *Journal of Biomolecular Structure and Dynamics* 0 (0): 1–20. https://doi.org/10.1080/07391102.2020.1860829.

Ridhwan, Mohamad Jemain Mohamad, Syahrul Imran Abu Bakar, Normala Abd Latip, Nurunajah Ab Ghani, and Nor Hadiani Ismail. 2022. "A Comprehensive Analysis of Human CYP3A4 Crystal Structures as a Potential Tool for Molecular Docking-Based Site of Metabolism and Enzyme Inhibition Studies." *Journal of*

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Computational Biophysics and Chemistry 21 (03): 259–85. https://doi.org/10.1142/S2737416522300012.

Rittle, Jonathan, and Michael T. Green. 2010. "Cytochrome P450 Compound I: Capture, Characterization, and C-H Bond Activation Kinetics." *Science (New York, N.Y.)* 330 (6006): 933–37. https://doi.org/10.1126/science.1193478.

Rodrigues, Carlos HM, Douglas EV Pires, and David B Ascher. 2018. "DynaMut: Predicting the Impact of Mutations on Protein Conformation, Flexibility and Stability." Nucleic Acids Research 46 (W1): W350–55. https://doi.org/10.1093/nar/gky300.

Ruppé, Etienne, Amine Ghozlane, Julien Tap, Nicolas Pons, Anne-Sophie Alvarez, Nicolas Maziers, Trinidad Cuesta, *et al.* 2019. "Prediction of the Intestinal Resistome by a Three-Dimensional Structure-Based Method." *Nature Microbiology* 4 (1): 112–23. https://doi.org/10.1038/s41564-018-0292-6.

Saba, Nikhat, and Alpana Seal. 2018. "Identification of a Less Toxic Vinca Alkaloid Derivative for Use as a Chemotherapeutic Agent, Based on in Silico Structural Insights and Metabolic Interactions with CYP3A4 and CYP3A5." *Journal of Molecular Modeling* 24 (4): 1–14.

UNIVERSITY of the
WESTERN CAPE

Sabiu, Saheed, and Kehinde Idowu. 2022. "An Insight on the Nature of Biochemical Interactions between Glycyrrhizin, Myricetin and CYP3A4 Isoform." *Journal of Food Biochemistry* 46 (3): e13831. https://doi.org/10.1111/jfbc.13831

Sahay, Archna, Anil Piprodhe, and Mashitha Pise. 2020. "In Silico Analysis and Homology Modeling of Strictosidine Synthase Involved in Alkaloid Biosynthesis in Catharanthus Roseus." *Journal of Genetic Engineering and Biotechnology* 18 (1): 44. https://doi.org/10.1186/s43141-020-00049-3.

Sahoo, Susrita, Satya Narayan Sahu, Subrat kumar Pattanayak, Namrata Misra, and Mrutyunjay Suar. 2020. "Biosensor and Its Implementation in Diagnosis of

http://etd.uwc.ac.za/

Infectious Diseases." In *Smart Biosensors in Medical Care*, 29–47. Elsevier. https://doi.org/10.1016/B978-0-12-820781-9.00002-4.

Saiz-Rodríguez, Miriam, Susana Almenara, Marcos Navares-Gómez, Dolores Ochoa, Manuel Román, Pablo Zubiaur, Dora Koller, *et al.* 2020. "Effect of the Most Relevant CYP3A4 and CYP3A5 Polymorphisms on the Pharmacokinetic Parameters of 10 CYP3A Substrates." *Biomedicines* 8 (4): 94. https://doi.org/10.3390/biomedicines8040094.

Salmaso, Veronica, and Stefano Moro. 2018. "Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview." *Frontiers in Pharmacology* 9: 923. https://doi.org/10.3389/fphar.2018.00923.

Salo-Ahen, Outi M. H., Ida Alanko, Rajendra Bhadane, Alexandre M. J. J. Bonvin, Rodrigo Vargas Honorato, Shakhawath Hossain, André H. Juffer, *et al.* 2021. "Molecular Dynamics Simulations in Drug Discovery and Pharmaceutical Development." *Processes* 9 (1): 71. https://doi.org/10.3390/pr9010071.

Samuels, Eric R., and Irina F. Sevrioukova. 2021. "Rational Design of CYP3A4 Inhibitors: A One-Atom Linker Elongation in Ritonavir-Like Compounds Leads to a Marked Improvement in the Binding Strength." *International Journal of Molecular Sciences* 22 (2): 852. https://doi.org/10.3390/ijms22020852.

Sankar, Kannan, Stanley R. Krystek, Stephen M. Carl, Tyler Day, and Johannes K. X. Maier. 2018. "AggScore: Prediction of Aggregation-Prone Regions in Proteins Based on the Distribution of Surface Patches." *Proteins: Structure, Function, and Bioinformatics* 86 (11): 1147–56. https://doi.org/10.1002/prot.25594.

Sata, F. 2000. "CYP3A4 Allelic Variants with Amino Acid Substitutions in Exons 7 and 12: Evidence for an Allelic Variant with Altered Catalytic Activity." *Clinical Pharmacology & Therapeutics* 67 (1): 48–56. https://doi.org/10.1067/mcp.2000.104391.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Sellés Vidal, Lara, Ciarán L. Kelly, Paweł M. Mordaka, and John T. Heap. 2018. "Review of NAD(P)H-Dependent Oxidoreductases: Properties, Engineering and Application." *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1866 (2): 327–47. https://doi.org/10.1016/j.bbapap.2017.11.005.

Selvam, K., D. Senbagam, T. Selvankumar, C. Sudhakar, S. Kamala-Kannan, B. Senthilkumar, and M. Govarthanan. 2017. "Cellulase Enzyme: Homology Modeling, Binding Site Identification and Molecular Docking." *Journal of Molecular Structure* 1150 (December): 61–67. https://doi.org/10.1016/j.molstruc.2017.08.067.

Selzer, Paul M., Richard J. Marhöfer, and Oliver Koch. 2018. "Sequence Comparisons and Sequence-Based Database Searches." In *Applied Bioinformatics*, 35–50. Springer.

Sétáló Jr, György. 2013. "Problem-Based Test: Replication of Mitochondrial DNA during the Cell Cycle." *Biochemistry and Molecular Biology Education* 41 (4): 273–77. https://doi.org/10.1002/bmb.20712

Sevrioukova, Irina F., and Thomas L. Poulos. "Current approaches for investigating and predicting cytochrome P450 3A4-ligand interactions." Monooxygenase, peroxidase and peroxygenase properties and mechanisms of cytochrome P450 (2015): 83-105. https://doi.org/10.1007/978-3-319-16009-2_3

Sevrioukova, Irina F., and Thomas L. Poulos. "Structural basis for regiospecific midazolam oxidation by human cytochrome P450 3A4." Proceedings of the National Academy of Sciences 114, no. 3 (2017): 486-491. https://doi.org/10.1073/pnas.1616198114

Sevrioukova, Irina F., and Thomas L. Poulos. 2013. "Understanding the Mechanism of Cytochrome P450 3A4: Recent Advances and Remaining Problems." *Dalton Trans.* 42 (9): 3116–26. https://doi.org/10.1039/C2DT31833D.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Sharn, Hari Om, Dev Bukhsh Singh, and Satendra Singh. 2021. “Comparative Structural Analysis of Bifunctional Methylenetetrahydrofolate Dehydrogenase/Methenyltetrahydrofolate Cyclohydrolase from Bordetella Pertussis and Bordetella Parapertussis: A Drug Target against Pertussis.” In Silico Pharmacology 9 (1): 7. https://doi.org/10.1007/s40203-020-00069-4.

Shin, Woong-Hee, Xuejiao Kang, Jian Zhang, and Daisuke Kihara. 2017. “Prediction of Local Quality of Protein Structure Models Considering Spatial Neighbors in Graphical Models.” *Scientific Reports* 7 (1): 40629. https://doi.org/10.1038/srep40629.

Shukla, Samvedna, Bhawana Mishra, Himanshu Avashthi, and Muktesh Chandra. 2022. “Biological Sequence Analysis.” In *Bioinformatics*, 33–47. Elsevier.

Singh, Ravi, Ankit Ganeshpurkar, Devendra Kumar, Dileep Kumar, Ashok Kumar, and Sushil Kumar Singh. 2020. “Identifying Potential GluN2B Subunit Containing N-Methyl-D-Aspartate Receptor Inhibitors: An Integrative *in Silico* and Molecular Modeling Approach.” *Journal of Biomolecular Structure and Dynamics* 38 (9): 2533–45. https://doi.org/10.1080/07391102.2019.1635530.

Singh, Suresh B., Lucy Q. Shen, Matthew J. Walker, and Robert P. Sheridan. 2003. “A Model for Predicting Likely Sites of CYP3A4-Mediated Metabolism on Drug-like Molecules.” Journal of Medicinal Chemistry 46 (8): 1330–36. https://doi.org/10.1021/jm020400s.

Singh, Suresh B., Lucy Q. Shen, Matthew J. Walker, and Robert P. Sheridan. 2003. “A Model for Predicting Likely Sites of CYP3A4-Mediated Metabolism on Drug-like Molecules.” *Journal of Medicinal Chemistry* 46 (8): 1330–36. https://doi.org/10.1021/jm020400s.

Smith, Aubrey A., and Amanda Caruso. 2013. “In Silico Characterization and Homology Modeling of a Cyanobacterial Phosphoenolpyruvate Carboxykinase

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Enzyme." Research Article. Structural Biology. Hindawi. March 30, 2013. https://doi.org/10.1155/2013/370820.

Smith, Robert L., and Stephen C. Mitchell. 2020. "Pharmacogenetics and Drug Metabolism: Historical Perspective and Appraisal." *Xenobiotica* 50 (1): 3–8. https://doi.org/10.1080/00498254.2019.1668579.

Smith, Sherri A., Sandra Gagnon, and Nigel J. Waters. 2017. "Mechanistic Investigations into the Species Differences in Pinometostat Clearance: Impact of Binding to Alpha-1-Acid Glycoprotein and Permeability-Limited Hepatic Uptake." *Xenobiotica* 47 (3): 185–93. https://doi.org/10.3109/00498254.2016.1173265

Sobolev, Oleg V., Pavel V. Afonine, Nigel W. Moriarty, Maarten L. Hekkelman, Robbie P. Joosten, Anastassis Perrakis, and Paul D. Adams. 2020. "A Global Ramachandran Score Identifies Protein Structures with Unlikely Stereochemistry." *Structure* 28 (11): 1249-1258.e2. https://doi.org/10.1016/j.str.2020.08.005.

Sohaib Shahzan, M., A.S. Smiline Girija, and J. Vijayashree Priyadharsini. 2019. "A Computational Study Targeting the Mutated L321F of ERG11 Gene in C. Albicans, Associated with Fluconazole Resistance with Bioactive Compounds from Acacia Nilotica." Journal de Mycologie Médicale 29 (4): 303–9. https://doi.org/10.1016/j.mycmed.2019.100899.

Soni, Neelesh, and M.S. Madhusudhan. 2017. "Computational Modeling of Protein Assemblies." *Current Opinion in Structural Biology* 44 (June): 179–89. https://doi.org/10.1016/j.sbi.2017.04.006.

Šrejber, Martin, Veronika Navrátilová, Markéta Paloncýová, Václav Bazgier, Karel Berka, Pavel Anzenbacher, and Michal Otyepka. 2018. "Membrane-Attached Mammalian Cytochromes P450: An Overview of the Membrane's Effects on Structure, Drug Binding, and Interactions with Redox Partners." *Journal of Inorganic Biochemistry* 183 (June): 117–36. https://doi.org/10.1016/j.jinorgbio.2018.03.002.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Sripriya Akondi, Vishnu, Vineetha Menon, Jerome Baudry, and Jana Whittle. 2022. "Novel Big Data-Driven Machine Learning Models for Drug Discovery Application." *Molecules* 27 (3): 594. https://doi.org/10.3390/molecules27030594.

Stefaniu, Amalia, Azhar Rasul, and Ghulam Hussain. 2020. Cheminformatics and Its Applications. BoD – Books on Demand. 10.5772/intechopen.83236

Stephens, C., M. I. Lucena, and R. J. Andrade. 2018. "2.26 - Idiosyncratic Drug-Induced Liver Injury: Mechanisms and Susceptibility Factors☆." In *Comprehensive Toxicology (Third Edition)*, edited by Charlene A. McQueen, 625–50. Oxford: Elsevier. https://doi.org/10.1016/B978-0-12-801238-3.64089-8.

Stillemans, Gabriel, Leila Belkhir, Dennis A. Hesselink, Vincent Haufroid, and Laure Elens. 2018. "Pharmacogenetic Associations with Cytochrome P450 in Antiretroviral Therapy: What Does the Future Hold?" *Expert Opinion on Drug Metabolism & Toxicology* 14 (6): 601–11. https://doi.org/10.1080/17425255.2018.1478964.

Stone, Tracy A., Gregory B. Cole, Dorna Ravamehr-Lake, Huong Q. Nguyen, Farheen Khan, Simon Sharpe, and Charles M. Deber. 2019. "Positive Charge Patterning and Hydrophobicity of Membrane-Active Antimicrobial Peptides as Determinants of Activity, Toxicity, and Pharmacokinetic Stability." *Journal of Medicinal Chemistry* 62 (13): 6276–86.

Stroet, Martin. 2018. "Improving the Accuracy of Molecular Dynamics Simulations: Parameterisation of Interaction Potentials for Small Molecules." PhD Thesis, The University of Queensland. https://doi.org/10.14264/uql.2018.432.

Studer, Gabriel, Christine Rempfer, Andrew M Waterhouse, Rafal Gumienny, Juergen Haas, and Torsten Schwede. 2020. "QMEANDisCo—Distance Constraints Applied on Model Quality Estimation." Edited by Arne Elofsson. *Bioinformatics* 36 (6): 1765–71. https://doi.org/10.1093/bioinformatics/btz828.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Su, Chen-Ying, Lung-Kun Yeh, Chi-Chun Lai, Kuan-Yi Li, Ching-Li Tseng, and Hsu-Wei Fang. 2020. “Effects of Lysosomal Deposition on the Friction Coefficient of Hydrogel Contact Lenses.” Contact Lens and Anterior Eye 43 (2): 144–48. https://doi.org/10.1016/j.clae.2019.09.007

Suhaibun, Sherman & Elengoe, Asita & Poddar, Ruma. (2020). Technology Advance in Drug Design Using Computational Biology Tool. Malaysian Journal of Medicine and Health Sciences. 16. 2636-9346.

Sun, Guo, Chaoqun Hu, Qing Mei, Minghe Luo, Xu Chen, Zhengyuan Li, Yuanzhen Liu, Zixin Deng, Zhengyu Zhang, and Yuhui Sun. 2020. “Uncovering the Cytochrome P450-Catalyzed Methylenedioxy Bridge Formation in Streptovaricins Biosynthesis.” *Nature Communications* 11 (1): 4501. https://doi.org/10.1038/s41467-020-18336-5.

Tan, Qi, Limin Duan, YanLing Ma, Feng Wu, Qi Huang, Kaimin Mao, Wenjing Xiao, *et al.* 2020. “Is Oseltamivir Suitable for Fighting against COVID-19: In Silico Assessment, in Vitro and Retrospective Study.” *Bioorganic Chemistry* 104 (November): 104257. https://doi.org/10.1016/j.bioorg.2020.104257.

Tang, Lloyd Wei Tat, Ravi Kumar Verma, Ren Ping Yong, Xin Li, Lili Wang, Qingsong Lin, Hao Fan, and Eric Chun Yong Chan. 2021. “Differential Reversible and Irreversible Interactions between Benzbromarone and Human Cytochrome P450s 3A4 and 3A5.” *Molecular Pharmacology* 100 (3): 224–36. https://doi.org/10.1124/molpharm.121.000256

Tang, Peng-fei, Xiang Zheng, Xiao-xia Hu, Cheng-cheng Yang, Zhe Chen, Jian-chang Qian, Jian-ping Cai, and Guo-xin Hu. 2020. “Functional Measurement of CYP2C9 and CYP3A4 Allelic Polymorphism on Sildenafil Metabolism.” *Drug Design, Development and Therapy* Volume 14 (November): 5129–41. https://doi.org/10.2147/DDDT.S268796.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Tartaglia, Gian Gaetano, Andrea Cavalli, Riccardo Pellarin, and Amedeo Caflisch. 2004. “The Role of Aromaticity, Exposed Surface, and Dipole Moment in Determining Protein Aggregation Rates.” *Protein Science* 13 (7): 1939–41. https://doi.org/10.1110/ps.04663504

Taylor, Christopher, Ian Crosby, Vincent Yip, Peter Maguire, Munir Pirmohamed, and Richard M. Turner. 2020. “A Review of the Important Role of CYP2D6 in Pharmacogenomics.” *Genes* 11 (11): 1295. https://doi.org/10.3390/genes11111295.

Teama, Salwa. 2018. “DNA Polymorphisms: DNA-Based Molecular Markers and Their Application in Medicine.” In *Genetic Diversity and Disease Susceptibility*, edited by Yamin Liu. InTech. https://doi.org/10.5772/intechopen.79517.

Thapa, Bishnu, and Krishnan Raghavachari. 2019. “Energy Decomposition Analysis of Protein–Ligand Interactions Using Molecules-in-Molecules Fragmentation-Based Method.” *Journal of Chemical Information and Modeling* 59 (8): 3474–84. https://doi.org/10.1021/acs.jcim.9b00432

Thapa, Bishnu, Daniel Beckett, Jon Erickson, and Krishnan Raghavachari. 2018. “Theoretical Study of Protein–Ligand Interactions Using the Molecules-in-Molecules Fragmentation-Based Method.” *Journal of Chemical Theory and Computation* 14 (10): 5143–55. https://doi.org/10.1021/acs.jctc.8b00531

Thomaz, Douglas Vieira, Edson Silvio Batista Rodrigues, Fabio Bahls Machado, and Isaac Yves Lopes De Macedo. "Investigation of Cyclobenzaprine Interactions with P450 Cytochromes CYP1A2 and CYP3A4 through Molecular Docking Tools." Traektoriâ Nauki= Path of Science 5, no. 2 (2019): 4001-4006.doi.10.22178/post.43-1

Throngnumchai, Boonyalit, Sarocha Jitrakorn, Pakkakul Sangsuriya, Sasimanas Unajak, Pongsak Khunrae, Ha Thanh Dong, Vanvimon Saksmerprome, and Triwit Rattanarojpong. 2021. “Refolded Recombinant Major Capsid Protein (MCP) from

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Infectious Spleen and Kidney Necrosis Virus (ISKNV) Effectively Stimulates Serum Specific Antibody and Immune Related Genes Response in Nile Tilapia (Oreochromis Niloticus).” *Protein Expression and Purification* 184: 105876. https://doi.org/10.1016/j.pep.2021.105876

Tokunaga, Yuji, Thibault Viennet, Haribabu Arthanari, and Koh Takeuchi. 2020. “Spotlight on the Ballet of Proteins: The Structural Dynamic Properties of Proteins Illuminated by Solution NMR.” *International Journal of Molecular Sciences* 21 (5): 1829. https://doi.org/10.3390/ijms21051829.

Tornio, Aleksi, and Janne T. Backman. 2018. “Cytochrome P450 in Pharmacogenetics: An Update.” In *Advances in Pharmacology*, 83:3–32. Elsevier. https://doi.org/10.1016/bs.apha.2018.04.007.

Torres, Pedro H. M., Ana C. R. Sodero, Paula Jofily, and Floriano P. Silva-Jr. 2019. “Key Topics in Molecular Docking for Drug Design.” *International Journal of Molecular Sciences* 20 (18): 4574. https://doi.org/10.3390/ijms20184574.

Townshend, Raphael, Rishi Bedi, Patricia Suriana, and Ron Dror. 2019. “End-to-End Learning on 3d Protein Structure for Interface Prediction.” *Advances in Neural Information Processing Systems* 32.

Tramontano, Anna. 2017. “The Computational Prediction of Protein Assemblies.” *Current Opinion in Structural Biology* 46 (October): 170–75. https://doi.org/10.1016/j.sbi.2017.10.006.

Tverdohleb, Tatiana, Bora Dinc, Ivana Knezevic, Kenneth D. Candido, and Nebojsa Nick Knezevic. 2016. “The Role of Cytochrome P450 Pharmacogenomics in Chronic Non-Cancer Pain Patients.” *Expert Opinion on Drug Metabolism & Toxicology* 12 (11): 1303–11. https://doi.org/10.1080/17425255.2016.1209482.

Tyukhtenko, Sergiy, Girija Rajarshi, Ioannis Karageorgos, Nikolai Zvonok, Elyssia S. Gallagher, Hongwei Huang, Kiran Vemuri, *et al.* 2018. “Effects of Distal Mutations on the Structure, Dynamics and Catalysis of Human Monoacylglycerol

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Lipase." *Scientific Reports* 8 (January): 1719. https://doi.org/10.1038/s41598-017-19135-7.

Urban, Philippe, Thomas Lautier, Denis Pompon, and Gilles Truan. 2018. "Ligand Access Channels in Cytochrome P450 Enzymes: A Review." *International Journal of Molecular Sciences* 19 (6): 1617. https://doi.org/10.3390/ijms19061617.

Vatansever, Sezen, Avner Schlessinger, Daniel Wacker, H. Ümit Kaniskan, Jian Jin, Ming-Ming Zhou, and Bin Zhang. "Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: State-of-the-arts and future directions." Medicinal Research Reviews 41, no. 3 (2021): 1427-1473. https://doi.org/10.1002/med.21764

Verma, Anukriti, Vinay Kumar Singh, and Smriti Gaur. 2016. "Computational Based Functional Analysis of Bacillus Phytases." Computational Biology and Chemistry 60 (February): 53–58. https://doi.org/10.1016/j.compbiolchem.2015.11.001.

Vidal-Limon, Abraham, José E. Aguilar-Toalá, and Andrea M. Liceaga. 2022. "Integration of Molecular Docking Analysis and Molecular Dynamics Simulations for Studying Food Proteins and Bioactive Peptides." *Journal of Agricultural and Food Chemistry* 70 (4): 934–43. https://doi.org/10.1021/acs.jafc.1c06110

Vihinen, Mauno. 2021. "Functional Effects of Protein Variants." *Biochimie* 180 (January): 104–20. https://doi.org/10.1016/j.biochi.2020.10.009.

Wang, D, Y Guo, SA Wrighton, GE Cooke, and W Sadee. 2011. "Intronic Polymorphism in CYP3A4 Affects Hepatic Expression and Response to Statin Drugs." *The Pharmacogenomics Journal* 11 (4): 274–86. https://doi.org/10.1038/tpj.2010.28.

Wang, Ercheng, Weitao Fu, Dejun Jiang, Huiyong Sun, Junmei Wang, Xujun Zhang, Gaoqi Weng, Hui Liu, Peng Tao, and Tingjun Hou. 2021. "VAD-MM/GBSA: A Variable Atomic Dielectric MM/GBSA Model for Improved

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Accuracy in ProteiProtein-Ligandng Free Energy Calculations." *Journal of Chemical Information and Modeling* 61 (6): 2844–56. https://doi.org/10.1021/acs.jcim.1c00091

Wang, Xiang, Xiaoli Zhang, Xian-ping Dong, Mohammad Samie, Xinran Li, Xiping Cheng, Andrew Goschka, *et al.* 2012. "TPC Proteins Are Phosphoinositide-Activated Sodium-Selective Ion Channels in Endosomes and Lysosomes." Cell 151 (2): 372–83. https://doi.org/10.1016/j.cell.2012.08.036.

Wang, Xiaolei, Qing Liu, Wenjue Zhong, Liping Yang, Jing Yang, Adrian Covaci, and Lingyan Zhu. 2020. "Estimating Renal and Hepatic Clearance Rates of Organophosphate Esters in Humans: Impacts of Intrinsic Metabolism and Binding Affinity with Plasma Proteins." *Environment International* 134: 105321. https://doi.org/10.1016/j.envint.2019.105321

Wang, Ying, Qiu-Geng Ou-Yang, Wan-li Huang, Huan-le Huang, Xin-lei Zhuang, Qian-meng Lin, and Da-li Zeng. 2020. "Investigation of the Inhibitory Effect of Simvastatin on the Metabolism of Lidocaine Both in Vitro and in Vivo." Drug Design, Development and Therapy 14: 1739. https://doi.org/10.2147/dddt.s241022

UNIVERSITY of the
WESTERN CAPE

Wang, Ying, Qiu-Geng Ou-Yang, Wan-li Huang, Huan-le Huang, Xin-lei Zhuang, Qian-meng Lin, and Da-li Zeng. 2020. "Investigation of the Inhibitory Effect of Simvastatin on the Metabolism of Lidocaine Both in Vitro and in Vivo." *Drug Design, Development and Therapy* 14: 1739. https://doi.org/10.2147/dddt.s241022

Waterhouse, Andrew, Martino Bertoni, Stefan Bienert, Gabriel Studer, Gerardo Tauriello, Rafal Gumienny, Florian T. Heer, *et al.* 2018. "SWISS-MODEL: Homology Modelling of Protein Structures and Complexes." *Nucleic Acids Research* 46 (W1): W296–303. https://doi.org/10.1093/nar/gky427.

Webster, Kyria M., Mujun Sun, Peter J. Crack, Terence J. O'Brien, Sandy R. Shultz, and Bridgette D. Semple. 2019. "Age-Dependent Release of High-Mobility Group Box Protein-1 and Cellular Neuroinflammation after Traumatic Brain Injury

http://etd.uwc.ac.za/

in Mice." *Journal of Comparative Neurology* 527 (6): 1102–17. https://doi.org/10.1002/cne.24589.

Wei, Guo-Wei. "Protein structure prediction beyond AlphaFold." Nature Machine Intelligence 1, no. 8 (2019): 336-337. http://dx.doi.org/10.1038/s42256-019-0086-4

Williams, P. A. 2004. "Crystal Structures of Human Cytochrome P450 3A4 Bound to Metyrapone and Progesterone." *Science* 305 (5684): 683–86. https://doi.org/10.1126/science.1099736.

Wright, William C., Jude Chenge, and Taosheng Chen. 2019. "Structural Perspectives of the CYP3A Family and Their Small Molecule Modulators in Drug Metabolism." *Liver Research* 3 (3): 132–42. https://doi.org/10.1016/j.livres.2019.08.001.

Xiao, Fei, Xingyu Song, Peiyi Tian, Mi Gan, Gennady M. Verkhivker, and Guang Hu. 2020. "Comparative Dynamics and Functional Mechanisms of the CYP17A1 Tunnels Regulated by Ligand Binding." *Journal of Chemical Information and Modeling* 60 (7): 3632–47. https://doi.org/10.1021/acs.jcim.0c00447

Xu, Liang, and Liao Y. Chen. 2020. "Molecular Determinant of Substrate Binding and Specificity of Cytochrome P450 2J2." *Scientific Reports* 10 (1): 22267. https://doi.org/10.1038/s41598-020-79284-0.

Xue, Yuhan, Junhao Li, Zengrui Wu, Guixia Liu, Yun Tang, and Weihua Li. 2019. "Computational Insights into the Different Catalytic Activities of CYP 3A4 and CYP 3A5 toward Schisantherin E." *Chemical Biology & Drug Design* 93 (5): 854–64. https://doi.org/10.1111/cbdd.13475

Yang, Jinhua, and Junfeng Zhao. 2018. "Recent Developments in Peptide Ligation Independent of Amino Acid Side-Chain Functional Group." *Science China Chemistry* 61 (1): 97–112. https://doi.org/10.1007/s11426-017-9056-5.

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

Yang, Li-Quan, Peng Sang, Yan Tao, Yun-Xin Fu, Ke-Qin Zhang, Yue-Hui Xie, and Shu-Qun Liu. 2014. “Protein Dynamics and Motions in Relation to Their Functions: Several Case Studies and the Underlying Mechanisms.” *Journal of Biomolecular Structure and Dynamics* 32 (3): 372–93. https://doi.org/10.1080/07391102.2013.770372.

Yano, Jason K., Michael R. Wester, Guillaume A. Schoch, Keith J. Griffin, C. David Stout, and Eric F. Johnson. 2004. “The Structure of Human Microsomal Cytochrome P450 3A4 Determined by X-Ray Crystallography to 2.05-Å Resolution: Fig. 1.” *Journal of Biological Chemistry* 279 (37): 38091–94. https://doi.org/10.1074/jbc.C400293200.

Yao, Huan, Hua Ke, Xiaobin Zhang, San-Jiang Pan, Ming-Shuang Li, Liu-Pan Yang, Georg Schreckenbach, and Wei Jiang. 2018. “Molecular Recognition of Hydrophilic Molecules in Water by Combining the Hydrophobic Effect with Hydrogen Bonding.” *Journal of the American Chemical Society* 140 (41): 13466–77. https://doi.org/10.1021/jacs.8b09157.

Yarima, A. A., S. M. Sambo, M. D. Kwairanga, K. N. Sharbat, Z. Arifullah, and A. Y. Fardami. 2020. “Structural Characterization of Novel Luciferase from Bioluminescence Fungi Verticillium Longisporum.” *Biotechnology Journal International*, September, 49–73. https://doi.org/10.9734/bji/2020/v24i530117.

Yee, Adelinda A., Anthony Semesi, Maite Garcia, and Cheryl H. Arrowsmith. 2014. “Screening Proteins for NMR Suitability.” *Methods in Molecular Biology (Clifton, N.J.)* 1140: 169–78. https://doi.org/10.1007/978-1-4939-0354-2_13.

Yenenler, Aslı, Umut Gerlevik, and Ugur Sezerman. 2020. “Understanding the Impacts of Self-Shuffling Approach on Structure and Function of Shuffled Endoglucanase Enzyme via MD Simulations.” *Turkish Journal of Biochemistry* 45 (2). https://doi.org/10.1515/tjb-2018-0180.

http://etd.uwc.ac.za/

Zanger, Ulrich M., and Matthias Schwab. 2013. “Cytochrome P450 Enzymes in Drug Metabolism: Regulation of Gene Expression, Enzyme Activities, and Impact of Genetic Variation.” *Pharmacology & Therapeutics* 138 (1): 103–41. https://doi.org/10.1016/j.pharmthera.2012.12.007.

Zhang, Qiangfeng Cliff, Donald Petrey, Raquel Norel, and Barry H. Honig. 2010. “Protein Interface Conservation across Structure Space.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (24): 10896–901. https://doi.org/10.1073/pnas.1005894107.

Zhang, Rui Xue, Ken Dong, Zhigao Wang, Ruimin Miao, Weijia Lu, and Xiao Yu Wu. 2021. “Nanoparticulate Drug Delivery Strategies to Address Intestinal Cytochrome P450 CYP3A4 Metabolism towards Personalized Medicine.” *Pharmaceutics* 13 (8): 1261. https://doi.org/10.3390/pharmaceutics13081261.

Zhang, Yilue. 2019. “CYP3A4 Inhibition and Induction Studies Coupled to Parallel Artificial Membrane Permeability Assay (PAMPA) for Improved Prediction of in-Vitro Botanical-Drug Interactions,” July. https://etd.auburn.edu//handle/10415/6784.

Zhang, Zhoupeng, and Wei Tang. 2018. “Drug Metabolism in Drug Discovery and Development.” *Acta Pharmaceutica Sinica B* 8 (5): 721–32. https://doi.org/10.1016/j.apsb.2018.04.003.

Zhao, Hui, Senyao Xue, Qingzhen Meng, and Cui Zhou. 2021. “*In Vitro* Study on the Effect of Leonurine Hydrochloride on the Enzyme Activity of Cytochrome P450 Enzymes in Human Liver Microsomes.” *Xenobiotica* 51 (9): 977–82. https://doi.org/10.1080/00498254.2021.1947544.

Zhao, Yonghong, Mark A. White, B. K. Muralidhara, Ling Sun, James R. Halpert, and C. David Stout. 2006. “Structure of Microsomal Cytochrome P450 2B4 Complexed with the Antifungal Drug Bifonazole: Insight into P450

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Conformational Plasticity and Membrane Interaction." *The Journal of Biological Chemistry* 281 (9): 5973–81. https://doi.org/10.1074/jbc.M511464200.

Zhou, Diansong, Lovisa Afzelius, Scott W. Grimm, Tommy B. Andersson, Randy J. Zauhar, and Ismael Zamora. "Comparison of methods for the prediction of the metabolic sites for CYP3A4-mediated metabolic reactions." Drug metabolism and disposition 34, no. 6 (2006): 976-983. https://doi.org/10.1124/dmd.105.008631

Zhou, Shu-Feng. 2008. "Drugs Behave as Substrates, Inhibitors and Inducers of Human Cytochrome P450 3A4." *Current Drug Metabolism* 9 (4): 310–22. https://doi.org/10.2174/138920008784220664.

Zhou, Xiao-Yang, Xiao-Xia Hu, Chen-Chen Wang, Xiang-Ran Lu, Zhe Chen, Qian Liu, Guo-Xin Hu, and Jian-Ping Cai. 2019a. "Enzymatic Activities of CYP3A4 Allelic Variants on Quinine 3-Hydroxylation in Vitro." *Frontiers in Pharmacology* 10: 591. https://doi.org/10.3389/fphar.2019.00591

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

Appendix

A. The physicochemical and general protein properties of the selected CYP3A4 variants and the wild type.

The physicochemical properties of each of the selected mutated sequences of CYP3A4 and the wild-type sequence investigated in this study are shown in Table A.0.1.

Table A.0.1: The physicochemical and general protein properties of selected CYP3A4 variants and the wild type**.**

Protein Properties (Relative clearance)	CYP3A4*1 **(100%)**	CYP3A4*2 **(27.93%)**	CYP3A4*11 **(206.96%)**	CYP3A4*23 **(213.61%)**	CYP3A4*24 **(30.29%)**
Molecular Weight*(Da)	**57343.18**	**57353.22**	**57373.27**	**57373.20**	**57352.19**
Amino Acid(aa)	503	503	503	503	503
Isoelectric Point (pH)	8.27	8.27	8.27	8.00	8.27
Instability Index (Ii)	41.19	41.57	41.19	41.19	40.42
Aliphatic Index (Ai)	95.47	95.47	95.47	95.47	95.47
Gravy*	**-0.040**	**-0.042**	**-0.035**	**-0.033**	**-0.039**
Debye Screening Length(k)	9.62	9.62	9.62	9.62	9.62
Sequence-based Isoelectric point	7.62	7.62	7.62	7.40	7.64

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Prediction(pH)					
Structure-based pI Prediction(pH)	8.46	8.47	8.46	8.22	8.46
Extinction Coefficient (EC)M−1 cm−1)	46215	46215	46215	51715	46215
Frictional Coefficient(kg/s)	53e-011	5.3e-011	5.3e_011	5.3e_-011	5.3e_-011
Diffusion Coefficient(cm^2/s)	7.8e-007	7.8e-007	7.8e-007	7.8e-007	7.8e-007
Radius of gyration(Å)	22.76	22.76	22.76	22.77	22.76
Hydrodynamic radius(Å)	31.84	31.84	31.84	31.81	31.80
Sedimentation constant(s)	4.2e-013	4.2e-013	4.2e_013	4.2e_013	4.2e_013
Protein Eccentricity	0.74	0.74	0.74	0.74	0.74
Accessible surface Area (Water probe) *($Å^{2)}$	**20369.7**	**20407.5**	**20375.3**	**20419.1**	**20353.5**
Hydrophobic Surface Area* $Å^2$	**11941.1**	**11973.5**	**11925.1**	**12037.4**	**11944.4**

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Hydrophilic surface Area* $Å^2$	**7448.2**	**7448.2**	**74616.6**	**7433.2**	**7386.7**
Protein volume* $Å^3$	**50087.6**	**50033.6**	**50083.6**	**50113.4**	**50098.0**
Protein mobility* (cm^2/Vs)	**4.3**	**4.4**	**4.3**	**3.5**	**4.4**
Protein helix ratio (%)	51.9	51.5	51.9	51.9	51.9
Henry's Function f(ka)	1.11	1.11	1.11	1.11	1.11
Protein Net charge*	**5.56**	**5.63**	**5.57**	**4.60**	**5.66**
Sum positive surface area ($Å^2$) *	**12080.30**	**12090.82**	**12095.13**	**12254.23**	**12337.73**
Sum negative surface area ($Å^2$) *	**6996.98**	**6860.53**	**6858.70**	**6591.23**	**6606.73**
Sum donor surface area ($Å^2$) *	**3379.18**	**3296.91**	**3313.10**	**3134.33**	**3146.54**
Sum acceptor surface area ($Å^2$) *	**4015.71**	**4048.43**	**4062.33**	**3973.62**	**4001.24**
Protein Charge at Debye Length*	**1.43**	**1.45**	**1.44**	**1.18**	**1.46**

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za/

Sum of Aggregation score*	**630.13**	**621.64**	**611.90**	**687.59**	**687.17**
Debye screening length K^{-1}	**1056.67**	**1039.69**	**1056.96**	**1152.37**	**1048.99**
Hydrophobicity Moment*	**2989.8**	**2070.2**	**2069.9**	**2038.0**	**2089.3**
Zeta potential at Debye Length*(mV)	**7.48**	**7.58**	**7.49**	**7.62**	**7.62**
Zeta Dipole Moment(mV)	0.00	0.00	0.00	0.00	0.00
Zeta Quadrupole Moment(mV)	0.54	0.54	0.54	0.51	0.51

*-parameters with notable differences that are subjected to further correlation study. Properties with notable differences are written in coloured fonts and indicated with asterisks.

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za/

B. The Correlation plotted graph

The correlation graph from figure B.0.1 to B.0.33 shows the correlation between the reported functional difference (relative clearance) and the differences observed in the data obtained from the characterization investigation in Table 4.16.

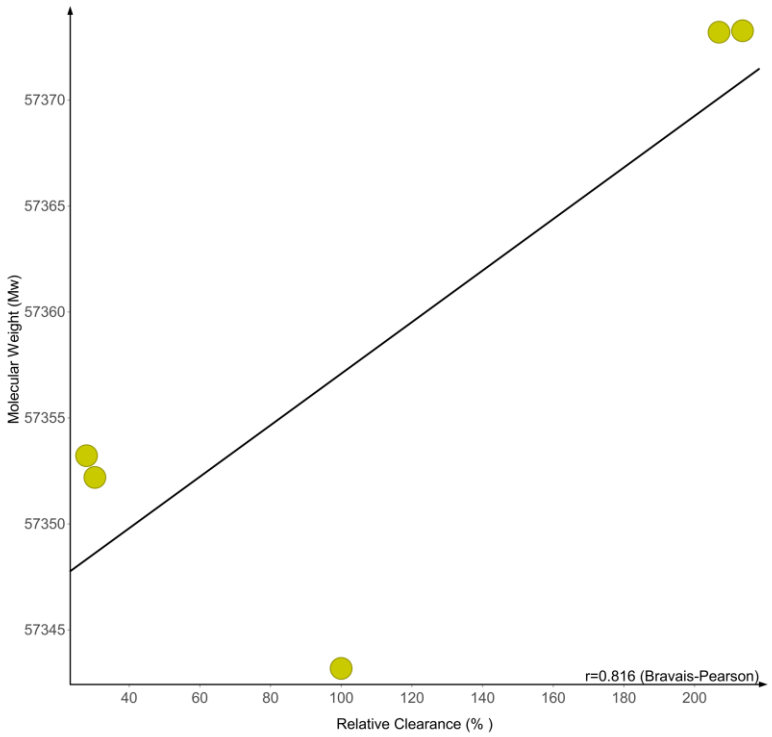


Figure B.0.1: The correlation plot of Molecular weight (Da) against the relative clearance (%)

The green dots represent the molecular weight of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as “r”

http://etd.uwc.ac.za/

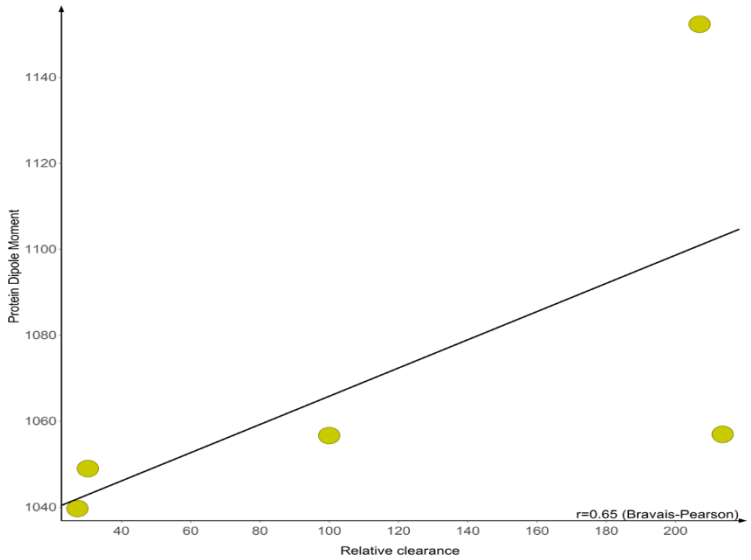


Figure B.0.2: The correlation plot of protein dipole moment (Debye) against the relative clearance (%)

The green dots represent the protein dipole moment of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

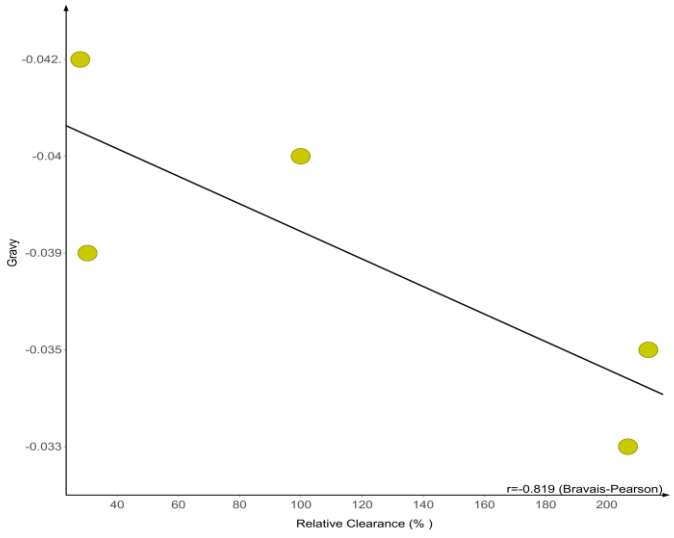


Figure B.0.3: The correlation plot of Gravy against the relative clearance (%)

The green dots represent the Gravy of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

http://etd.uwc.ac.za/

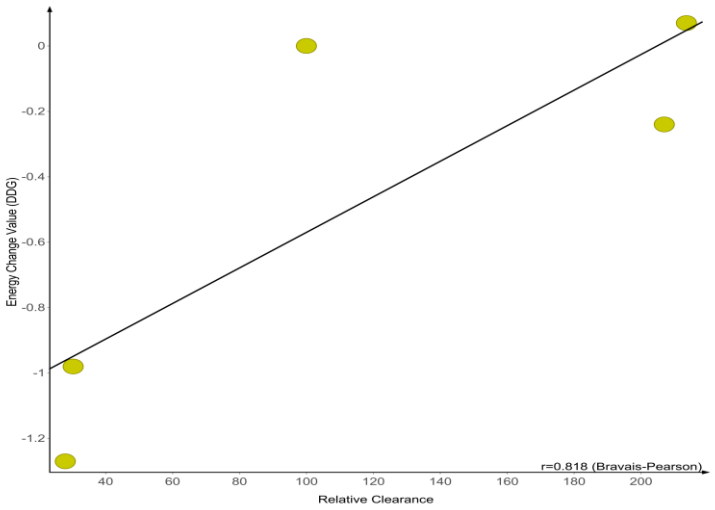


Figure B.0.4: The correlation plot of energy change value (Kcal/mol) against the relative clearance (%)

The green dots represent the energy change value of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravais-Pearson correlation value is assigned as "r"

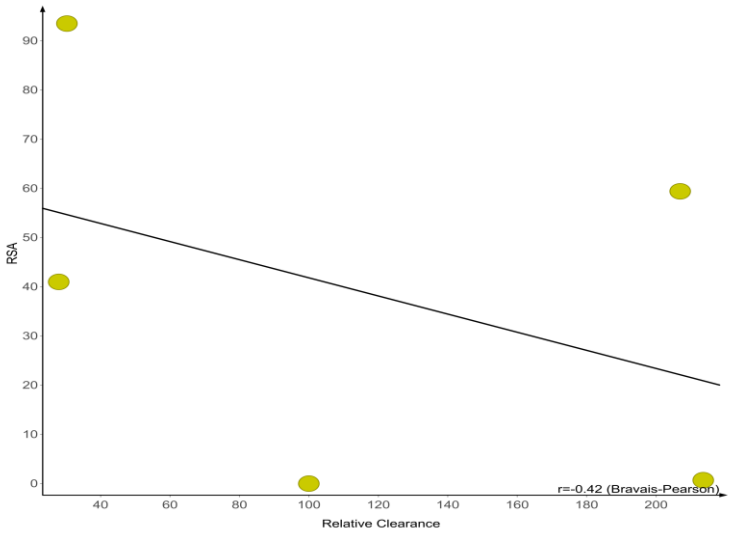


Figure B.0.5: The correlation plot of relative solvent area ($Å^2$) against the intrinsic relative clearance (%)

The green dots represent the relative solvent area of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

http://etd.uwc.ac.za/

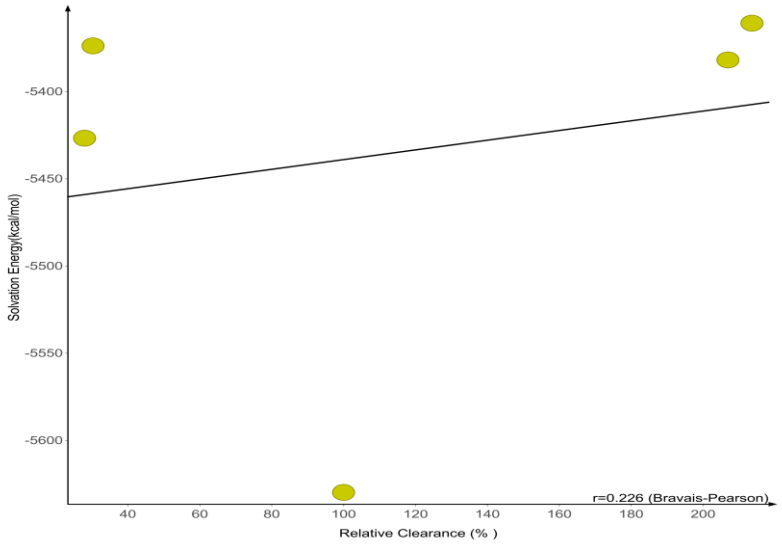


Figure B.0.6: The correlation plot of solvation energy (Kcal/mol) against the relative clearance (%)

The green dots represent the solvation energy of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

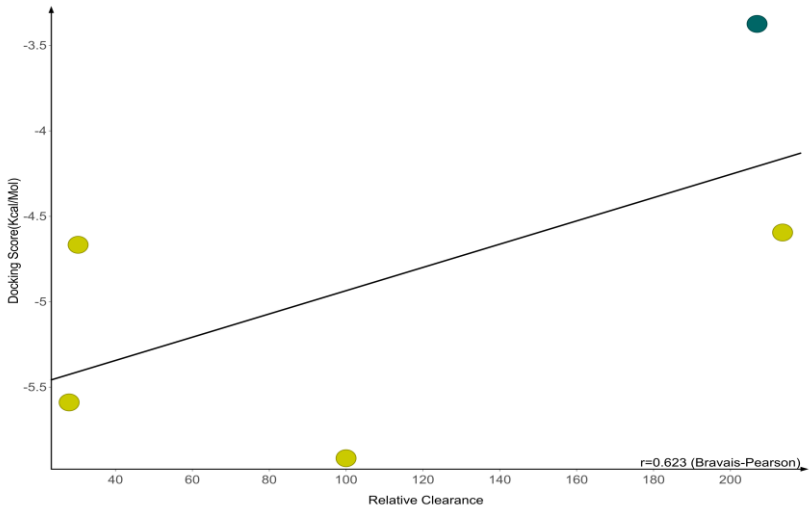


Figure B.0.7: The correlation plot of docking score (Kcal/mol) against the relative clearance (%)

The green dots represent the docking score of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

http://etd.uwc.ac.za/

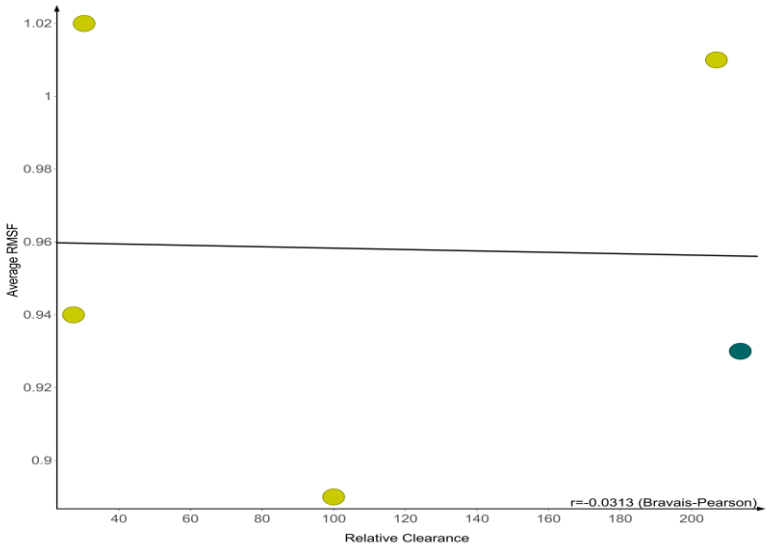


Figure B.0.8: The correlation plot of average root mean square fluctuation (Å) against the relative clearance (%)

The green dots represent the average root mean square fluctuation of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

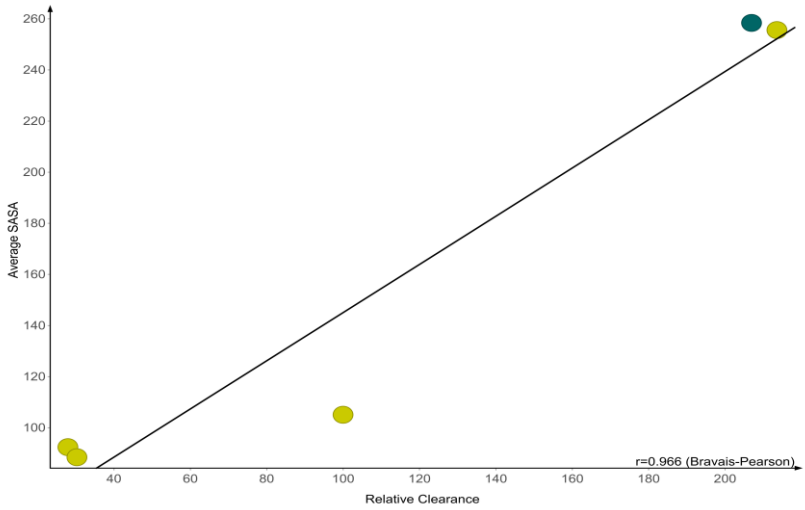


Figure B.0.9: The correlation plot of solvent accessible surface area (Å^2) against the relative clearance (%)

The green dots represent the solvent accessible surface area of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

http://etd.uwc.ac.za/

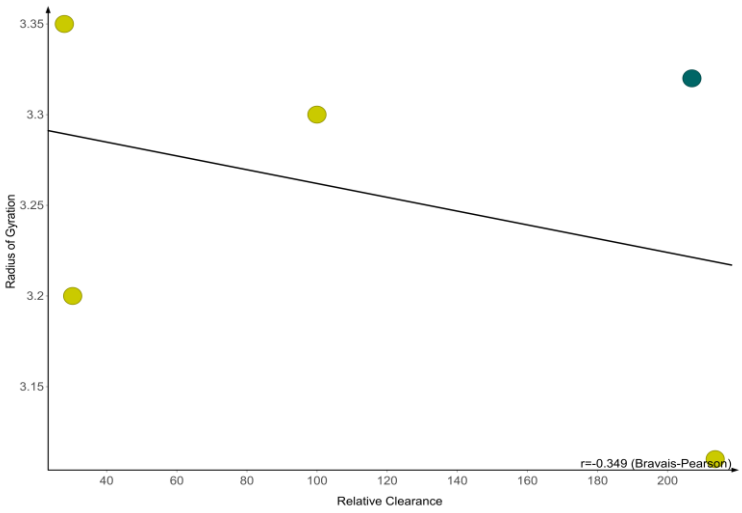


Figure B.0.10: The correlation plot of the radius of gyration (Å) against the relative clearance (%)

The green dots represent the radius of gyration of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

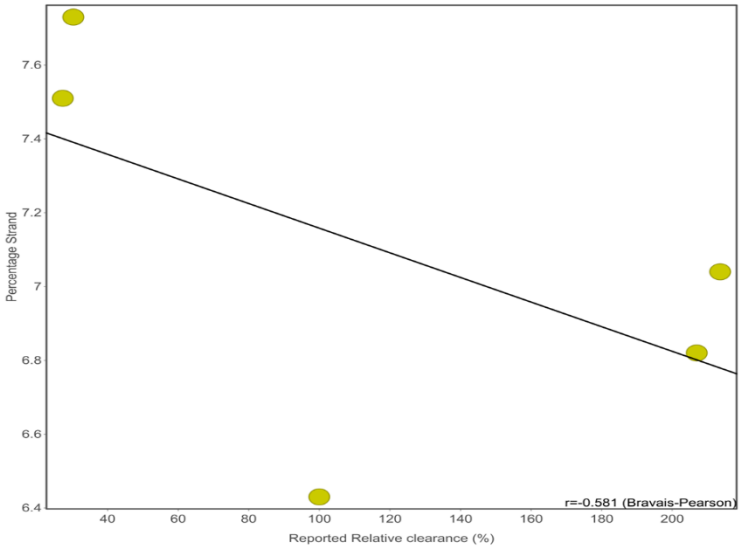


Figure B.0.11: The correlation plot of percentage strand (%) against the relative clearance (%)

The green dots represent the percentage strand of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

http://etd.uwc.ac.za/

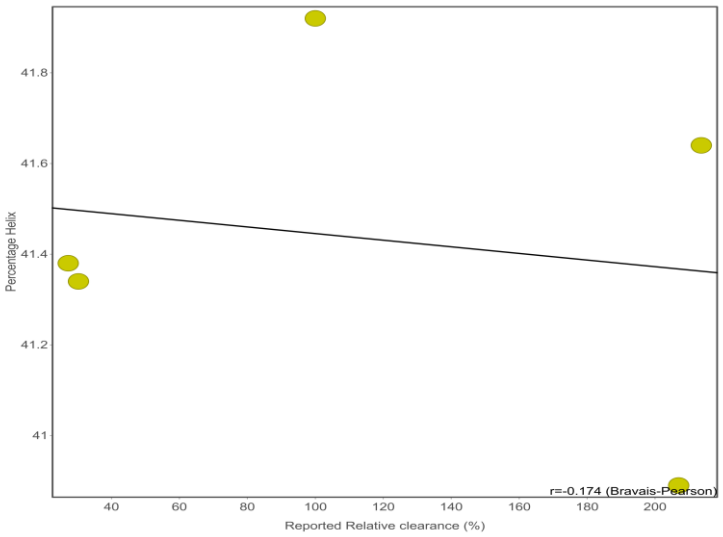


Figure B.0.12: The correlation plot of percentage helix (%) against the relative clearance (%)

The green dots represent the percentage helix of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

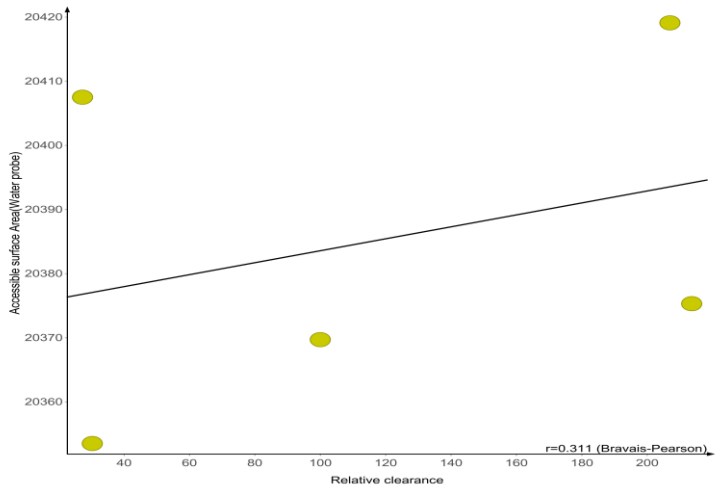


Figure B.0.13: The correlation plot of accessible surface area (water probe) ($Å^2$) against the relative clearance (%)

The green dots represent the accessible surface area (water probe) of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

http://etd.uwc.ac.za/

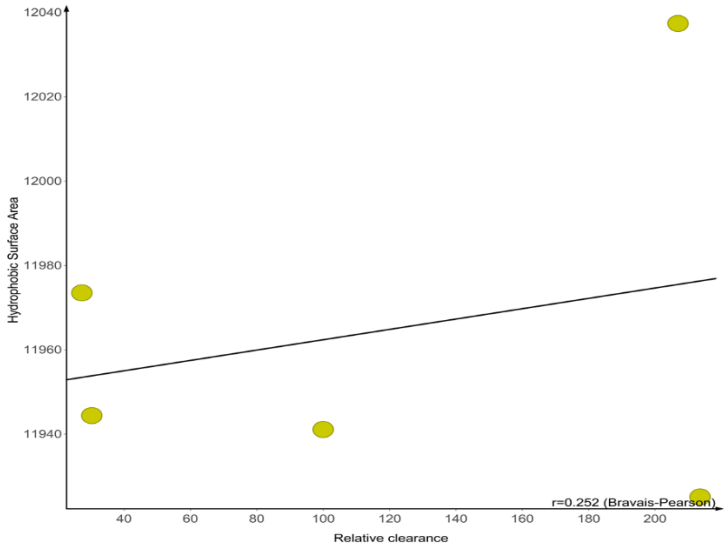


Figure B.0.14: The correlation plot of the hydrophobic surface area (Å^2) against the relative clearance (%)

The green dots represent the hydrophobic surface area of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

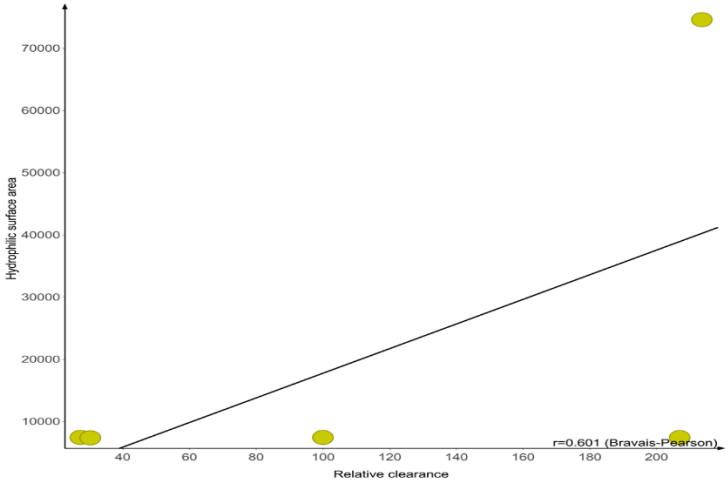


Figure B.0.15: The correlation plot of the hydrophilic surface area (Å^2) against the relative clearance (%)

The green dots represent the hydrophilic surface area of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

http://etd.uwc.ac.za/

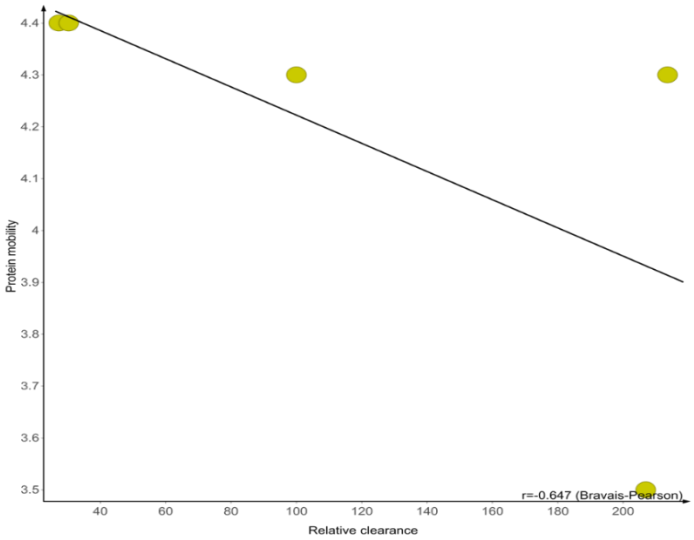


Figure B.0.16: The correlation plot of protein mobility (cm^2/vs) against the relative clearance (%)

The green dots represent the protein mobility of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

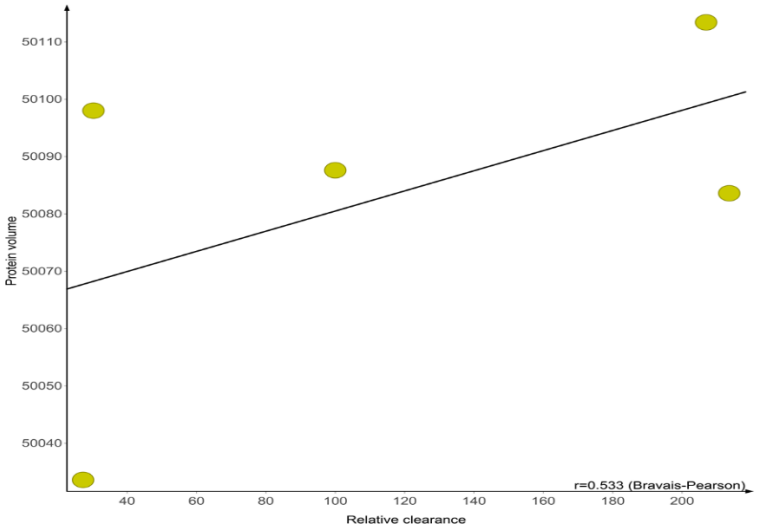


Figure B.0.17: The correlation plot of protein volume against the relative clearance (%)

The green dots represent the protein volume of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

http://etd.uwc.ac.za/

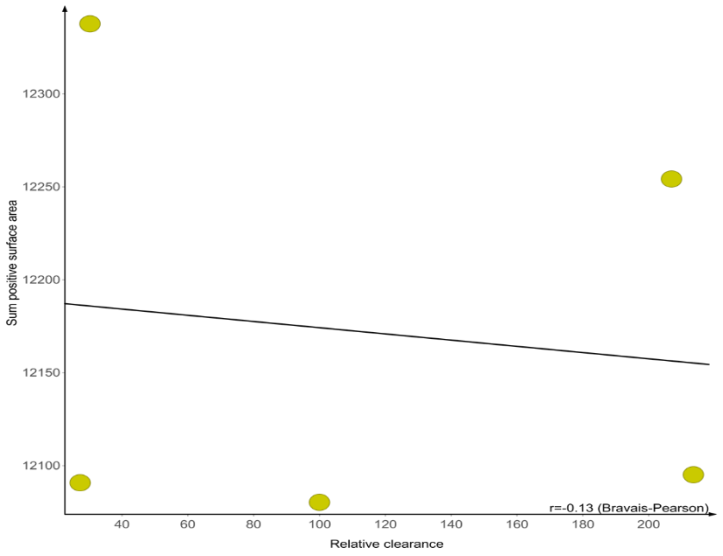


Figure B.0.18: The correlation plot of sum positive surface area ($Å^2$) against the relative clearance (%)

The green dots represent the sum positive surface area of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

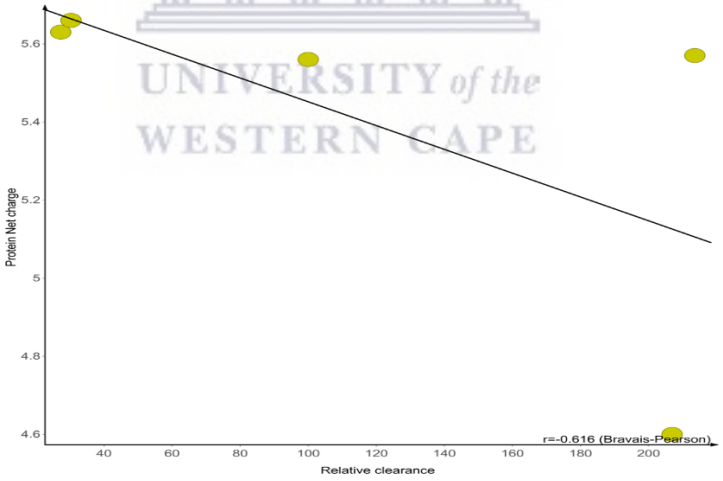


Figure B.0.19: The correlation plot of protein net charge against the intrinsic relative clearance (%)

The green dots represent the protein net charge of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

http://etd.uwc.ac.za/

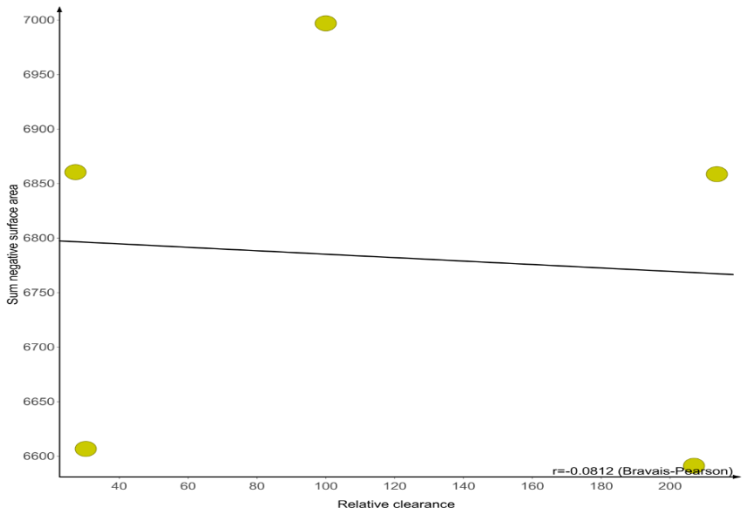


Figure B.0.20: The correlation plot of sum negative surface area (Å^2) against the relative clearance (%)

The green dots represent the sum negative surface area of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

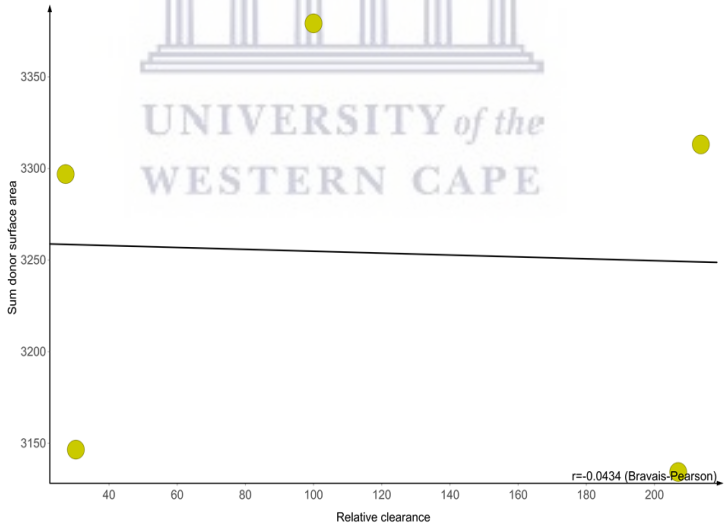


Figure B.0.21: The correlation plot of sum donors surface area (Å^2) against the relative clearance (%)

The green dots represent the sum donors' surface area of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

UNIVERSITY of the WESTERN CAPE
http://etd.uwc.ac.za/

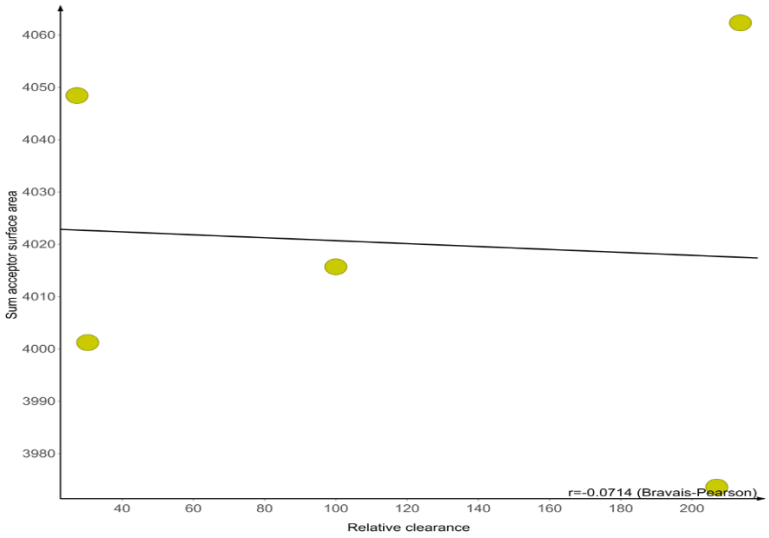


Figure B.0.22: The correlation plot of sum acceptor surface area(Å^2) against the relative clearance (%)

The green dots represent the sum acceptor surface area of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

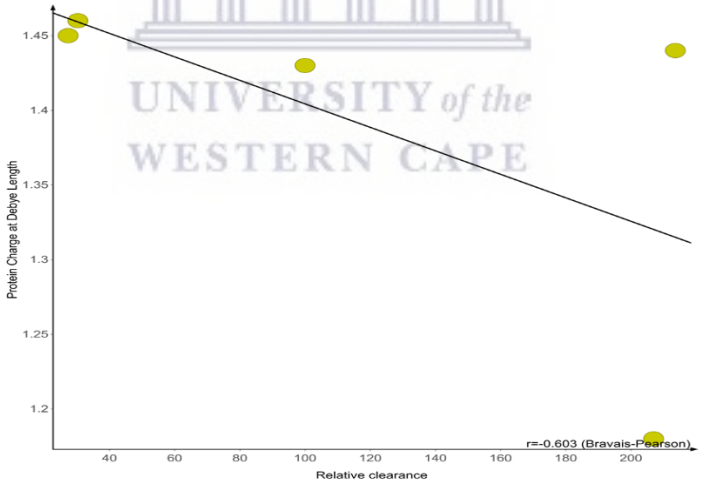


Figure B.0.23: The correlation plot of protein charge at Debye length against the relative clearance (%)

The green dots represent the protein charge at the Debye length of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

http://etd.uwc.ac.za/

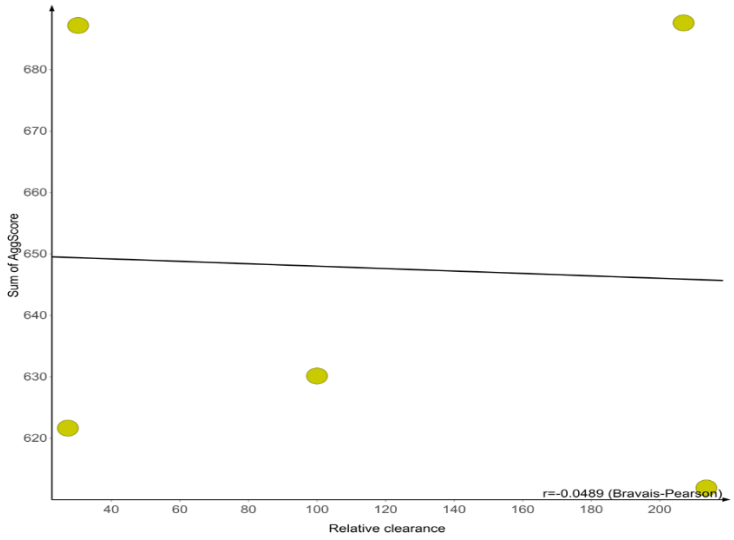


Figure B.0.24: The correlation plot of the sum of aggregation score against the relative clearance (%)

The green dots represent the sum of the aggrescore of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

UNIVERSITY of the
WESTERN CAPE

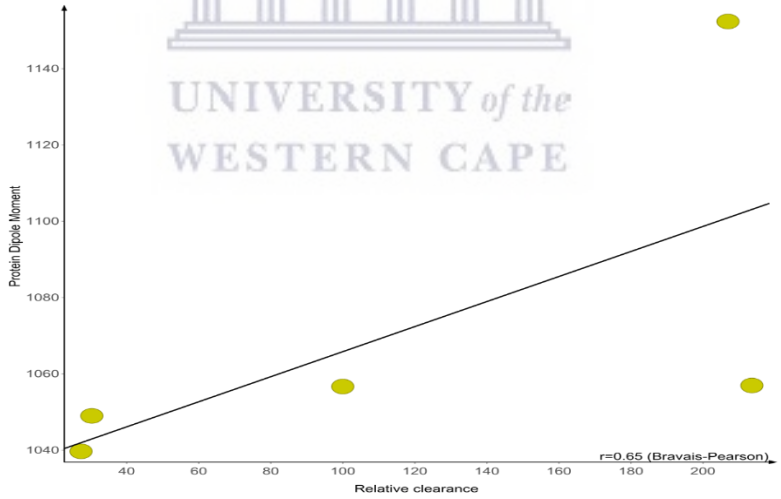


Figure B.0.25: The correlation plot of protein dipole (Debye) moment against the intrinsic relative clearance (%)

The green dots represent the protein dipole moment of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

http://etd.uwc.ac.za/

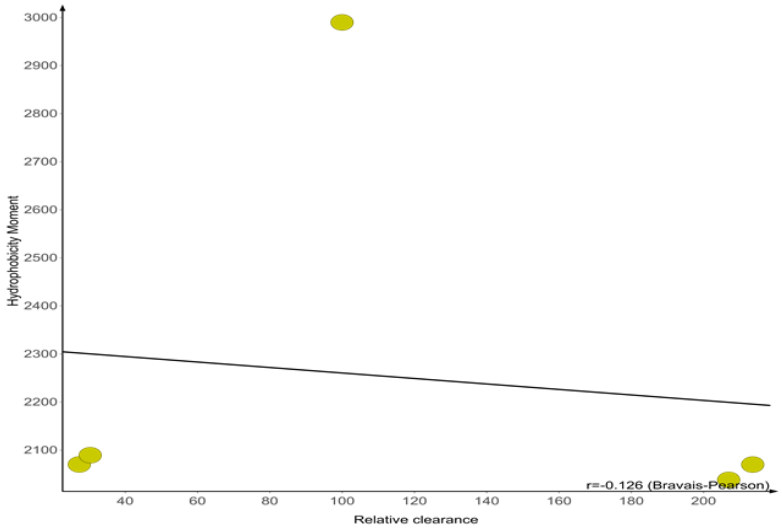


Figure B.0.26: The correlation plot of hydrophobicity moment against the relative clearance (%)

The green dots represent the hydrophobicity moment of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as “r”

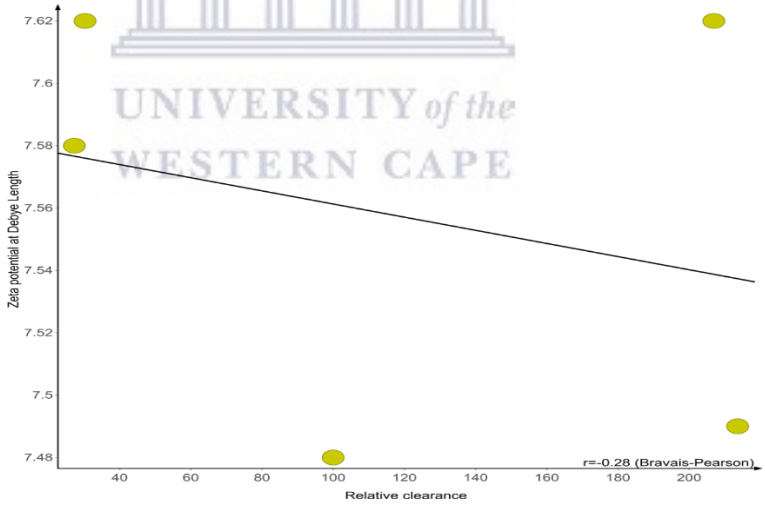


Figure B.0.27: The correlation plot of zeta potential at Debye length (mV) against the relative clearance (%)

The green dots represent the sum of Zeta potential at Debye length of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as “r”

http://etd.uwc.ac.za/

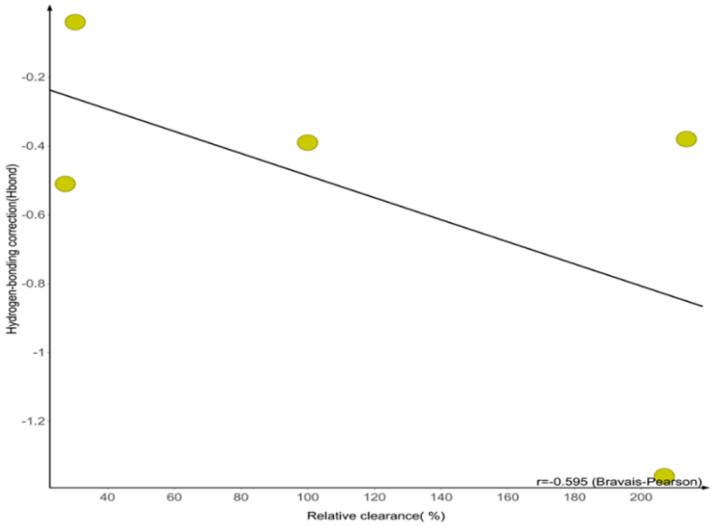


Figure B.0.28: The correlation plot of hydrogen bond correction (Kcal/mol)) against the relative clearance (%)

The green dots represent the hydrogen bond of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

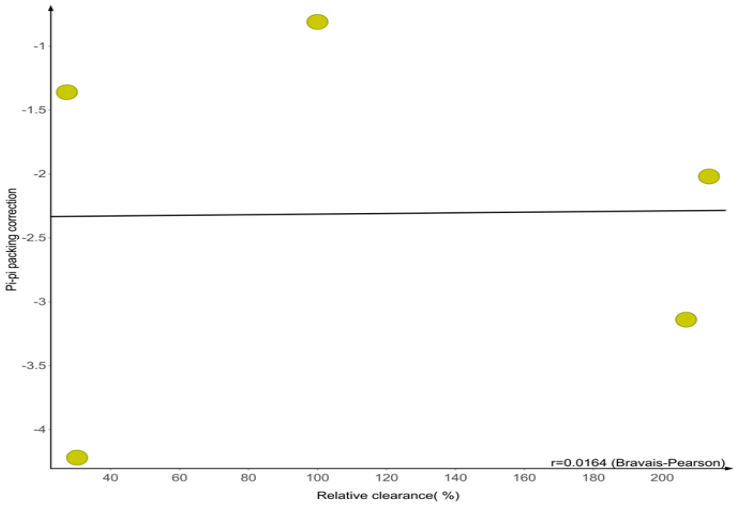


Figure B. 0.29: The correlation plot of pi-pi packing correction against the relative clearance (%)

The green dots represent the pi-pi packing of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

http://etd.uwc.ac.za/

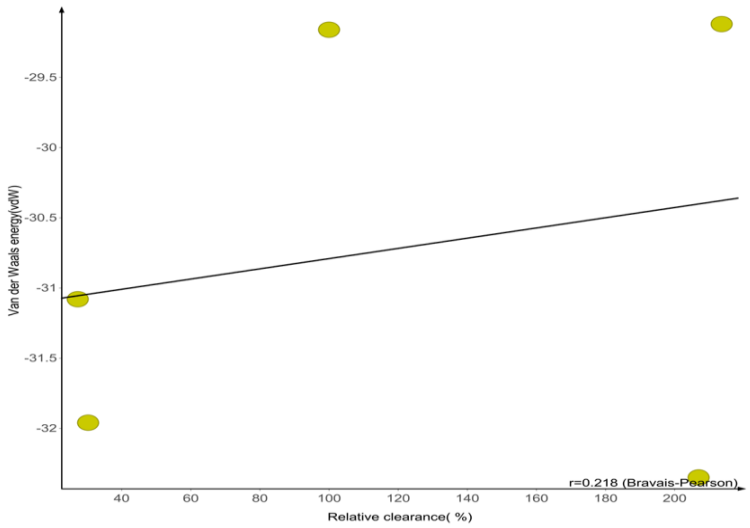


Figure B.0.30: The correlation plot of Van der Waals energy (Kcal/mol) against the relative clearance (%)

The green dots represent the Van der Waals energy of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

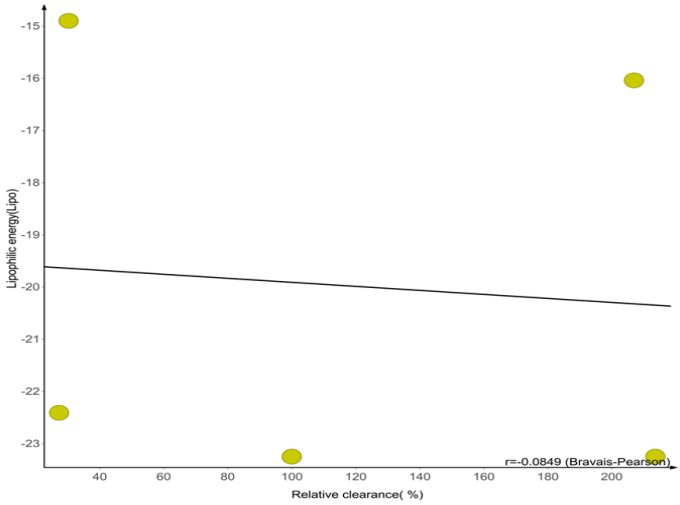


Figure B.0.31: The correlation plot of lipophilic energy (Kcal/mol) against the relative clearance (%)

The green dots represent the lipophilic energy of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

http://etd.uwc.ac.za/

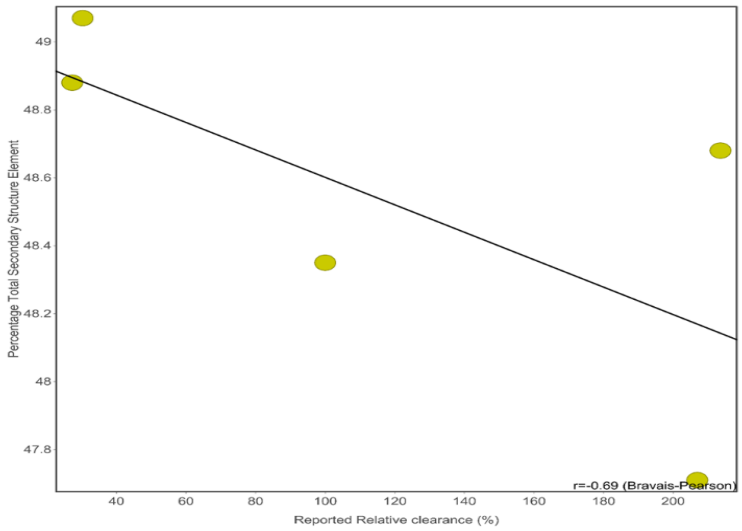


Figure B.0.32: The correlation plot of percentage total secondary structure element (%) against the relative clearance (%)

The green dots represent the percentage of total secondary structure elements of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

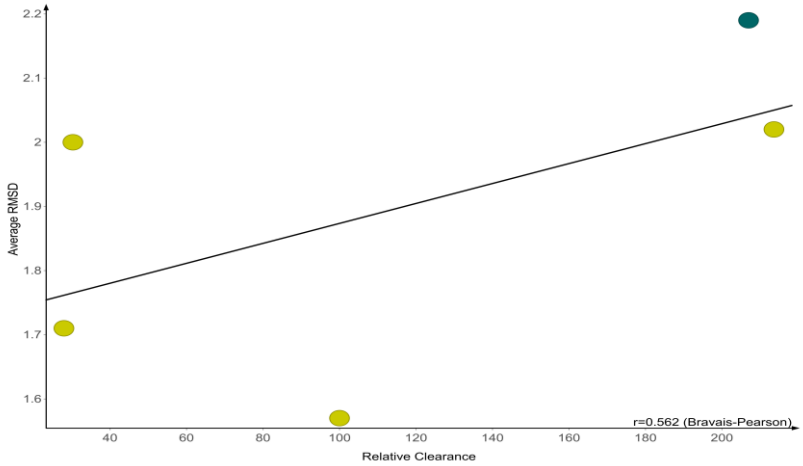


Figure B.0.33: The correlation plot of average RMSD (Å) against the relative clearance (%)

The green dots represent the average RMSD of the selected CYP3A4 variants/wildtype against the reported relative clearance. The Bravis-Pearson correlation value is assigned as "r"

http://etd.uwc.ac.za/