



**COMPUTATIONAL AND MICRO-ANALYTICAL
TECHNIQUES TO STUDY THE *IN VITRO* AND *IN SILICO*
MODELS OF NOVEL THERAPEUTIC DRUGS**

By

Njabulo Joyfull Gumede

(20204180)

Submitted in fulfilment of the requirements of the Doctor of Philosophy degree in Chemistry
in the Faculty of Applied Sciences at the Durban University of Technology

Declaration

I **Njabulo Joyfull Gumede** declare that the thesis submitted for the Doctor of Philosophy degree at the Durban University of Technology has not been submitted to any other University and that its only prior publication was in the form of conference papers, journal articles and the registration of a provisional patent.

Student Name: Mr. Njabulo J. Gumede

Student Signature:..... **Date:**.....

Supervisor Name: Prof. K. Bisetty

Signature:..... **Date:**.....

Co-Supervisor Name: Prof. S. Sagrado. Universidad de Valencia (Spain)

Signature:..... **Date:**

ACKNOWLEDGEMENTS

I would like start by extending my appreciation to Professors K. Bisetty (Durban University of Technology, DUT), and S. Sagrado (Universidad de Valencia, UV) for their support during the course of this journey and the entire UV group for their help during my study visits to Valencia, Spain. I would like to thank the Research and Technology Transfer Directorate at Mangosuthu University of Technology for financial support of the registered project (NSI/02/2010). I would like to further thank the National Research Foundation for financial support through the Thuthuka grant (UID80659). Not forgetting to thank the Technology Innovation Agency (TIA) for the 2015-2016 seed funding of cancer therapeutics project.

This has been a long and a hard journey for me to travel for the past six years. I have learned to be strong even though things are not going my way. There are two incidents that disturbed my progress in this PhD journey which are my car hi-jacking and the burglary in my house. Where all my computers and other gadgets with some results of my PhD work were stolen.

I would like to give thanks to all my peers from DUT & UV who have assisted me. I would also like to dedicate this thesis to my mother who have been diagnosed with meningitis and is now homebound. Lastly, this thesis is dedicated to everyone who is infected or affected by all forms of cancer in any way. I say to you stay strong believe in GOD the almighty, all things are possible with him.

Abstract

In drug discovery and development projects, metabolism of new chemical entities (NCEs) is a major contributing factor for the withdrawal of drug candidates, a major concern for other chemical industries where chemical-biological interactions are involved. NCEs interact with a target macro-molecule to stimulate a pharmacological or toxic response, known as pharmacodynamics (PD) effect or through the Adsorption, Distribution, Metabolism, and Excretion (ADME) process, triggered when a bio-macromolecule interacts with a therapeutic drug. Therefore, the drug discovery process is important because 75% of diseases known to human kind are not all cured by therapeutics currently available in the market. This is attributed to the lack of knowledge of the function of targets and their therapeutic use in order to design therapeutics that would trigger their pharmacological responses.

Accordingly, the focus of this work is to develop cost saving strategies for medicinal chemists involved with drug discovery projects. Therefore, studying the synergy between *in silico* and *in vitro* approaches maybe useful in the discovery of novel therapeutic compounds and their biological activities. In this work, *in silico* methods such as structure-based and ligand-based approaches were used in the design of the pharmacophore model, database screening and flexible docking methods. Specifically, this work is presented by the following case studies:

The first involved molecular docking studies to predict the binding modes of catechin enantiomer to human serum albumin (HSA) interaction; the second involved the use of docking methods to predict the binding affinities and enantioselectivity of the interaction of warfarin enantiomers to HSA. the third case study involved a combined computational strategy in order to generate information on a diverse set of steroidal and non-steroidal CYP17A1 inhibitors obtained from literature with known experimental IC_{50} values. Finally, the fourth case study involved the prediction of the site of metabolisms (SOMs) of probe substrates to Cytochrome P450 metabolic enzymes CYP 3A4, 2D6, and 2C9 making use of P450 module from Schrödinger suite for ADME/Tox prediction.

The results of case study I were promising as they were able to provide clues to the factors that drive the synergy between experimental kinetic parameters and computational thermodynamics parameters to explain the interaction between drug enantiomers and the

target protein. These parameters were correlated/converted and used to estimate the pseudo enantioselectivity of catechin enantiomer to HSA. This approach of combining docking methodology with docking post-processing methods such as MM-GBSA proved to be vital in estimating the correct pseudo binding affinities of a protein-ligand complexes. The enantioselectivity for enantiomers of catechin to HSA were 1,60 and 1,25 for site I and site II respectively.

The results of case study II validates and verifies the preparation of ligands and accounting for tautomers at physiological pH, as well as conformational changes prior to and during docking with a flexible protein. The $\log K_S = 5.43$ and $\log K_R = 5.34$ for warfarin enantiomer-HSA interaction and the enantioselectivity ($ES = K_S/K_R$) of 1.23 were close to the experimental results and hence referred to as experimental-like affinity constants which validated and verified their applicability to predict protein-ligand binding affinities.

In case study III, a 3D-QSAR pharmacophore model was developed by using 98 known CYP17A1 inhibitors from the literature with known experimental IC_{50} values. The starting compounds were diverse which included steroidal and non-steroidal inhibitors. The resulting pharmacophore models were trained with 69 molecules and 19 test set ligands. The best pharmacophore models were selected based on the regression coefficient for a best fit model with R^2 (ranging from 0.85-0.99) & Q^2 (ranging from 0.80-0.99) for both the training and test sets respectively, using Partial Least Squares (PLS) regression. On the other hand, the best pharmacophore model selected was further used for a database screening of novel inhibitors and the prediction of their CYP17A1 inhibition. The hits obtained from the database searches were further subjected to a virtual screening workflow docked to CYP17A1 enzyme in order to predict the binding mode and their binding affinities. The resulting poses from the virtual screening workflow were subjected to Induced Fit Docking workflow to account for protein flexibility during docking. The resulting docking poses were examined and ranked ordered according to the docking scores (a measure of affinity).

Finally, the resulting hits designed from an updated model from case study III were further synthesized in an external organic chemistry laboratory and the synthetic protocols as well as spectroscopic data for structure elucidation forms part of the provisional patent specification. A provisional patent specification has been filed (**RSA Pat. Appln. 2015/07849**). The case studies performed in this thesis have enabled the discovery of non-steroidal CYP17A1 inhibitors.

LIST OF PUBLICATIONS

1. Myalowenkosi I. Sabela, **Njabulo. J. Gumedede**, Laura. Escuder-Gilabert, Yolanda Martín-Biosca, Krishna Bisetty, María Jose Medina-Hernández and Salvador Sagrado. Connecting simulated, bioanalytical, and molecular docking data on the stereoselective binding of (\pm)-catechin to human serum albumin. *Anal. Bioanal. Chem.*, (2012), 402:1899-1909.
2. **N.J. Gumedede**, P. Singh, M.I. Sabela, K. Bisetty, L. Escuder-Gilabert, M.J. Medina-Hernández and S. Sagrado. Experimental-like affinity constants and enantioselectivity estimates from flexible docking. *J. Chem. Inf. Model*, (2012) 52 (10): 2754-2759.

Note: A patent application (**RSA Pat. Appln. 2015/ 07849**) culminating from this work has been filed. Therefore, the data, computational, experimental and synthetic protocols of novel inhibitors will not be reported in this thesis due to confidentiality reasons. However, once the complete patent has been filed and a licensing partner have been sought the results will be disseminated. In fact, manuscripts will be submitted to international journals for possible publication.

LIST OF CONFERENCE PRESENTATIONS

1. Oral presentations:

Presented a paper titled “**Enantioselective binding of xenobiotic enantiomers to Human Serum Albumin using a Molecular docking approach: Warfarin as a test case**”. 2nd South African Young Scientist Conference, DST-NRF-ASSAF. On the 26 – 28 September 2011 at Diep in Die Burg, Pretoria.

Presented a paper titled “**Enantioselective binding of (±)-Catechin enantiomers to Human serum albumin using a flexible Molecular docking and MMGBSA methods to study the synergy between *in silico* and *in vitro* data.**” CHPC National Conference, CHPC-CSIR-Department of Science and Technology. On the 6 – 9 December 2012 at CSIR Convention Centre, Pretoria.

2. Poster presentations:

N. J. Gumede, M. I. Sabela, K. Bisetty, L. Asensi-Bernardi, Y. Martín-Biosca, L. Escuder-Gilabert, M.J. Medina-Hernández, S. Sagrado, S. Torres-Cartas, Electrophoretic and Molecular Docking Data Connection on the Enantioselective Binding of (±)-Catechin to Human Serum Albumin, 39th International Symposium on High Performance Liquid Phase Separations and Related Techniques. HPLC 2013, Amsterdam 20 June 2013.

N. J. Gumede, P. Singh, M. I. Sabela, K. Bisetty, Y. Martín-Biosca, L. Escuder-Gilabert, M.J. Medina-Hernández, S. Sagrado, S. Torres-Cartas, Using Chromatographic Information to Validate Molecular Docking Estimations. A Case Study, 39th International Symposium on High Performance Liquid Phase Separations and Related Techniques, HPLC 2013, Amsterdam, 20 June 2013.

N. J. Gumede, K. Bisetty, L. Escuder-Gilabert, M.J. Medina-Hernández, S. Sagrado, 3D-QSAR pharmacophore modelling on CYP17A1 inhibitors: Targeting prostate cancer, 20th Euro QSAR Workshop, St. Petersburg, Russia, 29 August 2014 - 4 September 2014.

Table of Contents

Declaration.....	II
Acknowledgements.....	III
Abstract.....	IV-VI
List of publications & contribution report	VII
List of conference presentations	VIII
Table of contents.....	IX-XI
Table of figures.....	XII
List of Tables.....	XIII
List of abbreviations.....	XIV

CHAPTER 1

INTRODUCTION	1
1.1 Drug enantiomer-HSA interactions	1
1.2 Therapeutic drug metabolism in men.....	3
1.3 Enzyme kinetics.....	7
1.3.1 Enzyme Inhibitors.....	10
1.4 Target enzyme (CYP17A1) to cure prostate cancer.....	11
1.5 Problem Statement.....	13
1.6 Background of the problem.....	15
1.7 Aims and Objectives.....	20
1.9 Thesis outline.....	21

CHAPTER 2

LITERATURE REVIEW.....23

2.1 A historical background of computational models in drug discovery23

2.2 Molecular Modelling approaches.....25

2.2.1 Molecular Docking – Structure-based drug design (SBDD) method.....25

2.2.2 Pharmacophore modelling -Ligand-based Drug Design (LBDD) Method.....29

CHAPTER 3

METHODOLOGY.....31

3.1 Computational Methods.....31

3.2 In vitro bio-analytical assay methods.....32

CHAPTER 4

CASE STUDY I.....35

4.1 Introduction.....35

4.2 Materials and Methods.....36

4.2.1 Computational Details.....36

4.2.2 Ligand preparation.....37

4.2.3 Glide Grid generation.....37

4.2.4 Glide docking method.....37

4.2.5 MM-GB/SA methodology.....37

4.3 Results and Discussion.....38

4.3.1 Molecular docking on (\pm)-catechin enantioselectivity to HSA.....38

4.4 CONCLUSIONS.....41

CHAPTER 5

CASE STUDY II.....42

5.1 Introduction.....42

5.2 Materials and Methods.....44

5.2.1 Software Methods.....44

5.2.2 Protein selection and preparation.....	44
5.2.3 Ligands preparation.....	45
5.2.4. Docking Protocols.....	45
5.2.4.1 Glide XP procedure.....	45
5.2.4.2 QPLD procedure.....	45
5.2.4.3 IFD procedure.....	46
5.3 Results and Discussion.....	46
5.3.1 Conversion of docking scores to experimental quantitative affinity constants.....	46
5.4 Conclusions.....	51
CHAPTER 6	
CASE STUDY III.....	52
6.1 Introduction.....	52
6.2 Materials and Methods.....	53
6.2.1 Data Treatment.....	53
6.2.2 Generation of 3D multiple conformers.....	54
6.2.3 3D-QSAR pharmacophore model.....	54
6.2.4 Density Functional Theory (DFT) Calculations.....	55
6.2.5 Molecular Docking.....	56
6.3 Results and Discussion.....	56
6.3.1 Design of a pharmacophore model.....	56
6.3.2 DFT results.....	61
6.3.3 Molecular docking results.....	63
6.4 CONCLUSIONS.....	67

CHAPTER 7

CASE STUDY IV	68
7.1 Introduction.....	68
7.2 Experimental Methods.....	69
7.2.1 Computational details.....	69
7.3 Results and Discussion.....	69
7.3.1 Regiochemistry prediction of CYP450 isoforms.....	69
7.4 CONCLUSIONS.....	72

CHAPTER 8

CONCLUDING REMARKS	73
---------------------------------	----

CHAPTER 9

CURRENT DATA, FUTURE PLANS	76
REFERENCES	79
APPENDIXES	101
Table A1. Structures and systematic names of CYP 17A1 inhibitors.....	101
Table A2. Training set and test set structural data.....	121
Table A3. Results of two pharmacophore hypothesis with important features.....	125
Table A4. Electronic properties calculated by using DFT optimization.....	125
Fig. A1. Docking results for 3d structure one of highly active molecules.....	126

List of Figures

Fig. 1.1 Crystallographic measured crystal structures.....	2
Fig. 1.2 Structure of Ferric heme coordinated to Porphyrin IX.....	5
Fig. 1.3 Catalytic cycle involving CYP450 hydroxylation of a substrate.....	6
Fig. 1.4 Michaelis-Menten plot depicting the relationship between V_o and $[S]$	8
Fig. 1.5 Lineweaver Burk plot that satisfies Michealis-Menten equation.....	10
Fig 1.6 3D structural representation of TOK001.....	13
Fig 1.7 Stepwise process for the discovery and development of NCEs.....	15
Fig. 1.8 Drugs withdrawn in the market from 1950 – 2013 due to serious side-effects.....	18
Fig 3.1 Horizontal hierarchy depicting a combined SBDD & LBDD methods.....	31
Fig. 3.2 Vertical hierarchical process for the experiments planned.....	32
Fig. 4.1 Glide docking protocol for protein-ligand binding affinity approximation.....	35
Fig. 4.2 Experimental and docking synergy for enantioselectivity.....	39
Fig. 5.1 3-Dimensional representation of the structure of warfarin.....	42
Fig. 5.2 Predicted interaction between warfarin enantiomers in Sudlow site I of HSA.....	49
Fig. 6.1 Best common pharmacophore hypothesis AADHRR.82.....	57
Fig. 6.2 Results for the most active and least active structures.....	58
Fig. 6.3 Validation plots of pIC_{50} (estimated vs. experimental) for AADHRR.82.....	60
Fig. 6.4 DFT results for the active molecule (+)-3c.....	62
Fig. 6.5 Docking screen for an overlay of a co-crystallized ligand.....	65
Fig. 6.6 Docking pose of the protein-ligand complex of the most active molecule (+)-3c.....	66
Fig. 7.1 CYP3A4 prediction of SOMs and the intrinsic reactivity.....	70
Fig. 7.2. CYP2D6 prediction of SOMs and the intrinsic reactivity.....	71
Fig. 7.3. CYP2C9 prediction of SOMs and the intrinsic reactivity.....	71

List of Tables

Table 3.1. Probe substrates, type of metabolites formed, and selective inhibitors.....	34
Table 4.2 Molecular docking results for the best pose predicted.....	40
Table 5.1 <i>In vitro</i> bioanalytical assay and <i>in silico</i> docking simulation.....	47
Table 5.2 Molecular docking results at molecular level.....	50
Table 6.1 Statistical results for the 3D-QSAR model.....	59
Table 6.2 Root-mean-square-deviations (RMSDs).....	64

List of Abbreviations

ADME/Tox	Absorption, Distribution, Metabolism and Excretion/Toxicity
BS	Biological Space
Cl _{int}	Intrinsic clearance
CPR	Cytochrome P450 Reductase
CS	Chemical Space
CYP	Cytochrome P450
DDI	Drug-Drug Interactions
DMPK	Drug Metabolism and Pharmacokinetics
FEP	Free Energy Perturbation
GLIDE	Grid-based ligand docking with energetics
GOLD	Genetic Optimization of Ligand Docking
GPCRs	G-Protein Coupled Receptors
HLM	Human Liver Microsomes
HPLC	High Performance Liquid Chromatography
HSA	Human Serum Albumin
IFD	Induced Fit Docking
IC ₅₀	Inhibitor concentration causing 50% reduced enzyme activity
IUPAC	International Union of Pure and Applied Chemistry
K _i	Inhibition constant
K _{inact}	Maximum rate of inactivation
K _m	Michaelis-Menten constant
LBDD	Ligand Based Drug Design
LID	Ligand Interaction Diagram
MBI	Mechanism Based Inhibition
MDS	Molecular Dynamics Simulation
MIC	Metabolite Intermediate Complex
MIF	Molecular Interaction Field
MS	Mass Spectrometry
MM-GB/SA	Molecular Mechanics Generalised Born Solvation Approximation
MM-PB/SA	Molecular Mechanics with Poisson-Boltzmann Surface Area
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NCE	New Chemical Entity
NME	New Molecular Entity
PCA	Principal Component Analysis
PC	Prostate Cancer
PLS	Partial Least Squares
QM	Quantum Mechanics
QM/MM	Quantum Mechanical/Molecular Mechanics
QPLD	Quantum Polarized Ligand Docking
3D-QSAR	3D-Quantitative Structure Activity Relationship
REST	Replica Exchange with Solute Tempering
RMSD	Root Mean Square Deviation
SBDD	Structure Based Drug Design
SMR	Structure Metabolism Relationship
SOM	Site of Metabolism
SRS	Substrate Recognition Sites
TDI	Time Dependent Inhibition
V _{max}	Maximum velocity of enzyme catalyzed reactions

CHAPTER 1

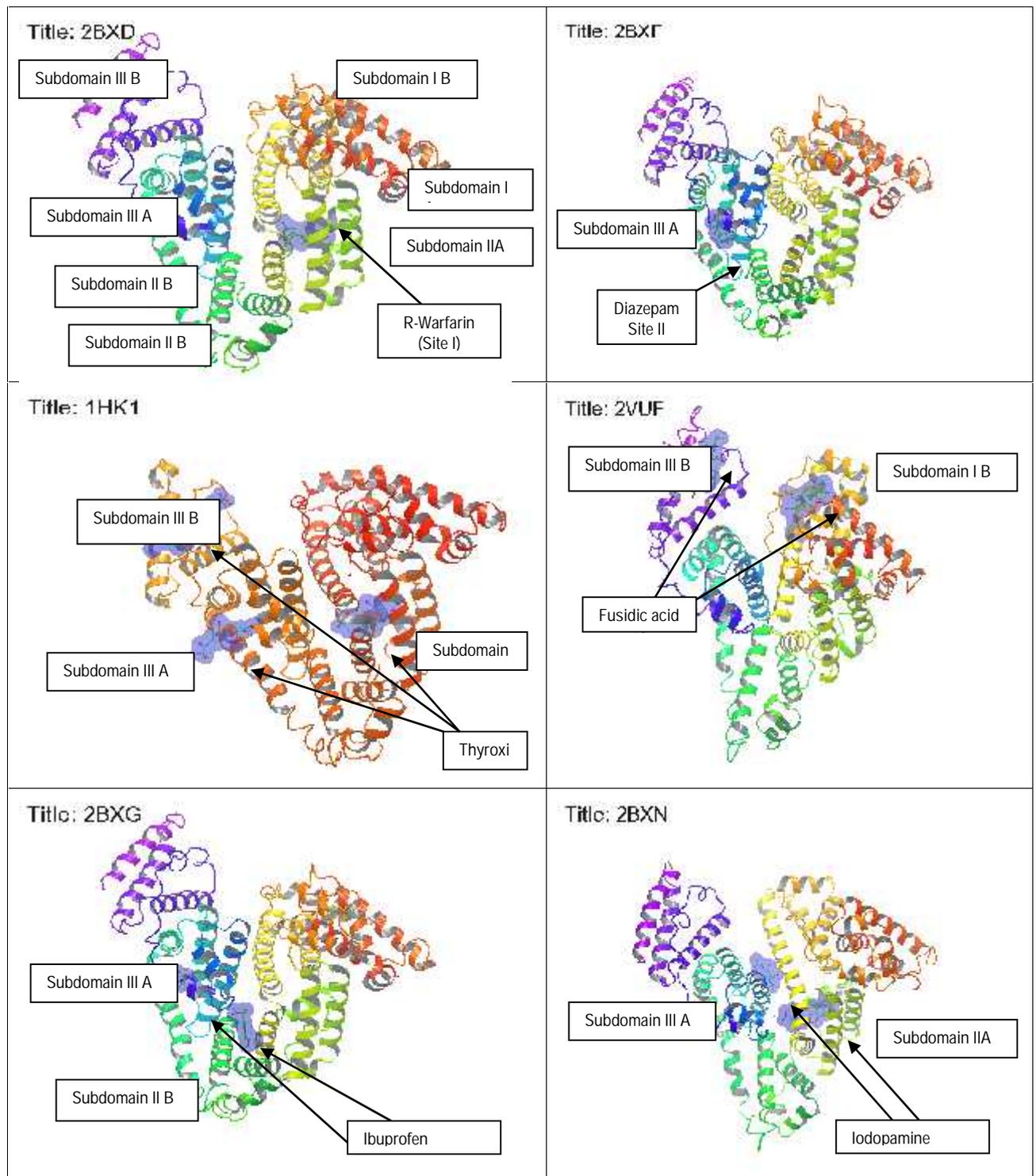
INTRODUCTION

1.1 Drug enantiomer-HSA interactions

Human serum albumin (HSA) is the most abundant carrier protein in plasma and is able to bind a wide variety of therapeutic drugs (Petitpas, 2001). It is in fact most abundant in the circulatory system (i.e., it has the biggest complexation potential), and displays a high degree of enantioselectivity among plasmatic proteins. HSA plays a pivotal role in the pharmacokinetic characterization of chiral xenobiotics including therapeutic drug enantiomers (Sabela, 2012). The main binding sites includes site I (warfarin site) and site II (diazepam site), therefore, binding of therapeutics in one of these sites can have a significant impact on their pharmacokinetic and pharmacodynamics properties (Gumede, 2012). However, there are several low affinity binding sites of HSA that bind amino acids and other drugs (see Fig. 1.1.) below.

Fig. 1.1 as shown below outlines different PDB structures of HSA with co-crystallized ligands in different binding sites available in the literature. In fact, the protein binds a number of relatively insoluble endogenous compounds such as unesterified fatty acids, bilirubin, and bile acids; and thus facilitates their transport throughout the circulation system (Kragh-Hansen, 1990; Peters, 1995; Petitpas, 2001). HSA is also capable of binding a wide variety of exogenous systems, and much interest on this protein stems from the fact that it facilitates drug delivery (Carter, 1994). Therefore, drug action in living organisms is controlled by a series of pharmacological processes including binding to carrier proteins such as HSA in order to reach the target in order to stimulate pharmacological effects. Most of these processes present a higher degree of enantioselectivity resulting in differences between the activities of drug enantiomers to carrier proteins and enzymes (Gumede, 2012). Among others, interactions with plasma proteins are critical features describing the biological activity of therapeutic drugs (Escuder-Gilabert, 2009).

Figure 1.1. Crystallographic measured crystal structures showing different binding sites regions of HSA with site markers in their respective subdomains and clefts.



Studying drug–human serum albumin binding is a very attractive undertaking in the pharmaceutical industry (Andrisano, 2000). Since drug–protein interactions affects pharmacological activities, distribution and elimination of drugs. Therefore, bio-analytical

methods have been developed, validated extensively and have subsequently been used to quantify the drug-protein interactions, and more recently, enantioselectivity estimations. Furthermore, several critical reviews have been published on the most important analytical approaches to describe drug-protein binding (Vuignier, 2010). On the other hand, micro-analytical separation methods for estimating the enantioselective binding of drugs to plasma proteins have also been reviewed (Escuder-Gilabert, 2009). The general assumption is that quantitative predictions of binding affinities from bio-analytical methods provide estimates accurate enough; even though, there is a need to verify the quality of the results which is not normally performed in Research and Development (R&D) laboratories, and pose as a risk in published data (Asensi-Bernardi, 2010; Gumede, 2012). The main drawback related to bio-analytical methods of analysis is that they are unable to reveal much information related to the identification of molecular mechanisms involved in the affinity and enantiodiscrimination. Therefore, molecular modelling methods can be used jointly with bio-analytical experiments for these studies, as we have shown in Case study I and II.

1.2 Therapeutic drug metabolism in men

This section is based primarily on the therapeutic drug metabolizing enzymes such as the cytochrome P450 superfamily. Cytochrome P450 enzyme is an important enzyme as it is able to function in the majority of bio-organisms (Tian, 2009). Isoforms of cytochrome P450 enzyme are found in most living organisms and are also involved in the biosynthesis of steroids hormones in the body (Shaik, 2010). Furthermore, xenobiotics such as therapeutic drugs are metabolized by multiple enzymes in human organisms; the main enzyme responsible for this task is the cytochrome P450 superfamily (Rudik, 2014; de Groot, 2002 & 2007; and Moroy, 2012). Additionally, there are two types of drug metabolizing enzymes involved in the pharmacokinetic (PK) events i.e. Phase I and Phase II metabolic enzymes (Kamatagi, 2001).

Cytochrome P450 enzymes are accountable for the phase I metabolism of 75% of known endogenous and exogenous substances (Rittle, 2010). Phase I metabolism is the first route of the metabolism of xenobiotics by human cytochrome P450 isozymes (Sun, 2011; and Kamatagi, 2014). Phase I metabolism involves several enzyme-catalyzed redox reactions i.e. S-oxidation, N-reduction, and hydration reactions to mention just a few (de Groot, 2006 & 2007; Nawak, 2014; and Sun, 2010). In fact, cytochrome P450 enzyme is a heme-containing enzyme belonging to a superfamily of enzymes in the human genome. This superfamily contains a total of 70 families of heme-containing enzymes (Campagna-Slater, 2012, De Groot,

2004; Wrighton, 1992; Sun, 2010; and Sun, 2011), of which only 57 are known in human as CYP450 isoforms and only nine are responsible for metabolism of therapeutic drugs in men.

The nine CYP450 isoforms referred to above are as follows: CYP1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1, and 3A4. Other CYP450 isoforms such as CYP11, CYP17, CYP19, and CYP21 are involved in steroid biosynthesis (Hayes, 2014; Nelson, 1999; de Groot, 2007; Wang, 2014). In fact, CYP is an abbreviation for cytochrome P450. The P450 part stems from the wavelength of maximum absorption (λ_{max}) of Fe^{3+} -porphyrin complex. The active site of CYP450 enzyme consist of ferric heme and porphyrin groups, and is mainly reactive to molecular oxygen (Rittle, 2010; Margareta, 2014). Therefore, the CYP450 enzyme is known to be a mono-oxygenase enzyme for this reason. Furthermore, the enzyme facilitates the attachment of molecular oxygen in its active site to break the non-reactive functional groups of organic compounds in the presence of co-factors such as NADPH like for example to yield an alcohol as a soluble metabolite (Tian, 2009; Shaik, 2010). In fact, more specifically CYP450 phase I enzymes catalyzes substrate metabolism through oxidation, reduction and hydrolysis of organic compounds. The reaction in equation 1 below is a typical example of cytochrome P450 catalyzed hydroxylation of a substrate (R-H) in the presence of a co-factor (NADPH) to yield a metabolite (R-OH) and $NADP^+$.



Attempts to study how cytochrome P450 enzymes catalyses the mechanism leading to the metabolism of new molecular entities (NMEs) are very much interesting to the scientific community (Shaik, 2010). Understanding the mechanisms of drug metabolism can be viewed from three extremes of PK events. Firstly, to alter the target compounds into active metabolites that is reactive to the target to cure diseases. Secondly, by the conversion of NMEs into soluble and non-toxic metabolites. Thirdly, through the conversion of NMEs into toxic metabolites and drug-drug interactions that can lead to therapeutic drug withdrawals in later phases of the drug discovery process (Li, 2009; and Li, 2011).

The prediction of mechanisms of CYP catalysed metabolism of new chemical entities (NCEs) as well as the regio-chemistry of possible metabolites of NCEs that could be generated by CYPs are difficult to comprehend. It is therefore of utmost importance to understand the underlying mechanisms of metabolisms in order to prevent late-stage withdrawals of NCEs in clinical trials (Olsen, 2015). Cytochrome P450 enzyme is a superfamily of isoforms which is

involved in the catalysis of biosynthesis of steroid hormones and metabolism of xenobiotics. Therefore, it is an exciting and yet an important undertaking to study the structure of cytochrome P450 on how it performs its functions in metabolism reactions (Shaik, 2010). Cytochrome P450 enzymes contain the ferric heme porphyrin group in its' active site. It is involved in the oxidation reactions like for example of aliphatic, aromatic hydroxylation, hetero atom oxidation, and N- or O-dealkylation reactions. These reactions thus yield soluble metabolites that can easily be excreted and eliminated from the body (Olsen, 2015).

Most of the proteins in their active sites contain Ferric heme, which is composed of a porphyrin ring that is coordinated to Iron as shown in Fig. 1.2 below. The heme group's function is to facilitate electron transfer, transportation of oxygen molecule, and catalysis. The ability of heme to undergo a change in oxidation state from Fe (III) to Fe (II) aids the formation of an active species (compound 1) which activates C-H bonds of the substrates for metabolism to occur (Blomberg, 2014).

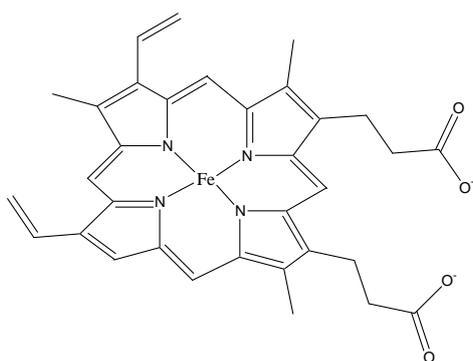


Fig. 1.2 Structure of Ferric heme coordinated to Porphyrin IX.

The catalytic cycle for the biotransformation of inactivated C-H bonds of the substrate, in this case the hydroxylation of the aromatic moiety of an NCE is shown in Fig. 1.3 below.

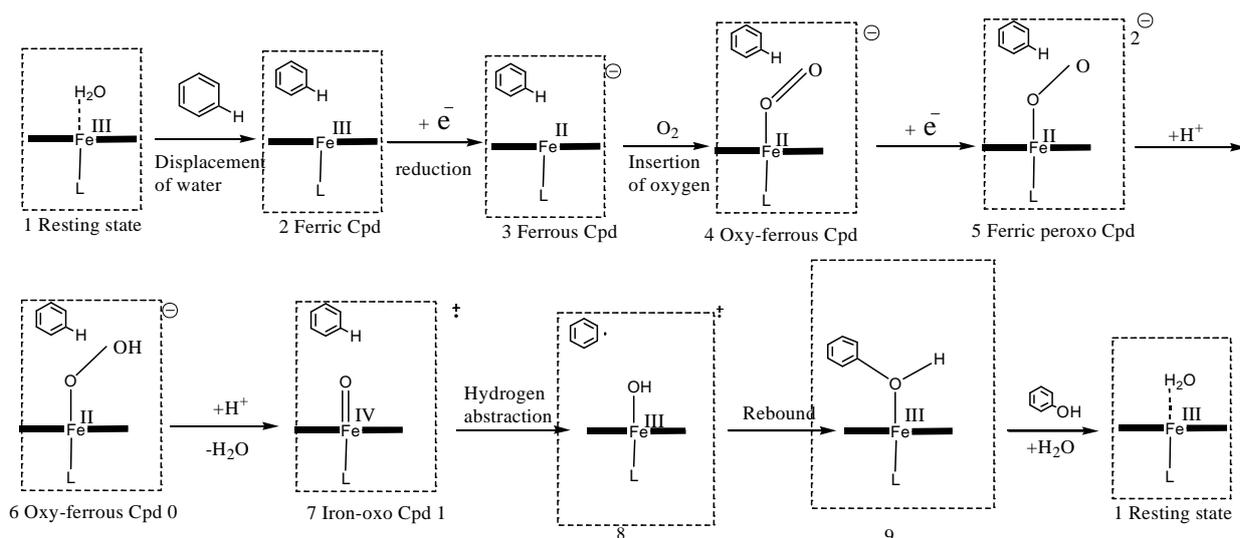


Fig. 1.3 Catalytic cycle involving CYP450 hydroxylation of a substrate.

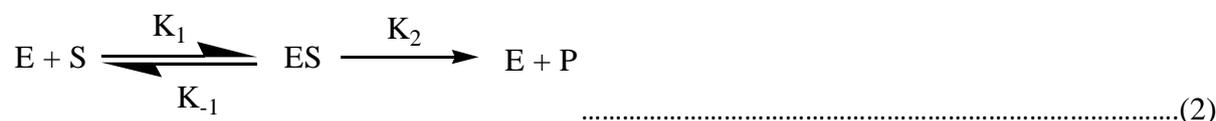
The heme is shown with two bold horizontal lines, and the cysteine proximal ligand indicated as L. In the resting state (**1**), heme is hexacoordinated with the proximal ligand L and water molecule in a low-spin doublet state (Shaik, 2010). In the resting state, the enzyme is not reactive; the reactivity of the enzyme is facilitated by the change in oxidation state, ligand composition, and the changes in spin states of ferric heme and is most common among CYP450 enzymes (Guallar, 2004). In Fig. 1.3, the first step of the catalytic cycle involves water that is coordinated to ferric heme in the active site of the enzyme (**1**); with a low-spin resting state, which is displaced by the substrate to form the penta-coordinated ferric porphyrin (**2**); with a high-spin state and high electron affinity (Guallar, 2004; Shaik, 2010; and Bloemberg, 2014). The enzyme is stimulated by the reduction of the intermediates using two electrons coming from co-factors such as Cytochrome b5 and NADPH-P450 reductase (Bloemberg, 2014). The ferric porphyrin complex then accepts an electron from the co-factors to yield a ferrous complex anion (**3**); Molecular oxygen then binds with the ferrous compound to yield an intermediate oxy-ferrous compound (**4**); The oxy-ferrous complex has a singlet spin-state, and hence it is a good electron acceptor. Accepting a second electron yields a ferric-peroxo anion species (**5**) (Olsen, 2015). Protonation of the ferric-peroxo complex yields compound 0 (ferric-hydroperoxide) (**6**); Compound 0 abstracts a proton (Somersault O-O cleavage) to form a high-valent compound 1 iron-oxo species (**7**) (Shaik, 2010). Compound 1 is believed to have two close-lying spin-states, which are the quartet and doublet states. Hence, it has triplet coupled electrons which are either coupled ferromagnetically or antiferromagnetically to the porphyrin radical (Bloemberg, 2014). During C-H hydroxylation of a substrate, compound 1 abstracts

one electron from a substrate to yield a radical intermediate (**8**) (Tian, 2009). The iron-bound hydroxyl then reacts with the radical intermediate to yield a ferric alcohol complex (**9**) (Schöneboom, 2004; Altun, 2006). The alcohol is then formed and the water molecule re-enters and regenerates the resting state. A review of the enzyme kinetics with particular emphasis on the factors influencing the rates of substrate-enzyme recognition patterns is presented in the next section.

1.3 Enzyme Kinetics

In the 1890s the German chemist Emil Fischer (1852–1919) proposed a lock-and key approach when enzymes bind with the substrates. According to Fischer, the active site of the enzyme is a rigid body where a substrate binds with an enzyme and fits snugly in it's active site like a key in a lock. This theory, however, has been extended in order to allow for flexibility of proteins in solution. Enzymes that are flexible undergo induced-fit effects when they bind with substrates in order to alter the conformation of the active site.

A very important discovery in enzyme kinetics was made by the German biochemist Leonor Michaelis and his Canadian assistant Maud Leonora Menten (Michaelis and Menten, 1913). This theory builds on the work of the French Chemist Victor Henri (1872–1940), who proposed a mechanism to explain the dependence of the initial rate of enzyme-catalysed reactions on concentration of the substrate. The Michaelis-Menten equation was developed 102 years ago and it is still applicable nowadays in a quest to study the rates of enzyme-substrate kinetics (Xie, 2013). The M-M theory describes how a substrate (S) binds with an enzyme (E) in order to form an Enzyme-substrate (ES) complex, which subsequently yields the product P, as can be seen in equation 2 below (Zhang, 2005).



Substrate-enzyme binding event occurs in the active site pocket of the enzyme. Enzymes accelerate reactions by lowering the activation free energy change ($\Delta G^\#$). The equilibrium of the reaction remains unaffected by the enzyme. Where k_1 and k_{-1} are the forward and reverse rate constants for substrate binding and k_2 is the catalytic rate constant. The binding of enzyme and substrate to form the enzyme-substrate complex (E-S) is in fact a fast process, and is a rate

limiting step. Whereas the catalysis of the enzyme-substrate complex (ES) to form the enzyme and the product is a slow process, and is a rate determining step. Enzyme-substrate interactions are predominantly non-covalent i.e. governed by ionic, hydrogen bonds, π - π , and hydrophobic interactions. The conformation of the substrate to position itself and be accessible to Fe (+3) moiety of ferric-porphyrin coordination system is important for hydroxylation to occur for example in Cytochrome P450 metabolism. Accordingly, the Michaelis-Menten equation is used to measure the relationship between the reaction velocity and substrate concentration.

In the formation of an ES complex, the forward rate is given by: $V_1 = K_1[S][E]$. While the rate of the reverse reaction is given by: $V_{-1} = K_{-1}[S][E]$. Accordingly, the relationship between the reaction velocity, V and substrate concentration [S] is hyperbolic as shows in Fig. 1.4 below.

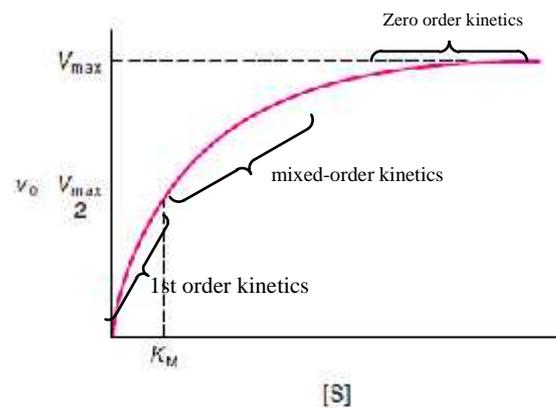


Fig. 1.4 Michaelis-Menten plot depicting the relationship between V_0 and $[S]$ for an enzyme-catalysed reaction.

The Michaelis-Menten equation is the rate equation for a one-substrate enzyme catalysed reaction. It quantitatively relates the initial rate, the maximum rate, and the initial substrate concentration to the Michaelis constant K_M as shown in equation 3 below.

$$V = \frac{V_m [S]}{[S] + K_M} \dots\dots\dots(3)$$

We then get: $K_M = \frac{K_{-1} + K_2}{K_1}$.

The Michaelis-Menten constant, K_M is a constant with units' (M) and a constant derived from rate constants. The K_M value is, under true Michaelis-Menten conditions, an estimate of the dissociation constant of E from S. Therefore, a small K_M value means tight binding; while a

high K_M means weak binding between an enzyme and a substrate (Zhang, 2005). On the other hand, V_{max} is a constant with units of s^{-1} . On the other hand, V_{max} is the theoretical maximal rate of the reaction which has not been achieved in reality. In order to reach V_{max} , it would require that all enzyme molecules are tightly bound with the substrate. Therefore, as $[S]$ is increased V_{max} is asymptotically moved upward toward the maximum value as shown in Fig. 1.5 above.

The Michealis-Menten equation follows zero and first order kinetics in a sense that when $[S]$ is low, the equation for rate is 1st order in $[S]$. Whereas, when $[S]$ is high, the equation for rate is zero-order in $[S]$. Accordingly, if $[S] \gg [E]_{total}$ the enzyme is saturated with the substrate in its' active site. Therefore, $[ES]$ is equal to $[E]_{total}$ the maximum rate of distribution V_{max} can be defined as: $V_{max} = K_2 [E]_{total}$. Therefore, The Michaelis-Menten equation is the rate equation for a one-substrate enzyme catalysed reaction. It quantitatively relates the initial rate, the maximum rate, and the initial substrate concentration to the Michaelis-Menten constant, K_M (Xie, 2013).

It has been observed experimentally, that the plot of V versus $[S]$ is not essentially valuable in determining the value of V_{max} because finding the asymptotic value of V_{max} at very high substrate concentrations has proved to be difficult. In 1934, Hans Lineweaver and Dean Burk published a paper which introduced a double-reciprocal plot of $1/v$ v.s. $1/[S]$, by rewriting equation 3 above to give equation 4 presented below (Lineweaver, 1934).

$$\frac{1}{V} = \frac{K_M}{V_m [S]} + \frac{1}{V_m} \dots\dots\dots(4)$$

The Lineweaver Burk plot is useful when used to determine the type of inhibition i.e. competitive, non-competitive and uncompetitive inhibition (Wilkinson, 1961). The Lineweaver-Burk plot as shown in Fig. 1.5 below satisfies Michealis-Menten equation, where both K_m and V_{max} can be obtained from the slope and the intercept of the straight-line graph.

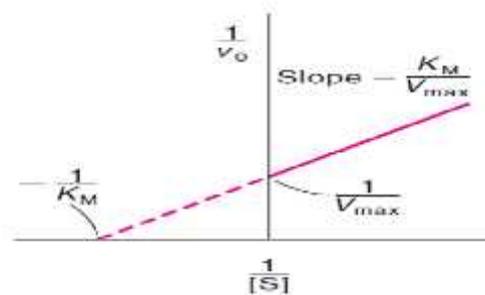


Fig. 1.5 Lineweaver-Burk plot that satisfies Michaelis-Menten equation.

The main drawback of Lineweaver-Burk plot is that it tends to compress the data at high concentration of the substrate into a somehow small region, which tends to emphasise points at low concentrations which has proved to be less accurate (Wilkinson, 1961). The evolution of linear and non-linear regression techniques nowadays has changed the way we measure the IC₅₀, K_i, V_{max} and K_M for enzyme-substrate inhibition. However, the Lineweaver Burk plot is still applicable in enzyme kinetics.

1.3.1 Enzyme Inhibitors

Inhibitors can interact with an enzyme via covalent and non-covalent interactions. Therefore, the inhibitors that binds enzymes via covalent interactions are called irreversible inhibitors. While inhibitors that binds enzymes via non-covalent interactions are called reversible inhibitors. In therapeutic drug design our interest lies mostly on reversible inhibitors. Therefore, in reversible inhibition there is a competition between the inhibitor, [I] and a substrate, [S] for binding in the active site. Since, the inhibitor binds to the enzyme and not the enzyme-substrate complex. The V_{max} is not affected by the competition between the substrate and an inhibitor for the active site of the enzyme. Therefore, K_M becomes $K_M = \frac{K_M + K_M [I]}{K_i}$, where K_i is the equilibrium constant for E + I → EI. The bond between the enzyme and a substrate becomes weaker, while K_M becomes large.

In reversible inhibition, non-competitive inhibition occurs, where the inhibitor can either bind to the enzyme only or the enzyme-substrate complex. The binding of the inhibitor, I to the enzyme does not affect the binding of the substrate, S into the enzyme. This makes sense because the inhibitor does not bind in the active site of the enzyme. Therefore, the V_{max} cannot be recovered by raising the concentration of [S] and the K_M remains unchanged. In mixed non-competitive inhibition, the binding of the inhibitor into the enzyme influences the

binding of the substrate. Therefore, K_M and V_{max} are changed since the inhibitor binds close to the active site which has an influence on binding of the substrate. In uncompetitive inhibition when the inhibitor binds to the enzyme-substrate complex. The K_M and V_{max} are altered since the inhibitor binds close to the active site.

1.4 Target Enzyme (CYP17A1) inhibition to cure Prostate Cancer

It is well documented in the literature that Prostate Cancer (PC) is among the most prevalent diseases among men in industrialised countries (Purushottamachar, 2012; Hu, 2010; Beltran, 2011; Gianti, 2012; Hille, 2008; Mendieta, 2008; Clement, 2003; and Jagusch, 2008). In fact, PC is the second largest cancer related disease which causes deaths each year (Hu, 2010; Haidar, 2003). The deaths caused by prostate cancer are as a result of androgen dependent or hormone refractory diseases (Attard, 2009; Clement, 2003; Jagusch, 2008; and Hu, 2010). The development and progression of PC is influenced by the involvement of androgens and androgen receptor (AR) (Purushottamachar, 2012).

PC develops in the prostate tissues, where normal cells starts by forming malignant tumours which subsequently spreads all over the body, which then gives rise to metastases (Gianti, 2012;). The first treatment that is used to diagnose the disease includes surgery or radiation. However, androgen deprivation therapy remains as the method of choice for the treatment of advanced or metastatic form of prostate cancer (Purushottamachar, 2012; Hu, 2010). The level of prostate-specific antigen (PSA) rises after the first therapy for patients with metastatic PC (Yamaoka, 2012). Therefore, the next step is the use of androgen deprivation therapy in conjunction with luteinizing hormone–releasing hormone (LHRH) (Bryce, 2012; Hu, 2010). The PC will eventually progress into castration resistant prostate cancer (CRPC) which is only treated with chemotherapy. However, chemotherapy comes with its own side effects which are undesirable (Beltran, 2011; Yamaoka, 2012). Therefore, attempts have been made in trying to find alternative treatments to combat the disease such as surgical and medical castration in combination with antiandrogens (Hu, 2010; Krug, 2013; Hille, 2008; Bryce, 2012). These attempts were futile and were later replaced by orchidectomy which was later replaced by administration of gonadotropin-releasing hormone (GnRH) analogues. Therefore, Androgen Receptor (AR) antagonists are used in combination with GnRH analogues in order to prevent mutations of adrenal androgens which leads to castration resistant prostate cancer

(CRPC). However, long-term use of this combined androgen blockage (CAB) therapy results in drug resistance to CAB (Hu, 2010; Krug, 2013; Jagusch, 2008; Haidar, 2003).

The inhibition of CYP17A1 enzyme is the route that is thought to prevent castration resistance prostate cancer (CRPC). The CYP17A1 enzyme catalyses two major routes involved in steroid biosynthesis viz. 17 α -hydroxylase and C_{17,20}-lyase activities. In the first step, the substrates progesterone and pregnenolone are hydroxylated in the 17 α - position of the substrates to form 17 α -hydroxyprogesterone and 17 α -hydroxypregnenolone, respectively. While in the second step, the C_{17,20}-lyase activities breaks the C₁₇-C₂₀ bonds of 17 α -hydroxyprogesterone to yield dehydroepiandrosterone (DHEA) (Krug, 2013; Clement, 2003; Hakki, 2006; Jagusch, 2008; Hu, 2010; Bryce, 2012; Hille, 2008; Yamaoka, 2012; Haidar, 2003). The main advantage of CYP17A1 inhibitors is that they hinder androgen biosynthesis in the testicles and adrenals as well as the formation of intracellular androgens in cancer cells (Hu, 2010; Moreira, 2007).

The group of CYP450 enzymes involved in steroid biosynthesis includes the following: CYP11A1 which is involved in cholesterol side chain cleavage, CYP21 involved in steroid-21-hydroxylase, CYP17 catalyses the 17 α -hydroxylase-C_{17,20}-lyase activities, CYP19 involved in aromatase activities, CYP11B1 involved in steroid-11 β -hydroxylase activities and CYP11B2 in aldosterone-synthase activities. However, inhibition of CYP11A1 and CYP21 is not suitable as a drug target. Since, the former affects the biosynthesis of all steroid hormones while the latter is involved in the biosynthesis of gluco- and mineralocorticoids (Hille, 2009). The CYP 17A1 steroidogenic enzyme has a molecular mass of 56 kDa. Fig. 1.6 below shows a 3D structure for CYP17A1 bound with TOK001 (Galeterone). Recently, DeVore et al. reported two crystal structures for CYP17A1 co-crystalized with CYP17 inhibitors Abiraterone (3RUK) and TOK001 (3SWZ), and were deposited into the Protein Data Bank (PDB) at a resolution of 2.6 Å and 2.4 Å respectively (DeVore, 2012). New research incorporating such new information could reveal more consistent results from molecular modelling techniques.

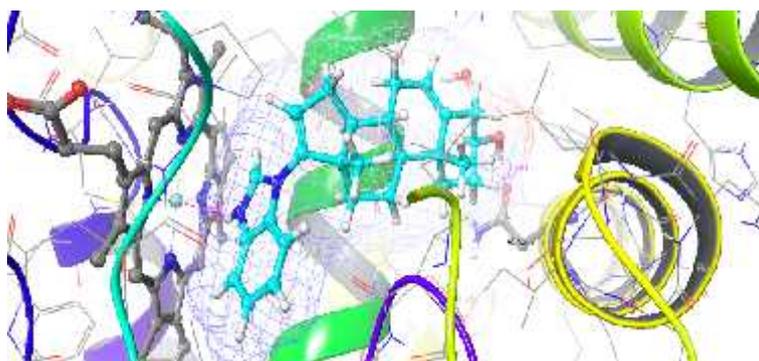


Fig. 1.6 3D structural representation of TOK001 (electrostatic potential) on the active site cavity of CYP17A1 enzyme showing metal coordination with Ferric heme.

The catalytic cycle for the hydroxylation and cleavage activities is similar to the catalytic cycle for hepatic metabolising enzymes, as shown in Fig 1.3 of the previous section above. However, both the hydroxylation and cleavage activities occurs in the same active site. The reaction is facilitated by two electrons from NADPH through its redox partner, cytochrome P450 reductase (CPR) and cytochrome b5 (Akhtar, 2005). Ketoconazole is a selective inhibitor for CYP17A1 and has been used clinically for PC treatment. However, Ketoconazole is a weak inhibitor and suffers from off-target inhibition, which pose as a threat with side effects due to drug-drug interactions (Hu, 2010; Hille, 2008; Jagusch, 2008). The steroidal inhibitor Abiraterone acetate has been shown to be an effective inhibitor of CYP17A1. However, steroidal inhibitors have a tendency to show affinity to steroid receptors resulting in side-effects. This reason prompted the development of non-steroidal CYP17A1 inhibitors (Hu, 2010; Bryce, 2012; Yamaoka, 2012).

1.5 Problem Statement

One of the thought provoking questions usually asked is how many disease entities are known to human kind? How many are cured by therapeutics? How many drug targets have been characterised through the genome project? A disease ontology resource:

<http://disease-ontology.org/>, has reported that 8000 diseases are currently been known world-wide. One begins to wonder whether prescribed medications do cover the spectrum of all diseases currently known to human kind. In fact, almost 75% of the diseases known to human kind have not been addressed by therapeutics currently in the market. This is attributed to the lack of knowledge of the function of targets and their therapeutic use in order to design therapeutics that would trigger their pharmacological responses.

There is a known synergy between biological and chemical sciences leading to their joint importance in life-sciences. More especially, since medicinal chemists are tasked with a job of designing novel therapeutic compounds to cure different types of diseases. In rationalizing the discovery of novel therapeutic compounds, a major focus is on understanding of disease biology such as genes, enzymes, proteins and receptors which will thereby explain their interactions and functions. From a medicinal chemistry perspective, there is much interest in studying the relationship between chemical structures and molecular descriptors. These chemical structures could be mapped onto some chemical space (CS) to identify the drug-like molecules using a series of parameters towards the profile of a clinical candidate. The chemical descriptor information can in some instances predict the distribution and metabolism of drug candidates (Ekins, 2003). On the other hand, from a biological perspective crystal structural information can be easily accessible from the RCSB protein data bank (PDB) www.rcsb.org/. Then native docking and cross-docking can be performed to validate the molecular docking algorithm and reproduce the native conformer of the ligand in the protein active site. Therefore, Protein-ligand interaction studies can then be performed to study the binding modes and binding affinities of novel ligands to the target receptor.

The last two decades or so, has seen the Pharmaceutical Research and Development (R & D) companies been faced with a challenge of an increase in the therapeutic drug development costs as well as a decline in the registration of new therapeutics. The root cause for the former is that drug discovery is a very long process and involves expensive *in vivo* and *in vitro* experiments. Starting from the synthesis of new chemical entities to *in vivo* and *in vitro* biological activity study prior to clinical studies which usually succumbs to safety issues being detected before clinical studies and the drug then is abandoned and the life-cycle begins again until a safe drug is synthesized. The reason for the latter might be largely due to high failure rates in post-clinical studies because of toxicities and drug-drug interactions. Pharmaceutical companies are then forced to assess drug safety issues earlier in the drug discovery and development pipeline in order to reverse this current trend. Furthermore, the experiments using animals as models in the drug development process needs to be done away with because of ethical issues involved.

A report issued by United States Food and Drug Administration (FDA) [FDA Critical Path Initiative White Paper, 2004], aims to address the abovementioned short falls as it states that: “*there is a great need for the pharmaceutical industry to consider adopting and using*

computer-based predictive models to improve the predictability and efficiency of developing a new chemical entity (NCE) from a developmental stage in the laboratory to the commercialisation of the new product after clinical trials". Accordingly, in complementing computationally developed models, it is also important to develop high-throughput *in vitro* systems to test the applicability of computational models and thereby validate them. In this way the drug discovery process could be shortened which will in turn save money, time and effort. This approach is able to eliminate or reduce the use of animals in experiments since, which forms part of the aims and objectives of the Registration, Evaluation, Authorisation and restriction of chemical substances, CE n° 1907/2006) (REACH) normative (Huynh, 2009) and the Reduction, Refinement and Replacement of experiments related to animals as models (3R approach) (Russel, 1959).

1.6 Background to the problem

The drug discovery pipeline at its very core includes target identification, target validation, lead identification, lead optimization, pre-clinical studies, synthetic routes, Phase 0, Phase I, II, III clinical trials, or phase IV, licensing, and registration of the new product after approval as shown in the process diagram in Fig. 1.7 below. The entire drug discovery and development process shown in Figure 1.7 usually takes about approximately 14 years to complete (Baranczewski, 2006; Gashaw, 2011). Even after approval, the new drug is monitored in order to ensure the safety of patients using the medications. The expenses that are connected with drug discovery and development of a new pharmaceutical product totals to US\$ 2 billion.



Fig 1.7. Stepwise process for the discovery and development of NCEs.

In target identification, the only important information that will drive the drug discovery process is the knowledge of the disease mechanism as well as the active role played by the enzyme, protein or receptor (Gashaw, 2011). This refers to the identification of a biological pathway that can be used to treat or cure a disease. Following that would be target validation, which is a crucial step, since this step is thought to be the major factor that contributes to drug discovery failures. This is because the only way to test whether a new drug is able to inhibit

the activity of a target protein, enzyme or receptor is through tests on humans. The validation process is aimed at defining a part that a target must accomplish in order to design drugs that would be able to cure a disease. The regulation of the target is also important because the knowledge of a novel target needs to be disseminated to the scientific community. This might also help in identifying new chemical pathways that might be followed by potential therapeutics.

This process is then followed by lead identification where various approaches are used such as *in vitro* high throughput screening of large databases of compounds. More recently, *in silico* approaches such as pharmacophore modelling, molecular docking, shape-based screening, de novo drug design, and virtual screening workflows have gained momentum in this research area. In many cases in lead identification the pharmacokinetics parameters i.e. Adsorption, Distribution, Metabolism, and Excretion (ADME/Tox) as well as toxicity; physicochemical properties are determined using physics-based scoring functions and Quantitative Structure Activity Relationships (QSAR) approaches. Lead optimization on the other hand, is a process where a lead compound identified from the previous step is optimised to improve its' potency to the target enzyme. *In silico* approaches such as e-pharmacophore, combinatorial screening, Quantitative Structure Activity Relationships (QSAR), functional group modification, shape matching, and molecular interaction fields are normally used in lead optimization. This process is then followed by the synthesis of hits and their derivatives which is followed by *in vitro* biological activities. Pre-clinical studies are then performed to determine whether there are any off-target interactions. Meaning that the study is intended to determine whether the inhibitor can undergo drug-drug interactions, or does it inhibit major Cytochrome P450 enzymes. On the other hand, is the drug metabolized by Cytochrome P450 isoforms and are there any toxic metabolites formed? The answer to these above questions is to determine whether the drug is fit for human consumption or not.

In most cases the inventors usually patent the drug they have discovered in order to protect their intellectual property (IP) at this stage of the discovery process. Furthermore, most of the inventors in academia after a patent is approved usually license the drug to big pharmaceutical companies. The academics then avoids the expensive and long process of clinical trials, as the licensee company would be conducting further clinical trials and subsequently launching the drug. Therefore, high margins of profit can be attained when following this model since inventors get a huge one-time payment for the new chemical entity (NCE) from entering into a revenue sharing agreement with the licensee company. If the

potential drug candidate passes Phase I, II and III clinical trials the new medicinal product is then registered for human use by the Food and Drug Administration (FDA).

The question that one can have when reading the previous paragraphs would be: why so many drug failures if the drug process is smooth and achievable? The high failure rate is the main bottleneck in the drug discovery process. Drug failures are caused by low potency, toxicity, as well as drug-drug interactions. Historically, the main route to the discovery of NCEs was through the synthesis of a series of compounds sharing a common skeleton structure, and then tests them for *in vivo* efficacy (Lowe, 2012). Further studies were often conducted to measure the selectivity and toxicity of the newly synthesized molecule and their derivatives. This is the stage where most of the problems related to toxicity and low potency were identified. Therefore, most of the drug discovery projects were halted at this stage since most of the funding money would have been used up in the synthetic steps. Further optimization of toxic structures would not commence since there would be no further funding. Obviously, companies spend billions of dollars to fund a potential drug candidate that sometimes pass clinical trials and then suddenly is withdrawn in the market. The loss incurred is normally very massive and investors usually do not want to be associated with companies that have their shares to drop because of withdrawals. The blockbuster drug Vioxx that was marketed by the drug discovery section of Merck was withdrawn from the market because of serious side effects. Merck's stock market plunged within a day when their top selling drug was withdrawn from the market. Another case that is also more prominent in the drug discovery history is that of latrepirdine where the drug was withdrawn in phase III clinical trials. Latrepirdine was a multi target drug but because of efficacy issues the drug was no longer of use to the pharmaceutical market. Fig. 1.8 shows a bar graph of new chemical entities that have been withdrawn in the market from 1950 – 2013. During this period, 211 drugs have been withdrawn from the market which means billions of dollars were lost when these drugs were developed (Kola, 2004).

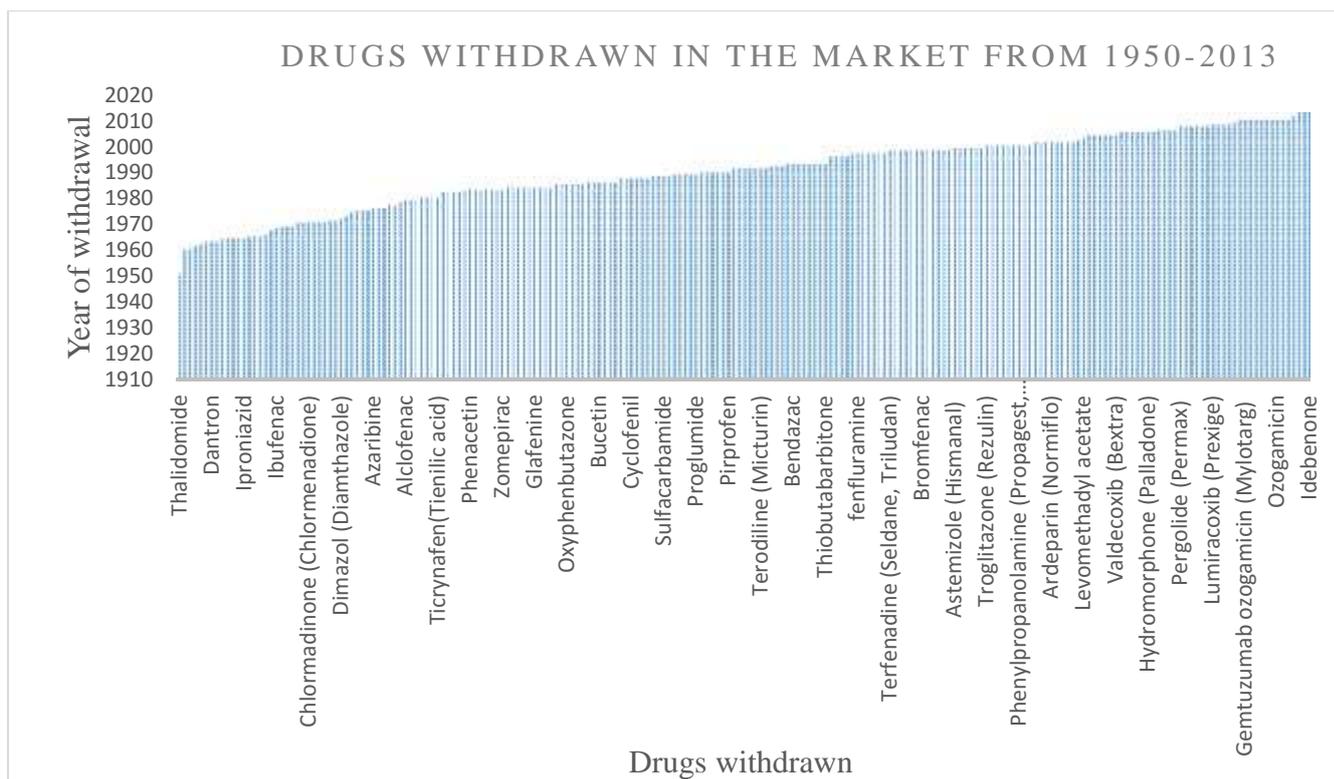


Fig. 1.8 Drugs withdrawn in the market from 1950 – 2013 due to serious side-effects.

As can be seen from Fig. 1.8 above significant withdrawals increased dramatically in the past two decades. It is worth noting that major withdrawals took place between 1998 and 2013. This crisis is assumed to be caused by the lack of understanding of deep biological mechanisms related to the flow of pathological conditions. Furthermore, the lack of knowledge of biological mechanisms related to toxicities in drugs is also a major obstacle. The end result to this so called ignorance in understanding the abovementioned factors is that 50% of prescribed medications do not have any therapeutic effect, since some of generic medications are counterfeit. This is more evident in third world countries that usually import generic drugs from other countries with no stringent rules to test them before they go to the shelves as corruption and monopoly is the norm in those countries. Furthermore, the mortality rate due to the drug's side effects is very high. In rural areas in third world countries people do not have access to health care facilities that are sophisticated in order to detect diseases early, and the fact that medications to treat diseases such as cancer and stroke are very expensive in the third world countries because they are exported from overseas.

When one is attempting to solve this problem, we need to look at two inter-related notions in chemical-biological interactions. In fact, the way in which NCEs interacts with a

target macro-molecule can be viewed from two extremes: On one hand, a drug can interact with a bio-macromolecule to stimulate a pharmacological or toxic response which is called a pharmacodynamics (PD) effect. On the other hand, when a bio-macromolecule interacts with a therapeutic drug by ADME the whole process is called a pharmacokinetic (PK) effect (Ekins, 2007). The PK and PD effects in an NCE are the major driving forces in its fate in the drug discovery and development process. The reason why PK and PD effects are determined early in the drug discovery process has a lot to do with identifying potency, toxicity and PK aspects early in the drug discovery paradigm to prevent late-stage withdrawals as is evident in Fig. 1.7 above.

To this end, the use of computational modelling in the design of NCEs is an attractive undertaking which will thus provide clues on the efficacy and safety issues related to NCEs earlier on in the discovery process (Lin, 1997; Van de Waterbeemd, 2003; Ekins, 2007; and Moitessier, 2008). **Note:** the novel compound is not yet a drug when it is in the design stage and hence the term NCE is used to denote a potential drug. A new chemical entity (NCE), which is sometimes referred to as a new molecular entity (NME), is according to the U.S. Food and Drug Administration, “*a drug that contains no active moiety that has been approved by the FDA in any other application submitted under section 505(b) of the Federal Food, Drug, and Cosmetic Act*” . Hence, NCEs and NMEs can be used interchangeable because they hold the same meaning. In this thesis we aim to use structure-based-drug design techniques coupled with ligand-based-drug design techniques to simulate PK and PD effects on NCEs when they bind with bio-macromolecules.

1.7 Aims and Objectives

The main aim of this work is to design NCEs using structure-based and ligand-based drug-design techniques. The primary goal is to study the NCE-macro-molecule interaction, inhibition to target enzymes, and metabolic enzymes respectively. Furthermore, the applicability of the models developed computationally would be tested on the methodology developed experimentally by making use of high-throughput *in vitro* micro-analytical assay to study the inhibition, distribution and metabolism of NCEs.

Objectives:

- 1) To perform molecular docking to predict the binding modes of catechin enantiomer–HSA interaction. Furthermore, Prime MM-GBSA was used as a docking post-processing method to predict the correct estimates of relative binding affinities and enantioselectivity of catechin enantiomers to HSA as case study I.
- 2) Validate the combined strategy involving a rigid and flexible docking method was optimized in order to predict the binding affinities and enantioselectivity of the interaction of warfarin enantiomers to HSA. To also predict the correct binding modes of warfarin-HSA explaining chiral recognition at atomic level as case study II.
- 3) To propose for the first time a combined computational strategy in order to generate information on the CYP17A1 inhibition where; (i) a 3D-QSAR pharmacophore model was performed on a diverse set of steroidal and non-steroidal CYP17A1 inhibitors obtained from literature with known experimental IC_{50} values; (ii) A Density Functional Theory (DFT) calculation was then used for evaluating electronic properties of selected inhibitors, which reflects their reactivity; and (iii) A Flexible ligand-protein Molecular docking method was validated by calculating the RMSDs of poses obtained and overlaid with conformers of co-crystallized ligands, and was then used on candidate compounds to confirm their agreement with the pharmacophore hypothesis as case study III.
- 4) To predict the site of metabolisms (SOMs) of probe substrates to Cytochrome P450 metabolic enzymes CYP 3A4, 2D6, and 2C9 making use of P450 module from Schrödinger suite for ADME/Tox prediction as case study IV.

- 5) To use the best fit model with good statistics from our 3D-QSAR pharmacophore model to perform a 3D-database search of molecules sharing similar pharmacophore features exhibited by the model which can be considered as potential inhibitors of CYP17A1 (Data not reported for confidentiality issues).

1.4 Thesis Outline

After giving a historical perspective of the computational models in drug discovery along with the role of chemists in the drug discovery process in this chapter, further chapters in this thesis are organized as follows:

1. Chapter 1: This chapter deals with the literature review based primarily on pharmacokinetic (PK) events with carrier protein (HSA) and therapeutic drug metabolizing enzymes such as cytochrome P450 superfamily. A discussion on enantioselective binding of therapeutic drug enantiomers to HSA. A further discussion involves enzyme kinetics with particular emphasis on the factors influencing the rates of substrate-enzyme recognition patterns is also presented in this chapter.
2. Chapter 2: This chapter deals with the state of the art in molecular modelling techniques, with special focus on molecular docking and pharmacophore modeling strategies of structure based and ligand based drug design.
3. Chapter 3: Involves the computational methodology used in this thesis via a combination of structure-based and ligand-based drug design methods.
4. Chapter 4: Deals with case study I which focuses on the prediction of binding affinities and enantioselectivity of catechin enantiomers to HSA using docking and MM-GB/SA.
5. Chapter 5: Deals with case study II which focuses on quantitative affinity prediction of warfarin enantiomers to HSA using structure based drug design approaches.
6. Chapter 6: Deals with case study III which focuses on the development of a pharmacophore model and database screening in order to design novel inhibitors.
7. Chapter 7: Deals with case study IV which focuses on the prediction of the site of metabolism of known substrates for known CYP isoforms.
8. Chapter 8: The concluding remarks summarize the novelty and achievement of this thesis.

9. Chapter 9: Deals with the current results, future outlook and advanced computational methods to address some shortcomings shown by structure-based and ligand-based drug design methods.

CHAPTER 2

LITERATURE REVIEW

2.1 A historical background of computational models in drug discovery

The advent and evolution of computers in drug discovery and development has been a game changer during the last two decades. Therefore, the role of a chemist involved in drug discovery has been broadened to include the evolution of *in silico* approaches. This is evident because *in silico* computational modelling is capable of speeding up the drug discovery process, and hence reducing the need for expensive laboratory synthesis, biological activities and pre-clinical studies which normally turn out to fail in clinical trials. The term *in silico* was first used at a workshop in 1989 at Los Alamos, New York by Pedro Miramontes on a paper entitled: “*DNA and RNA physicochemical constraints, cellular Automata and Molecular evolution*” (Miramontes, 1989). Later on in 1990, the word *in silico* appeared in his further work; where *in silico* was used to characterise biological experiments using a computer to model the *in vivo* or *in vitro* behaviour of new chemical entities (NCEs) (Miramontes, 1992). Currently, *in silico*/computational models have successfully been used by chemists, in medicinal chemistry research to better understand the way in which small organic molecules bind with macro-molecules to elicit their pharmacological/toxic responses at a molecular level.

The assumption made when using *in silico* approaches is based on a premise that the activity of a molecule and its’ properties is directly related to its’ biological activity (Chohan, 2008). This is advantageous because *in vitro/in vivo* experiments are not able to reveal the types of functional groups of the ligands and target proteins, enzymes or receptors responsible for that particular inhibition or affinity. On the other hand, *in vitro/in vivo* experiments are not able to model/simulate the correct orientation/conformation of a therapeutic agent in the active sites of the target enzymes, receptors or proteins. Therefore, improvements in software and hardware have enabled computational chemists to use data derived from *in silico* approaches to make informed decisions in the drug discovery process (Chohan, 2008). This is done before performing experiments, which is then followed by *in vitro* and *in vivo* biological tests. Even though molecular modelling methods are widely used in other fields such as Materials Sciences; the focus in this study is based specifically on the life-sciences field.

Earlier organic chemists used to experience problems when attempting to explore the chemical space (CS) when visualising small molecules. On the other hand, it was also difficult

to visualise macro-molecules, which means the biological space (BS) was difficult to explore as well. This problem persisted up until August Wilhelm von Hoffman developed a first physical molecular model in 1860 (Christoph, 1992). Stereochemistry was not yet discovered at that particular point in time. In 14 years after Hoffman's discovery, Jacobus Henricus van't Hoff and Joseph Le Bel discovered the tetrahedral geometry and stereochemistry of molecules in space represented in three dimensions (3D) (Le Bel, 1891; van't Hoff, 1874). Furthermore, van't Hoff established that most of the organic compounds exist in different conformational space (van't Hoff, 1875). In 1953, Derek Harold and Richard Barton further confirmed the theory on conformation suggested by van't Hoff (Ley, 2002).

Barton played a crucial role in making conformational analysis an essential part of organic chemistry. This was evident in his paper entitled: "*The conformation of the Steroid Nucleus*" in 1950 published in the journal of Experientia (Baron, 2001). These discoveries of molecular representations such as ball and stick models for molecular structures up to the point when computer generated models in 1970 revolutionized organic chemistry. Modern day computer models generated from 2D or 3D structures are now used in the drug discovery and development processes. In 1924, Louis de Broglie proposed that electrons show wave-like properties (de Broglie, 1924). Later on in 1926, Erwin Schrödinger, Weigner Heisenberg and Paul Dirac derived mathematical equations independently to further shed more light on the wave-like properties of electrons (Schrödinger, 1926; Born, 1926; Dirac, 1963). The theory of Wave Mechanics by Schrödinger and Quantum Mechanics by Heisenberg gave birth to the fundamental principle of bonding between the molecules. The basic principles of Quantum Mechanics (QM) in studying atomic and molecular orbitals is used in molecular modelling to calculate electronic properties that explains molecular reactivity such as Molecular Electrostatic Potential (MESP), Highest Occupied Molecular Orbital (HOMO), Lowest Unoccupied Molecular Orbital (LUMO), sampling of partial charges, and dipole moment. Furthermore, ligand conformations are required for most 3-D ligand-based and structure-based methods. The former involves pharmacophore modelling, shape-based screening, and 3D-QSAR model building. Whereas, the latter includes molecular docking, virtual screening workflows and e-pharmacophore models. Therefore, the work done by earlier organic chemists paved the way forward for computational and medicinal chemists to explore the chemical, conformational and biological spaces.

2.2. Molecular Modelling approaches

2.2.1 Molecular Docking – Structure-based drug design (SBDD) method

The invention of the DOCK method in 1982 was an early step in using computational software programme to study protein-ligand interactions, and this proved to be a game-changer in this field (Kuntz, 1982). The DOCK method initially adopted a geometric scoring criteria and later was extended to energy based scoring function for ranking binding poses (Mobley, 2009). The updated version of the DOCK method was followed by the development of very fast molecular docking methods in recent years such as GLIDE, AutoDock, FlexX, ICM, PMF, and GOLD (Zhong, 2007). Docking is a computational technique that places a small molecule in the binding/active site of a receptor/protein and then estimates its' binding affinity (Kroemer, 2007; Yuriev, 2009). Docking in its' broadest sense entails the change in conformation of the ligand resulting in the ligand's ability to orient itself in order to bind with the amino acid residues in the active site of the protein/enzyme/receptor. Docking can essentially be subdivided into three different categories:

- a) Rigid body docking, where both the receptor and the ligand are held rigid,
- b) Flexible ligand docking, where the receptor is held rigid and the rotatable groups of the ligand are allowed to be flexible,
- c) Flexible docking, where both the receptor and ligand flexibility are taken into account.

Docking approaches have proved to possess high speed and are therefore methods of choice for hit/lead identification in computer-aided-drug design (Mobley, 2009). Reviews published recently and comparative studies thereof; have revealed that Glide and Surflex outperforms other docking programs (Jain, 2003; Jain, 2007). A comparative study made by Zhou et al. on three widely used docking programs i.e. Glide, Gold, and Dock for virtual database screening when they are applied to the same target and ligand set. Glide Extra Precision (XP) in this study was shown to consistently yield enrichments superior to the two alternative methods, while GOLD outperforms DOCK on average (Zhou, 2007). In terms of docking accuracy several conclusions can be drawn like for example in a fair and exhaustive comparative study undertaken by Warren et al. In this study Warren et al. discovered that the accuracy of the docking program is highly dependent on the protein under study (Warren, 2006). Furthermore, according to their assessment they revealed that MVP, Glide, GOLD and Flexx may be considered as the best four. On another comparative study Englebienne et al. recently looked at the α -mannosidase and reported Glide as the best of the seven programmes studied (Englebienne, 2007). However, as it has been pointed out earlier docking programs are both

target- and ligand-dependant and should always be evaluated and validated for the protein understudy (Yuriev, 2009).

Docking algorithms uses a scoring function in order to predict the binding pose preferable after a ligand molecule is bound to the target receptor (Li, 2014). Therefore, the scoring function applied then selects the best pose for a given molecule and rank-order ligands according to their docking scores (Yuriev, 2009). In order to identify the energetically most favourable pose, each pose is evaluated or scored based on its fit to the target in terms of shape and properties such as hydrogen bonding, π -bonding and electrostatic interactions. In Glide, the binding free energy estimation is given as a docking score as shown in equation 5 below.

$$\text{GScore} = 0.05 * \text{vdW} + 0.15 * \text{Coul} + \text{Lipo} + \text{Hbond} + \text{Metal} + \text{Rewards} + \text{RotB} + \text{Site} \dots \dots \dots (5)$$

Where GScore is the Glide Score. In fact, the ligand's docking score is the sum of the Glide Scores plus the state penalty for a given protonation or tautomeric state for a ligand. The vdW is the Van der Waals energy term and is calculated with reduced net ionic charges on groups with formal charges. Coul is the Coulomb energy term, it is calculated with reduced net ionic charges on groups with formal charges, in a similar way as the van der Waals energy. Lipophilic term rewards favourable hydrophobic interactions. While the HBond term measures the number of hydrogen bonds between the ligand and the receptor. Metal-binding term includes only the interactions with anionic or highly polar acceptor atoms. Rewards and penalties terms for various features, such as buried polar groups, hydrophobic enclosure, correlated hydrogen bonds, and amide turns. RotB term refers to the penalty for freezing rotatable bonds. Lastly, the site term refers to polar interactions in the active site. Where polar but non-hydrogen-bonding atoms in a hydrophobic region are rewarded.

Therefore, a good score for a given molecule indicates that it is a good binder to the protein of interest. The pose is then rejected or accepted based on the score for that pose (Kroemer, 2007). The correlation of experimental binding affinities with their analogous docking scores appears to be more challenging than searching for the best ligand conformation and their orientations (Murcko, 1995; Tame, 1999; & Bohm, 1999). Docking score's ability to distinguish between binders and non-binders varies widely depending on the target protein (Graves, 2008). Furthermore, binding affinity prediction using docking methods is a very difficult undertaking because all docking programs currently in the market are still struggling to address this problem. It is an area in which software developers are still working on. The challenges facing this very promising field include the following: receptor quality, side chain

and backbone flexibility of the receptor, structural waters, ligand representation, solvation and entropy of binding.

On the other hand, another method of note developed to predict correct relative binding affinities is Molecular Mechanics with Poisson-Boltzmann Surface Area (MM-PB/SA) calculations, pioneered by Kollman and co-workers. This method uses a combination of molecular mechanics and continuum solvation to compute average binding energies for configurations extracted from MD simulations of the unbound and bound states (Kuhn, 2000). The impressive results obtained by this methodology encouraged some authors to use molecular mechanics based scoring functions with Generalised Born Surface Area (GB/SA) as the implicit solvent model in the rescoring process (Cheatham, 1998; Srinivasan, 1998; Vorobjev, 1999; Kollman, 2000). MM-GB/SA is more computationally demanding than docking methods, it is aimed at estimating the change in enthalpy of bound and unbound states of the ligands to the target. The binding free energy estimates includes the enthalpy change and the change in solvation free energy from the implicit solvent model (Mobley, 2009). The main drawback of MM-PB/SA or GB/SA are that both of the methods, are occasionally not predictive and they need prior knowledge of a likely bound complex as a starting point, although such starting conformations can be taken from docking generated poses as a post-processing stage (Kuntz, 1982).

In addressing these challenges molecular dynamics simulations (MD) and Monte Carlo simulation (MC) coupled with free energy perturbation (FEP) or thermodynamic integration (TI) calculations have proved to be most rigorous computational approaches used to estimate relative binding affinities of protein-ligand complexes as docking post-processing methods (Jorgensen, 1989; Kollman, 1993; Jorgensen, 1998; Tame, 1999; Pearlman, 2001; Simonson, 2002; Guimarães; 2005). Even though these methods have yielded impressive results for several protein-ligand systems, they are however, computationally exhaustive and have generally been applied to study a small number of ligands in congeneric series (Guimarães, 2008). Current docking programs provides good-quality binding poses, an energy function with a more physically reasonable description of binding contributions is needed as a docking post-processing step to accurately predict the binding affinities of protein-ligand complexes. This initiative will enable the design of tight-binding ligands in hit-lead identification/lead optimisations.

The Free-Energy Perturbation (FEP)/Replica Exchange with solute tempering (REST) algorithm has recently been developed by scientist at Schrödinger. Wang et al. reveals that the FEP/REST protocol is able to accurately predict the binding affinities of a diverse range of ligands and targets (Wang, 2015). This work was further validated with a project where Scientists at Schrödinger used 1000 molecules on different targets such as Kinases, Proteases, Bromo domains, GPCRs, and Protein-Protein Interactions to accurately predict the binding affinities using the FEP/REST protocol (Sherman, 2015). Wang et al. further used the FEP/REST method to predict the correct binding affinities of NCEs which were then sent for synthesis and were further tested for their biological activities. The experimental binding affinities predicted by using an *in vitro* assay agree with the binding affinities obtained by using FEP/REST protocol (Wang, 2015). Furthermore, a retrospective comparison between FEP/REST with Glide SP, Glide XP and MM-GB/SA revealed that FEP/REST accurately predicts the binding affinities of protein-ligand complexes and outperforms other Schrödinger structure-based drug design methods. In fact, FEP/REST obtained an R^2 of 0.75, followed by MM-GB/SA with 0.35, and Glide SP with 0.29 when the predicted binding affinities were correlated with experimental binding affinities (Wang, 2015; & Sherman, 2015). Further study to extend on the ground breaking work by Wang et al. has recently been reported by Steinbrecher et al. on a diverse set of seven targets i.e. Lysozyme, DNA Ligase, Mcl-1, MUP-I, LDH, JAK-2, HSP90, and p38 Map Kinase. In this validation study FEP+ protocol was used, where relative binding affinities were directly estimated and subsequently correlated with known experimental binding affinities. The results of this study revealed that FEP+ method correctly predicts the binding affinities of diverse set of target (target-binding selectivity). Further results reveals that FEP+ method is able to correctly estimate the binding affinity of strong, medium and weak binders. This therefore, means that the FEP+ method proposed in this study can be transferable to any target system because of its ability to distinguish between binders and non-binders (Steinbrecher, 2015).

A comparative study on the performance of FEP+ when compared with other methods such as docking and MM-GB/SA reveals that FEP+ outperforms the two molecular modelling methods (Steinbrecher, 2015). The accuracy of the FEP/REST method or FEP+ method can be attributable to the improved force-field with an updated version of OPLS-2005 to the use of OPLS-2.1 force-field by a correct depiction of organic compounds during conformational search which has led to the improvement of results. The FEP+ or FEP/REST protocol has been implemented to run on a graphics processing unit (GPU), instead of a central processing unit (CPU) which has improved the speed and performance (Wang, 2015; Sherman, 2015;

Steinbrecher, 2015). This approach is relatively new and was not used in this study because it was reported after we had already developed the models for the design of novel inhibitors.

2.2.2 Pharmacophore modelling -Ligand-based Drug Design (LBDD) Method

Pharmacophore modelling is a ligand based drug design (LBDD) method that has recently proved to be one of the most powerful and fast methods in the design and subsequent discovery of new chemical entities (Elumalai, 2012). According to IUPAC, Pharmacophore modelling can be defined as “an ensemble of steric and electronic features that is necessary to ensure the optimal intermolecular interactions with a specific biological target and to trigger or block its biological activity”, (Sakkia, 2012, pg. 67). An extension to this definition is made by Lu et al. where they define a pharmacophore as a set of structural features that are common to a set of compounds and are important for the increase in the binding affinity of that particular compound to the target receptor (Lu, 2011). Pharmacophore modelling seek to identify common structural features that aid in high affinity of hits and which are absent in low affinity structures (Sakkia, 2012). Therefore, a good pharmacophore model should possess all the chemical features of the molecule represented in a 3D-space. Which in turn provide a representation of a conformational space that yield pharmacophore features important for biological activities (Lu, 2011).

The pharmacophore model is developed by searching a 3D space of pharmacophore features in the training set. The pharmacophore model is validated by searching a 3D space of test molecules in the data set (John, 2010). The use of 3D-QSAR pharmacophore models in the design of new chemical entities using a database search in order to predict their biological activities of a set of compounds using statistical correlation methods such as Partial Least Squares (PLS) regression, has been a game changer in LBDD methods (John, 2010; Tanwar, 2013). Furthermore, Pharmacophore modelling together with regression methods such as Partial Least Squares (PLS), Principal Component Analysis (PCA), Multiple Linear Regression (MLR) (Elumalai, 2012; Prasad, 2013; Vyas, 2013; Khanfar, 2013; Lokwani, 2013; Tanwar, 2013; Jain, 2013). Furthermore, statistical methods such as Comparative Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA) have also been used successfully (Sachin, 2007; Telvekar, 2008, Zhang, 2013). These multiple regression methods are applied for QSAR methods in this research area and have been used extensively in the design and prediction of biological activities of new chemical entities in computer aided drug design initiatives (Lu, 2011; Elumalai, 2012; Khanfar, 2013).

The joint use of 3D-QSAR pharmacophore model and molecular docking has proved to be a powerful predictive tool for new chemical entities prior to synthesis and this aid in

understanding the binding modes of protein ligand complexes. This helps to further gain more insight for further lead optimization of protein ligand complexes (Lu, 2011; Kirubakaran, 2013). The ability of the 3D-QSAR pharmacophore model to accurately predict the binding affinities of new chemical entities is a very interesting undertaking. This is because NCEs can then be synthesized in an organic laboratory and subsequently be tested for their PK and PD effects which minimises time and costs of the drug-discovery process due to late-stage failures as has been pointed out in the statement of the problem section above.

CHAPTER 3

METHODOLOGY

3.1 Computational Methods

The computational approaches used in this thesis include the combination of structure-based drug design (SBDD) and ligand-based drug design (LBDD) methods discussed in Chapter 2. A combination of these methods is very powerful in the design of new chemical entities prior to synthesis and assay. The specific protocols implemented in this study, are explained in the form of case studies. However, as a test case, Fig. 3.1 shown below illustrates the process followed in the design of NCEs for the treatment of prostate cancer (PC). Firstly, molecular docking methods such as Glide-SP/XP, QPLD, MM-GB/SA, and IFD were validated in Paper I and II (Sabela, 2012; Gumede, 2012). Secondly, cross-docking methods were applied on different chemical targets of pharmaceutical relevance aimed at reproducing the bound conformations of the ligands to the targets in X-ray crystal structures (unpublished data). A Pharmacophore model was then developed for molecules with available IC_{50} 's obtained literature. A 3D-QSAR model was then developed from the hypothesis using PLS regression and predicting the pIC_{50} values of the training and test set molecules in the data set. Finally, the 3D-QSAR model was then used for a database screening of drug-like molecules from the database of 2 million compounds (Unpublished data). The site of metabolisms was predicted using P450 SOM module.

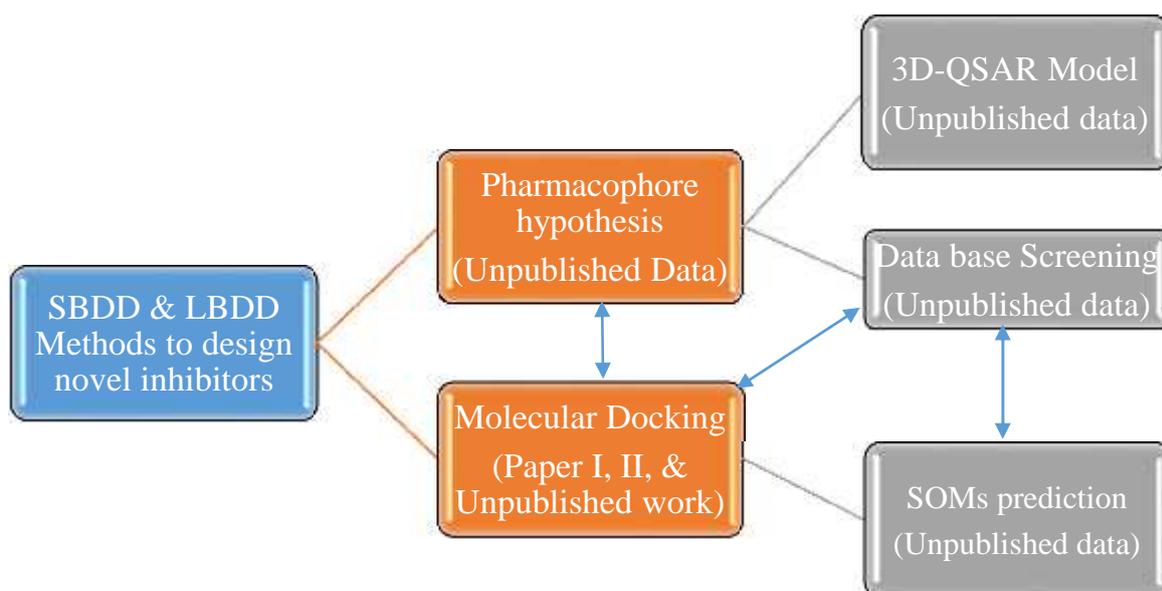


Fig. 3.1 Horizontal hierarchy depicting combined SBDD & LBDD methods.

3.2 *In vitro* Bio-analytical Assay Methods

For the *in silico* computational methods illustrated in Fig. 3.1 above, the steps followed in the synthesis, structure elucidation, and biological activities for the target and off-target interactions of the hits and their derivatives are shown as a vertical hierarchical process depicted in Fig. 3.2 below. However, the *in vitro* experimental results from the protocols as well as the synthetic schemes and structure determination of the hits are illustrated in Fig. 3.2 below will not be disclosed in this thesis because the aspects of this methodology form part of the invention which is in the provisional patent.

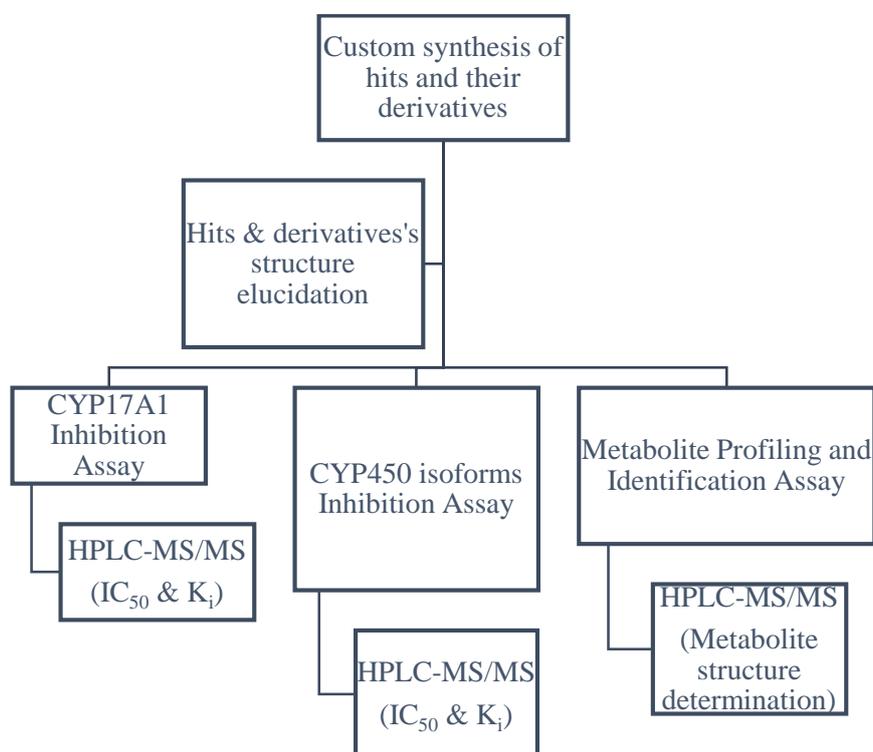


Fig. 3.2 Vertical hierarchical process for the experiments planned for the synthesis, structure determination, and *in vitro* bio-analytical assays for NMEs target and off-target interactions.

In order to test the biological activities of the hits with CYP17A1 enzyme an HPLC-MS/MS method was developed in order to estimate the IC_{50} of CYP17A1-Hit inhibition. In fact, thirteen hits were obtained from the database screening using the pharmacophore model as the search query. The candidate compounds/hits, at seven concentrations, were incubated with microsomes containing heterologously expressed CYP17A1 and the relevant probe substrate (progesterone) at eleven different concentrations. Following an appropriate incubation period at 37 °C, the reactions were terminated by the addition of an organic solvent

and the production of metabolite quantified by LC-MS/MS. Solvent controls, indicating the maximum metabolite produced in the absence of any inhibition, were included in the experimental design. The Michaelis-Menten (V_{max} and K_m) parameters for the production of probe metabolite were determined for each assay condition via non-linear curve fitting and the IC_{50} and K_i determined using an appropriate model of inhibition (competitive, noncompetitive or uncompetitive). The percent inhibitions versus Log_{10} compound concentration data were plotted and the IC_{50} determined using a sigmoidal dose response equation in GraphPad prism software.

Another HPLC-MS/MS methodology was developed to study the inhibition of CYP450 isoform-hit inhibition in the presence of probe substrates. In this methodology, test compounds at six concentrations were incubated at 37 °C with microsomes containing heterologously expressed specific Cytochrome P450 isoforms. The compound effects on metabolic capability were investigated by monitoring the production of metabolites of probe substrates for each isoform using LC-MS/MS analysis. Solvent controls were included to indicate the maximum amount of metabolite produced in the absence of any inhibition or substrate competition. The percent inhibition versus Log_{10} compound concentration data was plotted and the IC_{50} determined using a sigmoidal dose response equation in GraphPad prism. The six major human isoforms measured were, CYP1A2, CYP2C8, CYP2C9, CYP2C19, CYP2D6 and CYP3A4. The Michaelis-Menten (V_{max} and K_m) parameters for the production of probe metabolite were determined for each assay condition using non-linear curve fitting and the K_i and IC_{50} determined using an appropriate model of inhibition (competitive, noncompetitive or uncompetitive). Furthermore, a metabolite profiling and identification assay was developed in order to establish the metabolic profiles of the metabolites formed using selective inhibitors to measure the extent of metabolism, using hepatocytes of different species such as human, dog, rat and mouse. Phase 1 and 2 metabolisms for each species was subsequently measured. The results of these developed methods are important in predicting the pre-clinical behaviour of the molecules to humans which could avoid late-stage withdrawals.

Table 3.1 shown below depicts the probe substrates, selective inhibitors and the mode of metabolism of the substrates to CYP450 isoforms. Ideally, the probe substrates and selective inhibitors were used to develop the methods for the metabolism of the hits to the target and metabolic enzymes in this thesis. The results will be shown in Case study IV.

Table 3.1. The probe substrates, the type of metabolites formed, selective inhibitors and the mode of metabolism that occur when designing and validating the methods for target and off-target interactions of hits.

Enzyme	Probe Substrates	Metabolite	Selective Inhibitor	Mode of metabolism
1. CYP17A1	Testosterone	6 -hydroxytestosterone	Ketoconazole	6 -hydroxylation
	Progesterone	17 -hydroxyprogesterone		17 -hydroxylation
2. CYP3A4	Midazolam	1'-hydroxymidazolam	Ketoconazole	1'-hydroxyzylation
	Testosterone	6 -hydroxytestosterone		6 -hydroxylation
3. CYP1A2	Phenacetin	Acetaminophen	-Naphthaflavone	O-deethylation
4. CYP2C8	Paclitaxel	6 -hydroxypaclitaxel	Montelukast	6 -hydroxylation
5. CYP2C9	Diclofenac	4'-hydroxydiclofenac	sulfaphenazole	4'-hydroxylation
6. CYP2C19	S-Mephenytoin	4'-hydroxy-S-mephenytoin	S-benzylnirvanol	4'-S-hydroxylation
7. CYP2D6	Dextromethorphan	Dextrorphan	Quinidine	Demethylation

CHAPTER 4

CASE STUDY I

The use of molecular docking and MM-GB/SA to estimate the enantioselectivities of (+/-)-catechin enantiomers to Human Serum Albumin (Sabela, 2012).

4.1. INTRODUCTION

The Grid-based Ligand Docking with Energetics (Glide) protocol uses hierarchical filters to search for conformers of ligands in the active site of the receptor. The Glide protocol can be performed on the rigid or flexible modes. The rigid mode follows a lock and key approach. While the flexible mode, allows for conformational changes for each ligand in the data set. A summary of the steps that Glide docking follows in its workflow is illustrated in Fig. 4.1.

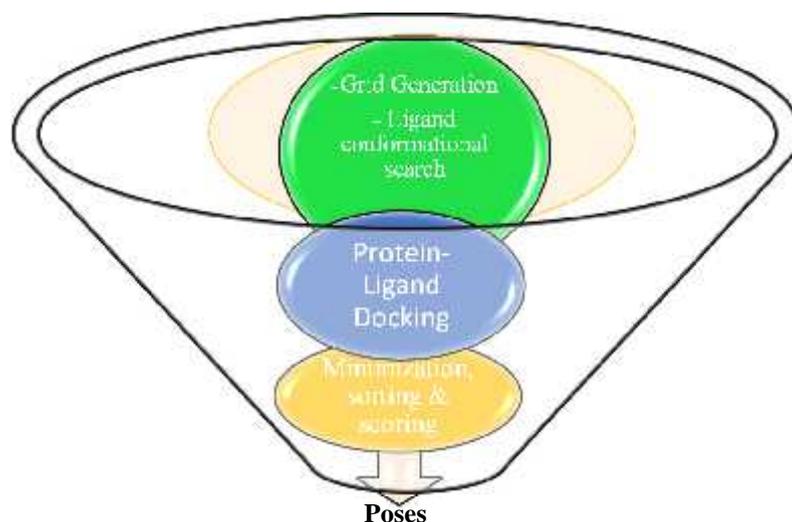


Fig. 4.1 Glide docking protocol for protein-ligand binding affinity approximation.

In fact, Glide uses hierarchical filters to evaluate the binding of the ligand to the receptor (Friesner, 2004; Halgren, 2004; Friesner, 2006). The grid file for the protein co-crystallized with the ligand to mark the active site of the ligand, presents shape and properties of the receptor that offers a progressive increase in scoring accuracy of protein-ligand poses (Glide v5.0, 2010; Friesner, 2004). Glide program has been able to predict the binding modes and binding affinities of protein-ligand complexes in the past. However, the docking scores are far from being able to predict the relative binding affinities that correlates with experimental binding affinities so far, as discussed in Section 2.

In order to account for the shortfall of Glide, for not being able to predict the relative binding affinities of protein-ligand poses, Molecular mechanics (MM) generalized born surface area (MM-GB/SA) method was used as a docking post-processing method, to calculate ligand binding free energies and ligand strain energies for a set of ligands and a single receptor (Prime v2.2, 2010).

As a case study, a polyphenolic compound catechin which falls under natural products called flavonoids was chosen for this study. Recently, more attention has been devoted to catechin due to their beneficial health effects, mainly as antioxidants (Soares, 2007). Catechin exists as four stereoisomers because they have two chirality centers. However, the (+)-catechin (2R-3S) and (-)-catechin (2S-3R) enantiomers are more biological active among these stereoisomers. Therefore, they are appropriate as a test subject for this case study. Since, the bioavailability and ADME profile for these flavonoids remain sketchy (Ishii, 2010). Molecular modelling methods are very attractive in that they have the ability to reveal the forces giving rise to the binding between the ligand and a receptor at molecular level. Accordingly, in this case study the main aim was to perform molecular docking to predict the binding modes of catechin enantiomer–HSA interaction. Furthermore, Prime MM-GB/SA was used as a docking post-processing method to predict the correct estimates of relative binding affinities and enantioselectivity of catechin enantiomers to HSA.

4.2 MATERIALS AND METHODS

4.2.1 Computational Details

Schrödinger's Maestro 9.1 was used as the primary graphical user interface (GUI) (Maestro v9.1, 2010) where all computational calculations were made. The Ligands were prepared using Ligprep (Ligprep v2.4, 2010). The proteins were prepared using the Protein preparation wizard (Schrödinger Suite 2010, 2010). All docking calculations were undertaken using Glide 5.0 (Friesner, 2004; Halgren, 2004; Friesner, 2006). The docked poses were further used for the calculations of relative binding free energies by using Prime MM-GB/SA module (Prime v2.2, 2010).

4.2.2 Ligand preparation

Ligprep was used to prepare 3D coordinates for catechin enantiomers. Protonation and tautomeric states were generated by using Epik a module from Schrödinger suite at pH 7.4 in order to mimic physiological conditions. All possible chiralities were generated on the structures. One low energy ring conformation of each of the ligands under study was generated using OPLS 2005 force field.

4.2.3 Glide Grid generation

The Constraints tab of the Receptor Grid Generation panel was used to define Glide constraints for the receptor grids to be generated. Glide constraints are receptor-ligand interactions that are believed to be important to the binding mode, based on structural or biochemical data. Setting constraints enables Glide to screen out ligands, conformations, or poses that do not meet these criteria early on in their evaluation for docking suitability (Glide v5.0, 2010). Positional constraints were selected by looking at a functional group in the ligand where a hydrogen bond can likely to occur. We used a prepared crystal structure of a receptor ligand complexes i.e. 2BXD for site I and 2BXF for site II. The amino acid residues responsible for hydrogen bonding for site I is Tyr150 and Tyr411 for site II in HSA. These two amino acids were chosen as hydrogen bond constraints for the grid files to mark active site I & II in HSA. The grid files for 2BXD and 2BXF where then prepared.

4.2.4 Glide Docking method

The 3D conformers of catechin enantiomers that were previously prepared in section 4.2.2 above were used as starting structures. The grid files for protein-ligand complexes prepared in section 4.2.3 above were added into the Glide panel. The prepared ligands were docked flexible into the active site of the protein using Glide Extra Precision protocol (Friesner, 2004; Halgren, 2004; Friesner, 2006).

4.2.5 MM-GB/SA methodology

Molecular mechanics generalized born surface area (MM-GB/SA) method was used to predict the free energy of ligand binding to protein. In this study, MM-GB/SA calculations were carried out using the Prime MM-GB/SA module (Prime v2.2, 2010). This procedure was

employed as a docking post-processing step. Therefore, docked ligand poses generated with GlideXP in the previous section were used to calculate ligand binding energies and ligand strain energies for both (-)-Catechin and (+)-Catechin and a single protein HSA, using the MM-GBSA technology available with Prime.

4.3 RESULTS AND DISCUSSION

4.3.1 Molecular Docking on (±)-Catechin Enantioselectivity to HSA

Some preliminary tests were performed computationally in order to explore the potential of molecular docking methodologies with an aim to measure their synergy with the experimental estimation of enantioselectivity (chiral recognition) (Sabela, 2012). In this work Schrödinger molecular modelling software most specifically Glide, IFD, and MM-GB/SA were used. Docking programs are able to map the binding mode of protein-ligand poses. Therefore, it was important to implement a combined strategy to complement docking scores with relative binding affinities obtained from MM-GB/SA to predict the enantioselectivity and binding affinities of enantiomers of catechin to HSA. Therefore, a semi-quantitative study that aims to estimate the *ES* (a measure of a relative extent of affinity data involving two enantiomers; with the same molecular structure) could be valuable as a starting point for other researchers, for comparison purposes (e.g. GLIDE protocols, or using other docking programs etc.) (Sabela, 2012). In this case study, the relative binding free energies (ΔG°), were measured, which provided us with a parameter related to the affinity constant K_a that is obtained by bioanalytical methods. Equation 6 shown below depicts the direct relationship between ΔG and K_a .

$$G^\circ = - RT \ln K_a \dots\dots\dots(6)$$

In this study, it has been assumed, as a hypothesis, that the ratio between ΔG° for enantiomers should represent a relative enantioselective binding quantity, implying that $\Delta G^\circ (-)\text{-C} / \Delta G^\circ (+)\text{-C}$ representing a pseudo estimate of enantioselectivity (*ES*), which can be compared with the experimental *ES* estimation (1.5 ± 0.2) (Sabela, 2012).

The Human Serum albumin protein has several binding sites in different subdomains binding endogenous, exogenous and fatty acids (Ghuman, 2005). Table 4.1 as shown below, outlines different PDB structures of HSA with co-crystallized ligands in different binding sites. Several subdomains in the HSA were investigated with the GLIDE-Prime MM-GB/SA

protocol through the complexes available in the Protein Data Bank (subdomains; IIA from 2BXD, IIIA from 2BXF, IB from 2VUF, IIIB from 1HK1, IIA - IIB from 2BXG and cleft from 2BXN). Only in the first two cases, a pseudo- $ES > 1$, consistent with the experimental ES value, was encountered (in the remaining cases, values close or below unity were observed).

Another hypothesis, related to the (\pm)-catechin features, was assumed. Since the molecule show low hydrophobicity with a $\log P$ (o/w) = 0.4 using Qikprop module at pH = 7.4 which is almost neutral (Herrero-Martinez, 2005), we have assumed that the observed ES to HSA (i.e. the slight affinity differences between enantiomers, in favour of (-)-C) could be mainly related to the hydrogen bonding interaction (rather than to other effects such as hydrophobic contacts). The H-bond interactions predicted by Glide-Prime MM-GB/SA are included in Table 4.2 (see Fig. 4.2 below for structure of catechin with atom numbering).

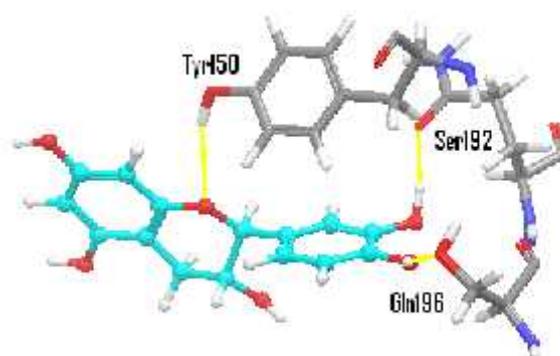


Fig. 4.2 Experimental and docking synergy for enantioselectivity: H-bonding interactions between (-)-catechin and human serum albumin (HSA, site I) residues, predicted by molecular docking. The results are consistent with the experimental enantioselectivity data for (\pm)-catechin-HSA interaction using capillary electrophoresis and a novel mathematical approach allowing statistical advantages (Sabela, 2012).

Table 4.2 below shows some detailed GLIDE-Prime MM-GB/SA results for subdomains IIA and IIIA (sites I and II, respectively).

Table 4.2. Molecular docking results for the best pose predicted by Glide-Prime MM-GB/SA^[a] consistent with the experimental *ES* value (1.5 ± 0.2). The hydrogen bonding (H-bond) interaction (ranked according to bond radii) is indicated. (Sabela, 2012).

HSA site ^[b] (subdomain; PDB ID)	Enantiomer	Group (Ring, atom)	HSA residues (contact)	H-bond distance (Å)	UG (kcal/mol)	Pseudo-ES ^[c]
I (IIA; 2BXD)	(-)-C	OH (B3')	Gln196 (NH)	1.738	-27.25	1.60
		OH (B4')	Ser192 (CO)	1.971		
		O (Glycoside)	Tyr150 (OH)	2.430		
		O (C3)	Lys199 (NH)	1.905		
	(+) -C	O (A5)	Tyr150 (OH)	2.003		
		OH (B3')	Glu292 (O)	2.091	-17.01	
		O (C3)	His242 (NH)	2.134		
		O (A5)	Arg257 (NH)	2.153		
II (IIIA; 2BXF)	(-)-C	OH (B3')	Ser489 (CO)	1.762	-25.47	
		OH (A3)	Ans391 (CO)	1.808		
	(+) -C	O (Glycoside)	Ser489 (OH)	2.153		1.25
		OH (A3)	Ser489 (O)	1.815		-20.41
		O (B3')	Lys414 (NH)	2.054		

^[a] Schrödinger's Maestro 9.1 software. (±)-catechin enantiomers and HSA were prepared at pH 7.4 to mimic physiological conditions. Docking calculations were undertaken by using Glide 5.6. (Glide v5.6, 2010) The docked poses by GlideSP (Halgren, 2004) were re-docked (GlideXP (Friesner, 2006). The resulting poses were post-processed by using a molecular mechanics (MM) based scoring function with the Generalized Born (GB) model as the implicit solvent model (MM-GB/SA), to calculate relative binding free energies, UG°, Prime MM-GB/SA (Kawatkari, 2009). The HSA flexible region was chosen as any residue within 12 Å of the ligand in each active site.

^[b] PDB database (<http://www.rcsb.org>) was used to obtain the computational information for HSA complexes. Site I was set from the PDB Warfarin-HSA complex (2BXD), Site II from the PDB Diazepam-HSA complex (2BXF) (Ghuman, 2005).

^[c] The ratio of the free energy change, UG (-)-C/UG (+)-C, was used as a pseudo-enantioselectivity approximation.

Site I appears in the subdomain with a greater degree of enantioselectivity with a (pseudo-*ES* = 1.6). Docking results for site I reveals that (-)-Catechin undergoes a change in conformation during docking, more specifically on the B ring. The change in conformation suggests a stronger H-bond between GLN196 and the B4' ring with the shortest bond radii of 1.738 Å; resulting in the strengthening of the interactions. This is a significant result and could be vital in explaining the favourable pseudo-*ES* for this enantiomer, in agreement with the proposed hypothesis. However, the number of H-bonds seemed to be less relevant, as well as the number of hydrophobic contacts predicted by GLIDE-Prime MM-GB/SA although in the case of site I, it is consistent with the high pseudo-*ES* associated to this subdomain (Sabela, 2012).

4.4 CONCLUSIONS

The results of this study suggest that the reliability of the GLIDE-estimations are target-dependent and still deserves more validation and verification using different types of targets. However, this preliminary study suggests that the interaction of enantiomers in site I and site II of HSA can be postulated as the most plausible in reality. The stronger hydrogen bond interactions between the hydroxyl group of the B-ring of (-)-catechin, after a conformational change due to the flexibility of HSA residues, is key in explaining the moderate *ES* observed experimentally (Sabela, 2012). Although the results have to be viewed with caution, they however provide an initial finding, to be compared with other molecular modelling software programs/protocols. Molecular docking is still far to be considered as an accurate or fully-validated methodology to estimate binding affinities. However, the study of relative chiral-recognition such as pseudo-*ES* values, in comparison to experimental *ES* values is able to shed more light on the forces giving rise to binding affinities at atomic level. This strategy could further broaden the possibilities of a synergy between experimental and computational methods in two extremes where: (i) in one extreme, docking could help to explain at the molecular level the *ES* results found experimentally and (ii) on another extreme, where experimental *ES* values could serve to validate docking approximations.

CHAPTER 5

CASE STUDY II

Enantioselective binding of Warfarin enantiomers to Human Serum Albumin using Molecular Modelling approaches

5.1 INTRODUCTION

Warfarin is widely used as an anticoagulant and frequently used as a rodenticide (Hirsh, 1998; Porter, 2010). Warfarin was firstly synthesized in 1950 by Seidman et al and was later commercialised under the trade name Coumadin Sodium (Seidman, 1950; Link, 1959). A three dimensional structure of warfarin is shown in Fig. 5.1 below. Warfarin is commercially prescribed as a racemic mixture in a 1:1 ratio of R- and S- enantiomers (Kaminsky, 1997; Zou, 1998; Jones, 2010). Both of warfarin enantiomers elicit their therapeutic effect by inhibiting the reduction of vitamin K 2,3-epoxide to vitamin K hydroquinone by vitamin K epoxide reductase (VKOR) (Suttie, 1987; Rost, 2004).

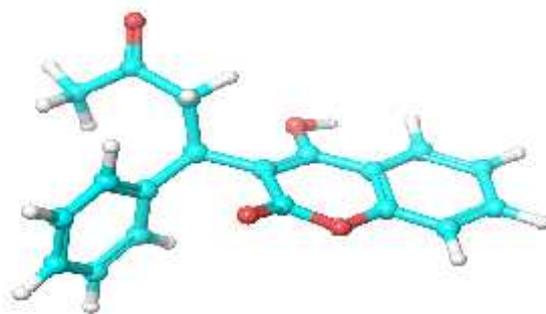


Fig. 5.1 A 3-Dimensional representation of the structure of warfarin.

The inhibition mechanism of vitamin K by warfarin occurs after the epoxide form of vitamin K is reduced to vitamin K quinone (Tie, 2008). The S-enantiomer of warfarin is more potent as an anticoagulant than the R-enantiomer in both rats and in men, with a potency of 2-5 times more than that of its mirror image (Breckenridge, 1974; Yacobi, 1974; Zou, 1998). Most of the pharmacological processes in living organisms responsible for drug action in the body, presents a higher degree of enantioselectivity resulting in a difference between the activities of drug enantiomers (Gumede, 2012). Since, the pharmacological processes in the body give rise to a high degree of enantioselectivities resulting from the differences between the activities of drug enantiomers. More specifically, the eutomers which elicit a major

therapeutic effect and to a lesser extent the distomers, which possesses no/minor therapeutic effects or eliciting the toxic effects (Chuang, 2006; Brooks, 2008). Therefore, HSA binding is important in solubilizing compounds that can aggregate and be poorly absorbed or distributed to their targets. HSA has up to now been a main focus of attention in the pharmaceutical industry because of its ability to bind a variety of endogenous and exogenous compounds (Gumede, 2012).

The use of molecular modelling methods to study the binding modes, binding affinity, and enantioselectivity of warfarin enantiomers to warfarin enantiomers to HSA is aimed at answering the following questions that experimental methods are failing to answer: (1) which enantiomer of warfarin binds with high affinity to HSA? (2) Which tautomeric state of warfarin is responsible for its binding to HSA under physiological conditions? (3) To establish whether warfarin enantiomers when bound to HSA undergoes some conformational changes. In fact, it has been postulated in the literature that the binding of warfarin to HSA follows a two-step binding model, since the reaction is reversible (see equation 5 below). This model assumes that the binding of warfarin in the first step is fast and follows a lock and key approach (Bos, 1989). The second step occurs by a change in conformation in HSA in order to accommodate warfarin in its active site (Kremer, 1982).

Equation 6 shown below indicates that ΔG°_b is directly related to the experimentally determined binding constant K_a . Where R is the universal gas constant, T is the temperature in Kelvin.

$$G^{\circ}_{\text{bind}} = - RT \ln K_a \dots\dots\dots(6)$$

The enantioselectivity of enantiomers is related by equation (7), when one enantiomer shows high affinity to the receptor than the other enantiomer, enantioselectivity gives rise to the ratio of the binding affinity for the two enantiomers (Haeffner, 1998).

$$r_{R,S} = \frac{K_R}{K_S} \dots\dots\dots(7)$$

The magnitude of the $r_{R,S}$ can be related to the free-energy difference of the enantiomeric association equilibria between chiral enantiomers and the protein, as shown in equation 5 above. Therefore, this can be given by equation (8) below

$$G^{\circ}_{\text{bind}} = - RT \ln \dots\dots\dots(8)$$

The difference in the free-energy of binding between the two enantiomers ($G^{\circ}_R - G^{\circ}_S = G^{\circ}$) and enantioselectivity can be represented according to equation (9) below.

$$G^{\circ}_{\text{bind (R, S)}} = -RT \ln \dots\dots\dots(9)$$

The above equation demonstrates that there is a direct relationship between the difference in the free-energy of binding for enantiomers and enantioselectivity. In this work, a combination of rigid and flexible SBDD methods were optimized in order to address the following two objectives: (i) to predict the affinity/enantioselectivity binding parameters that can be directly compared to *in vivo/in vitro* binding affinities for the binding of warfarin enantiomers to HSA as a test case and (ii) to predict the correct binding modes of warfarin-HSA explaining chiral recognition at atomic level (Gumede, 2012).

5.2. MATERIALS AND METHODS

5.2.1 Software Methods

Computational experiments were performed by using Maestro 9.2 (Maestro, 2011) GUI in the Schrödinger 2011 suite. A series of modules from Schrödinger, such as Glide 5.7 (Glide, 2011) and Macro Model 9.9 (MacroModel, 2011) were used in this study (Gumede, 2012).

5.2.2. Protein Selection and Preparation

The X-ray crystal structure of HSA bound to R-warfarin (PDB code 2BXD) with a resolution of 3.05 Å, was uploaded from the protein databank (PDB). The three-dimensional (3D) crystallographic HSA structure was obtained from the RCSB Protein Data Bank (PDB). Specifically, the crystal structure of HSA co-crystallized with R-warfarin in the active site I (PDB code 2BXD; subdomain IIA), with a resolution of 3.05 Å, was selected. The downloaded structure was subjected to Maestro's protein preparation module (Impact, 2011). The following steps were accomplished: (i) Hydrogen atoms were added to the crystal structure. (ii) The side-chain residues of Glycine and Aspartic acid were allowed to rotate in order to maximize hydrogen bond interactions. (iii) The Schrödinger's Prime 3.0 module (Prime, 2011) was used to fill in the missing side chains. (iv) Water molecules within 5 Å of the co-crystallized ligand were removed. (v) The pH of the entire system was adjusted to 7.4 using Epik (Epik, 2011). (vi) Hydrogen bonding network was optimized, and finally a geometry optimization was

performed to a maximum Root-Mean-Square Deviation (RMSD) of 0.30 Å using OPLS_2005 force-field (Gumede, 2012).

5.2.3. Ligand Preparation

Structures of warfarin enantiomers were prepared by using Schrödinger's Ligprep v2.5 (Ligprep, 2011) in order to generate 3D coordinated of the ligands and tautomeric states at pH 7.4. For this purpose, enantiomeric pairs for warfarin were generated. Additionally, the ionization states for the tautomeric forms of warfarin were predicted using Schrodinger's Epik (Epik, 2011) module at pH 7.4 (ionization constant, $pK_a = 5.1$ for warfarin enantiomers was predicted). The extensive conformational search was performed using the Monte Carlo Multiple Minimum (MCMM) search algorithm implemented in Schrodinger's MacroModel 9.9 program coupled with OPLS_2005 force field under implicit solvent conditions using the GB/SA approximation. Thereafter, the ligands were subjected to a series of energy minimization steps, using the Molecular Mechanics (MM) minimization with PRCG (Polak-Ribiere-type conjugate gradient) method (Gumede, 2012).

5.2.4. Docking Protocols

A series of docking methods were used in this study. Firstly, GlideXP was used to generate initial poses for warfarin-HSA complex.

5.2.4.1 Glide XP procedure

Schrödinger's docking algorithm implemented as Glide (Grid based ligand docking with energetics) was used to predict the binding affinities and binding geometries for the HSA-ligand complexes, marking the hydrogen bond and the position of the ligand in the active site I. Initial docking calculations were performed in extra precision mode using GlideXP (Glide, 2011) with standard van der Waals scaling of 0.8 to include modest 'induced fit' effects. The two conformers with the lowest estimated potential energy were selected for further calculations in the next procedures that follow below.

5.2.4.2 Quantum Polarized Ligand Docking (QPLD) procedure

The two selected conformers from GlideXP (section 5.2.4.1) were used as starting geometries for QM-Polarized Ligand Docking approach, QPLD (Quantum Polarised Ligand Docking, 2011; Cho, 2005). The first step involves the generation of Quantum Mechanical charges for the free ligand performed at the density functional theory (DFT) level using the

B3LYP functional and the 6-31 G* basis set within the Jaguar module (Jaguar, 2011). The protein was subsequently modelled with the Molecular Mechanics methods using the OPLS_2005 force-field. Ligands were re-docked using GlideXP included into QPLD to generate poses.

5.2.4.3. Induced Fit docking (IFD) procedure

The two QPLD conformers (section 5.2.4.2) were submitted as starting geometries to IFD (Induced Fit Docking, 2011) calculations using Glide 5.7 (Glide, 2011). In the first IFD stage, a softened-potential docking was performed using GlideSP (Standard Precision) mode generating twenty initial poses. For each one, a full circle of protein refinement was performed using Prime 3.0 (Prime, 2011). Residues of the protein within 5.0 Å of ligand poses were refined and side chains were conformationally changed and subsequently minimized. Ligands were re-docked with GlideXP included into IFD to generate poses. The resulting structure from GlideXP were subjected to QPLD to account for the QM charges during docking. The resulting structures from this procedure were used as starting structures for an IFD job that accounts for backbone movements of the receptor and conformational changes during docking (Gumede, 2012).

5.3. RESULTS AND DISCUSSION

5.3.1. Conversion of docking scores to experimental quantitative affinity constants

A combined molecular docking protocol was performed in order to establish the binding mechanism of warfarin enantiomers to HSA as a case study. Computational affinity data to establish the binding affinity of warfarin enantiomers as well as enantioselectivity is still missing in the literature. The unavailability of this data has prompted us to use bioanalytical experimental affinity data currently available in the literature for comparison purposes. However, docking and molecular dynamics methods to study the binding affinity of racemic warfarin to HSA are reported in the literature (Deeb, 2010). A direct comparison between the data obtained from this case study and the results reported by Deeb et al. is impossible because they used racemic warfarin instead of specific enantiomers as we did in this work and a different computational approach to the one we performed in this case study. Therefore, a more direct comparison in our context is that of experimental bioanalytical assay reported in the literature. The docking scores as a thermodynamic parameter measured in (kcal

mol⁻¹) are used as standard Gibbs free energy change, G° ; for binding affinity between a protein and a ligand (Jozwiak, 2008; Mobley, 2009; Lammerhofer, 2010; Li, 2012). This should be the apparent free energy change in reality, G° , since the experiments were undertaken close to physiological conditions (Alberty, 1998; Alberty, 1999; Li, 2010). In this case study, the docking scores were converted to affinity constants for both enantiomers, K_a (M⁻¹), using equation 7 above.

The *in vitro* bioanalytical binding affinity constants (K_a), reported here in this case study as Log $K_{(R/S)}$ (see table 5.1 below) were obtained from the literature. Furthermore, the docking scores were converted to K_a by using equation 4 above (Gumede, 2012). All the experimental results for *in vitro* bioanalytical assay sourced from the literature suggest that S-warfarin has high affinity to HSA than R-warfarin, enantioselectivity > 1.20. The results obtained by using Glide XP reveal that the S-warfarin has a higher binding affinity to HSA than the R-warfarin, enantioselectivity greater than 1).

Table 5.1. *In vitro* bioanalytical assay^a and *in silico* docking simulation^b of affinity constants converted from K_i to the logarithmic form and enantioselectivity (ES). Some aspects of this data table are part of Paper II (Gumede, 2012).

Journal	Method of analysis	Log K_S	Log K_R	ES
1 ^a (O'Reilly, 1971)	Equilibrium dialysis.	5.39	5.31	1.18
2 ^a (Miller, 1977)	Equilibrium dialysis	5.76	5.40	2.28
3 ^a (Lagercrantz, 1983)	Capillary electrophoresis (Zonal elution)	5.64	5.52	1.33
4 ^a (Loun, 1994)	Capillary Electrophoresis (Frontal Analysis).	5.41	5.32	1.24
5 ^a (Zou, 1998)	Equilibrium dialysis.	5.29	5.01	1.92
6 ^b (Gumede, 2012)	Glide XP	4.47	3.89	3.86
		4.53	4.49	1.10
7 ^b (Gumede, 2012)	QPLD	4.87	4.86	1.01
		4.85	4.84	1.03
8 ^b (Gumede, 2012)	IFD	5.43	5.34	1.23

The QPLD binding affinity estimates further reveals that the S-enantiomer binds with high affinity than the R-enantiomer, enantioselectivity < 1.05. The IFD binding affinities also confirms the same observation that the S-enantiomers binds with high affinity than the R-enantiomer of warfarin to HSA, enantioselectivity > 1.20. The Log K_a values obtained from the IFD method are close to the experimental Log K_a values. Hence, the experimental enantioselectivity is similar to the computational enantioselectivity (Gumede, 2012). Furthermore, the (G°) values which ranges from -6.93 to -7.78 kcal mol⁻¹ for racemic

warfarin-HSA binding were obtained from the literature (Deeb, 2010). These results were generated by the use of molecular docking in combination with molecular dynamics (MD) simulations. Equation 8 was used to calculate the resultant $\log K_a$ values which are ranging from 4.88 to 5.48 and are comparable with our IFD binding affinity estimates. The accuracy of docking scores obtained by using IFD can be attributable to the ability of the docking software to allow backbone and side-chain amino acid residues to move freely in order to adopt different conformations during the docking process.

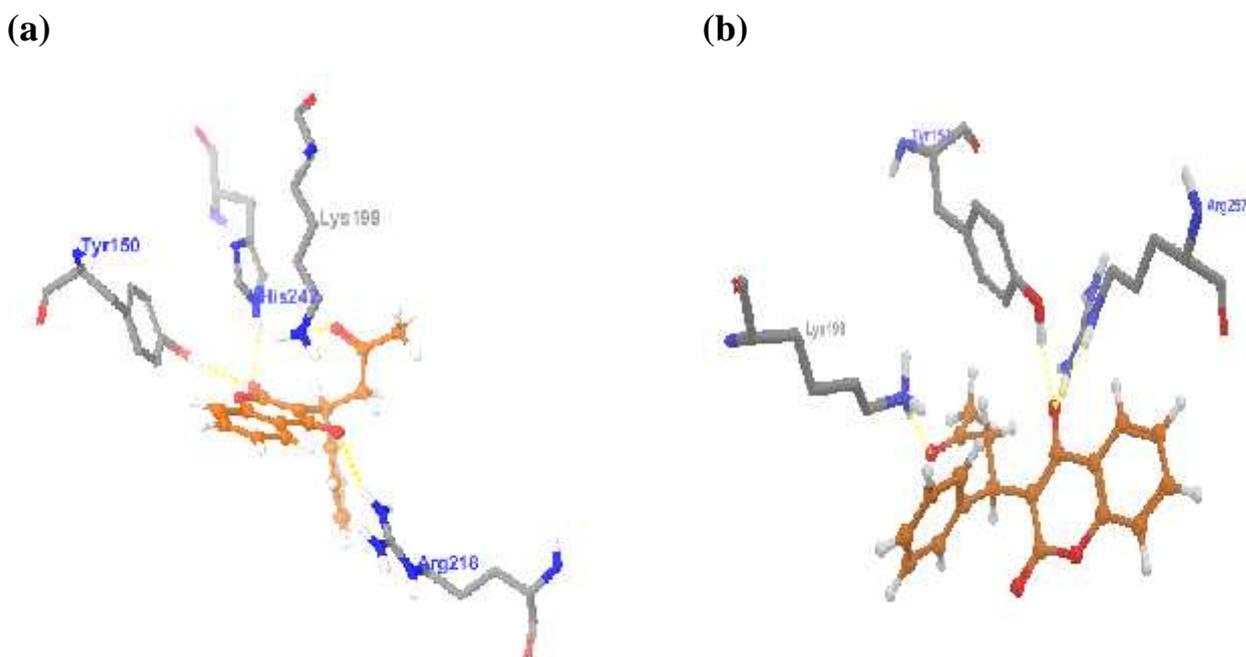


Fig. 5.2 Predicted interaction between warfarin enantiomers and residues in Sudlow site I binding pocket of HSA, corresponding to Table 5.2 below, with H-bond interactions (highlighted). (a) S-pose. (b) R-pose.

The binding process of Warfarin enantiomers to HSA involves a change in the conformation of the protein's active site to accommodate the ligand in the binding pocket, and the subsequent change in the conformation of the ligand to adapt to its new host (see Fig 5.2 a) and b) above). The high affinity of S-Warfarin to HSA as observed in Table 5.2 can be explained by a strong hydrogen bond between the N-H group of Arg218 as a hydrogen bond donor and a negatively charged oxygen group of the coumarin ring as a hydrogen bond acceptor with a bond radius of 1.811 Å. This is followed by an -OH group of Tyr150 as a hydrogen bond donor and the carbonyl group of the coumarin ring as a hydrogen bond acceptor with a bond distance of 1.915 Å. The third hydrogen bond involved the hydrogen bond between the N-H group of Lys199 and the carbonyl group of the acetyl group of Warfarin with a bond radius of 1.928 Å. Furthermore, the N-H group of His242 showed a hydrogen bond with the carbonyl group of the coumarin ring with a bond radius of 2.190 Å. The aromatic rings of His242 and Tyr411 showed a strong π -stacking with the coumarin ring along with some hydrophobic interactions as shown in Fig. 5.2. On the other hand, the R-Warfarin enantiomer's affinity to HSA in site I is due to the hydrogen bond formed between the N-H group of Lys199 and the carbonyl group of the acyl group with a bond distance of 1.868 Å. The N-H group of Arg257 and an OH group of Tyr150 showed another hydrogen bond formation with the negatively

charged oxygen atom of the coumarin ring with a bond distance of 1.877 Å, 2.033 Å and 2.192 Å, respectively (see Table 5.2 below).

Table 5.2. Molecular docking results at molecular level^a for the selected pose predicted by IFD (ID8 in Table 1), with an *ES* = 1.23, similar to the experimental one (*ES* = 1.24) (Gumede, 2012).

Chirality	Functional Groups.	Hydrogen bonds.	H-bond distance (Å)	Electrostatic interactions	IFD Docking Score (kcal.mol ⁻¹)	K _a (M ⁻¹)
S	Coumarin	Arg218 (N-H)	1.811		-7.71	2.69x10 ⁵
	Coumarin	Tyr150 (O-H)	1.915			
	Acetonyl	Lys199 (N-H)	1.928			
	Coumarin	His242 (N-H)	2.190			
	Coumarin	Tyr150 (Ar-H)	-	-		
	Coumarin	His242 (Ar-H)	-	-		
R	Acetonyl	Lys199 (N-H)	1.868		-7.58	2.20x10 ⁵
	Coumarin	Arg257 (N-H)	1.877			
	Coumarin	Arg257 (N-H)	2.033			
	Coumarin	Tyr150 (O-H)	2.192			
	benzyl	His242 (Ar-H)	-	-		

^a H-bonding (and its distance) and - interactions are indicated as the main driving forces involved in enantioselectivity. Ar refers to an aromatic group of amino acids. Hydrophobic contacts explained the adoption of the conformations of ligands in the binding cavity, for S- and R-poses, respectively.

The results obtained in this case study are consistent with the trend observed by (Deeb, 2010) on docking and molecular dynamics simulation of racemic warfarin, where they observed the side chains Lys199, Arg257 and His242 were in direct contact with the ligand, even though the extent of binding was not revealed. Furthermore, Ghuman et al. use X-ray crystallography to analyze the binding sites of HSA co-crystalized with warfarin enantiomers, and they revealed that Tyr150 is important in binding in site I. While in site II, Tyr411 plays a major role in hydrogen bonding (Ghuman, 2005). Furthermore, Petitpas et al. pointed out that steric hindrance between Trp214, Arg218 and the benzyl ring of warfarin decreases the binding affinity of warfarin to HSA protein (Petitpas, 2001). The results in Table 5.2 points out that the Arg218 is important in the hydrogen bond with the Coumarin ring, which is only observed in S-warfarin, which justifies the observed higher affinity of this enantiomer to HSA.

5.4 CONCLUSIONS

The synergy between quantitative *in vitro* bioanalytical estimates such as $\log K_S$, $\log K_R$ and ES for warfarin-HSA interaction and *in silico* molecular docking simulations is important in establishing molecular forces important in binding. Evidence has been presented in this case study that these combined approaches are able to reveal important binding kinetics/thermodynamics parameters. In fact, these binding parameters obtained in this case study can now be observed at macroscopic, microscopic, submicroscopic, and atomic levels for protein-ligand complexes, resulting from this synergy. The importance of this synergy is based on the fact that *in vitro* approaches could validate and verify the results obtained by *in silico* approaches. Since, the conformational space of organic compounds is very wide and it is difficult to correctly rank the binding poses from a docking calculation with molecules that exhibit different binding modes. Therefore, *in vitro* approaches can be used for decision making purposes, when deciding about the most plausible pose to use and report. On the other hand, *in silico* methods can further be used to explain the important factors giving rise to the binding event at molecular/atomic level i.e. the types of bonds formed, the functional groups involved in the binding event, a picture depicting the conformational space of both the ligand and the protein in its active site gives more insight on a detailed view of the target's ability to accommodate oncoming ligands in its binding pocket of the active site.

CHAPTER 6

CASE STUDY III

Structure Based Drug Design and Ligand Based Drug Design methods in the design of NCEs: CYP17A1 inhibitors as a test case

6.1 INTRODUCTION

Molecular docking techniques can be used as a structure-based drug design (SBDD) strategy to reveal the binding modes and binding affinities of ligand structures in the active site of a target receptor structure (Giangreco, 2013), in order to gain insights on the extent of intermolecular forces that drives the binding event. LBDD and SBDD approaches complement each other in both approaches, prior conformational search becomes essential (Schuster, 2011). Even though there is no evidence suggesting that the lowest energy conformer of the ligand is the correct conformer in a real context (Dror, 2009; Günther, 2006). Therefore, it is important to generate an ensemble of low energy conformers that have reached their convergence, and at the end of the process, those selected should overlay over co-crystallized ligands on PDB crystal structures (Giangreco, 2013; Lemmen, 2000).

Pharmacophore modelling can be used as a ligand-based drug design (LBDD) approach, as an abstract description of molecular features that are necessary for molecular recognition of a ligand by a biological macromolecule (Giangreco, 2013; Wermuth, 1998), thanks to the ensemble of steric and electronic features necessary to ensure optimal supramolecular interactions (Wermuth, 1998). The outputs could allow further computational calculations, such as Density Functional Theory (DFT) (Tawari, 2010) to predict electronic properties explaining the reactivity. On the other hand, *in Silico* computational techniques have the ability to explain the interactions between the ligand and the receptor at molecular level, and also predicting biological activities of molecules from their structural properties (Alzate-Morales, 2010). In the literature it has been reported that Purushottamachar et al performed a qualitative 3D pharmacophore model for well-known natural androgen receptor down-regulating agents, which was subsequently followed by a database search and synthesis of novel AR inhibitors (Purushottamachar, 2008). Furthermore, Gianti et al have used induced-fit docking on AR inhibitors based in homology models, since the X-ray crystal structure of the CYP17A1 enzyme was unavailable in that point in time (Gianti, 2012). Recently, two available crystal structures for CYP17A1 co-crystallized with CYP17 inhibitors Abiraterone (3RUK) and

TOK001 (3SWZ) were resolved and subsequently deposited into the Protein Data Bank (PDB) at a resolution of 2.6 Å and 2.4 Å respectively, (DeVore, 2012). New research incorporating such new information could reveal more consistent results from molecular modelling techniques.

Accordingly, in this work a combined computational strategy is proposed for the first time to generate information on the CYP17A1 inhibition where: (i) a 3D-QSAR pharmacophore model was performed on a diverse set of steroidal and non-steroidal CYP17A1 inhibitors obtained from literature with known experimental IC_{50} values. The pharmacophore hypothesis obtained from the more potent ones, were validated by comparing the prediction ability on the training set (model calibration) and a test set (excluded from the model calibration). (ii) A Density Functional Theory (DFT) calculation was then used for evaluating electronic properties of selected inhibitors, which reflects their reactivity. (iii) A Flexible ligand-protein Molecular docking was first validated against the available co-crystallised complex with X-ray available structures (TOK001-CYP17A1 complex), and then used on selected structures to confirm the agreement with the pharmacophore hypothesis (an approach not yet tested up to now). This combined strategy has enabled us to explore the synergy between SBDD & LBDD methods to present new information in the design of novel inhibitors targeting PC.

6.2 MATERIALS AND METHODS

6.2.1. Data Treatment

The information on a set of 98 steroidal and non-steroidal molecules with different core structures and broad inhibition activity to CYP17A1 enzyme (*in vitro* experimental IC_{50} between 13 to 20000 nM) was collected from literature (see Table 6.1 below) (Nnane, 1999; Zhuang, 2000; Handratta, 2005; Jagusch, 2008; Budha, 2008; Pinto-Bazurco Mendieta, 2008; Hu, 2010a; Hu, 2010b; Vasaitis, 2011; Kaku, 2011; Kaku, 2011;). Structures in 2D representation and their systematic names are also included (see Table A1, appendices). The *in vitro* experimental IC_{50} values, in molar (M) units, were converted into pIC_{50} (i.e. $-\log IC_{50}$) data (see Table A2, appendices). A similar coding of the inhibitors as appearing in the original publications was retained.

6.2.2. Generation of 3D Multiple Conformers

The ‘build panel’ of Maestro (v9.3.5) (Maestro, 2012), a Schrödinger suite 2012 graphical user interface, was used to build starting molecular structures for the 98 compounds which were energetically minimized in order to adjust bond length, bond orders as well as formal charges. Ligprep (v2.5) (Ligprep, 2012) was used to create tautomeric 3D low-energy structures at pH 7.4, to protonate the ionisable groups of tautomers. The stereochemistry for chiral compounds was retained (see Table A2 in the Appendices). The adjusted 3D structures were then subjected to a conformational search method using a Mixed Monte Carlo Multiple Minimum Low Mode (MCMMLMOD) conformational search method using MacroModel (v9.9) (MacroModel, 2012). OPLS-2005 force-field with GB/SA implicit solvation model was used to generate low-energy multiple conformers with a constant dielectric constant of 1.0. The number of minimization steps was set to 100. The maximum relative energy difference of 10 Kcal/mol was set for saving multiple conformers. A Root-Mean-Square-Deviation (RMSD) cut-off of 1.0 Å was set to eliminate redundant conformers. The number of resulting conformers per compound is shown in Table A2 in the appendices.

6.2.3. 3D-QSAR Pharmacophore Model

Pharmacophore modelling was developed by using PHASE (v3.4) (Phase, 2012), a module of Schrödinger 2012 product suite. Pharmacophore sites (variants) available from PHASE were used. They include hydrogen bond acceptor (A), hydrogen bond donor (D), negatively charged groups (N), positively charged group (P), hydrophobic groups (H) and aromatic rings (R). The 3D-contours representing the pharmacophore sites of the ligand, depicts the potential of non-covalent bonds between the ligands and the hypothetical target receptor.

The procedure was applied over the conformational space of structures with the highest- pIC_{50} values, generating a common pharmacophore hypothesis (CPHs) from their 3D conformations (see Table A2 in the Appendices) for the two CPH that exhibited modest statistical correlation with experimental data for the model’s predictive power. The CPH groups together with similar structural features/variants that are common in the training data set could be yielded using the procedure we have adopted (Dixon, 2006; Jain, 2013; Zhang, 2013; Tanwar, 2013). PHASE was employed to find common pharmacophores using 6 sites (the maximum number). The number of sites matched by all the structures is included in Table 6.

Scoring was performed in order to identify the best hypothesis, rationally ranking them for further investigation (Tawari, 2010; Durdagi, 2011; Tawari, 2011).

A 3D QSAR (PLS) model was generated by making use of the experimental ligand activities that matches the hypothesis from the previous step. A total of 60 hypothesis retrieved from the previous step with a 3D conformation of molecular structures in the data set were used for model generation. However, Table A2 only shows the four best pharmacophore hypothesis the rest of the models are not included. The ‘Atom-based pharmacophore model’ an option in PHASE was preferred (over pharmacophore-based alignment), since it has been described as adequate for structures that contains a small number of rotatable bonds with common structural framework (Dixon, 2006; Jain, 2013; Zhang, 2013). The PLS models were obtained and tested after randomly dividing the datasets into training and test (approx. 20% of the data) sets. A leave-n-out cross-validation on the training data set was used. A maximum of three PLS factors were fixed to prevent over-fitting. The elimination of the identified outliers was decided (and indicated in Table 1), taking into account that the *in vitro* experimental activities were measured with different assay methods (heterogeneous response variable). A new PLS was built with the remaining compounds. This would aid in obtaining a low-factor PLS-structure with balanced combination of predictive ability on the training and test compounds (R^2 and Q^2 values, respectively).

6.2.4. Density Functional Theory (DFT) Calculations

All Quantum Mechanical/Molecular Mechanics (QM/MM) calculations on the non-outlier molecular structures shown in Table A2 in the appendix section were performed with Jaguar (v7.9) (Jaguar, 2012). All geometry optimizations were carried out at the B3LYP level of density functional theory with the 6-31G* basis set (Tawari, 2010; Tawari, 2011). Electronic properties related to the reactivity of molecules in the pharmacophore model were computed. It was then followed by a single-point energy calculation at the optimum geometries to obtain aqueous solution phase energies using a continuum treatment of solvation Poisson-Boltzmann (PBF) model (Tawari, 2010). The electronic properties of interest included molecular electrostatic potential (MESP), highest occupied and lower unoccupied molecular orbital (HOMO and LUMO, respectively) and Interaction Strength (IS) (Jaguar, 2012).

6.2.5. Molecular Docking

A flexible ligand-protein molecular docking procedure was performed on the molecular structures optimized by DFT calculations in section 2.4. The Glide/IFD protocols described elsewhere (Tawari, 2011; Gumede, 2012) were used here, instead of Quantum Polarised Ligand Docking calculation, as in our previous work, since the charges for a free ligand has been previously calculated by a hybrid Quantum Mechanical calculation (DFT optimization). A cross-docking procedure was implemented to validate the docking method was performed on a series of different targets for selectivity of the method. The TOK001 pose estimated by docking is superimposed over the co-crystallized TOK001 to the enzyme CYP17A1 for comparison, as a way of validating the docking process.

6.3. RESULTS AND DISCUSSION

6.3.1. Design of a Pharmacophore Model

The data set in Table A2 in the appendices shows that the common pharmacophore hypothesis (CPHs) were constructed from the 11 highly active molecules (ID 1-5, 7-9, 11, 12 and 14) in the training set. The main idea was to match the pharmacophore features of the 11 highly active molecules in the hypothesis. A total of 60 different 6-point CHPs were generated (see Table A3 in the appendices). All CPHs were examined and scored to identify pharmacophores that yields the best alignment of the active compound. A 3D-QSAR pharmacophore model was generated by using the entire hypothesis. However, our best pharmacophore model (AADHRR.82) shown in Fig. 6.1 below consists of two hydrogen bond acceptors, one hydrogen bond donor, one hydrophobic group and two aromatic rings with point vectors pointing on the direction in which hydrogen bonds would come from. The statistical significance of a QSAR model measures the reliability of a selected model (Deora, 2013).

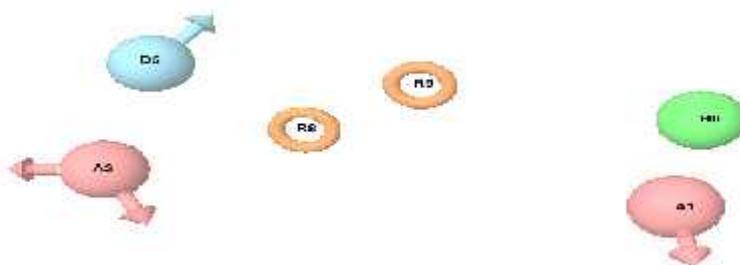
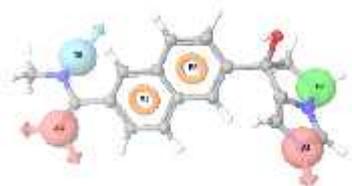


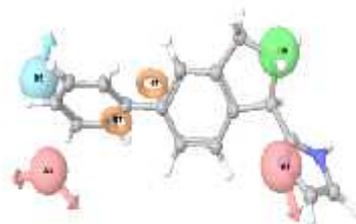
Fig. 6.1 Best common pharmacophore hypothesis AADHRR.82 showing the point vectors for hydrogen bond acceptor (A1) and (A3), hydrogen bond donor (D5), aromatic groups (R8) and (R9), as well as the hydrophobic group (H6) for potential hydrogen bonding, hydrophobic and π - π interactions, respectively when bound to the receptor.

The results in Table A2 in the appendix show the number of sites matched (A, D, N, P, H or R) by chemical structures studied in the model. This means that structures with high fitness scores represents the ligands that exhibits a greatest overlay with the CPHs. Furthermore, CPHs are modelled on a molecule that overlays with the hypothesis, where (+)-3c is the reference compound in our model because the structure exhibits a fitness score of 3.0 as shown (see Table A2 ID 4 and Fig. 6.2 as depicted below). Therefore, most of non-steroidal inhibitors have pharmacophore features that are common in their core structures. While steroidal inhibitors were outliers because their pharmacophore features were not similar to the rest of the molecules in the data set. Hence, they did share the same chemical scaffold as the non-steroidal inhibitors, and were not properly aligned with the pharmacophore model.

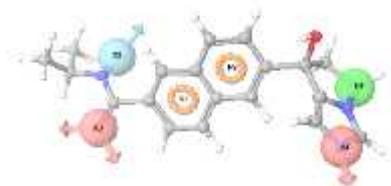
+3c



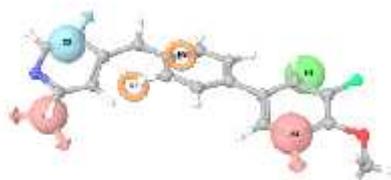
5bx



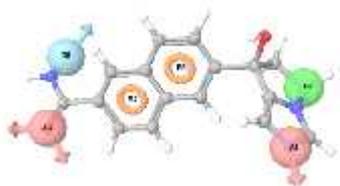
3g



16



3b



5

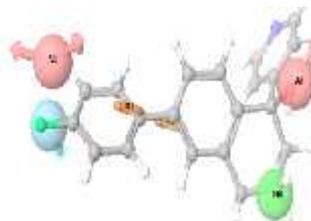


Fig. 6.2 Results for the most active (left part: +3c, 3b and 3g; fitness scores of 3.0, 2.9 and 2.8, respectively) and least active/consistent (right part: 16, 5bx and 5; fitness scores of 0.87, 0.93 and 0.98, respectively) ligands, mapped onto the pharmacophore hypothesis AADHRR.82.

Statistical results from the best pharmacophore hypotheses showing the best predictive ability, (i.e. good combination of R^2 and Q^2 statistics for training and test sets, respectively), were ADHRRR.116 ($R^2 = 0.88$ and $Q^2 = 0.75$; with a PLS model consisting of 4 latent variables-) and AADHRR.82 ($R^2 = 0.81$ and $Q^2 = 0.78$; with a PLS model consisting of 4 latent variables) are shown in Table 6.1 below. The ADHRRR and AADHRR part of the name refers to the variants and 116 and 82, respectively refers to the maximum number of hypotheses present in the pharmacophore model, which is unique for all highly active molecules in the data set.

Table 6.1. Statistical results for the 3D-QSAR model (from 88 compounds; after eliminating outliers) corresponding to the pharmacophore hypotheses in Table A3 in the appendices section.

Hypothesis	PLS Factors	R^2	Q^2	Reference ^a Ligand Conformer
ADHRRR.116	1	0.4939	0.2913	+3c
	2	0.7544	0.6564	
	3	0.8432	0.6061	
	4	0.8999	0.6706	
AADHRR.82	1	0.5404	0.4090	+3c
	2	0.7574	0.5900	
	3	0.8133	0.7756	
	4	0.8773	0.7470	

^a The reference ligand is the ligand conformer that provides the pharmacophore that matches the hypothesis.

The pharmacophore hypothesis AADHRR.82 was considered more consistent, since it requires a simpler latent structure-PLS model providing a better Q^2 value and balanced combination. As a provision, we can consider this 3D-QSAR pharmacophore hypothesis consisting of two hydrogen bond acceptors, one hydrogen bond donor, one hydrophobic group, and two hydrophobic groups (AADHRR.82) as satisfactory for our model (see Table 6.1).

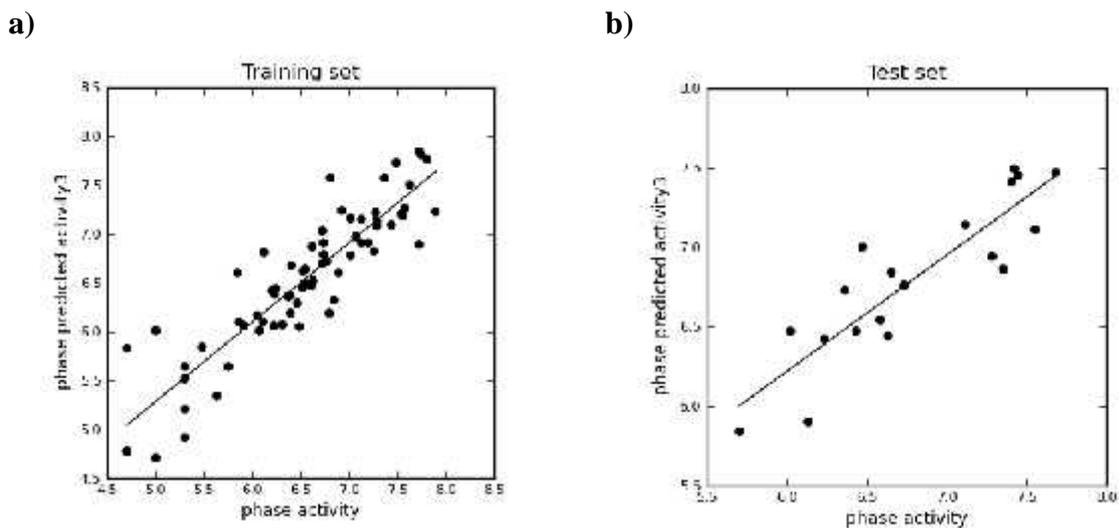


Fig. 6.3 Validation plots of pIC_{50} (estimated vs. experimental) for hypothesis AADHRR.82. The QSAR model corresponds to a 3 Latent variables-PLS model. (a) Training set; 69 compounds. (b) Test set, 19 compounds.

Fig. 6.3 shows the validation plots of pIC_{50} (estimated vs. experimental) for the training and test set structures (after outlier detection and removal), related to the hypothesis AADHRR.82 assumed as the correct pharmacophore model. The results suggest that this pharmacophore model was able to distinguish between high, medium and low active inhibitors on the data set under study. The three most active ligands (+)-3c, 3b and 3g, ID 4, 12 and 64, respectively, in Table A2 and Fig. 6.2) and least active ligands (16, 5bx and 5, ID 10, 53 and 57, respectively, in Table A2 and Fig. 6.2) were selected based on the fitness score parameter (see Table A2) in the appendix section.

The pharmacophore model may be mapped onto a reference ligand (+)-6-(7-Hydroxy-6,7-dihydro-5H-pyrrolo[1,2-c] imidazol-7-yl)-N-methyl-2-naphthamide, (+)-3C with a fitness score of 3.0 which fits the model. This alignment symbolizes a good match of features present in the reference ligand to the pharmacophore hypothesis comprising of training set ligands. Accordingly, in a further aspect of the database search after model building there is provided the use of (+)-6-(7-Hydroxy-6,7-dihydro-5H-pyrrolo[1,2-c] imidazol-7-yl)-N-methyl-2-naphthamide, (Compound (+)-3C in Table A2) in a training set as a reference ligand in a method of identifying inhibitors of an enzyme selected from the group consisting of CYP17A1 inhibitors. In fact, the reference molecule (+)-3c in Fig. 6.2 which is mapped onto the pharmacophore hypothesis (AADHRR.82) and shows point vectors for hydrogen bond acceptors (A3) and (A1) which are the carbonyl group of N-methyl-2-carboxamide (A3) and the Nitrogen of the imidazole ring (A1) are pointed in the direction where the amino acid groups

will more likely form hydrogen bonds with the amino acid residues of the enzyme when bound to the target enzyme in its active site. While the hydrogen bond donor group (N-H group) of N-methyl-2-carboxamide (D5) has a point vectors that point in the direction of the incoming hydrogen bond acceptor of the amino acid residue. The Naphthalene rings (R8) and (R9) respectively are properly aligned on the position where π - π interactions are most likely to occur with aromatic rings of the amino acids of the enzyme. The 5H-pyrrole ring (H6) is part of the hydrophobic group where hydrophobic contacts are more likely to occur for the reference ligand. Similar observations are evident on strong inhibitors 3b and 3g (on the left side in Fig.6.2). Similar observations to confirm the functional groups responsible for the tight binding of (+)-3C the reference molecule are evident in Fig. 6.5 (b) from docking outputs. The Ligand Interaction Diagram shows the carbonyl group which is mapped as (A3) in the pharmacophore model as a hydrogen bond acceptor binds with the N-H group of Arg239 as a hydrogen bond donor. While on the other hand, the N-H group (D5) on the pharmacophore model as a hydrogen bond donor shows a hydrogen bond with the carbonyl group of Asp298 as a hydrogen bond acceptor. The pyrrole ring exhibits some π - π stacking interactions with the pyrrole rings in the porphyrin moiety of ferric heme.

In sharp contrast, weak inhibitors such as 16, 5bx and 5 (on the right side in Fig. 6.2) are not overlaid with the pharmacophore hypothesis which clearly explains their weak *in vitro* experimental inhibition. Furthermore, it must be noted that the pharmacophore features are not properly aligned with the reactive functional groups for these weak inhibitors. The point-vector features (pharmacophoric sites) have clearly shown how a 3D-QSAR pharmacophore model is able to identify important characteristic features between the ligand and the target receptor (Jain, 2013).

6.3.2. DFT Results

DFT calculations were used to illustrate the electronic features that are important in the reactivity of the molecules. The most active and least active conformers from the previous section which are (+3c, 3b and 3g, ID 4, 12 and 64, respectively in Table A2.) with respect to the pharmacophore model were used as starting structures for DFT geometry optimization. Additionally, studying a molecule's highest occupied molecular orbital (HOMO) and the lowest occupied molecular orbital (LUMO) orbitals can be used to clearly explain the drug-receptor interactions as well as molecular reactivity. The orbital energies indicate the ability of

molecules to accept or donate electrons. Whereas orbital distribution in the molecule indicates the ability of a molecule to either be an electrophile or a nucleophile because of the reactive functional groups that might react with functional groups of the receptor in its' active site (Tawari, 2011). Fig. 6.4 shows some results for the most consistent structure, +3c (ID 4; Table 6.1). Figures 6.4a) and b) shows molecular orbital diagrams for HOMO and LUMO mapped onto the structure. The HOMO sites are mapped onto the aromatic rings indicates the ability of the molecule to donate electron pairs to appropriate acceptor amino acid residues of the receptor.

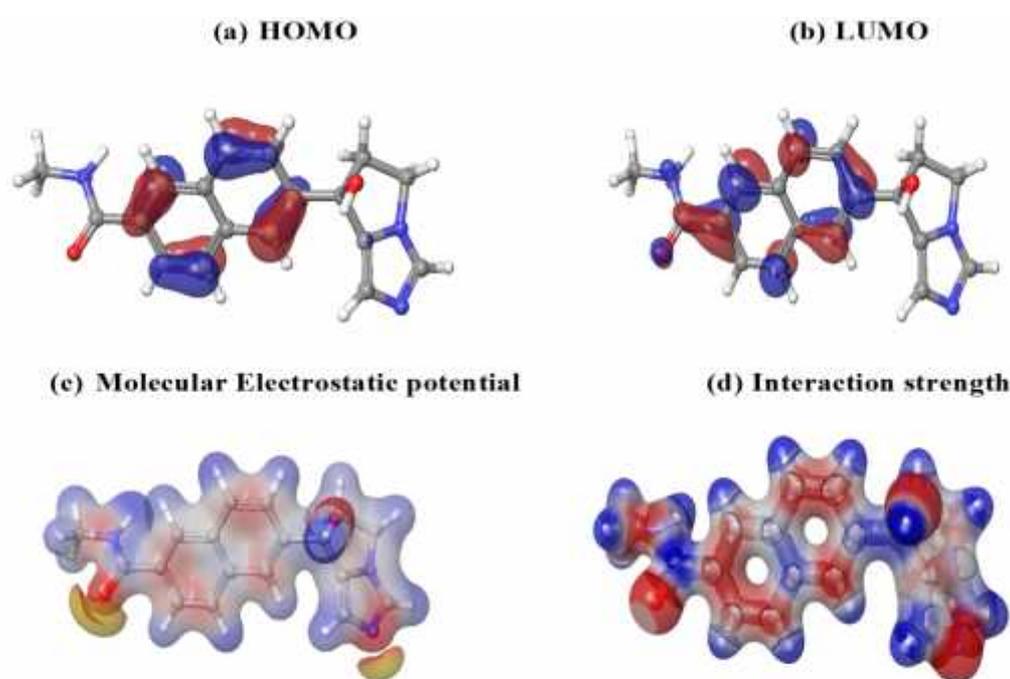


Fig. 6.4 DFT results for the active molecule (+)-3c. Orbital diagrams of (a) HOMO and (b) LUMO, mapped onto the structure. (c) 3D-contours of molecular electrostatic potential maps at -30kcal/mol. Regions: high electronic density (negative potential) in red; low electronic density (positive potential) in dark blue, electronegative groups in yellow. (d) Interaction strength contours mapped onto the structure. Groups that are susceptible to substitution (e.g. C=O, N-H and O-H) are visible.

The energies observed for HOMO and LUMO orbitals indicate that these molecules are reactive (see Table A4 in the appendices). The HOMO-LUMO band gaps for the most active molecules from the model were -0.171, -0.170, and -0.171 for (+)-3c, 3b, and 3g respectively. While for the least active molecules from the pharmacophore models the HOMO-LUMO energy gaps were -0.147, -0.178, and -0.188 for molecule 5, 16, and 5bx, respectively. It is

evident that there is a direct relationship between the HOMO-LUMO gaps and the reactivity of the molecules. Since the energy gaps for the most active compounds is consistent. Whereas, with the least active molecules in the model the energy gaps are variable and smaller than the energy difference for the most active molecules. Except for molecule 5 which has an energy gap which is higher than the most active molecules. The reactivity is caused by a rapid transfer of electrons and their exchange from HOMO to LUMO. A small change in energy between HOMO and LUMO is observed, suggesting that there is charge distribution facilitated by electron transfer. The HOMO orbital mapped onto the hydrogen bond donor *N*-methyl group is absent in weak inhibitors shown in Figures 6.4 (a-b). The LUMO sites mapped onto the carbonyl group of the *N*-methyl-2-carboxamide and the aromatic rings suggests that it is susceptible to nucleophilic attack, which is consistent with a pharmacophore feature for hydrogen bond acceptor (A3) in Fig.6.2. This feature is also absent in the least active molecules.

Figure 6.3 (c) shows 3D-contour maps of molecular electrostatic potential at -30 kcal/mol. It can be seen from Figure 6.3 (c) that: (i) regions of high electronic density (negative potential; in red) showing the distribution of electron clouds around the atoms of the molecule; (ii) a region of low electronic density (positive potential; in dark blue), showing the functional groups that are more electronegative and the functional groups that are less electronegative; and (iii) the most electronegative functional groups (yellow potential contours), locating the more reactive areas responsible for the interaction strength projected towards the enzyme (see Fig. 6.3 d). Additionally, Fig. 6.4 (d) illustrates interaction strength 3D-contour mapped onto the conformers. Groups that are susceptible to nucleophilic or electrophilic substitution (e.g. C=O, N-H and O-H) are visible with big openings on the 3D-conformational space. They also were absent on the least active structures. Therefore, these electronic properties were consistent with the predicted reactive sites of the molecules from the pharmacophore hypothesis.

6.3.3. Molecular Docking Results

In this study the ability of molecular docking to reproduce the co-crystallized ligand conformation in the active sites of the enzyme/protein have been tested. Furthermore, the molecular docking methodology was validated in a diverse set of targets important in the drug discovery pipeline. Table 6.2 shown below illustrates the cross-docking procedure on X-ray crystal structures co-crystallized with ligands in their active sites to validate and measure the

selectivity of docking methods. We obtained HSA X-ray crystal structures co-crystallized with site I and site II ligands from the protein data bank. We have further uploaded the X-ray crystal structure of SULT2A1 enzyme important in phase II clinical trials. We also downloaded X-ray crystal structures of CYP17A1 enzyme with moderate resolution from the protein data bank.

Table 6.2. Root-mean-square-deviations (RMSDs) in cross-docking and native docking approaches performed on different types of PDB X-ray crystal structures to validate and measure the selectivity of the docking methodology.

PDB ID	Resolution Å	Docking Score (CD) ⁱ kcal/mol	RMSD (CD) ⁱ Å	Docking Score (IFD) kcal/mol	RMSD (ND) ^j Å	Receptor	ligand	SITE	Type of Protein/ Enzyme
2bxa	2,35	-9,03	0,63	-10,61	0,18	HSA	CMPF ^a	I	Carrier protein
1HA2	2,50	-6,67	0,40	-7,31	1,12	HSA	S-WRF ^b	I	Carrier protein
2bxf	2,95	-8,50	0,20	-9,05	0,20	HSA	DZP ^c	II	Carrier protein
2bxg	2,70	-7,68	0,79	-7,75	0,24	HSA	IBPF ^d	II	Carrier protein Phase II Metabolic enzyme
1EFH	2,40	-14,62	0,14	-16,28	0,16	Sult2A1	ADP ^e	I	Phase II Metabolic enzyme
1OV4	2,70	-9,13	0,82	-9,99	0,31	Sult2A1	AET- SO ₄ ^f	I	Phase II Metabolic enzyme
2qp3	2,60	-9,07	0,27	-9,68	0,18	Sult2A1	AET ^f	I	Phase II Metabolic enzyme
3ruk	2,60	-8,86	1,31	-9,65	0,29	CYP17A1	ABT ^g	I	Target protein
3swz	2,40	-10,22	0,17	-10,02	0,17	CYP17A1	TOK001 ^h	I	Target protein

^aCMPF - 3-carboxy-4-methyl-5-propyl-2-furanpropanoic acid, ^bS-WRF – (S)-Warfarin, ^cDZP – Diazepam, ^dIBPF – Ibuprofen, ^eADP - Adenosine-3',5'-diphosphate, ^fAET – Aetiocholanolole, ^gABT – Abiraterone, ^hTOK001 – Galeterone, ⁱCD – Cross-docking, ^jND – Native Docking.

The results in Table 6.2 reveal that cross-docking and IFD (native docking) are able to reproduce the correct conformations of the co-crystallized ligands. However, the RMSDs resulting from both methods reveal that the accurate overlay of structures to co-crystallized ligands is system dependent (i.e. the results depend on the target enzyme/protein. Furthermore, it appears that both CD and IFD have been able to reproduce RMSD results less than 0.30 Å. This information validates docking methods and the RMSD data also reveals that the docking methods are selective and applicable to different types of targets.

Results in Fig. 6.4 (a) shows a cross-docking pose for the ligand TOK001, superimposed onto the available complex of the native ligand (TOK001) co-crystallized on the active site of CYP17A1 enzyme (PDB ID: 3SWZ). Both structures are greatly overlaid, suggesting that the docking approach accurately models the binding modes and its ability to correctly simulate the conformer obtained by X-ray crystallography with a root-mean-square-deviation (RMSD) of

0.17 Å. Additionally, the ligand interaction diagram in Fig. 6.5 (b) shows the binding mode of TOK001 on the receptor, indicating a hydrogen bond between 3 -OH (hydrogen bond donor) of the ligand and a carbonyl group of ASN202 amino acid (1.69 Å).

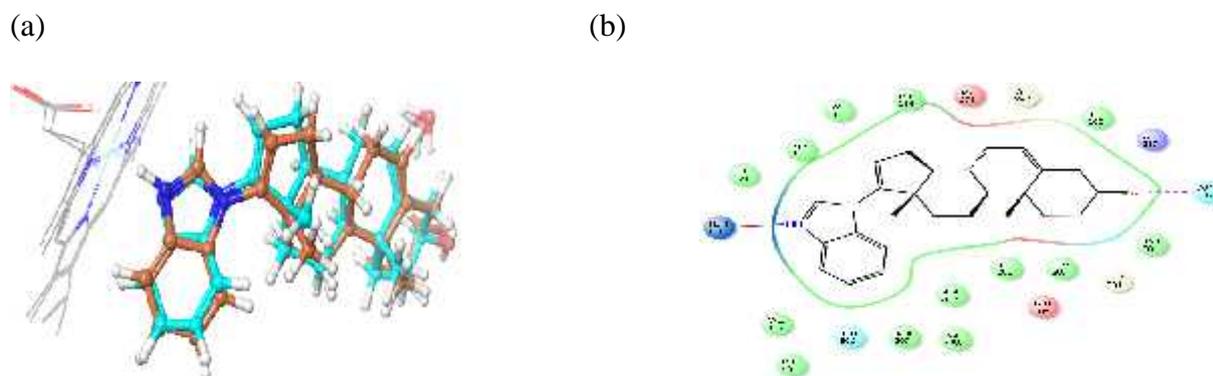


Fig. 6.5 (a) Docking screen for an overlay of a co-crystallized ligand TOK001(in cyan colour) on the active site of the enzyme (PDB ID: 3SWZ) with a docking pose of TOK001(in brown colour) on the native active site of the enzyme. (b) Ligand Interaction diagram of the native docking pose for TOK001 on the active site cavity of CYP17A1 enzyme.

Ferric heme shown in Fig. 6.5 (b) illustrates a salt-bridge interaction with an -NH group of a benzimidazole ring (bond radii, 2.17 Å). The ligand TOK001 is lined within hydrophobic amino acids of the enzyme such as ALA113, PHE114, ILE206, LEU209, and VAL483 which means that the active site is hydrophobic. In sharp contrast, there are basic (ASP298) and acidic groups (ARG239) of the receptor aligned towards carbon 6 of the ligand, respectively. These observations are consistent with the X-ray crystallographic information obtained by DeVore et al; and therefore provide reasonable evidence to the validation of the process and also give a reasonable guarantee over other estimated docking poses without X-ray crystal structures such as homology models obtained for the target CYP17A1 enzyme.

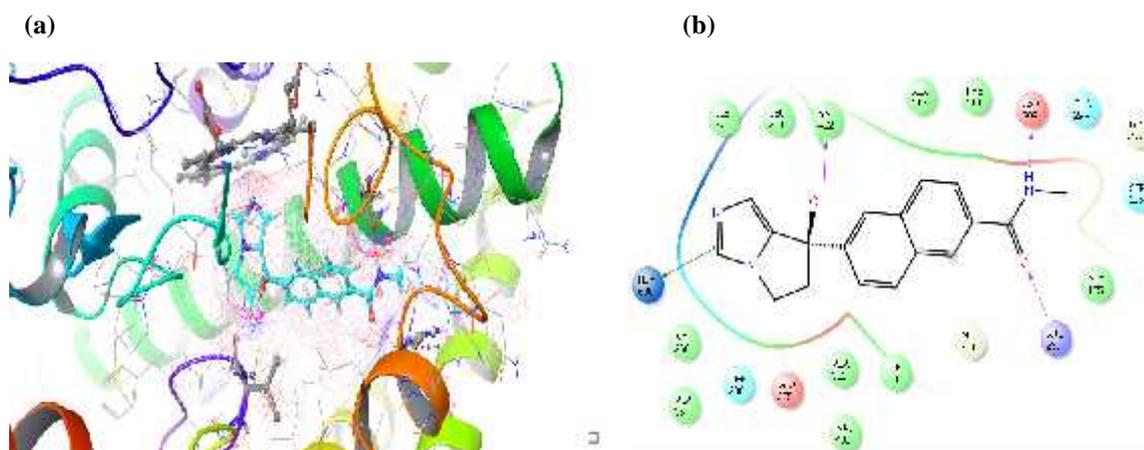


Fig. 6.6 (a) Docking pose of the protein-ligand complex of the most active molecule (+)-3c from the 3D-QSAR Pharmacophore model on the active site of the enzyme (in cyan colour) overlaid with electrostatic potentials showing the distribution of electrons for enzyme-ligand interaction. (b) Ligand interaction diagram to show the binding mode and the type of bonds that the ligand exhibits on the active site of the enzyme.

The non-steroidal reference molecule (+)-3c from the 3D-QSAR pharmacophore model shown in cyan in colour in Fig. 6.6 (a) is bound on the active site of CYP17A1 enzyme. The electrostatic potentials are mapped on both the ligand and some reactive functional groups of the enzyme. In fact, the electrostatic potential shows the distribution of electrons for the bound enzyme-ligand complex. The blue region indicates a region of low electron density whereas the red part shows the region of high electron density. There is a π - π interaction between the imidazole rings of (+)-3c and Ferric heme (see Fig. 6.6b). There is a strong hydrogen bond interaction between the N-H group of the ligand (hydrogen bond donor) and the carbonyl group of ASP298 (bond radii, 1.71 Å). Furthermore, hydrogen bonds between the carbonyl group of N-methyl carboxamide (hydrogen bond acceptor) with N-H group of ARG239 (hydrogen bond donor) as well as the hydrogen bond between the OH group of the ligand with the carbonyl group of VAL482 are visible; with a bond radius of 2.12 and 2.19 Å, respectively (see Fig 6.6 a and b). The ligand is aligned on the surface with the presence of hydrophobic amino acids such as ILE371, LEU214, VAL482, ALA113, PHE114, ALA105, ILE206, ALA302 and VAL366 this implies a similar binding mechanism to TOK001 a steroidal inhibitor. However, the acidic and basic groups such as ASP298 and ARG239 are present but in this case, they are involved in hydrogen bonding interaction with the functional groups of the ligand. Furthermore, the electrostatic potential maps onto the ligand and the receptor clearly explains what functional groups are important in the binding mechanism between CYP17A1 inhibitors and the amino acids residues of the receptor.

6.4. CONCLUSIONS

Docking calculations have confirmed that the pharmacophore sites (see Fig. 6.1 and 6.2) predicted by the 3D-QSAR model for the most active ligand (+)-3c really participates in hydrogen bonding as hydrogen bond donors or acceptors, π - π interaction between the benzimidazole ring and the hydrophobic interactions as well (see Fig. 6.6a and b). Furthermore, this structure fits snugly into the active site of the enzyme, which is populated by hydrophobic amino acids residues as shown in Fig. 6.6 (b). It can further be observed on Fig. A1. (a), and (b) (appendices) the other two highly active ligands 3d and 3g have shown a similar binding mechanism to the one for the reference ligand (+)-3c. The differences in their binding modes are facilitated by the functional groups in the core structure and also the fact that the active site is highly flexible.

The findings of this case study will help in further elucidating more potent CYP17A1 inhibitors since one can alter the functional groups of a core-structure and try to use molecular modelling techniques to simulate the binding affinity of the altered structure until a hit compound is achieved. Molecular modelling techniques such as docking coupled with pharmacophore modelling are very promising techniques in ligand-based and structure-based drug design. The molecular orbital theory enabled us to explore the reactive functional groups like for example we predicted the LUMO orbitals that are overlaid on reactive functional groups such as the carbonyl group and N-H group of N-methylcarboxamide on (+)-3c (reference molecule) see Fig. 6.6 (b). Docking calculations therefore confirmed this by showing hydrogen bonds between carbonyl group and N-H group of N-methylcarboxamide and the amino acid residues of the enzyme. The above mentioned approaches complement each other in a sense that if the reactivity of the novel inhibitor is unknown one can be able to predict the molecule's behaviour before even starting to synthesize molecules in the laboratory; this is a cost saving strategy to both organic chemists and medicinal chemists in involved in drug discovery and development projects.

CHAPTER 7

CASE STUDY IV

Validation of P450 Site of Metabolism suite from Schrödinger for Site of Metabolism prediction of probe substrates to Cytochrome P450 isoforms

7.1. INTRODUCTION

The Pharmaceutical, cosmetic, nutritional, and agrochemical industries, as well as academia, involved in the discovery of xenobiotics are faced with the major challenge of developing new chemical entities that have high efficacy and are safe for their intended use. Metabolism of xenobiotics is a major contributing factor to the withdrawal of drug candidates and also poses a major concern for other industries where chemical-biological interactions are involved. Cytochrome P450 (CYP450) major isoforms are greatly involved in the metabolism of xenobiotics and endogenous compounds.

The use of computational techniques for accelerating the discovery of new chemical entities and identifying substrates, inhibitors and inducers has enabled the understanding of xenobiotic-enzyme recognition to prevent drug-drug interactions, potential metabolic toxins and poor clearance of some xenobiotics. In the literature it has been reported that there have been some positive success results in the use of pharmacophore modelling, docking and quantum mechanics/molecular mechanics (QM/MM) approaches to enable the elucidation of site of metabolisms (SOMs) and reaction mechanisms for metabolism of NMEs. An account of the breakthroughs and challenges in this field has been discussed in section 2.1 of the literature review.

The Physics-Based ADME/Tox suite is a set of tools for evaluating ADME/Tox properties. The P450 Site of Metabolism suite from Schrödinger contains two tools viz. a tool for identifying likely sites of metabolism and exploring a 3D-space in some P450 isoforms i.e. CYP3A4, 2D6, 2C9. The in-house data from Schrödinger have revealed that the predictive ability of SOMs by CYP3A4 ligand-based reactivity models have been very enormous. The predictive ability of this model is due to the flexibility of CYP3A4's binding site. The use of metabolism workflow, which is combined with induced-fit docking (IFD), addresses the problem of CYP450 isoforms with regioselectivity issues and for ligands that undergoes orientation of chirality centers (Physics-Based ADME/Tox, 2015). Therefore, this case study is aimed at using a physics-based scoring function for the prediction of cytochrome P450 site

of metabolism of known probe substrates. The idea behind using this approach is that this case study will assist in decision making during the discovery process to elucidate structures of potential metabolites of NMEs.

7.2 EXPERIMENTAL METHODS

7.2.1 Computational Details

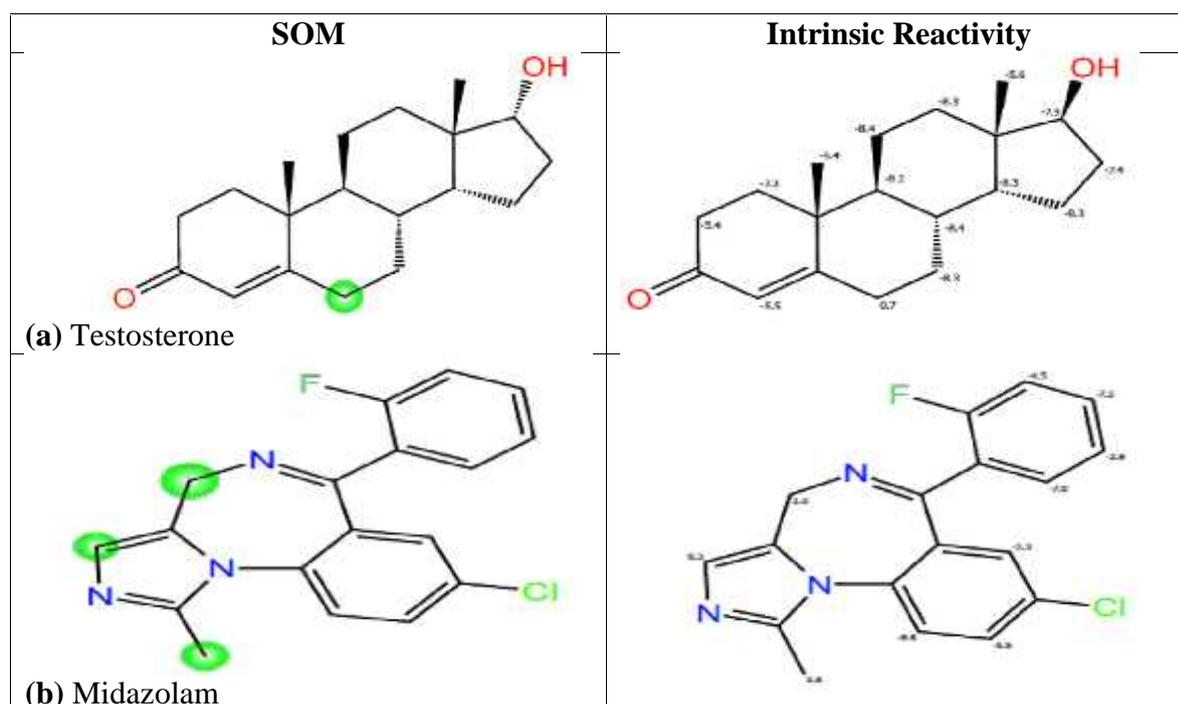
The 2D coordinates of the ligands were prepared using Ligprep (Ligprep, 2012) using the procedure reported in Case study II. The prepared 3D coordinates of the ligands were incorporated into the P450 SOM calculation panel from Maestro (Maestro, 2011). The isoform of interest was chosen among the available isoforms i.e. CYP 2C9, 2D6, and 3A4. However, for CYP3A4 metabolism, only the intrinsic reactivity is calculated because this protocol has already been patented by other companies. In the rest of the isoforms, an induced-fit docking calculation is performed.

7.3 RESULTS AND DISCUSSIONS

7.3.1 Regio-chemistry Prediction of CYP450 Isoforms

The results of the computational prediction of the site of metabolism of known probe substrates for CYP3A4 are shown in Table 7.1 below. The P450 SOM module has been able to correctly model the reactive sites on the substrates that are more likely to be involved in the metabolism of testosterone and midazolam by CYP3A4. The results (see Table 7.1 a below) have been able to confirm that testosterone is hydroxylated at the allylic hydrogen to yield an alcohol. Furthermore, the results for the SOM and intrinsic reactivity for midazolam (see Table 7.1 b below) have confirmed that the benzylic hydroxylation occurs. However, the software picked other site as well. This means that the conformational space of this structure exists in different conformations. Therefore, different conformations are accessible to ferric heme which facilitates metabolism. This means that the computational program samples all conformations of the structure that are possible and yields results based on them. Docking outputs facilitates the decision making in establishing the docking score and the binding mode of the enzyme-substrate complex. Therefore, for the prediction of SOMs of an unknown or NMEs it is important to complement the SOM outputs with intrinsic reactivity and Fe accessibility estimates accompanied by docking outputs. The results are consistent with the metabolic route mentioned in Table 3.1 in the methodology section.

Fig. 7.1. Prediction of SOMs and the intrinsic reactivity of probe substrates for CYP3A4.



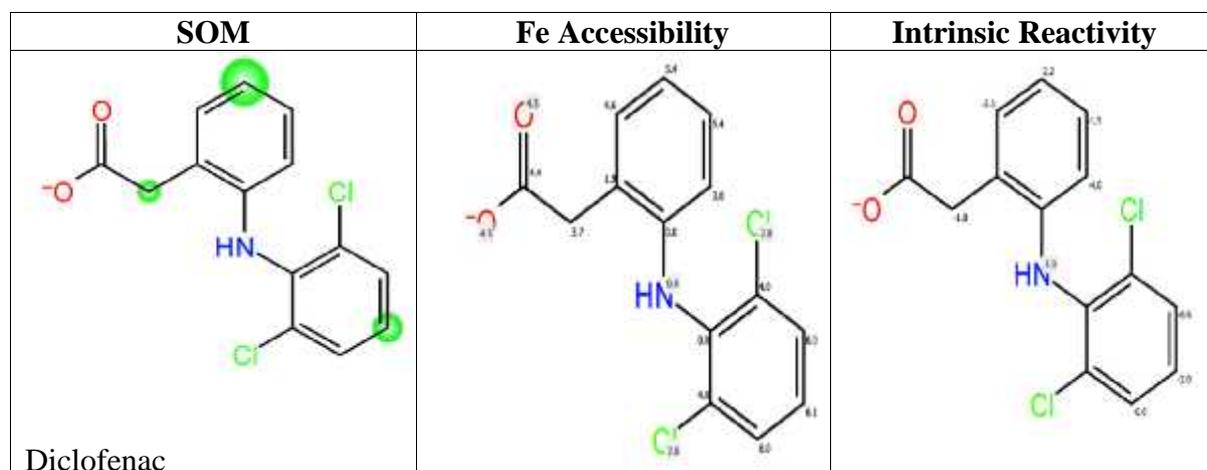
The results shown in Table 7.2 below shows the SOMs, Fe accessibility and the intrinsic reactivity of dextromethophine a probe substrate for CYP2D6. The SOM prediction has not been able to correctly sample the site of metabolism for this substrate. The correct mode of metabolism for this substrate is through the removal of the methyl group (demethylation) to form a hydroxyl group. The reason for the software programme's failure to sample the SOMs might be due to the conformational space of this molecule. The force-field that we used did not account for macro-cycles. Since, the methoxy group is not accessible to heme to facilitate the demethylation process. However, the software has been able to predict the intrinsic reactivity of this substrate. It is hoped that the new force-field OPLS-3.0 would be able to sample the real conformation of this substrate that is available in the new release of Schrödinger software.

Fig. 7.2. Prediction of SOMs and the intrinsic reactivity of Dextromethorphan a probe substrate for CYP2D6.



The P450 SOM prediction results for the metabolism of diclofenac by CYP2C9 are shown in Table 7.3 below. The software was able to predict the SOM for the p-hydroxylation of the aryl group. It has to be noted that the software predicted three SOM sites and the SOM results needs to be complemented by HPLC-MS/MS assay to elucidate the structures of metabolites in the drug discovery pipeline prior to clinical studies.

Fig. 7.3. Prediction of SOMs and the intrinsic reactivity of Diclofenac a probe substrate for CYP2C9.



7.4 CONCLUSIONS

The prediction of the site of metabolisms (SOMs) is a very recent but important step in the drug discovery pipeline. However, the methodology has not yet been validated spanning a wide range of substrate available in the clinic. However, the model presented in this case study could be used to predict the site of metabolism of novel inhibitors, complemented with docking studies. The results of this case study were able to provide insight into the importance of SOM predictions and validation thereof. However, the prediction of SOMs for NMEs should be handled with caution because the programs samples all possible conformations and gives the SOMs results based on them. In fact, the idea of SOM prediction is premised on the prediction of possible metabolism routes which can further be confirmed by metabolic profiling and identification by HPLC-MS/MS assays.

CHAPTER 8

8.1 CONCLUDING REMARKS

Computational drug-HSA interactions performed in this Thesis have been connected with experimentally derived data in view of a synergic aim between these two strategies. In addition, studies performed in this Thesis have enabled the discovery of non-steroidal CYP17A1 inhibitors. The proposed scheme in Fig. 3.1 and 3.2 highlights a steps-by-step guide in the design of novel inhibitors using computational and experimental methods. Fig. 3.1 in the methodology section highlights the *in silico* methods performed in the design stage. While Fig. 3.2 in the methodology section highlight the steps followed after the NMEs have been designed. In fact, this step entails the experimental process from the retrosynthesis analysis of the target molecules, synthesis, structure determination, and *in vitro* bioanalytical assays.

The results of case study I and II gave insights about the ability of docking and post-processing docking methods to reproduce experimental binding affinities and enantioselectivities of enantiomer-protein interactions. For this purpose, the use of known therapeutic drugs with estimated or known biological activities to predict their activities via computational methods served as a validation of *in silico* methods.

The results of case study I were promising to an extent that they were able to provide clues on the factors that drive the synergy between experimental kinetic parameters and computational thermodynamics parameters to explain the interaction between drug enantiomers and the target protein. These parameters were correlated or converted and used to estimate the pseudo enantioselectivity of catechin enantiomers to HSA. This approach of combining docking methodology with docking post-processing methods such as MM-GB/SA proved to be suitable in estimating the correct pseudo binding affinities of protein-ligand complexes. The enantioselectivity for enantiomers of catechin to HSA were 1.60 for site I, close to the value found experimentally, *ES* value (1.5 ± 0.2), and 1.25 for site II. Results of case study I can easily be transferred to novel drug targets provided that the docking outputs are complemented with experimental affinity estimations using validated experimental assays.

In case study II, a combination of *in silico* molecular docking strategies viz. Glide XP, QPLD & IFD have been explored. This strategy involved the following step-wise protocols i.e. pose generation, mathematical conversion from (G°) to (K_a), user defined criteria for the selection of poses. This protocol enabled us to obtain enhanced numerical docking outputs, which are close to *in vitro* experimental affinities, i.e. the Log K_a values obtained from the IFD

method, $\text{Log } K_S$ and $\text{Log } K_R$ of 5.43 and 5.34, respectively, are close to the experimental $\text{Log } K_a$ values (e.g. 5.41 and 5.32 from Capillary Electrophoresis). The docking results obtained have further explained the enantioselectivity of warfarin enantiomers to HSA (e.g. 1.23 and 1.24 from IFD and Capillary Electrophoresis, respectively), which has never been explored before using *in silico* molecular docking approaches.

In case study III, a 3D-QSAR pharmacophore model was developed for the first time using steroidal and non-steroidal inhibitors of CYP17A1 enzyme ranging from high, medium and low affinity inhibitors. This process enabled the exploration of a wide structural space of the interacting ligands with the target enzyme/receptor. A strong connection existed between a series of approaches explored in this work using 3D-QSAR pharmacophore modelling. This enabled the prediction of the functional groups that are common for all molecules, that are thought to play a major role in the binding mechanism with the receptor. The pharmacophore features exhibited by this model were aimed at assisting medicinal chemists to identify functional groups in lead compounds that are important for inhibition of the enzymes. The use of QM/MM optimization on lead compounds from the pharmacophore model has also shed more light on the structural aspects that aid their reactivity. As an example, the reference molecule from the model showed that functional groups mapped onto electrostatic potentials i.e. the negative potential groups (red in Fig. 6.3 a and b), electronegative groups (yellow in Fig. 6.3 (c) and interaction strength (red in Fig. 6.3, 3d) played a crucial role in the inhibition effect of these inhibitors. The blue region indicates a region of low electron density whereas the red part shows the region of high electron density. Therefore, this work has been able to open more avenues in studying the synergies between ligand-based and structure-based molecular modelling approaches.

The molecular docking methods were validated by performing a comparative study on the RMSD of the different ligands and targets using cross-docking and Induced Fit Docking methods. The results revealed that docking methods are able to successfully reproduce the conformations of co-crystallized ligands from protein-ligand complexes. Therefore, attempts to predict the binding modes and binding affinities of steroidal and non-steroidal inhibitors were made. The results revealed that the binding modes for highly active molecules in the data set followed a similar binding mode hypothesis previously exhibited by TOK001 and Abiraterone on CYP17A1 enzyme in the paper published by Devore et al. This observation provides a better understanding about the important active sites of the enzymes and functional groups important in the binding affinities of the inhibitors.

Finally, case study IV was aimed at predicting the site of metabolisms for the known probe substrates of CYP3A4, 2D6 and 2C9. The results were in agreement with the mode of metabolisms depicted in the methodology section (Table 3.1). This module was further used to predict the SOMs of NMEs that are part of a provisional patent (**RSA Pat. Appln. 2015/07849**).

The results of the patented structures are summarised in Chapter 9 that deals with current data, future plans and advanced computational methods to address some shortcomings shown by LBDD & SBDD methods. In fact, LC-MS/MS *in vitro* assays were developed and used in order to determine the IC_{50} and K_i for the inhibitor-CYP17A1 inhibition in the presence of probe substrates and to measure whether the inhibitors undergoes competitive, non-competitive, or uncompetitive inhibition. Furthermore, another LC-MS/MS assay for inhibition and metabolic profiling and identification of novel metabolites of candidate compounds (novel inhibitors) were further developed and used for six major human isoforms, viz. CYP1A2, CYP2C8, CYP2C9, CYP2C19, CYP2D6 and CYP3A4.

The Michaelis-Menten (V_{max} and K_m) parameters for the production of probe metabolites were determined for each assay condition via non-linear curve fitting and the K_i determined using an appropriate model of inhibition (competitive, non-competitive or uncompetitive). The methods and approaches proposed in this work were aimed at shortening the drug discovery process in the design, synthesis, and pre-clinical development, which could avoid late-stage withdrawals of NCEs due to their ineffectiveness and side-effects.

CHAPTER 9

9.1 CURRENT DATA, FUTURE PLANS AND ADVANCED COMPUTATIONAL METHODS TO ADDRESS SHORTCOMINGS OF SBDD & LBDD METHODS

The hits (13 drug-like molecules) that were obtained from the database search using a pharmacophore model as a search query were synthesized in an independent laboratory. Two of the compounds were synthesized as a racemic mixture. These molecules were further tested for CYP19A1 inhibition, which is an enzyme responsible for catalysing aromatization reaction, which is the last step in estrogen biosynthesis. The active site of aromatase (CYP19A1) enzyme is similar to the one for CYP17A1, which we wanted to check whether the candidate compounds are dual inhibitors using docking methodology. Surprisingly, 8 molecules were dual inhibitors. These candidate compounds are new chemical entities and are not derivatives of known molecules with known targets. Hence, a provisional patent was lodged and filed in October 2015 in order to protect our Intellectual Property (IP). A decision was then made to test these molecules for their biological activities experimentally using bioanalytical assays.

The bio-analytical assay employed includes an HPLC-MS/MS method, where the candidate compounds at 6 concentration levels were tested for their inhibition to CYP1A2, 2D6, 2C9, 2C19 and CYP3A4. The percentage activity plot v.s. inhibitor concentration plot revealed that these candidate compounds are weak inhibitors of these metabolic enzymes, which are known to be involved in drug-drug interactions. The selective inhibitors for these enzymes were used as positive controls and the IC₅₀ values for these inhibitors were lower than those of candidate compounds. This therefore means that the candidate compounds are weak inhibitors than the selective inhibitors. The IC₅₀ values for these candidate compounds range from 1.1 μM to >50 μM. The majority of these candidate compounds are weak inhibitors of CYP3A4 enzyme, which is responsible for the metabolism of the majority of drugs.

Another bioanalytical assay for the inhibition of the candidate compounds to CYP17A1 and CYP19A1 enzymes were performed at 6 concentration levels. A total of eight compounds had an IC₅₀ value of >50 μM for the inhibition of CYP17A1, and one molecule had an IC₅₀ value of 43.1 μM. This molecule is a racemic compound and it is a very weak inhibitor of metabolic enzyme isoforms, which means that it is very selective. This is our potential lead compound. Therefore, we are now planning to develop an assay that will also incorporate a chiral selector in order to separate both enantiomers *in situ*. Thereby, measuring the IC₅₀ and

Ki values for each of the stereoisomers. This will enable us to establish the identity of a eutomer and a distomer or rather the most potent or less potent isomer if there is any. This will also help us to measure the Km and Vmax for the formation of probe metabolites of natural substrates.

In the CYP19A1 inhibition assay, some difficulties were experienced since stock solutions were prepared in 1% DMSO and DMSO inhibits CYP19A1. Therefore, a proper solvent needed to be sought and 4% acetonitrile was used in five candidate compounds that were soluble in this solvent. Surprisingly, in this assay three molecules were able to inhibit the enzyme at sub-micro molar range of 5 to 15 μM . The other two molecules inhibited CYP19A1 with an IC_{50} values of >50 and 44 μM . The other three molecules were not soluble to 4% acetonitrile and a mixture of 0.5% DMSO in 4% acetonitrile was prepared. The three molecules had an IC_{50} that is between 0.06, 6 and >25 μM . However, it was also evident that the IC_{50} for the selective inhibitor had a 10fold decrease. This means that DMSO also contributed to the inhibition effect. A proper solvent is going to be sought in future and we also aim to measure the Ki as well as the Vmax and Km to establish the extent of the formation of metabolites of natural substrates. Furthermore, the solubility issues of our NCEs will be enhanced by using FEP/REST technology as well as Metadynamics simulations.

The use of LBDD and SBDD methods in this thesis has indeed enabled the discovery of lead compounds that inhibits CYP17A1 and CYP19A1 enzymes at sub micro molar levels. These inhibitors are also selective as they are weak inhibitors of five major CYP450 metabolic isoforms. The future prospect of this research is based more entirely on improving the potency, solubility and selectivity of the lead compounds. This aim is envisaged to involve the use of combinatorial enumeration workflow. Where the core structure will be linked to substituents obtained from a commercial fragment database. The resulting new structures will be subjected to a molecular docking method via virtual screening work-flows. This will then be followed by DFT optimization of the resulting ligand conformers. This will then be followed by Induced Fit docking protocol. The resulting poses from IFD will be subjected to a Free Energy Perturbation (FEP+) or Free Energy Perturbation/Replica Exchange Molecular Dynamics with Solution Tempering (FEP/REST) method in order to measure the correct binding affinities of the hits to the respective enzymes. The FEP/REST method has recently been validated extensively by Schrödinger scientists. The REST methods have enabled the exploration of the conformational space of bioactive compounds in order to reach their global or local minima, which is advantageous in this research area. Several industries in the pharmaceutical sector have successfully used this technology for enhancing potency, selectivity and solubility of their

lead compounds. Therefore, since Graphic Processing Unit (GPU) computing is amenable to Molecular Dynamics, hence, the speed and accuracy has been enhanced in recent years. These methods can now be used for high-throughput screening of NCEs for lead to hit identification.

REFERENCES

Agovino, T., The Washington Post. (Oct 25, 2005). Generic drugs could have saved U.S. \$20 billion.

Akhtar, M.K.; Kelly, S.L.; Kaderbhai, M.A., Cytochrome b5 modulation of 17 α -hydroxylase and 17 β -HSD activities in steroidogenesis, *J. Endocrinol.*, (2005), 187: 267–274.

Alberty, R.A., Calculation of Standard Formation Properties of Species from Standard Transformed Formation Properties of Reactants in Biochemical Reactions at Specified pH, *J. Phys. Chem. B.*, (1999), 103: 261-265.

Alberty, R.A., Calculation of Standard Transformed Gibbs Energies and Standard Transformed Enthalpies of Biochemical Reactants, *Arch. Biochem. Biophys.*, (1998), 353:116–130.

Alzate-Morales, J.H.; Vergara-Jaque, A.; Caballero, J., Computational study on the interaction of N1 substituted pyrazole derivatives with B-Raf Kinase: An unusual water wire hydrogen-bond network and novel interactions at the entrance of the active site; *J. Chem. Inf. Model.*, (2010), 50: 1101-1112.

Andrisano, V.; Bertucci, C.; Cavrini, V.; Recanatini, M.; Cavalli, A.; Varoli, L.; Felix, G.; Wainer, I.W., Stereoselective binding of 2,3-substituted 3-hydroxypropionic acids on an immobilised human serum albumin chiral stationary phase: stereochemical characterisation and quantitative structure–retention relationship study, *J. Chromatogr. A.*, (2000), 876: 75 – 86.

Asensi-Bernardi, L.; Martín-Biosca, Y.; Villanueva Camañas, R. M.; Medina-Hernandez, M. J.; Sagrado, S., Evaluation of enantioselective binding of fluoxetine to human serum albumin by ultrafiltration and CE-Experimental design and quality considerations. *Electrophoresis.*, (2010), 31: 3268–3280.

Attard, G.; Reid, A.H.M.; A'Hern, R.; Parker, C.; Oommen, N. B.; Folkard, E.; Messiou, C.; Molife, L. R.; Maier, G.; Thompson, E.; Olmos, D.; Sinha, R.; Lee, G.; Dowsett, M.; Kaye, S. B.; Dearnaley, D.; Kheoh, T.; Molina, A.; de Bono, J.S., Selective Inhibition of CYP17 with Abiraterone Acetate Is Highly Active in the Treatment of Castration-Resistant Prostate Cancer, *J. Clin. Oncol.*, (2009), 27: 3742-3748.

Baranczewski, P.; Stańczak, A.; Sundberg, K.; Svensson, R.; Wallin, Å.; Jansson, J.; Garberg, P.; Postlind, H., Introduction to *in vitro* estimation of metabolic stability and drug interactions of new chemical entities in drug discovery and development. *Pharmacol. Reports.*, (2006), 58: 453–472.

Baron D.H.R., Conformational Analysis in 100 years with Nobel Laureates, Beltran, H.; Beer, T. M.; Carducci, M. A.; de Bono, J.; Gleave, M.; Blomberg, M. R. A.; Borowski, T.; Himo, F.; Rong-Zhen, L.; Siegbahn, P. E. M., Bohm H.J.; Stahl M., Rapid empirical scoring functions in virtual screening, *Med. Chem. Res.*, (1999), 9: 445-462.

Blomberg, M. R. A; Borowski, T; Himo, F; Rong-Zhen, L; Siegbahn, P. E. M., Quantum Chemical Studies of Mechanisms for Metalloenzymes, *Chem. Rev.*, (2014), 114: 3601–3658.

Born, M.; Heisenberg, W.; Jordan, P.; "Zur Quantenmechanik, II," *Zs. f. Phys.*, (1926), 35: 557-615.

Bos, O. J. M.; Fischer, M. J. E.; Wilting, J.; Janssen, L. H. M., Mechanism by which warfarin binds to human serum albumin. Stopped-flow kinetic experiments with two large fragments of albumin, *Biochem. Pharmacol.*, (1989), 38, 1979–1984.

Breckenridge, A.; Orme, M.; Wesseling, H.; Lewis, R.J.; Gibbons, R., Pharmacokinetics and pharmacodynamics of the enantiomers of warfarin in man, *Clin. Pharmacol. Ther.*, (1974) 15: 424–430.

Brooks, W. H.; Daniel, K. G.; Sung, S.; Guida, W. C., Computational Validation of the Importance of Absolute Stereochemistry in Virtual Screening, *J. Chem. Inf. Model.*, (2008), 48: 639-645.

Bryce, A.; Ryan, C.J., Development and Clinical Utility of Abiraterone Acetate as an Androgen Synthesis Inhibitor, *Clin. Pharmacol. Ther.*, (2012), 91: 101-108.

Budha, N.R.; Mehrotra, N.; Tangallapally, R.; Qi, R.J.; Daniels, A.J.; Lee, R.E.; Meibohm, B.; Pharmacokinetically-guided lead optimization of nitrofuranylamide anti-tuberculosis agents, *APPS. J.*, (2008), 10: 157-165.

Campagna-Slater, V.; Pottel, J.; Therrien, E.; Cantin, L.; Moitessier, N., Development of a computational tool to rival experts in the prediction of sites of metabolism of xenobiotics by P450s, *J. Chem. Inf. Model.*, (2012), 52: 2471-2483.

Carter, D. C.; Ho, J. X., Structure of Serum Albumin, *Adv. Protein Chem.* (1994), 45: 153–176.

Cheatham, T.E.; Srinivasan, J.; Case, D.A.; Kollman, P.A., Molecular dynamics and continuum solvent studies of the stability of polyG-polyC and polyA-polyT DNA duplexes in solution; *J. Biomol. Struct. Dyn.*, (1998), 16: 265-280.

Cho, A. E.; Guallar, V.; Berne, B.; Friesner, R. A.; Importance of Accurate Charges in Molecular Docking: Quantum Mechanical/Molecular Mechanical (QM/MM) Approach. *J. Comput. Chem.*, (2005), 26: 915-931.

Chohan, K.K.; Paine, S.W.; Waters, N.J., Advancements in Predictive *in Silico* Models for ADME, *Curr. Chem. Biol.*, (2008), 2: 215-228.

Christoph, M.; "August Wilhelm Hofmann—"Reigning Chemist-in-Chief," *Angew. Chem. Int. Edit.*, (1992), 31: 1265–1282.

Chuang, V.T.G.; Otagiri, M., Stereoselective binding of Human Serum Albumin, *Chirality*, (2006), 18: 159-166.

Clement, O. O.; Freeman, C.M.; Hartmann, R.W.; Handratta, V.D.; Vasaitis, T. S.; Brodie, A.M.H.; Njar, V.C.O., Three Dimensional Pharmacophore Modeling of Human CYP17

Inhibitors. Potential Agents for Prostate Cancer Therapy, *J. Med. Chem.*, (2003), 46: 2345-2351.

Danchin, A., *The Delphic Boat – What Genomes Tell Us*. Harvard University Press: Cambridge, MA. 2002.

Danchin, A.; Medigue, C.; Gascuel, O.; Soldano, H.; Henaut, A., From data banks to data bases, *Res. Microbiol.*, (1991) 142: 913-916.

de Broglie, L.; A tentative theory of light quanta, *Phil. Mag.*, (1924), 47: 446-458.

De Groot, M.J.; Ekins, S., Pharmacophore modelling of cytochrome P450, *Adv. Drug Del. Rev.*, (2002), 54: 367-383.

De Groot, M.J.; Kirton, S.B.; Sutcliffe, M.J., In silico methods for predicting ligand binding determinants of cytochrome P450, *Curr. Top. Med. Chem.*, (2004), 4: 1803-1824.

De Groot, M.J.; Lewis, D.F.V.; Modi, S., Molecular modelling and quantitative structure-activity relationship of substrate and inhibitors of drug metabolism enzymes, *Compr. Med. Chem. II.*, (2007), 5: 809-825.

Deeb, O.; Rosales-Hernández, M.C.; Gómez-Castro, C.; Garduño-Juárez, R.; Correa-Basurto, J., Exploration of human serum albumin binding sites by docking and molecular dynamics flexible ligand–protein interactions, *J. Biopol.*, (2010), 93: 161-170.

Deora, G.S.; Joshi, P.; Rathore, V.; Kumar, K.L.; Ohlyan, R.; Kandale, A., Pharmacophore modelling and 3D QSAR analysis of isothiazolidinedione derivatives as PTP1B inhibitors; *Med. Chem. Res.*, (2013), 22: 3478-3484.

DeVore, N.M.; Scott, E.E., Structures of cytochrome P450 17A1 with prostate cancer drugs abiraterone and TOK001, *Nature.*, (2012), 482: 116-120.

Dirac, P.A.M.; "The Evolution of the Physicist's Picture of Nature," *Sci. Amer.*, (1963), 208: 45-53.

Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A., PHASE: A New Engine for Pharmacophore Perception, 3D QSAR Model Development, and

3D Database Screening. Methodology and Preliminary Results, *J. Comput. Aided. Mol. Des.*, (2006), 20, 647-671.

Dror, O.; Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H.J., Novel approach for efficient pharmacophore-based virtual screening: Method and applications, *J. Chem. Inf. Model.*, (2009), 49: 2333-2343.

Durdagi, S; Duff, H.J; Noskov, S.Y., Combined receptor and ligand-based approach to the universal pharmacophore model development for studies of drug blockade to the hERG1 pore domain, *J. Chem. Inf. Model.*, (2011), 51: 463-474.

Ekins, S.; Metres, J.; Testa, B., *In silico* pharmacology for drug discovery: methods for virtual ligand screening and profiling, *Br. J. Pharmacol.*, (2009) 152: 9-20.

Ekins, S.; Stresser, D.M.; Williams, J.A., In vitro and pharmacophore insights into CYP3A enzymes, *Trends. Pharmacol. Sci.*, (2003), 24: 161-166.

Elumalai, P.; Liu, H. L.; Zhao, J.H.; Chen, W.; Lin, D. S.; Chuang, C.K.; Tsai, W.B.; Ho, Y., Pharmacophore modeling, virtual screening and docking studies to identify novel HNMT inhibitors, *J. Taiwan. Inst. Chem. Eng.*, (2012), 43: 493–503.

Englebienne, P.; Fiaux, H.; Kuntz, D.A.; Corbeil, C.R.; Gerber-Lemaire, S.; Rose, D.R., Evaluation of Docking programs for predicting binding Golgi alpha-mannosidase ii inhibitors: a comparison with crystallography, *Proteins. Struct. Funct. Bioinf.*, (2007), 69: 160-176.

Epik version 2.2, Schrödinger, LLC, New York, NY, 2011.

Escuder-Gilabert, L.; Martínez-Gómez, M. A.; Villanueva Camanas, R. M.; Sagrado, S.; Medina-Hernández, M. J., Microseparation techniques for the study of the enantioselectivity of drugplasma protein binding, *Biomed. Chromatogr.*, (2009), 23, 225–238.

Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.;

Repasky, M. P.; Knoll, E. H.; Shaw, D. E.; Shelley, M.; Perry, J. K.; Francis, P.; Shenkin, P.

S., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy, *J. Med. Chem.*, (2004), 47: 1739-1749.

Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrín, P. C.; Mainz, D. T., Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes., *J. Med. Chem.*, (2006), 49: 6177–6196.

Gashaw, I.; Ellinghaus, P.; Sommer, A.; Asadullah, K., What makes a good drug target? *Drug. Discov. Today.*, (2011), 16: 1037-1043.

Ghuman, J.; Zunszain, P. A.; Petitpas I.; Bhattacharya, A. A.; Otagiri, M.; Curry, S., Structural Basis of the Drug-binding Specificity of Human Serum Albumin, *J. Mol. Biol.*, (2005), 353: 38–52.

Giangreco, I.; Cosgrove, D.A.; Packer, M.J., An Extensive and Diverse Set of Molecular Overlays for the Validation of Pharmacophore Programs; *J. Chem. Inf. Model.*, (2013), 53: 852-866.

Gianti, E.; Zauhar, R.J., Modeling Androgen Receptor Flexibility: A Binding Mode Hypothesis of CYP17 Inhibitors/Antiandrogens for Prostate Cancer Therapy, *J. Chem. Inf. Model.*, (2012), 52: 2670–2683.

Glide, version 5.0, Schrödinger, LLC, New York, NY, 2010.

Glide, version 5.7, Schrödinger, LLC, New York, NY, 2011.

Glide, version 5.7, Schrödinger, LLC, New York, NY, 2011.

Graves A.P.; Shivakumar D.M.; Boyce S.E.; Jacobson M.P.; Case D.A.; Shoichet B.K., Rescoring docking hits lists for model cavity sites: Predictions and experimental testing, *J. Mol. Biol.*, (2008), 377: 914-934.

Graves, A.P.; Shivakumar, D.M.; Boyce, S.E.; Jacobson, M.P.; Case, D.A.; Shoichet, B.K., Rescoring docking hits lists for model cavity sites: Predictions and experimental testing, *J. Mol. Biol.*, (2008), 377: 914-934.

Guallar, V.; Friesner, R.A., Cytochrome P450_{CAM} enzymatic catalysis cycle: A Quantum Mechanics/Molecular Mechanics study, *J. Am. Chem. Soc.*, (2004), 126: 8501-8508.

Guengerich, F. P.; Zhong-Liu, W.; Bartleson, C.J., Function of human cytochrome P450s: Characterization of the orphans, *Biochem. Biophys. Res. Comm.*, (2005), 338: 465–469.

Guimarães, C.R.W.; Boger, D.L.; Jorgensen, W.L., Elucidation of Fatty Acid Hydrolase Inhibition by Potent α -Ketoheterocycle Derivatives from Monte Carlo Simulations, *J. Am. Chem. Soc.*, (2005), 127: 17377-17384.

Guimarães, C.R.W.; Cardozo, M., MM-GB/SA Rescoring of Docking Poses in Structure-Based Lead Optimization, *J. Chem. Inf. Model.*, (2008), 48: 958-970.

Gumede, N.J.; Singh, P.; Sabela, M.I.; Bisetty, K.; Escuder-Gilabert, L.; Medina-Hernández, M.J.; Sagrado, S., Experimental-like affinity constants and enantioselectivity estimates from flexible docking. *J. Chem. Inf. Model.*, (2012), 52: 2754-2759.

Günther, S.; Senger, C.; Michalsky, E.; Goede, E.; Preissner, R.; Representation of target-bound drugs by computed conformers: implications for conformational libraries; *BMC. Bioinf.*, (2006), 7: 293-303.

Haeffner, F.; Norin, T.; Hult, K., Molecular Modeling of the Enantioselectivity in Lipase-Catalyzed Transesterification Reactions, *Biophys. J.*, (1998), 74: 1251-1262.

Haidar, S.; Ehmer, P.B.; Barassin, S.; Batzl-Hartmann, C.; Hartmann, R.W., Effects of novel 17 α -hydroxylase/C17, 20-lyase (P450 17, CYP 17) inhibitors on androgen biosynthesis *in vitro* and *in vivo*, *J. Steroid. Biochem. Mol. Biol.*, (2003), 84: 555–562.

Hakki, T.; Bernhardt, R., CYP17- and CYP11B-dependent steroid hydroxylases as drug development targets, *Pharmacol. Therapeut.*, (2006), 111: 27 – 52.

Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Handratta, V.D.; Vasaitis, T.S.; Njar, V.C.; Gediya, L.K.; Kataria, R.; Chopra, P.; Newman, D.; Farquhar, R.; Guo, Z.; Qui, Y.; Brodie, A.M., Novel C-17-heteroaryl steroidal CYP17 inhibitors/antiandrogens: synthesis, in vitro biological activity, pharmacokinetics, and antitumor activity in the LAPC4 human prostate cancer xenograft model; *J. Med. Chem.*, (2005), 48: 2972-2984.

Hayes, C.; Ansbro, D.; Kontoyianni, M., Elucidating Substrate Promiscuity in the Human Cytochrome 3A4, *J. Chem. Inf. Model.*, (2014), 54: 857–869.

Herrero-Martínez, J.M.; Sanmartin, M.; Rosés, M.; Bosch, E.; Ràfols, C., Determination of dissociation constants of flavonoids by capillary electrophoresis, *Electrophoresis.*, (2005), 26: 1886–1895.

Hille, U. E.; Hu, Q.; Pinto-Bazurco Mendieta, M. A. E.; Bartels, M.; Vock, C. A.; Lauterbach, T.; Hartmann, R. W., Steroidogenic cytochrome P450 (CYP) enzymes as drug targets: Combining substructures of known CYP inhibitors leads to compounds with different inhibitory profile, *Comptes. Rendus. Chimie.*, (2009), 12: 1117 – 1126.

Hille, U.E.; Hu, Q.; Vock, C.; Negri, M.; Bartels, M.; Müller-Vieira, U.; Lauterbach, T.; Hartmann, R.W., Novel CYP17 inhibitors: Synthesis, biological evaluation, structure-activity relationships and modelling of methoxy and hydroxy-substituted methyleneimidazolyl biphenyls, *Eur. J. Med. Chem.*, (2008), 51: 5064–5074.

Hirsh, J.; Dalen, J.E.; Anderson, D.R.; Poller, L.; Bussey, H.; Ansell, J.; Deykin, D.; Brandt, J.T., Oral anticoagulants: mechanism of action, clinical effectiveness, and optimal therapeutic range. *Chest.*, (1998), 114: 445S–469S.

Hu, Q.; Jagusch, C.; Hille, U.E.; Hauptenthal, J.; Hartmann, R.W., Replacement of imidazolyl by pyridyl in biphenylmethylenes results in selective CYP17 and dual CYP17/CYP11B1 inhibitors for the treatment of prostate cancer; *J. Med. Chem.*, (2010a), 53: 5749-5758.

Hu, Q.; Yin, L.; Jagusch, C.; Hille, U.E.; Hartmann, R.W., Isopropylidene Substitution Increases Activity and Selectivity of Biphenylmethylene 4-Pyridine Type CYP17 Inhibitors, *J. Med. Chem.*, (2010b), 53: 5049–5053.

Hussain, M.; Kelly, W. K.; Saad, F.; Sternberg, C.; Tagawa, S.T.I.; Tannock, F., New Therapies for Castration-Resistant Prostate Cancer: Efficacy and Safety, *Eur. Urol.*, (2011), 60: 279 – 290.

Huynh, L.; Masereeuw, R.; Friedburg, T.; Ingelman-Sundberg, M.; Manivet, P., In silico platform for xenobiotics ADME-T pharmacological properties modeling and prediction. Part 1: beyond the reduction of animal and model use, *Drug. Discov. Today.*, (2009),14: 401-405.

Induced Fit Docking protocol. Schrödinger, LLC, New York, NY, 2011.

Ishii, T.; Minoda, K.; Bae, M.J.; Mori, T.; Uekusa, Y.; Ichikawa, T.; Aihara, Y.; Furuta, T.; Wakimoto, T.; Kan, T.; Nakayama, T., Binding affinity of tea catechins for HSA: Characterization by high-performance affinity chromatography with immobilized albumin column, *Mol. Nutr. Food. Res.*, (2010), 54: 816–822.

Jaguar, version 7.9, Schrödinger, LLC, New York, NY, 2012.

Jagusch, C.; Negri, M.; Hille, U. E.; Hu, Q.; Bartels, M.; Jahn-Hoffmann, K.; Pinto-Bazurco Mendieta, M. A. E.; Rodenwaldt, B.; Müller-Vieira, U.; Schmidt, D.; Lauterbach, T.; Recanatini, M.; Cavalli, A.; Hartmann, R. W., Synthesis, biological evaluation and molecular modelling studies of methyleneimidazole substituted biaryls as inhibitors of human 17 α -hydroxylase-17,20-lyase (CYP17). Part I: Heterocyclic modifications of the core structure, *Bioorgan. Med. Chem.*, (2008), 16: 1992–2010.

Jagusch, C; Negri, M; Hille, U. E; Hu, Q; Bartels, M; Jahn-Hoffmann, K; Müller-Vieira, U; Schimidt, D; Lauterbach, T; Recanatini, M; Cavalli, A.; Hartmann, R.W; Synthesis, biological evaluation and molecular modeling studies of methyleneimidazole substituted

biaryl as inhibitors of human 17 α -hydroxylase-17,20-lyase (CYP17). Part I: Heterocyclic modifications of the core structure; *Bioorgan. Med. Chem.*, (2008), 16: 1992-2010.

Jain, A.N., Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine, *J. Med. Chem.*, (2003), 46: 499-511.

Jain, A.N., Surflex-Dock 2.1.: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J. Comput. Aided. Mol. Des.*, (2007), 21: 281-306.

Jain, S.V.; Ghate, M.; Bhadoriya, K.S.; Bari, S.B.; Sugandhi, G.; Mandwal, P., 3D-QSAR pharmacophore modelling and in silico screening of phospholipase A₂ inhibitors, *Med. Chem. Res.*, (2013), 22: 3096-3108.

John, S.; Thangapandian, S.; Sakkiah, S.; Lee, K.W., Identification of potent virtual leads to design novel indoleamine 2,3-dioxygenase inhibitors: Pharmacophore modelling and molecular docking studies, *Eur. J. Med. Chem.*, (2010), 45: 4004-4012.

Jones, D.; Moran, J.H.; Miller, G.P., Warfarin and UDP-glucuronosyltransferases: writing a new chapter of metabolism, *Drug. Metab. Rev.*, (2010), 42: 53.

Jorgensen, W.L., Free Energy Calculations: A breakthrough for Modelling Organic Chemistry in Solution, *Acc. Chem. Res.*, (1989), 22: 184-189.

Jorgensen, W.L., Free energy changes in solution. In Encyclopedia of Computational Chemistry; Wiley: New York (1998), Vol 2: 1061-1070.

Jozwiak, K.; Moaddel, R.; Ravichandran, S.; Plazinska, A.; Kozak, J.; Patel, S.; Yamaguchi, R.; Wainer, I.W., Exploring Enantiospecific Ligand-Protein Interactions Using Cellular Membrane Affinity Chromatography: Chiral Recognition as a Dynamic Process, *J. Chromatogr. B.*, (2008), 875: 200-207.

Kaku, T.; Hitaka, T.; Ojida, A.; Matsunaga, N.; Adachi, M.; Tanaka, T.; Hara, T.; Yamaoka, M.; Kusaka, M.; Okuda, T.; Asahi, S.; Furuya, S.; Tasaka, A., Discovery of orteronel (TAK700), a naphthylmethylimidazole derivative, as a highly selective 17,20-lyase inhibitor

with potential utility in the treatment of prostate cancer; *Bioorgan. Med. Chem.*, (2011a), 19: 6383-6399.

Kaku, T.; Tsujimoto, S.; Matsunaga, N.; Tanaka, T.; Hara, T.; Yamaoka, M.; Kusaka, M.; Tasaka, A., 17,20-Lyase inhibitors. Part 3: Design, synthesis, and structure-activity relationships of biphenylmethylimidazole derivatives as novel 17, 20-lyse inhibitors; *Bioorgan. Med. Chem.*, (2011b), 19: 2428-2442.

Kamataki, T., Drug Metabolism Enzyme, Hirokawa Publ., Tokyo, 2001.

Kaminsky, L.S.; Zhang, Z.Y., Human P450 metabolism of warfarin, *Pharmacol. Ther.*, (1997), 73: 67-74.

Kawatkar, S.; Wang, H.; Czerminski, R.; Joseph-McCarthy, D., Virtual fragment screening: an exploration of various docking and scoring protocols for fragments using Glide, *J. Comput. Aided. Mol. Des.*, (2009), 23: 527–539.

Khanfar, M. A.; AbuKhader, M. M.; Alqtaishat, S.; Taha, M. O.; Pharmacophore modeling, homology modeling, and in silico screening reveal mammalian target of rapamycin inhibitory activities for sotalol, glyburide, metipranolol, sulfamethizole, glipizide, and pioglitazone, *J. Mol. Graphics Modell.*, (2013), 42: 39–49.

Kirubakaran, P.; Muthusamy, K.; Singh, K. D.; Nagamani, S., Pharmacophore modeling, 3D-QSAR, and molecular docking study on naphthyridine derivatives as inhibitors of 3-phosphoinositide-dependent protein kinase-1, *Med. Chem. Res.* (2013), 22: 3812–3822.

Kola, I.; Landis, J., Can the pharmaceutical industry reduce attrition rates? *Nat. Rev.: Drug Discov.*, (2004), 3: 711–715.

Kollman, P.A., Free Energy Calculations: Applications to Chemical and Biochemical Phenomena, *Chem. Rev.*, (1993), 93: 2395-2417.

Kollman, P.A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W., Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models, *Acc. Chem. Res.*, (2000), 33: 889-897.

Kragh-Hansen, U., Structure and ligand binding properties of human serum albumin, *Dan. Med. Bull.*, (1990), 37: 57–84.

Kremer, J.M.H.; Bakker, G.; Wilting, J., The role of the transition between neutral and basic forms of human serum albumin in the kinetics of the binding to warfarin, *Biochim. Biophys. Acta.* (1982), 708: 239–242.

Kroemer, R.T., Structure-based Drug Design: Docking and Scoring, *Curr. Protein. Pept. Sc.*, (2007), 8: 312-328.

Krug, S.J.; Hu, Q.; Hartmann, R. W., Hits identified in library screening demonstrate selective CYP17A1 lyase inhibition, *J. Steroid. Biochem. Mol. Biol.*, (2013), 134: 75–79.

Kuhn, B.; Kollman, P.A., Binding of a Diverse Set of Ligands to Avidin and Streptavidin: An Accurate Quantitative Prediction of their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models, *J. Med. Chem.*, (2000), 43: 3786-3791.

Kuntz, I.D.; Blaney, J.M.; Oatley, S.J.; Langridge, R.; Ferrin, T.E., A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.*, (1982), 161: 269-288.

Lagercrantz, C.; Larsson, T., Comparative studies of the binding of some ligands to human serum albumin non-covalently attached to immobilized Cibacron Blue, or covalently immobilized on Sepharose, by column affinity chromatography, *Biochem. J.*, (1983) 213: 387-390.

Lammerhofer, M., Chiral recognition by enantioselective liquid chromatography: Mechanisms and modern chiral stationary phases, *J. Chromatogr. A.*, (2010), 1217: 814-856.

Le Bele, J.A.; “Les relations du pouvoir rotatoire avec la structure moléculaire”, *Rev. Scientifique*, (1891), 48: 609-617.

Lemmen, C.; Lengauer, T., Computational methods for the structural alignment of molecules; *J. Comp. Aided Mol. Des.*, (2000), 14: 215-232.

Ley, S.V.; Myers, R.M.; "Sir Derek Harold Richard Barton 8 September 1918—16 March 1998". *Biograph. Memoir. Fell. Royal. Soc.*, (2002), 48: 1–23.

Li, J.; Schneebeli, S.T.; Bylund, J.; Farid, R.; Friesner, R.A., IDSite: An Accurate Approach to Predict P450-Mediated Drug Metabolism, *J. Chem. Theory. Comput.*, (2011), 7: 3829–3845.

Li, W.; Liu, C.; Tan, G.; Zhang, X.; Zhu, Z.; Chai, Y., Molecular Modeling Study of Chiral Separation and Recognition Mechanism of α -Adrenergic Antagonists by Capillary Electrophoresis, *Int. J. Mol. Sci.*, (2012), 13: 710-725.

Li, X.; Dash, R.K.; Pradham, R.K.; Qi, F.; Thompson, M.; Vinnakota, K.C.; Wu, F.; Yang, F.; Beard, D.A., A Database of Thermodynamic Quantities for the Reactions of Glycolysis and the Tricarboxylic Acid Cycle, *J. Phys. Chem. B.*, (2010), 114: 16068-16082.

Li, Y.; Han, L.; Liu, Z.; Wang, R., Comparative Assessment of Scoring Functions on an Updated Benchmark: 2. Evaluation Methods and General Results, *J. Chem. Inf. Model.*, (2014), 54: 1717–1736.

LigPrep, version 2.4, Schrödinger, LLC, New York, NY, 2010.

Lineweaver, H.; Burk, D.; The Determination of Enzyme Dissociation Constants, *J. Am. Chem. Soc.*, (1934), 56: 658-666.

Link, K.P., The Discovery of Dicumarol and its Sequels. *Circulation.*, (1959), 19: 97-107.

Lokwani, D. K.; Sarkate, A. P.; Shinde, D. B.; 3D-QSAR and docking studies of benzoyl urea derivatives as tubulin-binding agents for antiproliferative activity, *Med. Chem. Res.*, (2013), 22:1415–1425.

Loun, B.; Hage, D.S., Chiral Separation Mechanisms in Protein-Based HPLC Columns. 1 Thermodynamic Studies of (R)- and (S)-Warfarin Binding to Immobilized Human Serum Albumin, *Anal. Chem.*, (1994), 66:3814-3822.

Lowe, D.B.; Nowhere to Go But Up: The Return of Medicinal Chemistry, *ACS Med. Chem. Lett.*, (2012), 3: 3–4.

Lu, S. H.; Wu, J.W.; Liu, H. L.; Zhao, J.H.; Liu, K.T.; Chuang, C.K.; Lin, H.Y.; Tsai, W.B.; Ho, Y., The discovery of potential acetylcholinesterase inhibitors: A combination of pharmacophore modeling, virtual screening, and molecular docking studies, *J. Biomed. Science.*, (2011), 18: 1-13.

MacroModel, version 9.9, Schrödinger, LLC, New York, NY, 2011.

Maestro, version 9.1, Schrödinger, LLC, New York, NY, 2010.

Maestro, version 9.2, Schrödinger, LLC, New York, NY, 2011.

Miller, J. H. M.; Smail, G. A., Acute anti-inflammatory effects of aspirin and dexamethasone in rats deprived of endogenous prostaglandin precursors, *J. Pharm. Pharmacol.*, (1977), 29: 235–237.

Miramontes, P.; *DNA and RNA physicochemical constraints, cellular Automata and Molecular evolution*, (1989), workshop at Los Alamos, New York.

Miramontes, P.; Un modelo de autómatas celulares para la evolución de los ácidos nucleicos [A cellular automaton model for the evolution of nucleic acids]. Tesis de doctorado en matemáticas. UNAM. 1992.

Mobley, D.L.; Dill, K.A., Binding of Small-Molecule Ligands to Proteins: “What you See” Is Not Always “What you Get”, *Structure.*, (2009), 17: 489-498.

Moroy, G.; Martiny, V.Y.; Vayer, P.; Villoutreix, B.O.; Miteva, M.A.; Towards in silico structure-based ADMET prediction in drug discovery, *Drug. Discov. Today.*, (2012), 17: 44-55.

Murcko, M.A., Computational methods to predict binding free energy in ligand-receptor complexes. *J. Med. Chem.*, (1995), 38: 4953-4957.

Narkhede, S. S.; Degani, M. S.; Pharmacophore Refinement and 3D-QSAR Studies of Histamine H3 Antagonists, *QSAR Comb. Sci.*, (2007), 26: 744 – 753.

Nelson, D.R., Cytochrome P450 and the Individuality of Species, *Arch. Biochem. Biophys.*, (1999), 369:1–10.

New Delhi: Encyclopedia Britannica (India) Pvt. Ltd., and I.K. International Pvt. Ltd., New Delhi, 2001.

Nnane, I.P.; Kato, K.; Liu, Y.; Long, B.J.; Lu, Q.; Wang, X.; Ling, Y. Z.; Brodie, A., Inhibition of Androgen Synthesis in Human Testicular and Prostatic Microsomes and in Male Rats by Novel Steroidal Compounds; *Endocrinology.*, (1999), 140: 2891–2897.

Nowak, P.; Wozniakiewicz, M.; Koscielniak, P., Simulation of drug metabolism. *Trends. Anal. Chem.*, (2014), 59: 42–49.

Olsen, L.; Oostenbrink, C.; Jørgensen, F.S., Prediction of cytochrome P450 mediated metabolism, *Adv. Drug. Deliv. Rev.*, (2015), 86: 61-71.

O'Reilly, R. A., Interaction of several coumarin compounds with human and canine plasma albumin, *Mol. Pharmacol.*, (1971), 7: 209-218.

Pearlman, D.A.; Charifson, P.S., Are Free Energy Calculations Useful in Practice? A comparison with Rapid Scoring Functions for the p38 MAP Kinase Protein System, *J. Med. Chem.*, (2001), 44: 3417-3423.

Peters, T. All about Albumin: Biochemistry, Genetics, and Medical Applications, Academic Press, (1995), San Diego.

Petitpas, I.; Bhattacharya, A. A.; Twine, S.; East, M.; Curry, S., Crystal Structure Analysis of Warfarin Binding to Human Serum Albumin ANATOMY OF DRUG SITE I, *J. Biol. Chem.*, (2001), 276: 22804 –22809.

Physics-Based ADME/Tox, Schrödinger Software Release 2015-2, Schrödinger, LLC, New York, NY, 2015.

Pinto-Bazurco Mendieta, M. A. E.; Negri, M.; Hu, Q.; Hille, U. E.; Jagusch, C.; Jahn-Hoffmann, K.; Müller-Vieira, U.; Schmidt, D.; Lauterbach, T.; Hartmann, R. W., CYP17 Inhibitors. Annulations of Additional Rings in Methylene Imidazole Substituted Biphenyls: Synthesis, Biological Evaluation and Molecular Modelling, *Arch. Pharm. Chem. Life Sci.*, (2008), 341: 597–609.

Pinto-Bazurco Mendieta, M.A.E.; Negri, M.; Jagusch, C.; Müller-Vieira, U.; Lauterbach, T.; Hartmann, R.W., Synthesis, biological evaluation, and molecular modeling of abiraterone analogues: novel CYP17 inhibitors for the treatment of prostate cancer, *J. Med. Chem.*, (2008), 51: 5009-5018.

Pollard, W. T.; Banks, J. L., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.*, (2004), 47: 1750–1759.

Porter, W. R.; Warfarin: history, tautomerism and activity, *J. Comp. Aided. Mol. Des.*, (2010) 24: 553-573.

Prasad, V.V.S.R.; Reddy, G.D.; Appaji, D.; Peters, G.J.; Mayur, Y.C.; Chemosensitizing acridones: In vitro calmodulin dependent cAMP phosphodiesterase inhibition, docking, pharmacophore modeling and 3D QSAR studies, *J. Mol. Graphics Modell.*, (2013), 40:116–124.

Prime version 3.0, Schrödinger, LLC, New York, NY, 2011.

Prime, version 2.2, Schrödinger, LLC, New York, NY, 2010.

Protein Preparation Wizard; Impact version 5.7, Schrödinger, LLC, New York, NY, 2011.

Purushottamachar, P., Novel Approach: Androgen Receptor Down-Regulating Agents for the Treatment of All Stages of Prostate Cancer Disease, *Biochem. Pharmacol.* (2012), 1: 1-2.

Purushottamachar, P.; Khandelwal, A.; Vasaitis, T.S.; Bruno, R.D.; Gediya, L.K.; Njar, V.C.O., Potent anti-prostate cancer agents derived from a novel androgen receptor down-regulating agent, *Bioorg. Med. Chem.*, (2008), 16: 3519-3529.

QM-Polarized Ligand Docking protocol. Schrödinger, LLC, New York, NY, 2011.

Rittle, J.; Green, M.T., Cytochrome P450 Compound I: Capture, Characterization, and C-H Bond Activation Kinetics, *Green Science*. (2010), 330: 933-937.

Rost, S.; Fregin, A.; Ivaskevicius, V.; Conzelmann, E.; Hortnagel, K.; Pelz, H.J.; Lappegard, K.; Seifried, E.; Scharrer, I.; Tuddenham, E.G.D.; Muller, C.R.; Strom, T.M.; Oldenburg, J., *Nature*, (2004), 427: 537–541.

Rudik, A.V.; Dmitriev, A.V.; Lagunin, A.A.; Filimonov, D.A.; Poroikov, V.V., Metabolism site prediction based on xenobiotic structural formulas and pass prediction algorithm. *J. Chem. Inf. Model.* (2014), 54: 498-507.

Russel, W.M.S., Burch, R.L., The principles of humane experimental technique. London, UK: Methuen, 1959.

Sabela, M.I.; Gumede, N.J.; Escuder-Gilabert, L.; Martín-Biosca, Y.; Bisetty, K.; Medina-Hernández, M.J.; Sagrado, S., Connecting simulated, bioanalytical, and molecular docking data on the stereoselective binding of (\pm)-catechin to human serum albumin, *Anal. Bioanal. Chem.*, (2012), 402: 1899-1909.

Sakkiah, S.; Baek, A.; Lee, K. W., Pharmacophore modeling and molecular dynamics simulation to identify the critical chemical features against human sirtuin 2 inhibitors, *J. Mol. Struct.*, (2012), 1011: 66–75.

Schöneboom, J.C.; Cohen, S.; Lin, H.; Thiel, W., Quantum Mechanical/Molecular Mechanical investigation of the mechanism of C-H hydroxylation of camphor by cytochrome P450_{cam}: Theory supports a two-state rebound mechanism. *J. Am. Chem. Soc.*, (2004), 126: 4017-4034.

Schrödinger, E.; Mitteilung, E.; "Quantisierung als Eigenwert problem," *Ann. d. Phys.*, (1926), 79: 361-376.

Schuster, D.; Kowalik, D.; Kirchmair, J.; Laggner, C.; Markt, P.; Aebischer-Gumy, C.; Ströhle, F.; Möller, G.; Wolber, G.; Wilckens, T.; Langer, T.; Odermatt, A.; Adamski, J., Identification of chemically diverse, novel inhibitors of 17 α -hydroxysteroid dehydrogenase type 3 and 5 by pharmacophore-based virtual screening; *J. Steroid. Biochem. Mol. Biol.*, (2011), 125: 148-161.

Seidman, M., Studies on 3-(*a*-phenyl-*b*-acetylethyl)-4-hydroxycoumarin. PhD Thesis (1950), University of Wisconsin.

Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W., P450 Enzymes: their structure, reactivity, and selectivity-modeled by QM/MM calculations, *Chem. Rev.*, (2010), 110: 949-1017.

Sherman, W., Vice President Schrödinger Applications Science, FEP+ Free Energy Calculations in Drug Discovery, Schrödinger Webinar, Spring 2015.

Sieburg, H.B., Physiological studies in silico. *Studies Scie. Compl.*, (1990), 12: 321-342.

Simonson, T.; Georgios, A.; Karplus, M., Free Energy Simulations Come of Age: Protein-Ligand Recognition. *Acc. Chem. Res.*, (2002), 35: 430-437.

Soares, S.; Marteus, N.; Freitas, V., Interaction of Different Polyphenols with Bovine Serum Albumin (BSA) and Human Salivary α -Amylase (HSA) by Fluorescence Quenching, *J. Agric. Food. Chem.*, (2007), 55: 6726-6735.

Srinivasan J.; Cheatham T.E.; Cieplak P.; Kollman P.A.; Case D.A., Continuum Solvent Studies of the stability of DNA, RNA, and Phosphoramidate-DNA Helices, *J. Am. Chem. Soc.*, (1998), 120: 9401-9409.

Steinbrecher, T.B.; Dahlgren, M.; Cappel, D.; Lin, T.; Wang, L.; Krilov, G.; Abel, R.; Friesner, R.; Sherman, W., Accurate Binding Free Energy Predictions in Fragment Optimization, *J. Chem. Inf. Model.*, 2015, 55: 2411–2420.

Sun, H.; Scott, D.O., Structure-based drug metabolism prediction for drug design, *Chem. Biol. Drug. Design.*, (2010), 75: 3-17.

Sun, H.; Veith, H.; Xia, M.; Austin, C.P.; Huang, R., Predictive models for cytochrome P450 isozymes based quantitative high throughput screening data, *J. Chem. Inf. Model.*, (2011), 51: 2474-2481.

Suttie, J.W.; The Biochemical Basis of Warfarin Therapy, *Adv. Exp. Med. Biol.*, (1987), 214: 3-16.

Tame J.R., Scoring functions: a view from the bench, *J. Comp. Aided. Mol. Des.*, (1999), 13: 99-108.

Tanwar, O.; Marella, A.; Shrivastava, S.; Alam, M. M.; Akhtar, M.; Pharmacophore model generation and 3D-QSAR analysis of N-acyl and N-arylpiperazines for enzymatic and cellular B-Raf kinase inhibition, *Med. Chem. Res.*, (2013), 22: 2174-2187.

Tanwar, O.; Marella, A.; Shrivastava, S.; Alam, M.M.; Akhtar, M., Pharmacophore model generation and 3D-QSAR analysis of N-acyl and N-arylpiperazines for enzymatic and cellular B-Raf kinase inhibition, *Med. Chem. Res.*, (2013), 22: 2174-2187.

Tanwar, O.; Marella, A.; Shrivastava, S.; Alam, M.M.; Akhtar, M., Pharmacophore model generation and 3D-QSAR analysis of N-acyl and N-arylpiperazines for enzymatic and cellular B-raf kinase inhibition, *Med. Chem. Res.*, (2013), 22: 2174-2187.

Tawari, N.R.; Degani, M.S., Pharmacophore mapping and electronic feature analysis for a series of nitroaromatic compounds with antitubercular activity, *J. Comput. Chem.*, (2010), 31: 739-751.

Tawari, N.R.; Degani, M.S., Pharmacophore mapping and electronic feature analysis for a series of nitroaromatic compounds with antitubercular activity, *J. Comput. Chem.*, (2010), 31: 739-751.

Tawari, N.R.; Degani, M.S., Pharmacophore modelling and density functional theory analysis for a series of nitroimidazole compounds with antitubercular activity; *Chem. Biol. Drug Des.*, (2011), 78: 408-417.

Telvekar, V. N.; Kundaikar, H. S.; Patel, K. N.; Chaudhari, H. K.; 3-D QSAR and Molecular Docking Studies on Aryl Benzofuran-2-yl Ketoxime Derivatives as *Candida albicans* N-myristoyl transferase Inhibitors, *QSAR Comb. Sci.*, 2008, 27:1193 – 1203.

Tian, L.; Friesner, R. A., QM/MM Simulation on P450 BM3 Enzyme Catalysis Mechanism, *J. Chem. Theory. Comput.*, (2009), 5: 1421–1431.

Tie, J.K.; Stafford, D.W., Structure and function of vitamin K epoxide reductase. *Vit. Horm.*, (2008), 78:103–130.

van't Hoff, J.H.; “Sur les formules de structure dans l'espace”, *Arch. Néerlandaises des Scie. Exactes et Naturelles*, (1874), 9: 445-454.

van't Hoff, J.H.; “Sur les formules de structure dans l'espace”, *Bull. Soc. Chim.*, (1875), 23: 295-301.

Vasaitis, T.S.; Bruno, R.D.; Njar, V.C.O., CYP17 inhibitors for prostate cancer therapy; *J. Steroid. Biochem. Mol. Biol.*, (2011), 125: 23-31.

Vorobjev Y.N.; Hermans J., ES/IS: Estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model. *Biophys. Chemistry.*, (1999), 78: 195-205.

Vuignier, K.; Schappler, J.; Veuthey, J. L.; Carrupt, P. A.; Martell, S., Drug-protein binding: a critical review of analytical tools, *Anal. Bioanal. Chem.*, (2010), 398, 53–66.

Vyas, V. K.; Ghate, M.; Goel, A.; Pharmacophore modeling, virtual screening, docking and in silico ADMET analysis of protein kinase B (PKB) inhibitors, *J. Mol. Graphics Modell.*, (2013), 42:17–25.

Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R., Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field, *J. Am. Chem. Soc.*, (2015), 137: 2695–2703.

Wang, M.; Baaden, M.; Wang, J.; Liang, Z., A Cooperative Mechanism of Clotrimazoles in P450 Revealed by the Dissociation Picture of Clotrimazole from P450, *J. Chem. Inf. Model.*, (2014), 54: 1218–1225.

Warren, G.L.; Andrews, C.W.; Capelli, A.M.; Clarke, B.; LaLonde, J.; Lambert M.H.; Lindvall, M.; Nevins, N.; Semus, S.F.; Sender, S., A critical Assessment of Docking Programs and Scoring Functions, *J. Med. Chem.*, (2006), 49: 5912-5931.

Warren, G.L.; Andrews, C.W.; Capelli, A.M.; Clarke, B.; LaLonde, J.; Lambert, M.H.; Lindvall, M.; Nevins, N.; Semus, S.F.; Sender, S., A critical Assessment of Docking Programs and Scoring Functions, *J. Med. Chem.*, (2006), 49: 5912-5931.

Wermuth, C.G.; Ganellin, C.R.; Lindberg, P.; Mitscher, L.A., Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998), *Pure. Appl. Chem.*, (1998), 70: 1129–1143.

Wilkinson, G. N.; Statistical Estimations in Enzyme Kinetics, *Biochem. J.*, (1961), 80: 324 – 332.

Xie, X. S., Enzyme Kinetics, Past and Present, *Science.*, (2013), 342: 1457-1459.

Yacobi, A.; Levy, G., Pharmacokinetics of the warfarin enantiomers in rats, *J. Pharmacokinet. Biopharm.*, (1974), 2: 239–255.

Yamaoka, M.; Hara, T.; Hitaka, T.; Kaku, T.; Takeuchi, T.; Takahashi, J.; Asahi, S.; Miki, H.; Tasaka, A.; Kusaka, M., Orteronel (TAK-700), a novel non-steroidal 17,20-lyase

inhibitor: Effects on steroid synthesis in human and monkey adrenal cells and serum steroid levels in cynomolgus monkeys, *J. Steroid. Biochem. Mol. Biol.*, (2012), 129: 115– 128.

Yuriev, E.; Agostino, M.; Ramsland, P.A., Challenges and advances in computational docking: 2009 review, *J. Mol. Recognit.*, (2011), 24: 149-164.

Yuriev, E.; Agostino, M.; Ramsland. P.A., Challenges and advances in computational docking: 2009 review. *J. Mol. Recognit.*, (2011), 24: 149-164.

Zhang, W.; Xia, S.; Ye, J.; Tang, Y.; Li, Z.; Zhu, W.; Cheng, J., Structural features of GABA_A receptor antagonists: pharmacophore modelling and 3D-QSAR studies, *Med. Chem. Res.*, (2013), 22: 5961-5972.

Zhang, W.; Xia, S.; Ye, J.; Tang, Y.; Li, Z.; Zhu, W.; Cheng, J.; Structural features of GABA_A receptor antagonists: pharmacophore modeling and 3D-QSAR studies, *Med. Chem. Res.*, (2013), 22: 5961-5972.

Zhang, Z.; Wong, Y.N., Enzyme kinetics for clinically relevant CYP inhibition. *Curr. Drug. Metabol.*, (2005), 6: 241-257.

Zhong, S.; Macias, T.A.; MacKerell, A. D. Jr., Computational Identification of Protein-Protein Interactions, *Curr. Top. Med. Chem.*, (2007), 7: 63-82.

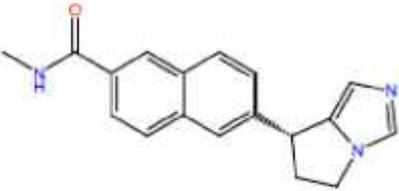
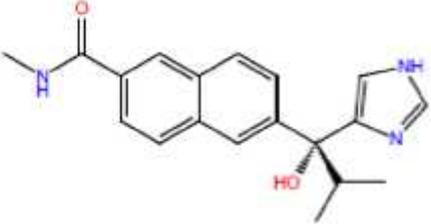
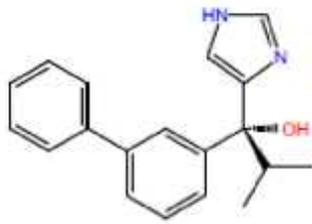
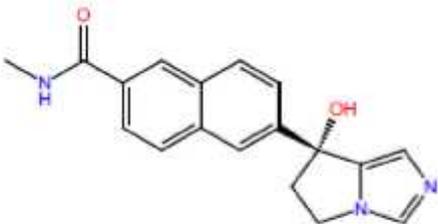
Zhou, Z.; Felts, A.K.; Friesner, R.A.; Levy, R.M., Comparative performance of several flexible docking programs and scoring functions: enrichment studies for a diverse set of pharmaceutically relevant targets, *J. Chem. Inf. Model.*, (2007), 47: 1599-1608.

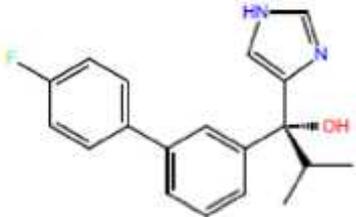
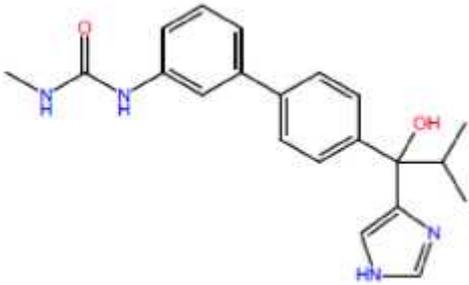
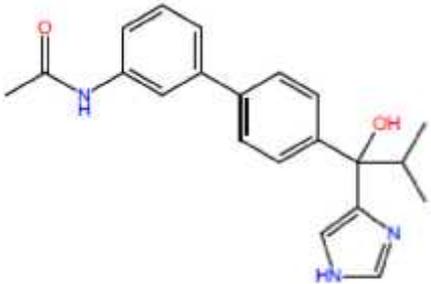
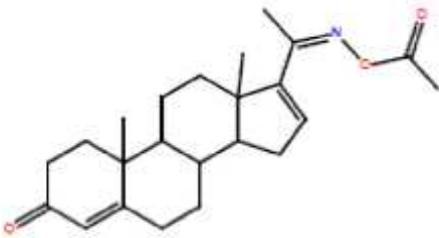
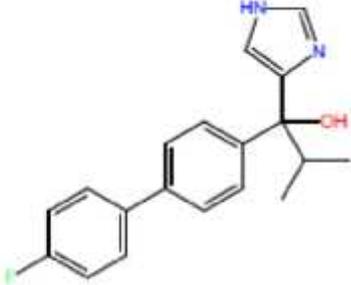
Zhuang, Y.; Wachall, B.G.; Hartmann, R.W., Novel imidazolyl and triazolyl substituted biphenyl compounds: synthesis and evaluation as nonsteroidal inhibitors of human 17 β -hydroxylase-C17, 20-lyase (P450C17), *Bioorg. Med. Chem.*, (2000), 8: 1245-1252.

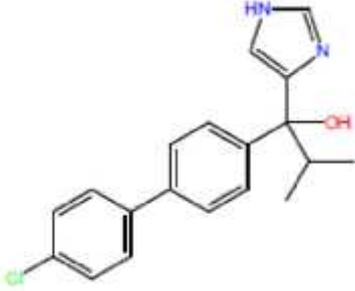
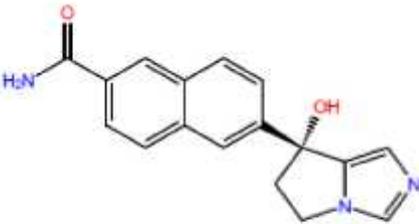
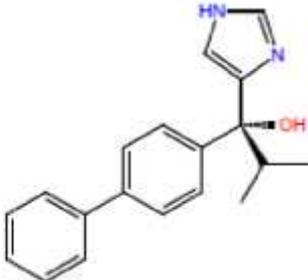
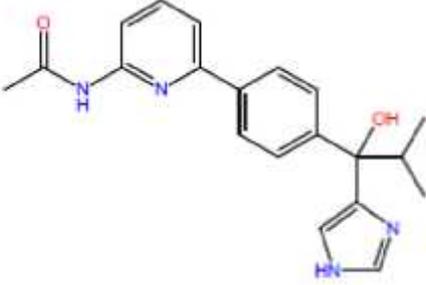
Zou, H.; Wang, H.; Zhang, Y., Stereoselective binding of warfarin and ketoprofen to Human Serum Albumin determined by Microdialysis combined with HPLC, *J. Liq. Chromatogr. Relat. Technol.*, (1998), 21: 2663–2674.

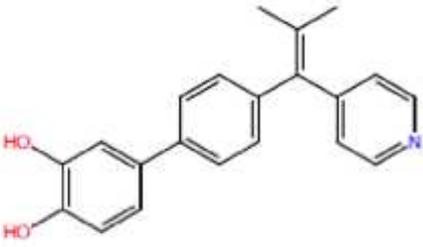
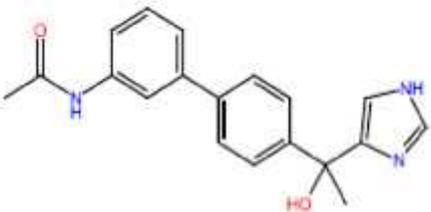
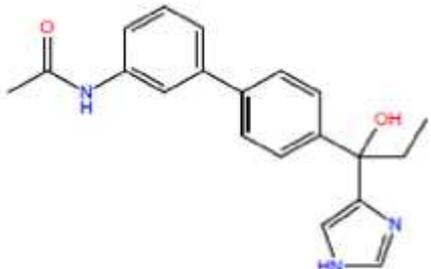
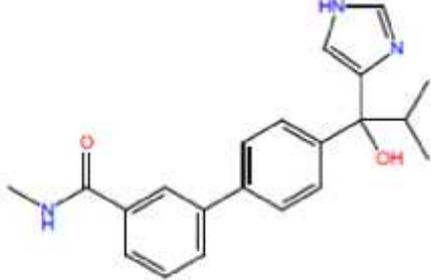
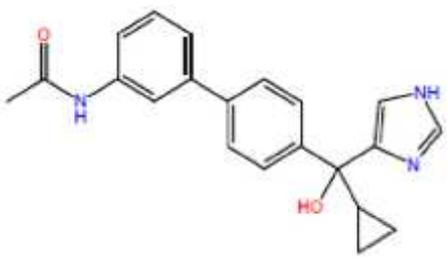
APPENDIXES

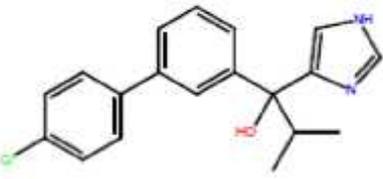
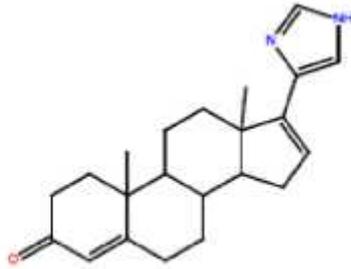
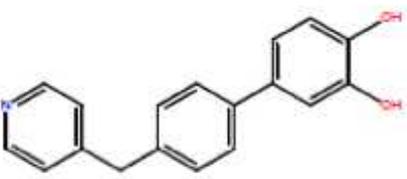
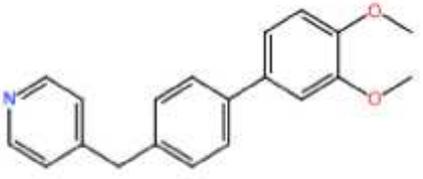
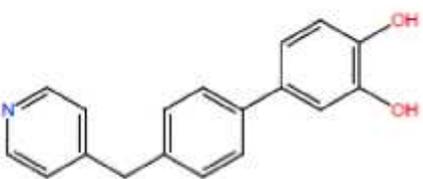
Table A1. Structures and systematic names of CYP 17A1 inhibitors with their codes as obtained from literature.

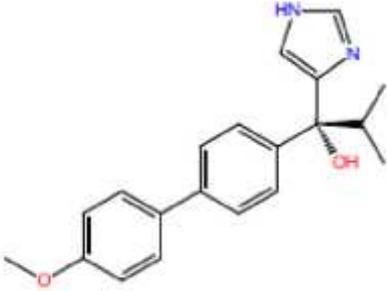
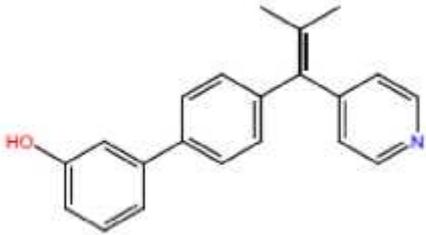
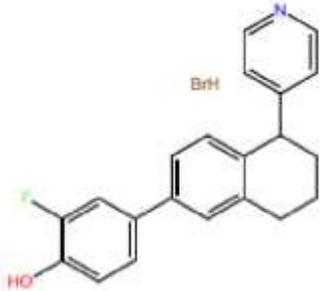
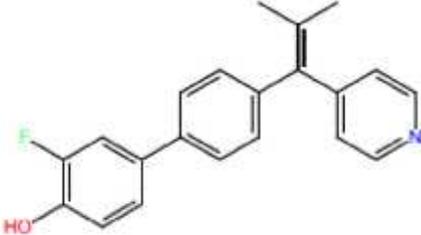
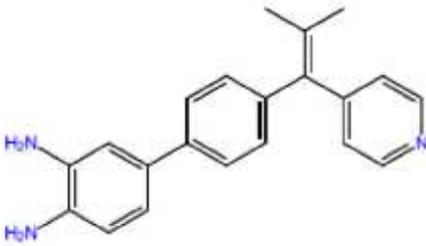
ID	Code [Ref]	2D Structure and Systematic name
1	3d (Kaku, 2011a)	 <p data-bbox="544 728 1436 763">6-(6,7-Dihydro-5H-pyrrolo[1,2-c]imidazol-7-yl)-N-methyl-2-naphthamide</p>
2	1 (Handratta, 2005)	 <p data-bbox="526 1075 1452 1144">6-[1-Hydroxy-1-isopropyl(1H-imidazol-5-yl)methyl]-N-methylnaphthalene-2-carboxamide</p>
3	5 (Budha, 2008)	 <p data-bbox="598 1467 1380 1496">1-[1,1'-biphenyl]-3-yl-1-(1H-imidazol-4-yl)-2-methyl-1-propanol</p>
4	(+)-3c (Handratta, 2005)	 <p data-bbox="534 1825 1444 1892">(+)-6-(7-Hydroxy-6,7-dihydro-5H-pyrrolo[1,2-c]imidazol-7-yl)-N-methyl-2-naphthamide</p>

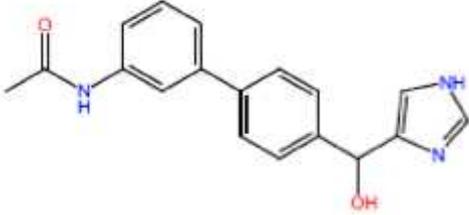
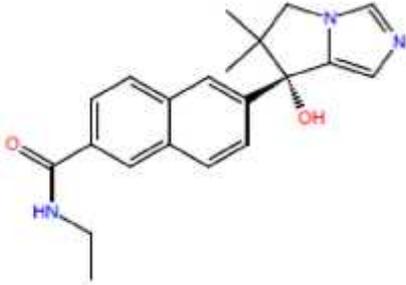
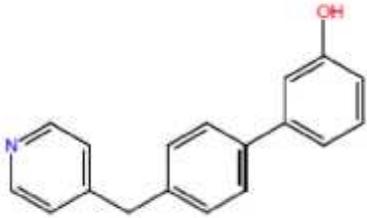
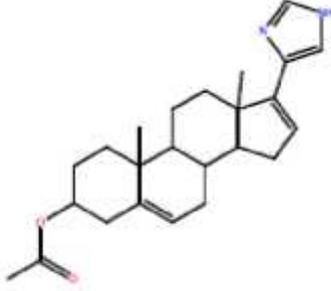
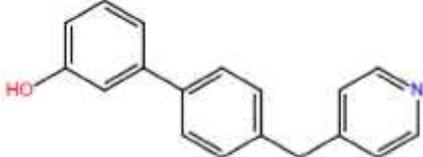
5	13 (Budha, 2008)	 <p>1-(4'-Fluoro[1,1'-biphenyl]-3-yl)-1-(1H-imidazol-4-yl)-2-methyl-1-propanol</p>
6	24 (Budha, 2008)	 <p>N-{4'-[1-Hydroxy-1-(1H-imidazol-4-yl)-2-methylpropyl][1,1'-biphenyl]-3-yl}-N'-methylurea</p>
7	17 (Budha, 2008)	 <p>N-{4'-[1-Hydroxy-1-(1H-imidazol-4-yl)-2-methylpropyl][1,1'-biphenyl]-3-yl}acetamide</p>
8	L26 (Nnane, 1999)	 <p>4,16-pregnadiene-3,20-dione-20-oxime acetate</p>
9	15 (Budha, 2008)	 <p>1-(4'-Fluoro[1,1'-biphenyl]-4-yl)-1-(1H-imidazol-4-yl)-2-methyl-1-propanol</p>

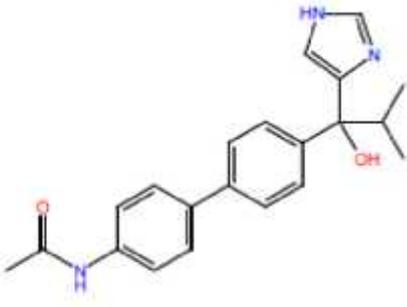
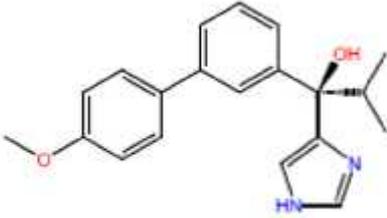
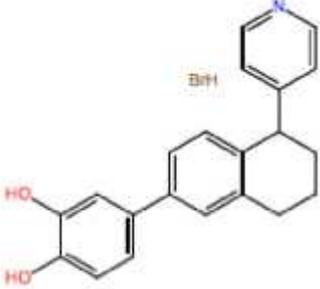
10	16 (Budha, 2008)	 <p>1-(4'-Chloro[1,1'-biphenyl]-4-yl)-1-(1H-imidazol-4-yl)-2-methyl-1-propanol</p>
11	TAK700a (Jagusch, 2008)	 <p>[(1S)-1-(6,7-dimethoxy-2-naphthyl)-1-(1H-imidazol-4-yl)-2-methylpropan-1-ol]</p>
12	3b (Kaku, 2011a)	 <p>6-(7-Hydroxy-6,7-dihydro-5H-pyrrolo[1,2-c]imidazol-7-yl)-2-naphthamide</p>
13	7 (Budha, 2008)	 <p>1-[1,1'-Biphenyl]-4-yl-1-(1H-imidazol-4-yl)-2-methyl-1-propanol</p>
14	26 (Budha, 2008)	 <p>N-{6-(4-[1-Hydroxy-1-(1H-imidazol-4-yl)-2-methylpropyl]phenyl)-2-pyridyl}acetamide</p>

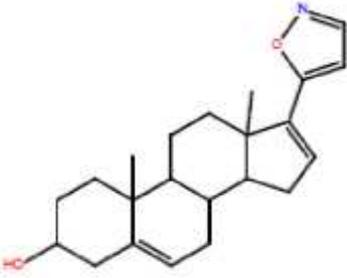
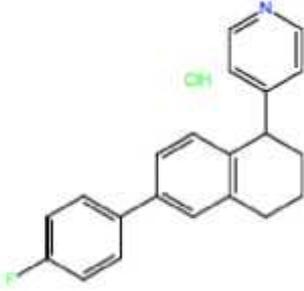
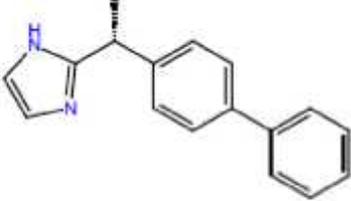
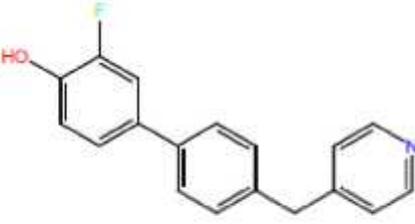
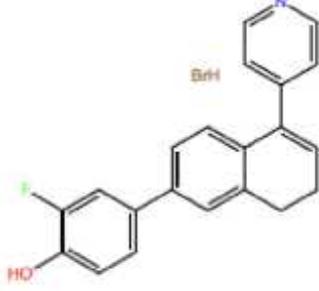
15	16 (Budha, 2008)	 <p data-bbox="486 492 1428 526">[(3'-Hydroxy-4'-Hydroxybiphenyl-4-yl)(pyridin-4-yl)1-isopropylidenemethane]</p>
16	32 (Budha, 2008)	 <p data-bbox="502 828 1412 862">N-[4'-[1-Hydroxy-1-(1H-imidazol-4-yl)ethyl][1,1'-biphenyl]-3-yl]acetamide</p>
17	33 (Budha, 2008)	 <p data-bbox="486 1198 1428 1232">N-[4'-[1-Hydroxy-1-(1H-imidazol-4-yl)propyl][1,1'-biphenyl]-3-yl]acetamide</p>
18	22 (Budha, 2008)	 <p data-bbox="478 1545 1436 1601">4'-[1-Hydroxy-1-(1H-imidazol-4-yl)-2-methylpropyl]-N-methyl[1,1'-biphenyl]-3-carboxamide</p>
19	34 (Budha, 2008)	 <p data-bbox="518 1915 1396 1971">N-[4'-[Cyclopropyl(hydroxy)-1H-imidazol-4-ylmethyl][1,1'-biphenyl]-3-yl]acetamide</p>

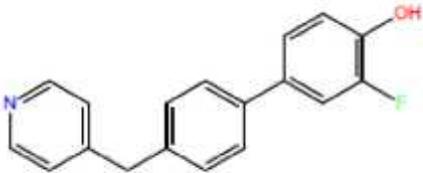
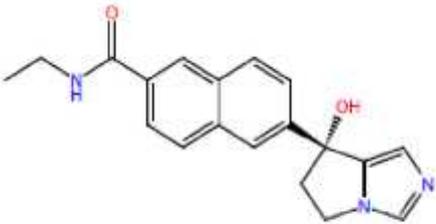
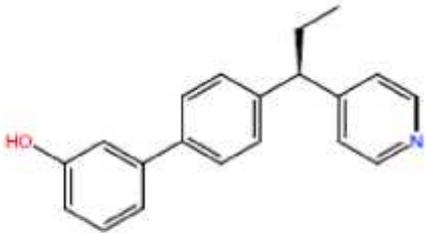
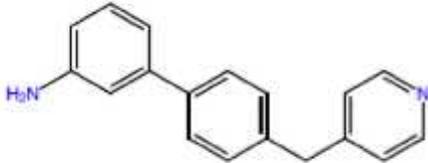
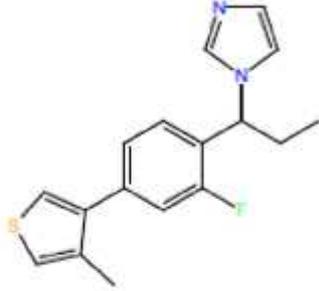
20	14 (Budha, 2008)	 <p>1-(4'-Chloro[1,1'-biphenyl]-3-yl)-1-(1H-imidazol-4-yl)-2-methyl-1-propanol</p>
21	L6 (Nnane, 1999)	 <p>17-(4'-imidazolyl)androsta-4,16-dien-3-one</p>
22	9 (Hu, 2010a)	 <p>4'-(Pyridin-4-ylmethyl)biphenyl-3,4-diol</p>
23	9a (Hu, 2010a)	 <p>4-[(3',4'-Dimethoxybiphenyl-4-yl)methyl]pyridine</p>
24	26 (Jagusch, 2008)	 <p>[(3'-Hydroxy-4'-Hydroxybiphenyl-4-yl)(pyridin-4-yl)methane]</p>

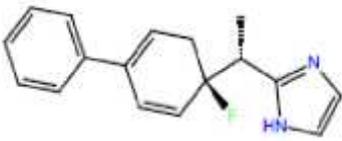
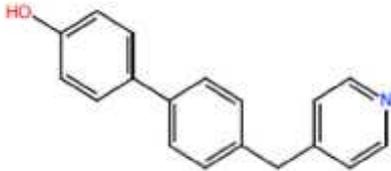
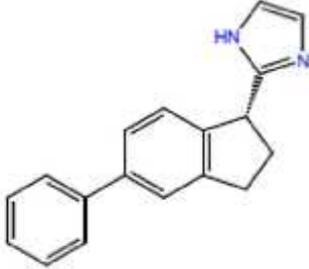
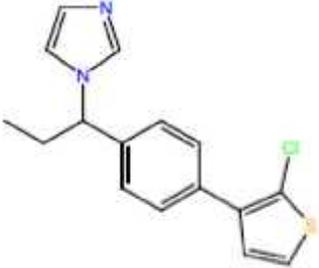
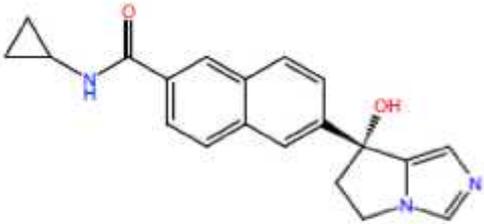
25	8 (Budha, 2008)	 <p data-bbox="491 495 1428 528">1-(1H-Imidazol-4-yl)-1-(4'-methoxy[1,1'-biphenyl]-4-yl)-2-methyl-1-propanol</p>
26	13 (Jagusch, 2008)	 <p data-bbox="560 853 1358 884">[(3'-Hydroxybiphenyl-4-yl)(pyridin-4-yl)1-isopropylidenemethane]</p>
27	20 (Kaku, 2011a)	 <p data-bbox="451 1193 1469 1227">2-Fluoro-4-(5-(pyridin-4-yl)-5,6,7,8-tetrahydronaphthalen-2-yl)phenol Hydrobromide</p>
28	15 (Jagusch, 2008)	 <p data-bbox="499 1503 1417 1532">[(3'-Fluoro-4'-Hydroxybiphenyl-4-yl)(pyridin-4-yl)1-isopropylidenemethane]</p>
29	22 (Jagusch, 2008)	 <p data-bbox="515 1861 1417 1890">[(3'-Amino-4'-Aminobiphenyl-4-yl)(pyridin-4-yl)1-isopropylidenemethane]</p>

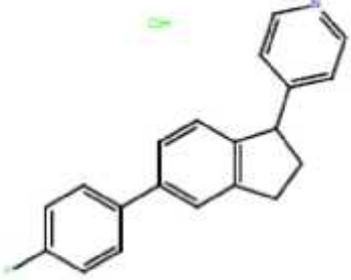
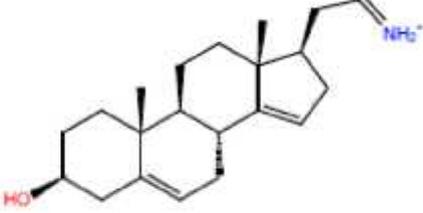
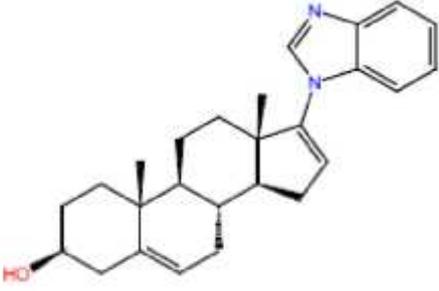
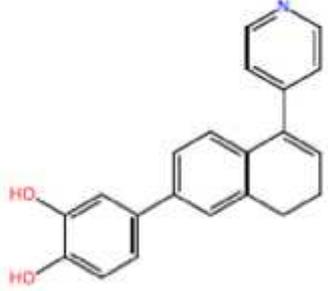
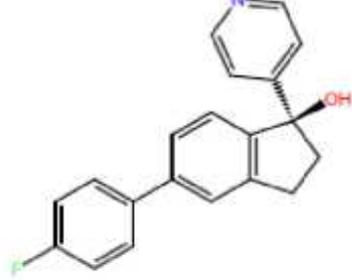
30	L12 (Nnane, 1999)	 <p>N-[4'-[1-Hydroxy(1H-imidazol-4-yl)methyl][1,1'-biphenyl]-3-yl]acetamide</p>
31	36 (Budha, 2008)	 <p>6-(7-Hydroxy-6,7-dihydro-6,6-dimethyl-5H-pyrrolo[1,2-c]imidazol-7-yl)-N-ethyl-2-naphthamide</p>
32	3i (Kaku, 2011a)	 <p>4-[(3'-Hydroxybiphenyl-4-yl)methyl]pyridine</p>
33	8 (Hu, 2010a)	 <p>3-acetoxy-17-(4'-imidazolyl)androsta-5,16-diene</p>
34	24 (Jagusch, 2008)	 <p>[(3'-Hydroxybiphenyl-4-yl)(pyridin-4-yl)methane]</p>

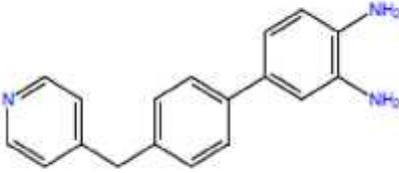
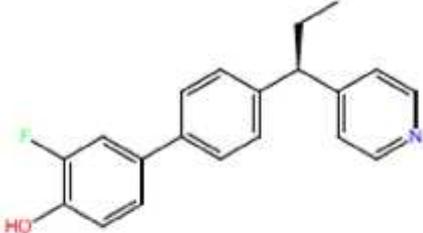
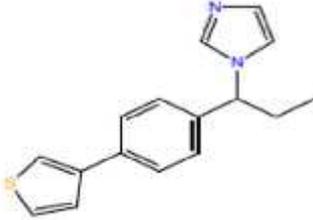
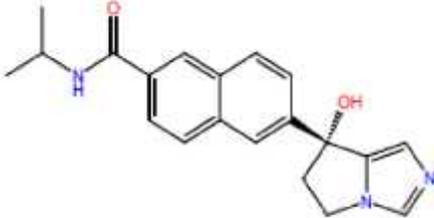
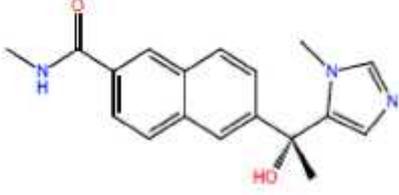
35	L38 (Nnane, 1999)	 <p data-bbox="507 524 1406 591">N-{4'-[1-Hydroxy-1-(1H-imidazol-4-yl)-2-methylpropyl][1,1'-biphenyl]-4-yl}acetamide</p>
36	18 (Budha, 2008)	 <p data-bbox="485 913 1433 949">1-(1H-Imidazol-4-yl)-1-(4'-methoxy-[1,1'-biphenyl]-3-yl)-2-methyl-1-propanol</p>
37	6 (Budha, 2008)	 <p data-bbox="443 1279 1469 1317">4-(5-(Pyridin-4-yl)-5,6,7,8-tetrahydronaphthalen-2-yl)benzene-1,2-diol Hydrobromide</p>
38	19 (Kaku, 2011a)	 <p data-bbox="475 1646 1442 1713">4'-[1-Hydroxy-1-(1H-imidazol-4-yl)-2-methylpropyl]-Nmethyl[1,1'-biphenyl]-3-sulfonamide</p>

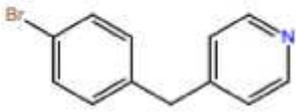
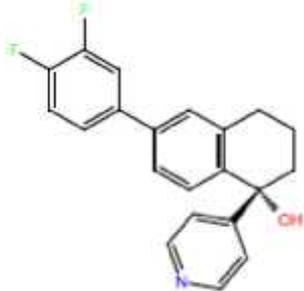
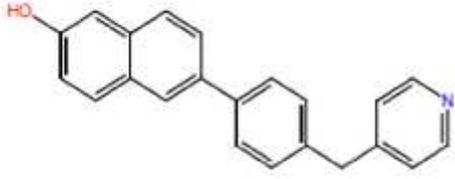
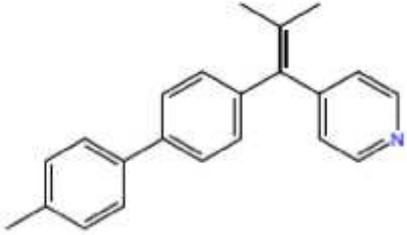
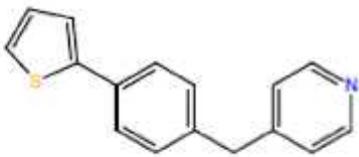
39	25 (Budha, 2008)	 <p>17-(5'-isoxazolyl)androsta-5,16-dien-3-ol</p>
40	23 (Budha, 2008)	 <p>4-(6-(4-Fluorophenyl)-1,2,3,4-tetrahydronaphthalen-1-yl)pyridineHydrochloride</p>
41	17 (Kaku, 2011a)	 <p>4-[1-(1Himidazolyl)ethyl]biphenyl</p>
42	5ax (Hu, 2010a)	 <p>3-Fluoro-4'-(pyridin-4-ylmethyl)biphenyl-4-ol</p>
43	10 (Hu, 2010a)	 <p>2-Fluoro-4-(5-(pyridin-4-yl)-7,8-dihydronaphthalen-2-yl)phenolHydrobromide</p>

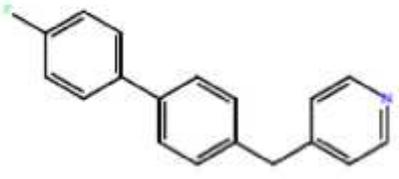
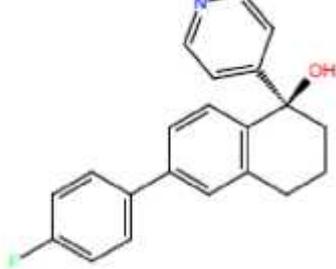
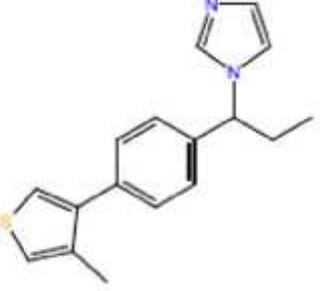
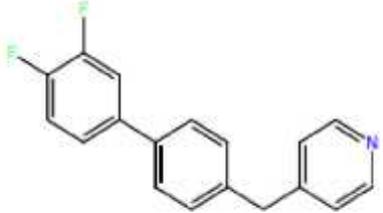
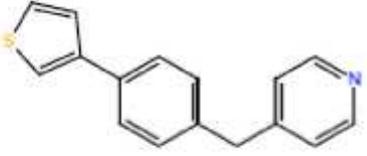
44	14 (Kaku, 2011a)	 <p>[(3'-Fluoro-4'-Hydroxybiphenyl-4-yl)(pyridin-4-yl)methane]</p>
45	25 (Jagusch, 2008)	 <p>N-Ethyl-6-(7-hydroxy-6,7-dihydro-5H-pyrrolo[1,2-c]-imidazol-7-yl)-2-naphthamide</p>
46	3e (Kaku, 2011a)	 <p>[(3'-Hydroxybiphenyl-4-yl)(pyridin-4-yl)1-ethylmethane]</p>
47	11 (Jagusch, 2008)	 <p>[(3'-Aminobiphenyl-4-yl)(pyridin-4-yl)methane]</p>
48	27 (Jagusch, 2008)	 <p>1-(1-(2-Fluoro-4-(4-methylthiophen-3-yl)-phenyl)propyl)-1 H-imidazole</p>

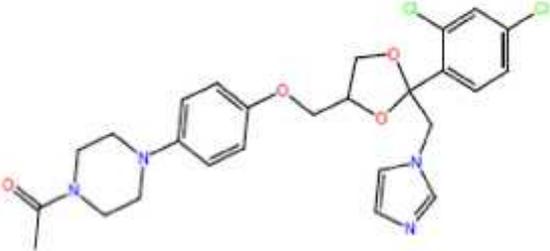
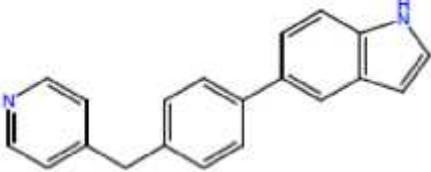
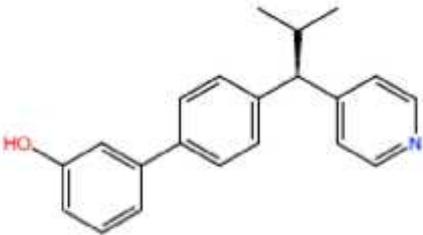
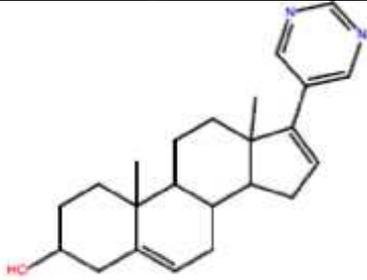
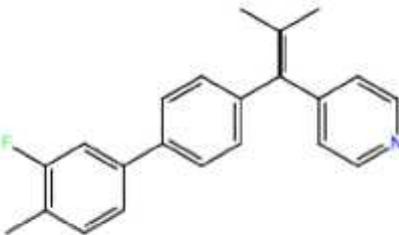
49	16 (Kaku, 2011a)	 <p>4-fluoro-4-[1-(1H-imidazolyl)ethyl]biphenyl</p>
50	30 (Budha, 2008)	 <p>4-[(4'-Hydroxybiphenyl-4-yl)methyl]pyridine</p>
51	5ay (Hu, 2010a)	 <p>1-(1H-imidazolyl)-5-phenylindane</p>
52	4 (Hu, 2010a)	 <p>1-(1-(4-(2-Chlorothiophen-3-yl)-phenyl)propyl)-1H-imidazole</p>
53	5bx (Hu, 2010a)	 <p>N-Cyclopropyl-6-(7-hydroxy-6,7-dihydro-5H-pyrrolo[1,2-c]imidazol-7-yl)-2-naphthamide</p>

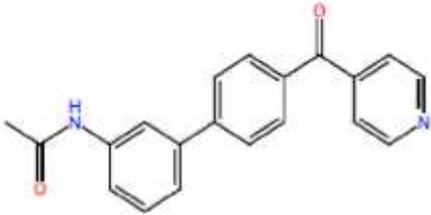
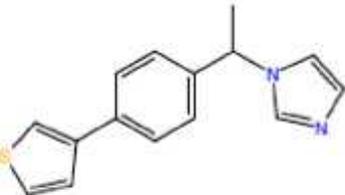
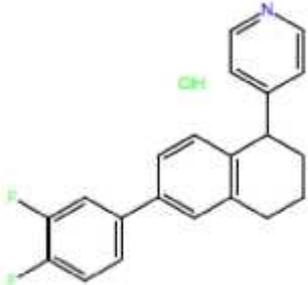
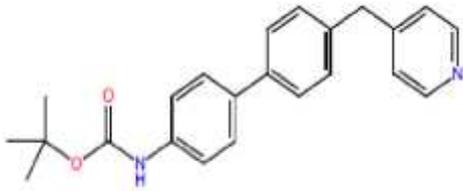
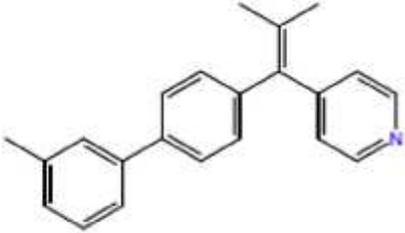
54	31 (Budha, 2008)	 <p>4-(5-(4-Fluorophenyl)-2,3-dihydro-1H-inden-1-yl)pyridine Hydrochloride</p>
55	3f (Kaku, 2011a)	 <p>(20S)-21-iminopregn-5,14-dien-3-ol</p>
56	5 (Handratta, 2005)	 <p>3-hydroxy-17-(1H-benzimidazole-1-yl)androsta-5,16-diene</p>
57	5 [Kaku, 2011b)	 <p>4-(5-(Pyridin-4-yl)-7,8-dihydronaphthalen-2-yl)benzene-1,2-diol Hydrobromide</p>
58	13 (Kaku, 2011a)	 <p>5-(4-Fluorophenyl)-1-(pyridin-4-yl)-2,3-dihydro-1H-inden-1-ol</p>

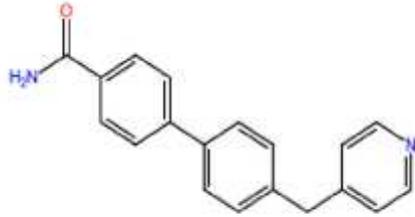
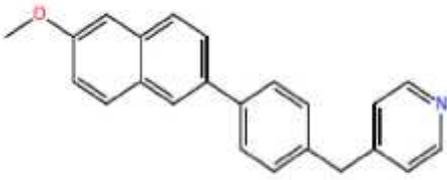
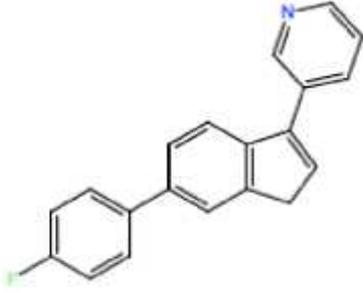
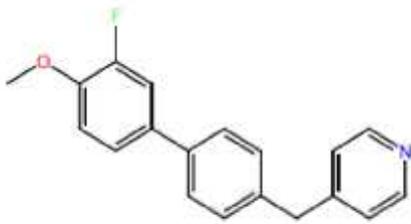
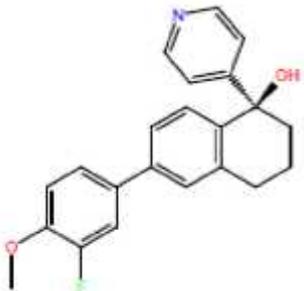
59	2 (Kaku, 2011a)	 <p data-bbox="608 488 1305 517">[(3'-Amino-4'-Aminobiphenyl-4-yl)(pyridin-4-yl)methane]</p>
60	28 (Jagusch, 2008)	 <p data-bbox="557 808 1358 837">[(3'-Fluoro-4'-Hydroxybiphenyl-4-yl)(pyridin-4-yl)1-ethylmethane]</p>
61	14 (Jagusch, 2008)	 <p data-bbox="647 1088 1270 1122">1-(1-(4-Thiophen-3-yl-phenyl)propyl)-1H-imidazole</p>
62	27 (Budha, 2008)	 <p data-bbox="512 1451 1401 1518">6-(7-Hydroxy-6,7-dihydro-5H-pyrrolo[1,2-c]imidazol-7-yl)-N-isopropyl-2-naphthamide</p>
63	14 (Hu, 2010a)	 <p data-bbox="501 1809 1412 1872">6-[1-Hydroxy-1-(1-methyl-1H-imidazol-5-yl)ethyl]-N-methylnaphthalene-2-carboxamide</p>

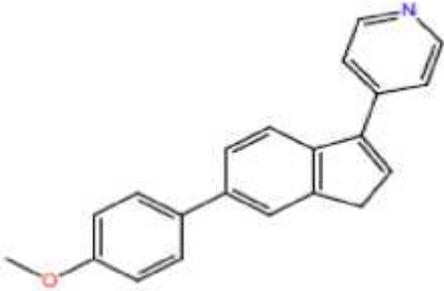
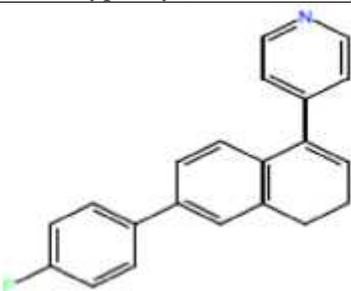
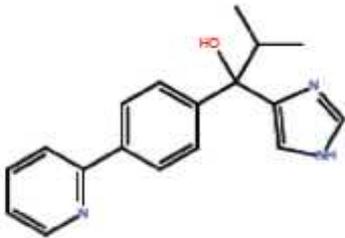
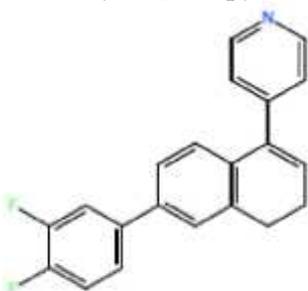
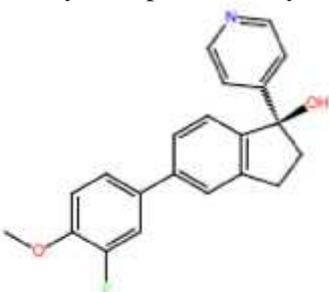
64	3g (Kaku, 2011a)	 <p data-bbox="790 470 1125 504">4-(4-Bromobenzyl)pyridine</p>
65	3j ((Kaku, 2011a)	 <p data-bbox="518 851 1396 887">6-(3,4-Difluorophenyl)-1-(pyridin-4-yl)-1,2,3,4-tetrahydronaphthalen-1-ol</p>
66	5a (Hu, 2010a)	 <p data-bbox="662 1232 1252 1270">6-[4-(Pyridin-4-ylmethyl)phenyl]naphthalen-2-ol</p>
67	6 (Kaku, 2011a)	 <p data-bbox="566 1556 1348 1590">[(4'-Methylbiphenyl-4-yl)(pyridin-4-yl)methylidene]methane</p>
68	25 (Hu, 2010a)	 <p data-bbox="734 1881 1181 1910">4-[4-(Thiophen-2-yl)benzyl]pyridine</p>

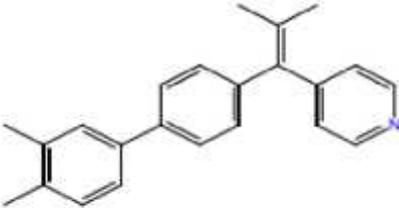
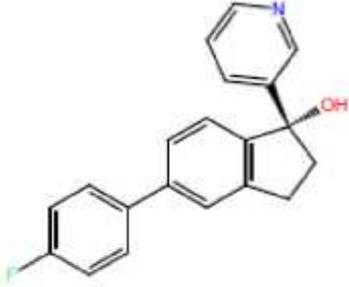
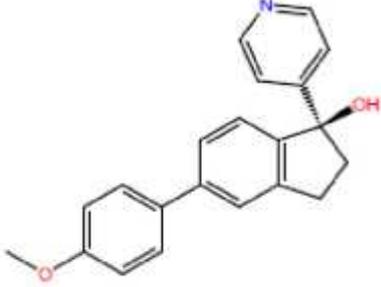
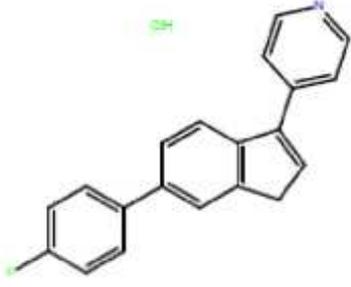
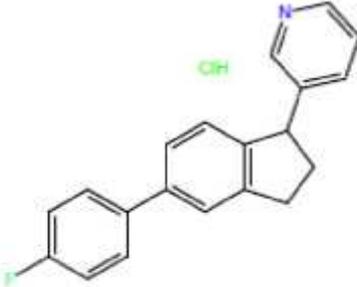
69	18 (Jagusch, 2008)	 <p>4-[(4'-Fluorobiphenyl-4-yl)methyl]pyridine</p>
70	15 [37]	 <p>6-(4-Fluorophenyl)-1-(pyridin-4-yl)-1,2,3,4-tetrahydronaphthalen-1-ol</p>
71	21 (Hu, 2010a)	 <p>1-(1-(4-(4-Methylthiophen-3-yl)phenyl)propyl)-1H-imidazole</p>
72	5 (Kaku, 2011a)	 <p>4-[(3',4'-Difluorobiphenyl-4-yl)methyl]pyridine</p>
73	29 (Budha, 2008)	 <p>4-[4-(Thiophen-3-yl)benzyl]pyridine</p>

74	15 (Hu, 2010a)	 <p>1-[4-(4-[(2R,4S)-2-(2,4-Dichlorophenyl)-2-(1H-imidazol-1-ylmethyl)-1,3-dioxolan-4-yl]methoxy}phenyl)piperazin-1-yl]ethan-1-one</p>
75	22 (Hu, 2010a)	 <p>5-[4-(Pyridin-4-ylmethyl)phenyl]-1H-indole</p>
76	KTZ (Jagusch, 2008)	 <p>[(3'-Hydroxybiphenyl-4-yl)(pyridin-4-yl)1-isopropylmethane]</p>
77	23 (Hu, 2010a)	 <p>3-hydroxy-17-(5(1H-pyrimidin-2-yl)androsta-5,16-diene</p>
78	12 (Jagusch, 2008)	 <p>[(3'-Fluoro-4'-Methylbiphenyl-4-yl)(pyridin-4-yl)1-isopropylidene methane]</p>

79	19 (Jagusch, 2008)	 <p data-bbox="699 488 1216 521">N-(4'-Isonicotinoylbiphenyl-3-yl)acetamide</p>
80	12 (Hu, 2010a)	 <p data-bbox="660 824 1257 853">1-(1-(4-Thiophen-3-ylphenyl)ethyl)-1H-imidazole</p>
81	28 (Budha, 2008)	 <p data-bbox="453 1200 1461 1234">4-(6-(3,4-Difluorophenyl)-1,2,3,4-tetrahydronaphthalen-1-yl)pyridine Hydrochloride</p>
82	18 (Kaku, 2011a)	 <p data-bbox="616 1503 1299 1536">tert-Butyl 4'-(Pyridin-4-ylmethyl)biphenyl-4-ylcarbamate</p>
83	17 (Hu, 2010a)	 <p data-bbox="568 1839 1347 1865">[(3'-Methylbiphenyl-4-yl)(pyridin-4-yl)1-isopropylidene]methane</p>

84	17 (Jagusch, 2008)	 <p>4'-(Pyridin-4-ylmethyl)biphenyl-4-carboxamide</p>
85	13 (Hu, 2010a)	 <p>4-[4-(6-Methoxynaphthalen-2-yl)benzyl]pyridine</p>
86	24 (Hu, 2010a)	 <p>3-(5-(4-Fluorophenyl)-3H-inden-1-yl)pyridine Hydrochloride</p>
87	8 (Kaku, 2011a)	 <p>4-[(3'-Fluoro-4'-methoxybiphenyl-4-yl)methyl]pyridine</p>
88	16 (Hu, 2010a)	 <p>6-(3-Fluoro-4-methoxyphenyl)-1-(pyridin-4-yl)-1,2,3,4-tetrahydronaphthalen-1-ol</p>

89	7 (Kaku, 2011a)	 <p>4-(5-(4-Methoxyphenyl)-3H-inden-1-yl)pyridine</p>
90	10 (Kaku, 2011a)	 <p>4-(6-(4-Fluorophenyl)-3,4-dihydronaphthalen-1-yl)pyridine</p>
91	11 (Kaku, 2011a)	 <p>1-(1H-Imidazol-4-yl)-2-methyl-1-[4-(2-pyridinyl)phenyl]-1-propanol</p>
92	12 (Kaku, 2011a)	 <p>4-(6-(3,4-Difluorophenyl)-3,4-dihydronaphthalendihydronaphthalen-1-yl)pyridine</p>
93	4 (Kaku, 2011a)	 <p>5-(3-Fluoro-4-methoxyphenyl)-1-(pyridin-4-yl)-2,3-dihydro-1H-inden-1-ol</p>

94	20 (Jagusch, 2008)	 <p data-bbox="507 450 1409 483">[(3'-Methyl-4'-Methylbiphenyl-4-yl)(pyridin-4-yl)1-isopropylidene methane]</p>
95	1 (Kaku, 2011a)	 <p data-bbox="580 831 1335 864">5-(4-Fluorophenyl)-1-(pyridin-3-yl)-2,3-dihydro-1H-inden-1-ol</p>
96	3 (Kaku, 2011a)	 <p data-bbox="564 1200 1351 1234">5-(4-Methoxyphenyl)-1-(pyridin-4-yl)-2,3-dihydro-1H-inden-1-ol</p>
97	9 (Kaku, 2011a)	 <p data-bbox="592 1570 1321 1603">4-(5-(4-Fluorophenyl)-3H-inden-1-yl)pyridine Hydrochloride</p>
98	15 (Kaku, 2011a)	 <p data-bbox="517 1939 1398 1973">3-(5-(4-Fluorophenyl)-2,3-dihydro-1H-inden-1-yl)pyridine Hydrochloride</p>

^aKTZ – Ketoconazole

Table A2. Illustrates the training set and test set structural data used for the development of a 3D-QSAR pharmacophore model. The predicted activity results and scoring data confirming the predictive ability of the pharmacophore model developed by PLS regression.

Entry ID	Name, Ref.	pIC ₅₀	qsar set	Type	# Conformers	# Sites Matched	Matched Ligand Sites	Align Score	Vector Score	Volume Score	Fitness Score	Predicted Activity1	Predicted Activity2	Predicted Activity3	Outliers
1	3d (Kaku, 2011a)	7,89	training	Nonsteroidal	9	6	A(1) A(2) D(3) H(4) R(6) R(7)	0,27	0,99	0,90	2,67	7,52	7,32	7,29	
2	1 (Handratta,2005)	7,80	training	Nonsteroidal	10	6	A(1) A(3) D(6) H(8) R(11) R(10)	0,60	0,80	0,77	2,07	7,79	7,71	7,70	
3	5 (Budha, 2008)	7,74	training	Nonsteroidal	10	4	A(1) A(-) D(-) H(6) R(9) R(8)	1,03	0,67	0,57	1,37	7,05	7,56	7,57	
4	(+)-3c (Handratta, 2005)	7,72	training	Nonsteroidal	7	6	A(1) A(3) D(5) H(6) R(8) R(9)	0,00	1,00	1,00	3,00	7,57	7,21	7,07	
5	13 (Budha, 2008)	7,72	training	Nonsteroidal	2	4	A(-) A(1) D(3) H(5) R(9) R(-)	0,97	0,84	0,42	1,46	6,88	7,36	7,83	
6	24 (Budha, 2008)	7,68	test	Nonsteroidal	4	6	A(2) A(3) D(5) H(8) R(12) R(11)	0,97	0,60	0,41	1,20	7,79	7,34	7,58	
7	17 (Budha, 2008)	7,62	training	Nonsteroidal	28	6	A(2) A(3) D(5) H(7) R(11) R(10)	0,98	0,60	0,42	1,20	7,83	7,40	7,63	
8	L26 (Nnane, 1999)	7,60	training	Nonsteroidal	2	3	A(1) A(3) D(-) H(10) R(-) R(-)	1,04	0,80	0,44	1,38	7,12	7,66	7,72	
9	15 (Budha, 2008)	7,57	training	Nonsteroidal	6	4	A(-) A(1) D(3) H(5) R(9) R(-)	0,82	0,81	0,51	1,64	7,00	7,43	7,93	
10	16 (Budha, 2008)	7,55	training	Nonsteroidal	4	4	A(-) A(1) D(3) H(5) R(9) R(-)	0,82	0,81	0,51	1,64	6,97	7,34	7,76	
11	TAK700 ^a (Jagusch, 2008)	7,55	training	Nonsteroidal	1	5	A(1) A(3) D(-) H(9) R(13) R(12)	0,79	0,65	0,62	1,61	7,48	7,83	7,78	
12	3b (Kaku, 2011a)	7,54	training	Nonsteroidal	4	6	A(1) A(3) D(6) H(7) R(9) R(10)	0,02	1,00	0,98	2,96	7,66	7,51	7,32	
13	7 (Budha, 2008)	7,48	training	Nonsteroidal	7	4	A(1) A(-) D(-) H(5) R(9) R(8)	0,96	0,58	0,58	1,36	6,95	7,36	7,35	
14	26 (Budha, 2008)	7,44	test	Nonsteroidal	1	6	A(3) A(4) D(6) H(9) R(12) R(11)	1,08	0,52	0,40	1,02	7,44	7,14	7,36	
15	16 (Budha, 2008)	7,43	training	Nonsteroidal	1	5	A(1) A(2) D(5) H(6) R(-) R(9)	0,89	0,85	0,69	1,80	6,92	7,19	7,26	
16	32 (Budha, 2008)	7,42	training	Nonsteroidal	1	5	A(2) A(3) D(5) H(-) R(9) R(8)	1,02	0,67	0,39	1,21	7,79	7,28	7,52	
17	33 (Budha, 2008)	7,40	test	Nonsteroidal	1	6	A(2) A(3) D(5) H(7) R(10) R(9)	0,99	0,59	0,42	1,18	7,76	7,35	7,56	
18	22 (Budha, 2008)	7,36	training	Nonsteroidal	1	6	A(2) A(3) D(5) H(7) R(11) R(10)	0,82	0,82	0,48	1,62	7,91	7,53	7,68	
19	34 (Budha, 2008)	7,35	training	Nonsteroidal	1	6	A(2) A(3) D(5) H(7) R(9) R(10)	0,93	0,65	0,47	1,34	7,66	7,20	7,36	
20	14 (Budha, 2008)	7,31	training	Nonsteroidal	1	5	A(-) A(1) D(3) H(5) R(9) R(10)	1,18	0,74	0,42	1,18	6,62	6,91	7,56	
21	L6 (Nnane, 1999)	7,30	training	Nonsteroidal	4	4	A(2) A(1) D(3) H(9) R(-) R(-)	1,14	0,56	0,47	1,08	7,27	7,22	7,33	
22	9 (Hu, 2010a)	7,28	training	Nonsteroidal	1	5	A(1) A(2) D(5) H(-) R(6) R(7)	0,95	0,83	0,58	1,61	7,04	7,28	7,27	
23	9a (Hu, 2010a)	7,28	training	Nonsteroidal	3	4	A(1) A(3) D(-) H(-) R(8) R(7)	1,05	0,83	0,63	1,58	6,94	7,15	7,25	

24	26 (Jagusch, 2008)	7,28	test	Nonsteroidal	1	5	A(1) A(2) D(5) H(-) R(6) R(7)	0,95	0,83	0,58	1,61	7,04	7,28	7,27	
25	8 (Budha, 2008)	7,27	training	Nonsteroidal	1	5	A(1) A(3) D(-) H(8) R(10) R(11)	0,98	0,66	0,64	1,47	6,90	7,05	7,11	
26	13 (Jagusch, 2008)	7,25	training	Nonsteroidal	1	4	A(1) A(-) D(3) H(4) R(-) R(7)	0,91	0,98	0,59	1,81	6,91	7,10	7,15	
27	20 (Kaku, 2011a)	7,19	test	Steroidal	5										yes
28	15 (Jagusch, 2008)	7,12	training	Nonsteroidal	1	4	A(1) A(-) D(3) H(5) R(-) R(8)	1,01	0,69	0,56	1,40	6,29	6,83	7,02	
29	22 (Jagusch, 2008)	7,12	training	Nonsteroidal	1	4	A(1) A(-) D(5) H(6) R(-) R(9)	0,92	0,96	0,57	1,76	7,12	7,33	7,35	
30	L12 (Nnane, 1999)	7,12	training	Nonsteroidal	1	3	A(1) A(3) D(-) H(5) R(-) R(-)	0,94	0,57	0,51	1,30	6,72	6,84	6,98	
31	36 (Budha, 2008)	7,11	test	Nonsteroidal	2	5	A(2) A(3) D(5) H(-) R(9) R(8)	1,03	0,72	0,54	1,40	7,33	7,09	7,11	
32	3i (Kaku, 2011a)	7,06	training	Nonsteroidal	1	6	A(2) A(3) D(5) H(8) R(11) R(12)	0,74	0,93	0,64	1,95	7,30	6,80	6,86	
33	8 (Hu, 2010a)	7,01	training	Nonsteroidal	1	4	A(1) A(2) D(-) H(-) R(6) R(5)	1,05	0,89	0,63	1,64	6,62	6,90	6,96	
34	24 (Jagusch, 2008)	7,01	test	Steroidal	1	4	A(1) A(2) D(-) H(-) R(4) R(5)	1,05	0,89	0,63	1,64	6,62	6,90	6,96	
35	L38 (Nnane, 1999)	6,97	training	Nonsteroidal	16	3	A(1) A(-) D(4) H(5) R(-) R(-)	1,15	0,75	0,52	1,31	6,74	7,05	6,98	
36	18 (Budha, 2008)	6,92	training	Nonsteroidal	1	5	A(2) A(3) D(5) H(7) R(11) R(-)	0,78	0,82	0,43	1,60	7,20	6,70	6,76	
37	6 (Budha, 2008)	6,89	test	Nonsteroidal	1	5	A(1) A(3) D(-) H(7) R(10) R(11)	1,19	0,78	0,55	1,34	6,88	6,92	6,90	
38	19 (Kaku, 2011a)	6,84	training	Nonsteroidal	1	5	A(1) A(3) D(4) H(-) R(7) R(8)	0,97	0,84	0,66	1,70	7,08	7,15	7,07	
39	25 (Budha, 2008)	6,82	training	Nonsteroidal	1	4	A(2) A(-) D(-) H(7) R(9) R(10)	0,96	0,58	0,62	1,39	6,64	7,03	6,88	
40	23 (Budha, 2008)	6,80	training	Nonsteroidal	1	6	A(4) A(3) D(6) H(9) R(12) R(11)	0,87	0,84	0,52	1,63	7,39	7,16	6,82	
41	17 (Kaku, 2011a)	6,79	training	Nonsteroidal	1	3	A(1) A(-) D(-) H(-) R(6) R(5)	1,10	0,59	0,51	1,19	6,05	6,46	6,62	
42	5ax (Hu, 2010a)	6,77	training	Nonsteroidal	18	4	A(1) A(-) D(-) H(3) R(5) R(6)	0,95	0,52	0,62	1,35	6,41	6,73	6,61	
43	10 (Hu, 2010a)	6,73	training	Nonsteroidal	1	4	A(1) A(-) D(3) H(-) R(5) R(6)	1,03	0,82	0,59	1,55	6,48	6,84	6,75	
44	14 (Kaku, 2011a)	6,73	training	Nonsteroidal	1	4	A(-) A(-) D(3) H(5) R(6) R(7)	1,10	0,76	0,39	1,24	6,49	6,58	6,62	
45	25 (Jagusch, 2008)	6,73	training	Nonsteroidal	1	4	A(1) A(-) D(3) H(-) R(5) R(6)	1,07	0,83	0,60	1,53	6,40	6,63	6,53	
46	3e (Kaku, 2011a)	6,72	training	Nonsteroidal	2	6	A(1) A(3) D(5) H(6) R(9) R(10)	0,65	0,99	0,81	2,25	7,22	6,89	6,70	
47	11 (Jagusch, 2008)	6,72	training	Nonsteroidal	1	5	A(1) A(2) D(-) H(4) R(5) R(6)	1,00	0,82	0,65	1,63	6,39	6,62	6,66	
48	27 (Jagusch, 2008)	6,65	test	Steroidal	25										yes
49	16 (Kaku, 2011a)	6,63	training	Nonsteroidal	1	3	A(-) A(-) D(-) H(3) R(6) R(5)	1,10	0,31	0,50	0,90	5,99	6,27	6,41	
50	30 (Budha, 2008)	6,63	training	Nonsteroidal	1	4	A(1) A(-) D(-) H(3) R(6) R(7)	0,94	0,88	0,65	1,75	6,43	6,84	6,51	
51	5ay (Hu, 2010a)	6,62	training	Steroidal	1	3	A(1) A(-) D(-) H(5) R(7) R(-)	0,89	0,64	0,58	1,48	6,47	6,82	6,68	
52	4 (Hu, 2010a)	6,61	training	Nonsteroidal	4	4	A(1) A(-) D(3) H(-) R(6) R(5)	1,03	0,82	0,58	1,54	6,48	6,83	6,73	
53	5bx (Hu, 2010a)	6,60	training	Nonsteroidal	1	4	A(1) A(-) D(-) H(3) R(5) R(6)	1,04	0,21	0,51	0,85	6,19	6,42	6,53	

54	31 (Budha, 2008)	6,58	training	Nonsteroidal	1	4	A(1) A(-) D(-) H(3) R(5) R(6)	0,95	0,91	0,65	1,77	6,33	6,60	6,22
55	3f (Kaku, 2011a)	6,54	training	Nonsteroidal	1	6	A(1) A(3) D(5) H(7) R(10) R(9)	0,13	1,00	0,91	2,80	7,41	6,95	6,70
56	5 (Handratta, 2005)	6,52	training	Nonsteroidal	1	3	A(-) A(-) D(-) H(6) R(11) R(10)	0,92	0,94	0,39	1,56	6,28	6,56	6,57
57	5 [Kaku, 2011b)	6,52	test	Steroidal	7									
58	13 (Kaku, 2011a)	6,51	training	Nonsteroidal	1	5	A(-) A(3) D(4) H(6) R(7) R(8)	0,97	0,84	0,43	1,46	7,02	6,90	6,82
59	2 (Kaku, 2011a)	6,48	training	Nonsteroidal	4	3	A(-) A(-) D(3) H(4) R(6) R(-)	0,92	0,91	0,46	1,60	6,18	6,03	6,41
60	28 (Jagusch, 2008)	6,47	test	Nonsteroidal	2	4	A(1) A(-) D(5) H(-) R(6) R(7)	1,04	0,79	0,57	1,49	6,71	6,96	6,86
61	14 (Jagusch, 2008)	6,46	training	Nonsteroidal	4	5	A(1) A(2) D(-) H(5) R(6) R(7)	1,10	0,83	0,63	1,55	6,49	6,59	6,50
62	27 (Budha, 2008)	6,43	training	Nonsteroidal	2	4	A(1) A(-) D(-) H(2) R(4) R(5)	0,94	0,94	0,68	1,83	6,35	6,64	6,32
63	14 (Hu, 2010a)	6,41	training	Nonsteroidal	1	3	A(1) A(-) D(-) H(-) R(5) R(4)	1,09	0,91	0,56	1,56	6,15	6,49	6,40
64	3g (Kaku, 2011a)	6,40	training	Nonsteroidal	1	6	A(1) A(3) D(5) H(6) R(9) R(10)	0,01	1,00	0,96	2,95	7,46	6,91	6,64
65	3j ((Kaku, 2011a)	6,39	training	Nonsteroidal	1	6	A(1) A(3) D(5) H(6) R(9) R(8)	0,74	0,97	0,69	2,04	7,26	7,11	6,88
66	5a (Hu, 2010a)	6,39	training	Nonsteroidal	1	3	A(-) A(1) D(-) H(2) R(-) R(4)	1,05	0,78	0,42	1,33	6,16	6,09	6,28
67	6 (Kaku, 2011a)	6,37	training	Steroidal	1	4	A(-) A(1) D(3) H(4) R(-) R(7)	1,10	0,53	0,44	1,05	6,13	6,07	6,36
68	25 (Hu, 2010a)	6,36	training	Nonsteroidal	1	3	A(1) A(-) D(-) H(-) R(5) R(6)	1,09	0,87	0,57	1,53	6,03	6,11	6,06
69	18 (Jagusch, 2008)	6,31	training	Nonsteroidal	1	3	A(1) A(-) D(-) H(2) R(-) R(6)	0,95	0,80	0,61	1,62	5,81	5,83	6,04
70	15 [37]	6,30	training	Nonsteroidal	1	3	A(1) A(3) D(-) H(9) R(-) R(-)	1,24	0,63	0,49	1,08	6,46	6,65	6,48
71	21 (Hu, 2010a)	6,24	training	Nonsteroidal	7	3	A(1) A(-) D(-) H(-) R(2) R(3)	1,06	0,95	0,56	1,63	6,11	6,39	6,21
72	5 (Kaku, 2011a)	6,23	training	Nonsteroidal	1	4	A(2) A(-) D(-) H(5) R(8) R(6)	1,03	0,41	0,49	1,03	5,96	5,95	5,83
73	29 (Budha, 2008)	6,23	training	Nonsteroidal	1	4	A(1) A(-) D(-) H(2) R(5) R(6)	0,95	0,89	0,65	1,74	6,33	6,70	6,31
74	15 (Hu, 2010a)	6,22	training	Nonsteroidal	1	3	A(1) A(-) D(-) H(-) R(6) R(5)	1,09	0,91	0,54	1,54	6,12	6,44	6,33
75	22 (Hu, 2010a)	6,20	test	Nonsteroidal	1	3	A(1) A(-) D(-) H(-) R(2) R(3)	1,06	0,91	0,56	1,59	6,33	6,71	6,50
76	KTZ ^b (Jagusch, 2008)	6,13	training	Nonsteroidal	1	5	A(5) A(3) D(-) H(8) R(10) R(11)	1,65	0,49	0,41	0,52	6,40	6,18	6,08
77	23 (Hu, 2010a)	6,12	training	Nonsteroidal	4	4	A(1) A(-) D(2) H(-) R(4) R(5)	1,07	0,77	0,59	1,47	6,34	6,40	6,32
78	12 (Jagusch, 2008)	6,11	training	Nonsteroidal	5	5	A(1) A(2) D(-) H(4) R(5) R(6)	1,10	0,75	0,61	1,45	6,16	6,28	6,24
79	19 (Jagusch, 2008)	6,07	training	Nonsteroidal	3	3	A(1) A(-) D(-) H(3) R(-) R(7)	0,95	0,80	0,61	1,62	5,77	5,78	5,97
80	12 (Hu, 2010a)	6,05	training	Nonsteroidal	6	5	A(3) A(2) D(4) H(-) R(5) R(6)	1,25	0,63	0,33	0,92	6,93	6,20	6,13
81	28 (Budha, 2008)	6,02	training	Nonsteroidal	3	4	A(1) A(-) D(-) H(2) R(4) R(5)	0,93	0,94	0,66	1,82	6,32	6,48	6,05
82	18 (Kaku, 2011a)	5,91	training	Nonsteroidal	2	3	A(-) A(1) D(-) H(2) R(-) R(6)	0,99	0,88	0,40	1,45	5,93	5,72	5,82
83	17 (Hu, 2010a)	5,86	training	Nonsteroidal	14	4	A(1) A(-) D(4) H(-) R(6) R(7)	1,08	0,86	0,56	1,52	6,13	5,81	5,69

yes

84	17 (Jagusch, 2008)	5,84	training	Nonsteroidal	3	3	A(1) A(-) D(-) H(2) R(-) R(6)	0,95	0,98	0,63	1,81	5,97	5,93	5,90	
85	13 (Hu, 2010a)	5,75	training	Nonsteroidal	6	4	A(1) A(-) D(4) H(-) R(5) R(6)	1,07	0,76	0,64	1,51	5,96	5,91	5,92	
86	24 (Hu, 2010a)	5,70	training	Nonsteroidal	6	3	A(-) A(2) D(-) H(-) R(7) R(6)	0,90	0,87	0,45	1,57	6,13	5,42	5,57	
87	8 (Kaku, 2011a)	5,63	training	Nonsteroidal	5	3	A(-) A(1) D(-) H(2) R(5) R(-)	1,08	0,68	0,51	1,29	5,78	5,45	5,53	
88	16 (Hu, 2010a)	5,48	training	Nonsteroidal	2	4	A(1) A(2) D(-) H(-) R(7) R(6)	1,11	0,57	0,57	1,21	5,68	5,27	5,31	
89	7 (Kaku, 2011a)	5,30	training	Nonsteroidal	2	5	A(2) A(3) D(-) H(7) R(10) R(8)	1,03	0,42	0,50	1,06	5,90	5,32	5,28	
90	10 (Kaku, 2011a)	5,30	training	Nonsteroidal	5	3	A(-) A(-) D(-) H(3) R(6) R(7)	1,06	0,82	0,49	1,43	5,44	4,91	5,07	
91	11 (Kaku, 2011a)	5,30	training	Nonsteroidal	1	3	A(-) A(-) D(-) H(3) R(6) R(5)	1,07	0,49	0,45	1,05	5,79	5,82	5,74	
92	12 (Kaku, 2011a)	5,30	training	Nonsteroidal	3	3	A(-) A(1) D(-) H(2) R(-) R(6)	1,03	0,90	0,47	1,52	5,93	5,61	5,58	
93	4 (Kaku, 2011a)	5,00	training	Nonsteroidal	2	5	A(3) A(1) D(4) H(5) R(-) R(8)	1,11	0,30	0,54	0,92	5,64	5,10	5,14	
94	20 (Jagusch, 2008)	5,00	training	Nonsteroidal	1	3	A(1) A(-) D(-) H(2) R(-) R(7)	0,95	0,98	0,63	1,81	5,75	5,47	5,36	
95	1 (Kaku, 2011a)	4,70	training	Nonsteroidal	3	4	A(-) A(1) D(3) H(4) R(6) R(-)	1,01	0,51	0,46	1,13	5,32	4,55	4,71	
96	3 (Kaku, 2011a)	4,70	training	Nonsteroidal	1	4	A(-) A(-) D(4) H(5) R(7) R(9)	1,01	0,86	0,47	1,49	5,55	4,90	4,91	
97	9 (Kaku, 2011a)	4,70	test	Nonsteroidal	2										
98	15 (Kaku, 2011a)	4,70	training	Nonsteroidal	4	4	A(1) A(-) D(-) H(3) R(6) R(5)	1,10	0,56	0,59	1,24	5,68	5,58	5,27	yes

^aTAK700 refers to Orteronel. ^bKTZ refers to ketoconazole.

Table A3. Results of two pharmacophore hypothesis with important features obtained from the 12 most active molecules in the data set.

Hypothesis	Survival	Survival inactive	Post- hoc	Site	Vector	Volume	Selectivity	Relative energy	Activity	Inactive
ADHRRR.116	3.310	1.852	5.626	0.640	0.884	0.790	2.549	0.571	7.72	1.457
AADHRR. 82	3.174	1.778	5.490	0.520	0.900	0.752	2.358	0.571	7.72	1.396

Table A4. Electronic properties calculated by using DFT optimization for the (active/inactive or) good- and poor-aligned structures to the best predictive hypothesis

Entry ID	Solvation Energy (kcal/mol)	HOMO (eV) ^a	LUMO (eV) ^b	E	QM Dipole (debye)
(+)-3c (Handratta, 2005)	-19.799	-0.226	-0.055	-0.171	9.66
3b (Handratta, 2005)	-20.117	-0.226	-0.056	-0.170	9.40
3g (Handratta, 2005)	-19.287	-0.226	-0.054	-0.171	9.62
5 (Mendieta, 2008)	-60.819	-0.225	-0.078	-0.147	28.78
16 (Zhuang, 2010)	-6.866	-0.213	-0.035	-0.178	0.00
5bx (Hu, 2010)	-10.362	-0.218	-0.030	-0.188	5.43

^aHOMO, Highest occupied molecular orbital. ^bLUMO, Lowest unoccupied molecular orbital.

