CYTOCHROME P450 PREDICTIONS IN SILICO

Henna Härkönen
M. Sc. Thesis
Master of Science in Pharmacy
University of Kuopio
Department of
Pharmaceutical Chemistry
June 2007

PREFACE

This M.Sc. Thesis has been a long-running process lasting from January 2006 until June 2007. The experimental part of this work was carried out in the Department of Pharmaceutical Chemistry in the University of Kuopio between January 2006 and May 2006 and the writing process lasted from October 2006 until today.

I would like to give thanks to my supervisor Ph.D. Carsten Wittekindt for guiding me throughout the experimental part and for answering my unending questions with true expertise. I am also grateful to my supervisor M.Sc. Tuomo Kalliokoski who patiently guided me throughout the thesis and gave valuable feedback all along the work. I would also like to thank him and Ph.D. Tuomo Laitinen for technical support and for reminding me to not to get carried away with the thesis. ”Sometimes you just have to stop writing and move on.”

I am also grateful to Professor Antti Poso, who caught my attention to molecular modeling and computational drug design with his utmost enthusiasm towards the subject. Without his encouragement to cling to this subject, I wouldn't be heading into the direction on my career where I'm going now.

Last – but not least – I would like to sincerely thank my partner Miikka who patiently encouraged me at every step of the work and who believed in me even when I didn't have the faith anymore. It has meant the world for me.

Kuopio 18.6.2007

Henna Härkönen

”Winners simply do what losers won't.”
From the movie Million Dollar Baby

THE UNIVERSITY OF KUOPIO, Faculty of Pharmacy
Master of Science in Pharmacy
Chemistry of Drugs and Adjuvants
HÄRKÖNEN, HENNA H.: Cytochrome P450 predictions in silico
M.Sc. Thesis, 110 pages
Supervisors: Ph.D. Carsten Wittekindt, M.Sc. Tuomo Kalliokoski

Keywords: cytochrome P450, ADMET, computational drug discovery and development, CYP2B6, molecular docking, CoMFA

ABSTRACT

Cytochrome P450 enzymes constitute a superfamily of enzyme proteins which have a significant part in the biotransformation of drugs and xenobiotics. Cytochrome P450s metabolize majority of currently known pharmaceutical agents and can cause drug-drug-interactions with co-administered drugs as well as unwanted adverse side effects. Following from this, cytochrome P450s represent a challenge for successful drug discovery and development. By exploring ADMET properties (absorption, distribution, metabolism, excretion, toxicity) of a potential drug candidate in the early phases of drug discovery process, predictions can be made concerning the pharmacokinetics and potential drug-drug-interactions of a drug molecule. As such, late-stage attrition of pharmaceutical agents can be reduced. Computational approach represents a powerful tool in predicting the ADMET profile of a potential drug molecule. It has the benefit over in vitro assays to possess less urgency for investments needed in resources, time and technology. The accurate prediction of the in vivo pharmacokinetics of a potential new drug, whilst existing merely as a virtual structure, is the ultimate goal of in silico ADMET screenings. If the prediction power of *in silico* models before chemical synthesis and expensive clinical trials becomes accurate and sophisticated enough, the models can replace some of the *in vitro* assays and *in vivo* experiments in the future and a great deal of time, money and resources can be saved. Consequently, computational approach is a widely accepted tool in early phases of drug discovery and an area of growing interest in drug research and development. The purpose of this literature review is to give the reader a comprehensive cross-section of different *in silico* approaches applied in cytochrome P450 predictions.

In the experimental part of this thesis, a database of 49 ligands was docked in the active site of CYP2B6 homology model. Based on conformations of ligands gained from molecular docking, comparative molecular field analysis (CoMFA) was performed by correlating the structural features of the ligands with experimentally determined inhibition potencies. Preliminary CoMFA model was generated for training set of 23 ligands and represented statistical values of good quality (q^2 = 0.564, S_{press} = 0.523, r^2 = 0.822, s = 0.345). Using the preliminary CoMFA model inhibition potencies were predicted for test set of 26 ligands (r^2 = 0.787, s = 0.712). Statistical values for CoMFA can be considered of good quality but further examining of the results is needed before the model could be successfully used in predicting potential substrates for CYP2B6.

KUOPION YLIOPISTO, farmaseuttinen tiedekunta
Proviisorin koulutusohjelma
Lääke- ja apuaineiden kemia
HÄRKÖNEN, HENNA H: Cytochrome P450 predictions *in silico*
Pro gradu-tutkielma, 110 s.
Ohjaajat: FT Carsten Wittekindt, proviisori Tuomo Kalliokoski

avainsanat: CYP-entsyymi, ADMET, tietokonepohjainen lääkeaineen suunnittelu, CYP2B6, molekyylitelakointi, CoMFA

TIIVISTELMÄ

Sytokromi P450-entsyymit eli CYP-entsyymit (cytochrome P450) ovat tärkein vierasaineita ja lääkeaineita metaboloiva entsyymiryhmä. CYP-entsyymien kautta metaboloituu suurin osa nykyään tunnetuista lääkeaineista, mikä aiheuttaa ei-toivottuja haittavaikutuksia johtuen lääkeaineiden farmakokineettisestä profiilista ja kliinisesti merkittäviä yhteisvaikutuksia annosteltaessa samanaikaisesti lääkeaineita, jotka käyttävät samoja, CYP-entsyymien katalysoimia metaboliareittejä. Mikäli jo lääkeaineiden tutkimus- ja kehitystyön aikaisissa vaiheissa pystytään tunnistamaan potentiaaliset CYP-entsyymien substraatit ja ennustamaan potentiaalisten lääkeainemolekyylien ADMET-ominaisuudet (absorption, distribution, metabolism, excretion, toxicity), vältytään tilanteelta, jossa kalliin ja aikaa vievän tutkimus- ja kehitystyön viimeisissä vaiheissa lääkeainemolekyyli joudutaan hylkäämään tai jo markkinoilla oleva lääkeaine joudutaan vetämään pois sen aiheuttamien yhteis- ja/tai haittavaikutusten takia. Tästä johtuen CYP-entsyymit ovat erittäin merkittävä ja haastava kohde uusien lääkeaineiden tutkimus- ja kehittämistyössä.

Yksi tärkeimmistä työkaluista, joilla potentiaalisten lääkeaineiden ADMET-ominaisuuksia tutkitaan, on tietokonepohjainen lääkeaineen suunnittelu. Tietokonemallien avulla lääkeainemolekyyliä ja sen ominaisuuksia voidaan tutkia virtuaalisesti ja niiden farmakokineettinen profiili *in vivo* voidaan ennustaa onnistuneiden *in silico*-mallien avulla. Verrattuna *in vitro*- ja *in vivo*- kokeisiin, tietokonepohjaisella lääkeaineen suunnittelulla säästetään aikaa, rahaa ja voimavaroja ennen potentiaalisen lääkeainemolekyylin synteesiä ja kallita, kliinisiä tutkimuksia. Tästä johtuen tietokonepohjaista lääkeaineen suunnittelua ja *in silico*-malleja käytetään laajasti lääkeaineen tutkimus- ja kehitystyön aikaisissa vaiheissa ja menetelmän suosio kasvaa yhä edelleen. Tämän kirjallisuuskatsauksen tarkoituksena on antaa lukijalle läpileikkaus erilaisista *in silico*-menetelmistä, joita käytetään tutkittaessa ja kehitettäessä uusia, CYP-entsyymien kautta metaboloituvia lääkeaineita.

Tämän työn kokeellisessa osassa 49 rakenteellisesti erilaista molekyyliä telakoitiin CYP2B6-entsyymin homologimalliin. Molekyylien telakoituneiden konformaatioiden perusteella niille tehtiin vertaileva molekyylikenttäanalyysi, CoMFA (comparative molecular field analysis) ja molekyylien biologinen aktiivisuus ennustettiin niiden rakenteellisten ominaisuuksien perusteella. Alustava CoMFA-malli, jossa käytettiin 23 CYP2B6-entsyymin ligandin muodostamaa testijoukkoa, kykeni onnistuneesti ennustamaan molekyylien IC_{50}-arvot (q^2 = 0.564, S_{press} = 0.523, r^2 = 0.822, s = 0.345). Tämän seurauksena alustavaa CoMFA-mallia käytettiin ennustettaessa IC_{50}-arvot testijoukon ulkopuoliselle, 26 ligandin muodostamalle joukolle (r^2 = 0.787, s = 0.712). Kokeellisen osan CoMFA-analyysia voidaan pitää onnistuneena ja tuloksia tilastollisesti merkittävinä, mutta jatkotutkimuksia tarvitaan ennen kuin luotua CoMFA-mallia voidaan käyttää uusien, CYP2B6-substraattien ennustamiseen.

ABBREVIATIONS

3D	three-dimensional
4-CPI	4-(4-chlorophenyl)-imidazole
7-EFC	7-ethoxy-4-trifluoromethylcoumarin
7-MFC	7-methoxy-4-trifluoromethylcoumarin
ADMET	absorption, distribution, metabolism, excretion and toxicity
Ala	alanine
AM1	Austin Model 1
ANN	artificial neural networks
Arg	Arginine
BIF	bifonazole
BLAST	Basic Local Alignment Search Tool
BNN	Bayesian neural networks
CART	classification and regression trees
CBS	common structural blocks
CC	combinatorial chemistry
CO	Carbon monoxide
CoMFA	Comparative Molecular Field Analysis
CPA	cyclophosphamide
CPU	central processing unit
CYP	Cytochrome P450
Cα	protein backbone carbon atom
DA	discriminant analysis
DFT	density function theory
FEP	free energy perturbation
GA	genetic algorithm
HTS	high throughput screening
Ic_{50}	the concentration of an inhibitory ligand which is required for 50% inhibition of its target
IFA	ifosfamide
Ile	isoleucine
IUPAC	International Union of Pure and Applied Chemistry

K_i	inhibitory constant; dissociation constant for an inhibitory ligand
kJ	kilojoule
K_m	Michaelis-Menten constant;the substrate concentration at which the rate of an enzymatic reaction is half its maximum
KNN	Kohonen neural networks
LD_{25}	indicator of lethality; a dose at which 25 % of test subjects will die
Leu	leucine
LFER	Linear Free Energy Relationships
log*P*	logarithm of the partition coefficient of the compound between 1-octanol and water
MC	Monte Carlo conformational search method
MD	molecular dynamics
MDMA	3,4-methylenedioxy-N-methylamphetamine
MEP	molecular electrostatic potential
MLR	multiple linear regression
MM	molecular mechanics
mRNA	messenger ribonucleic acid
NADH	reduced nicotinamide-adenine dinucleotide
NADPH	reduced nicotinamide-adenine dinucleotide phosphate
NNK	4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone
P450BM-3	bacterial cytochrome P450; CYP102
P450cam	bacterial cytochrome P450; CYP101
P450terp	bacterial cytochrome P450; CYP108
PAH	polycyclic aromatic hydrocarbon
PCA	principal component analysis
PCR	principal component regression
PDB	Protein Data Bank
Phe	phenylalanine
pIC_{50}	reverse logarithmic representation of Ic_{50}
pK_a	acid dissociation constant
PLS	partial least squares
PM3	Parameterized Model 3

Q^2	cross-validated squared correlation coefficient
QM	quantum mechanics
QSAR	quantitative structure-activity relationship
r^2	squared correlation coefficient
RMSD	root mean square deviation
RP	recursive partitioning
RP73401	3-cyclopentyloxy-N-(3,5-dichloro-4-pyridyl)-4-methoxybenzamide
SCR	structurally conserved region
SDEP	standard deviation of error of prediction
SEP	standard error of prediction
Ser	serine
SOM	Kohonen self-organizing map
S_{press}	standard error of prediction
SRS	substrate recognition site
SVM	support vector machine
SVR	structurally variable region
Thr	threonine
Val	valine
X-ray	Röntgen rays
Å	Ångström, 10^{-10} m
ΔG	free energy of binding
ΔH	enthalpy
ΔS	entropy

CONTENTS

I LITERATURE REVIEW:

Cytochrome P450 Predictions *in silico*

1 INTRODUCTION

1.1 Cytochrome P450 Superfamily

Cytochrome P450 proteins are one of the largest superfamilies of enzyme proteins (Werck-Reichhart and Feyereisen 2000). They were first discovered in the 1950´s–60´s with simultaneous discovery of atmospheric oxygen incorporating in substrate molecule during metabolism catalyzed by metalloprotein enzymes, which eventually lead to understanding cytochromes having a part in oxidative metabolism (Estabrook 2003).

Cytochrome P450 family is named by unique absorption peak at 450 nm with carbon monoxide-bound pigment, which was first found by Klingenberg in 1958 and later reported by Sato and Omura in 1962 (Omura 1999, Estabrook 2003). The absorption peak of reduced P450 is still used for the estimation of the P450 content of a probe (Bernhardt 2006). *P* stands for pigment and *450* refers the maximum absorbance at 450 nm. *Cytochrome* stands for hemoprotein and in general, cytochromes are considered as heme-containing membrane-bound proteins with covalently bound sulfur from a cysteine residue as a proximal ligand (Lewis 2006). The protein family is commonly referred as CYPs, shorthand for cytochrome P450s.

In 1965, during research of induction of drug-metabolizing enzymes present in endoplasmic reticulum of liver, Remmer observed an increase in the concentration of the CO-binding pigment of liver microsomes during induction (Estabrook 2003). This triggered several studies with cytochrome P450 having the key role in many reactions of drug and xenobiotic metabolism.

Nowadays CYPs are known to participate in biotransformation of drugs and bioconversion of xenobiotics, activation and metabolism of chemical carcinogens, degradation of herbicides and insecticides and biosynthesis of physiologically important compounds such as steroid hormones and vitamins (Omura 1999, Guengerich 2001a, Bernhardt 2006). In total of over 700 different CYP families have been found and the number of known genes encoding CYP proteins is over 6000 (Nelson 2006). The number is still increasing due to genome sequencing started in the 90´s (Denisov et al 2005).

CYPs are found both in mammalian tissues and plants and they are also present in fungi and bacteria (Omura 1999, Werck-Reinhardt and Feyereisen 2000, Denisov et al 2005). In mammalian tissues the majority of CYPs are present in the liver, but they are also located in extrahepatic tissues, for example in the lungs and on the skin (Guengerich 2001a). CYPs are said to have a post at the gates of the system throughout xenobiotics enter it (Raunio 2001). Accordingly, their role in biotransformation of drugs and xenobiotics makes CYPs one of the most important proteins in the field of modern drug development. However, only a relatively small set of cytochrome P450s contribute to drug metabolism although 90 % of the phase I metabolism of pharmaceutical agents is mediated by them (Lewis and Dicking 2002, Arimoto 2006, Lewis et al 2006). From the known CYP isoforms subfamilies 1A, 1B, 2A, 2B, 2C, 2D, 2E and 3A are known to be involved in drug metabolism (Guengerich 2001b, Li 2001, Bathelt et al 2002, Lewis and Dicking 2002, Lewis et al 2006). Figure 1 represents the most important isoforms participating in the biotransformation of pharmaceutical agents. Nonetheless, in literature there are differences in stating which ones of the cytochrome P450 isoforms have the greatest impact on drug metabolism.

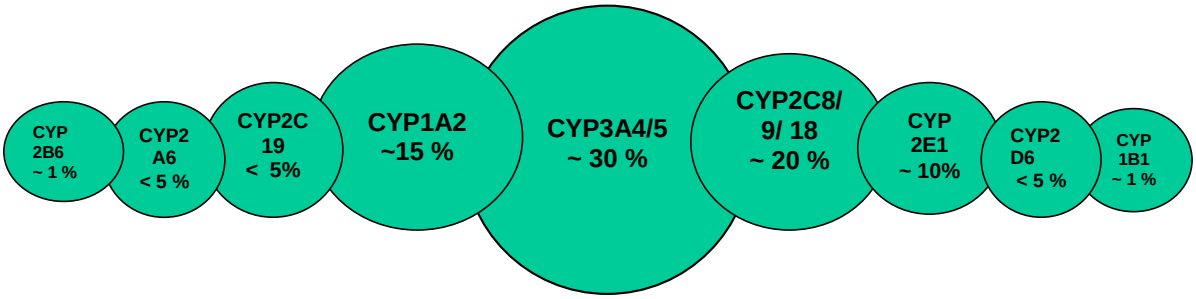


Figure 1: Chart representing the most important cytochrome P450 isoforms and their percentage of the total hepatic content. Adapted, constructed and compiled from Rendic and di Carlo (1997), Pelkonen and Raunio (1998), Lin and Lu (2001) and Danielson (2002)

In the following, only the most important CYP isoforms are presented. Introducing each individual CYP is beyond the scope of this research, and for further information concerning other isoforms the reader is referred to the literature. The CYP3A family comprises up to 30 % of the total hepatic content (Danielson 2002). One member of the family, CYP3A4 is supposed to metabolize nearly 60 % of all the known pharmaceuticals covering a broad variety of drugs, such as macrolide antibiotics, sedatives and immune

system modulators. In consequence, the CYP3 family and especially CYP3A4 is considered the most important isoform among cytochrome P450s. However, the CYP3 family consists of only four functional genes (CYP3A4, CYP3A5, CYP3A7, CYP3A43) compared to e.g. the CYP2 family, which comprises of 13 subfamilies consisting of 16 functional genes and 13 confirmed pseudogenes. As such, CYP2 family can be contemplated the largest family of cytochrome P450s comprising one third of the cytochrome P450s sequence. In addition, the total hepatic content of the CYP2 family enzymes is greater than the hepatic content of CYP3 family, comprising approximately 40 % (Rendic and di Carlo 1997, Danielson 2002). Furthermore, the polymorphism of CYP2 makes their research and discovery more difficult and time-consuming, and little is known of its inductive nature or transcriptional properties, albeit its participation in the metabolism of anti-cancer drugs (taxol, cyclofosfamide, ifosfamide), anticoagulant warfarin and pharmaceuticals used in Parkinson's disease (selegiline) is stated as. Since CYP2 family is not yet mapped through and through – as is not any other isoform –, the proposition of which isoform is truly the most important among cytochrome P450s, should weigh up carefully.

As mentioned before, CYPs have a significant part in oxidative metabolism. They are monooxygenases, and their primary catalytic function is the transfer of one oxygen atom from molecular oxygen (O_2) into various substrates (Bernhardt 2006). The general cycle is presented in Figure 2 (adapted from Guengerich 2001b). The reaction is considered to be a catalytic cycle which contains steps involving activation of oxygen molecule, substrate oxidation and product release. During this process, external reducing equivalents from NADPH (reduced nicotinamide-adenine dinucleotide phosphate) or NADH (reduced nicotinamide-adenine dinucleotide) are required (Bernhardt 2006). The reaction takes place after substrate binding to the active site of the CYP 450. The general formula of the reaction is:

$$\mathbf{RH + O_2 + NAD\,(P)H + H^+ \rightarrow ROH + H_2O + NAD(P)^+}$$

The cyclic oxidation reaction can be explained simplified as follows (Omura 1999, Raunio 2001, Denisov et 2005, Bernhardt 2006): After substrate binding, the first electron from NAD(P)H is transferred to heme-bound- $Fe^{3+\cdot}$ Molecular oxygen, O_2 binds to reduced heme-bound iron and forms an oxygenated complex, heme-Fe^{2+}-O_2. A second electron is

then transferred to the oxygenated complex, forming a substrate-bound heme-Fe^{2+}-O_2^-, which after protonation (2 H^+) degrades into substrate-bound $Fe^{3+}O$ and a water molecule, H_2O. Substrate-bound $Fe^{3+}O$ is further degraded leading into an oxidized substrate and original Fe^{3+}-P450-form. However, the reaction cycle can be explained in another way, which differs from the previously described by containing the one-electron transfer only once. The one-electron transfer from NAD(P)H occurs into substrate-bound heme-$Fe^{3+}O_2$- leading into ferric peroxo state (Fe^{3+}-O_2^{2-}), which is protonated leading first to Fe^{3+}-O_2H^- (hydroperoxo)-form. Second protonation then occurs to the distal oxygen atom of Fe^{3+}-O_2H^-, forming a very unstable intermediate Fe^{3+}-OOH_2, which is rapidly followed by splitting of the O-O-bond and the release of one water molecule, H_2O. The remaining porphyrin-metal-oxo-complex then attacks the substrate yielding oxygenated substrate and the original Fe^{3+}-P450-form, and dissociates to let the cycle start all over again.

Despite CYPs being considered mainly oxygenases, during their long period of evolution; 3.5 billion years, CYP family has become very versatile, and it has a broad field of chemical activity (Rendic and di Carlo 1997). Among CYPs there are isomerases, reductases, dehydrases and NO-synthases and in addition to oxidation, they contribute to various other reactions listed in Table 1 (Guengerich 2001b, Danielson 2003, Bernhardt 2006). The reactions mainly lead to the conversion of harmful xenobiotics less toxic by increasing their solubility, which is the first step in preparation of excretion (Denisov et al 2005). However, exeptions can be found among the listed reactions.

For example, CYP families CYP11, CYP17, CYP19 and CYP21 are involved in the biosynthesis of physiologically important steroids such as testosterone, progesterone and pregnenolone (Rendic and di Carlo 1997, Yamazaki and Shimada 1997, Bernhardt 2006, Storbeck et al 2007). CYP11, CYP17, CYP19 and CYP21 catalyze steroidogenesis mostly via hydroxylation reactions. Cytochrome P450s also activate non-toxic procarcinogens to toxic carcinogens (Rendic and di Carlo 1997). Cytochromes 1A1 and 1B1 oxidize e.g. enviromental chemicals such as polycyclic aromatic hydrocarbons (PAHs) (Shimada and Guengerich 2006). Oxidation of these compounds is the initial step in the activation of PAHs to carcinogenic products. Also CYP1A2, 2C9 and 3A4 participate in the activation of procarcinogens, although at much slower rates than CYP1A1 and 1B1 (Shimada et al 2001).

Table 1: Chemical reactions catalyzed by cytochromes P450 (Guengerich 2001b, Danielson 2003, Bernhardt 2006)

Reactions	
NO, NN- reduction	**Alkene epoxidation**
SO_2-reduction	**Epoxide reduction**
Hydroperoxide reduction	**Isomerizations**
1-electron oxidation	**Oxidative C-C bond cleavage**
Aldehyde deformylation	**Alkyne oxygenation**
Coupling reactions	**Arene epoxidation**
Ring formation, expansion	**N-, S-, O-dealkylation**
Hydrolysis	**N-hydroxylation**
1,2-shifts	**N-oxidation**
Mechanism-based heme inactivation	**S-oxidation**
Mechanism-based protein modification	**Oxidative deamination**
Reductive dehalogenation	**Oxidative dehalogenation**
Dehydratations	**Dehydrogenation**
Hydrocarbon hydroxylation	**Steroid hydroxylation**
Aromatic hydroxylation	

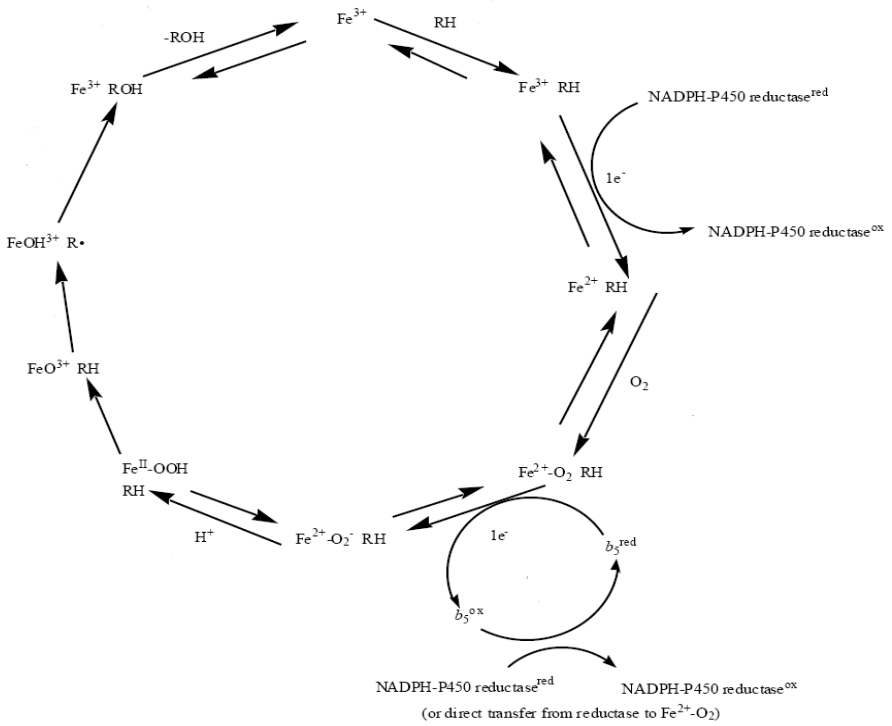


Figure 2: General cytochrome P450 reaction cycle

1.2 Fail early, fail cheap – the need of early stage ADME profiling

Drug discovery and development is a time-consuming and expensive process. The process takes up to average of 15 years, costing several hundred million dollars/euros (Yu and Adedoyin 2003). To avoid potential drug to be withdrawn from clinical studies or the market due to serious side effects, the need for lead compound to be studied in detail at early stage of drug development is highly important (Turpeinen et al 2006). Withdrawal of a compound usually occurs on clinical trials as a consequence of poor pharmacokinetic properties, animal toxicity or lack of efficacy (Ekins et al 2002). Adverse human side effects are also one reason and by far the most public, visible and expensive cause of failures. Notably, some drugs are withdrawn from the market after successful clinical trials due to drug-drug-interactions or severe side effects. At this point a lot of resources have already been invested to the drug, so it is quite understandable, why in modern pharmaceutical industry the strategy "fail early, fail cheap" is accepted (Li 2001). An example of the applicability of the strategy related to expensiveness of late-stage attrition can be made of Bayer and the cholesterol lowering drug cerivastatin (Lipobay®/Baycol®). After cerivastatin first reaching the market in 1997, Bayer had to withdraw the drug from market in 2001 after 31 deaths of severe rhabdomyolysis in the USA (Charatan 2001). The average costs for a new pharmaceutical agent to get to the market are estimated to be 800-900 million dollars/euros (Yu and Adedoyin 2003, Almarsdóttir and Traulsen 2005). In addition to this, Bayer faced up to approximately 5000 legal claims from patients, and the settlement payments ranged up to cost approximately 1.15 million dollars/euros in the end of March 2006 (Bayer 2006). The total amount of the costs is not yet available, since additional legal claims are possible to occur. The Lipobay story alone proves the “fail early, fail cheap”-strategy to be crucial and in its place in the field of successful drug design.

Despite the time and effort put in the process, the attrition rate of drugs discovered is still 76 % from target to IND (investigational new drug). 90 % of all tested compounds fail at the end of clinical trials (Yu and Adedoyin 2003). Majority of the unwanted adverse effects and drug-drug- interactions relate to measurable and to great extent predictable ADMET (absorption, distribution, metabolism, excretion and toxicity) properties. Arising from this, early consideration of ADMET properties has also become an essential tool in the field of successful discovery and development of new drugs and drug candidates (de

Graaf et al 2005a). Drug discovery as such is a multi-parameter optimization process, in which compounds are optimized for interaction with desired target, and to gain most favored drug-like properties of the candidate compounds (van de Waterbeemd and Gifford 2003). This parallel and iterative optimization approach includes also physicochemical properties (lipophilicity, solubility, pK_a, permeability, hydrogen bonding), selectivity, and potency of the compound, in addition to ADMET parameters. Since the optimization process on the whole is highly complex in nature, computational approach is more than suitable for the purpose, making the process more straight-forward, time-saving and as such, less expensive compared to more conventional methods (Zlokarnik et al 2005).

1.3 In vitro test systems – The traditional method in early stage drug discovery

Traditionally, *in vitro* assays have been used during early drug development, in assessing the potency of a compound to become a drug by studying its ADMET properties and providing with reasonable accuracy a preliminary prediction of the behavior of the compound *in vivo* (Yu and Adedoyin 2003). Over the time, *in vitro* assays have improved significantly via protocol simplification and technological advancement to keep up with the growing amount of information on ADMET data and the shortened cycle time characteristic to early drug discovery. Therefore various medium- and high-throughput *in vitro* ADMET screens using robotics and miniaturization are now available. However, the demand for preventing late-stage attrition by predicting ADMET properties is still increasing, and better results in screening the most promising compounds for further optimization are wanted. This creates a need for methods producing higher throughput with less investments which is characteristic for methods such as *in silico* modeling (van de Waterbeemd and Gifford 2003, Yu and Adedoyin 2003).

1.4 In silico – A challenging approach in ADMET profiling

The phrase *in silico* refers to an experiment or process performed virtually, as a computer simulation. Currently, computational tools are of great importance in modern drug discovery and development due to their ability to save time, resources and money and to produce rational and reliable information (Boobis et al 2002, Ekins 2003).

In silico methods are basically used in assisting in the selection of appropriate assays and compounds to go through the assays and further development (Boobis et al 2002). The benefit of using computational approaches over *in vitro* assays is less investment needed in resources, time and technology. The accurate prediction of the *in vivo* pharmacokinetics of potential new drug, whilst existing merely as a virtual structure, is the ultimate goal of *in silico* ADMET screenings. Obtaining merely virtual structures means avoidance of the need to synthesize and screen every molecule perused (Ekins et al 2002). If the prediction power of *in silico* models becomes accurate and sophisticated enough, the models can replace *in vitro* assays and *in vivo* experiments in the future (van de Waterbeemd and Gifford 2003). Nevertheless, at the moment best results in generating new and effective drugs are obtained in combination with experimental and computational approaches. Integration of both of these methods is valuable for generating more interpretable data, in order to identify the most promising compounds in drug discovery process for further optimization. The large databases containing information of the absorption, distribution, metabolism, excretion and toxicology of the compounds yielded by *in vitro* studies, are converted into more interpretable format by generating predictive computational algorithms, which are used to improve the process of drug discovery (Ekins 2003, 2006).

Computational approaches can be roughly divided into two main categories: ligand-based approaches and structure-based (protein-based) approaches (Yamashita and Hashida 2004, de Graaf et al 2005a, de Groot 2006), which are discussed in more detail in chapters 2 and 3 of this text. The approaches have been developed based on predicting potential substrates or inhibitors for target proteins, concentrating on either small molecules metabolized by certain protein or on the structures of proteins, and predicting ADMET properties of the compounds based on their chemical structure (de Groot 2006). The underlying need for this rests on the insufficiency of both combinatorial chemistry (CC) and high-throughput screening (HTS). Although having a crucial part in lead optimization and finding potential candidate drugs as being faster than conventional lead discovery, CC and HTS do not reduce the attrition rate of new drugs enough, nor increase the number of marketed new drugs as expected (Norinder and Bergström 2006). More robust and accurate computational tools are needed in order to design new compounds and libraries with desirable profile (O'Brien and de Groot 2005).

Ligand-based approach can be considered as an empirical method using statistical tools to explore linear and non-linear relationships between the structural descriptors of ligands, and observed parameters of a particular ADMET property (Ekins et al 2002). It includes methods such as pharmacophore modeling, quantitative structure-activity relationship (QSAR), three-dimensional (3D) QSAR and similarity searches (Ekins et al 2002, Yamashita and Hashida 2004). QSAR-approaches are derived from physicochemical and topological properties of compounds, and have proved to be pragmatic, rapid and rather successful approach in predicting drug metabolism by CYPs (de Graaf et al 2005a).

Structure-based approach endeavors to be insightful as well as predictive, including 3D molecular modeling techniques such as protein homology modeling and ligand-protein docking, as well as pharmacophore modeling based on x-ray or NMR structures of proteins (Ekins et al 2002, Yamashita and Hashida 2004, de Graaf et al 2005a, Wolber and Kosara 2006). Molecular modeling provides information of the behavior of molecules and molecular systems and their interaction with one another, using quantum mechanical methods and computational chemistry. Computational chemistry combines quantum mechanics, molecular mechanics, minimization, simulation, conformational analysis and other computer-based methods providing the possibility to understand and predict the behavior of target system and compound (Leach 2001). This information can be further exploited in predicting the ADMET properties of the drug candidates. Structure-based approach demands more computational power than ligand-based approach, although having lower throughput (Ekins 2003). The best ADMET models overall are consensus models, combining information of several computational models tailored to work in the necessary context (Yamashita and Hashida 2004, O'Brien and de Groot 2005).

This literature review presents computational approaches capable of producing predictive information concerning the target systems in the field of drug discovery and development. The main focus lies in *in silico* predictions of cytochrome P450 superfamily, considering both ligand- and protein-based methods. However, also *in silico* methods combining information from both of these methods are discussed.

2 LIGAND-BASED MODELS

Ligand-based models concentrate on information gained from the ligands only. Although ligand-based models can produce indirect information concerning the active site of the target protein, the focus is on a set of known substrates, inhibitors or metabolic products, and on their shape, electronic and physico-chemical properties, as well as conformations. Computational modeling utilizes mainly ligand-based models since the crystal structures of proteins are not as easily available as small molecules. Ligand-based models can be divided into two main categories: Pharmacophore models and quantum mechanical models (de Graaf et al 2005a).

The invention of pharmacophore *per se* occurred by Perkins in the middle of the nineteenth century, during his research concerning synthesis of coloring agents (Gund 2000). He discovered that the skeleton of a molecule left untouched but varying the functional groups, caused different colors and related properties. Near the end of the century, Paul Ehrlich discovered the coloring agent's ability to dye different tissues and thus, be able to cure certain diseases, which lead to the discovery of drugs having to bind to receptors for exerting their medicinal effect. This was followed by understanding the "lock and key" type of interactions, making way for fundamentals of pharmacophore modeling. The definition of *pharmacophore* was first offered in the early 1900s by Ehrlich, describing pharmacophore as a molecular framework that carries (*phoros*) the essential features responsible for a drug's (*pharmacon*) biological activity (Güner 2000; Erhlich 1909, Gund 1977). It was not until almost 70 years later that the definition was updated by Peter Gund in 1977 to its recent form: "a set of structural features in a molecule that is recognized at a receptor site and is responsible for that molecule's biological activity." Modern definition of pharmacophore is based on IUPAC recommendation: "A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optical supramolecular interactions with the specific biological target and to trigger (or block) its biological response" (Guerin et al 2006).

As Erlich discovered, in order to interaction occurring between a compound and the target protein, compounds have to bind to the receptor (Höltje et al 2003). Binding requires the compounds presenting certain structural elements of identical functionality in sterically

consistent locations to the target protein. For exerting qualitatively similar activities at the particular receptor, the compounds are assumed to possess similarity in these structural elements essential for the binding, producing generally similar biological properties.

Pharmacophoric groups are considered as representation of similar physical and chemical properties of the compounds, and specified spatial relationships of the groups, usually expressed as distances, distance ranges and geometric measures such as angles and planes, (Leach 2001). Generally used pharmacophoric features include hydrogen bond donor or acceptor, basic, acidic, aromatic or hydrophobic atoms (Mason et al 2001). However, pharmacophore does not merely represent particular molecular structure or association of functional groups. It can be considered providing a multiple pharmacophore descriptor, which after analyzing internal relationships between the descriptors, results a map representing the essential, steric and electronic features for a group of molecules exhibiting optimal interaction with the target structure, and which provides a certain pharmacological effect (Wermuth 2006). A representation of a pharmacophore is presented in Figure 2 (adapted from Ekins et al 1999c).

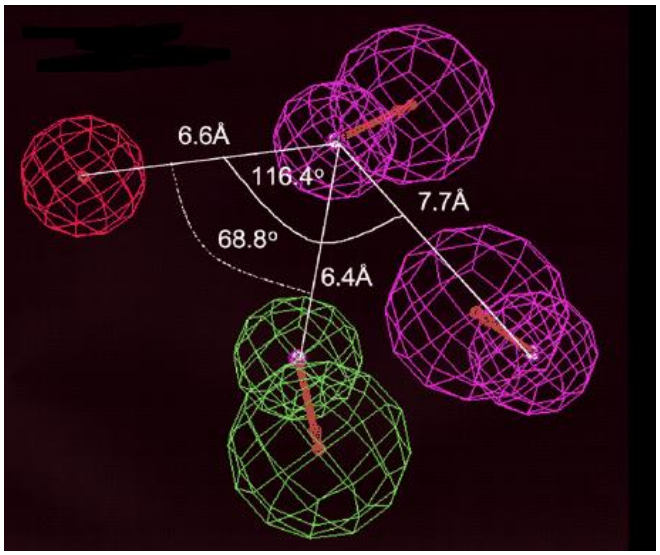


Figure 3: Pharmacophore model for CYP3A4 substrates. Represented are interbond angles and distances between pharmacophoric features. Red sphere represents hydrophobic region, purple regions are hydrogen bond acceptor features and green spheres are hydrogen bond donors

One of the major tasks using molecular modeling is to solve the pharmacophores for uniform groups of drug molecules. Information of the hypothetical pharmacophore for a group of molecules provides a very important source for understanding drug-receptor interactions at the same level (Höltje et al 2003).

According to de Graaf and coworkers (2005a), pharmacophore models can be further divided into two groups: classical pharmacophore models and 3D-QSAR models, which are discussed in the following chapters (de Graaf et al 2005a). Both methods are based on empirically derived information of substrates, inhibitors and metabolic products of the CYPs, concerning the shape, electronic properties and conformations of the compounds. The information can be further applied to produce knowledge about and on the active site of target protein and the features essential for ligands, in order to be biologically active and thus, potential lead compounds in drug discovery process.

Quantum mechanical (QM) models differ from pharmacophore models by focusing on elements of much smaller scale and consequently demanding more computer capacity due to complicated computational algorithms, which QM requires. Following from this, QM models provide more detailed information of the cytochrome P450s and their ligands. The QM models are discussed in chapter 2.3.

2.1 Classical pharmacophore models

Classical pharmacophore models are constructed by aligning the chemically similar structural elements essential for pharmacological activity of the molecule, *pharmacophoric groups*, onto each other (Yamashita and Hashida 2004, de Graaf et al 2005a). Molecular alignment can be addressed as approximating the binding geometry of a set of ligands. If the alignment is correct, the relevant chemical features of the ligand can be readily extracted in order to derive a pharmacophore, which can be used in searching for possible substrates or inhibitors from a ligand database (Lemmen and Lengauer 2000). As such, molecular alignment is the most important step concerning pharmacophore modeling in general, and it is discussed in detail in chapter 2.2.1.

Classical pharmacophore models can be applied as an input for 3D database searching of new substrates and inhibitors for potential drugs, molecular graphics, or automated 3D design and 3D-QSAR methods (Mason et al 2001). Using the information gained from the molecular alignment itself and the pharmacophore models, it is possible to derive complementary information of the structure and electrostatic distribution in the active site of target proteins (de Groot and Ekins 2001).

Still, classical pharmacophore models give an estimation of the activity of new compounds, not exact information. It should be noted, that models are based on a particular pharmacophore model, and another pharmacophoric model can give different kind of predictions (de Graaf et al 2005a). Despite recent classical pharmacophore models give predictions of the specificity and binding of the ligands in the active site based on their structural features, 3D-QSAR models are more advanced in producing quantitative parameters concerning ligand binding affinity towards the cytochromes.

Significantly, pharmacophore models can be constructed also based on available structures of ligand-protein complexes (Wolber and Langer 2005). In this approach, pharmacophores are derived based on chemical features and volume constraints of the bound ligand, as well as amino acids surrounding the ligand in the active site. The advantage of using ligand-protein complexes is producing selective pharmacophore models in screening of large compound databases for active molecules with described binding mode.

Classical pharmacophore models are quite useful when there is no structure available of the target enzyme (Lemmen and Lengauer 2000, Arimoto 2006). Nevertheless, they can also be combined with one another, when the structure of the protein is available (Griffith et al 2005). The combination of both models often produces more informative models of ligand-protein interaction. However, in most occasions protein structure of cytochromes P450 is not available due to complex and difficult nature of producing protein model, which actually works. Currently, there are few human cytochrome structures available (CYP2A6, CYP2B4, CYP2C5, CYP2C8, CYP2C9, CYP2D6 and CYP3A4), and the focus of new drug discovery is in the ligands of cytochromes (Williams et al 2000, Scott et al 2004, Lewis et al 2006, Marechal and Sutcliffe 2006, Rowland et al 2006, Yano et al 2006). Like Arimoto (2006) states in his article, “The ultimate goal is to predict the CYP specificity and the interactions for novel compound from its (ligand's) chemical structure”.

An overview of the vast and still growing number of pharmacophore models of cytochrome P450 substrates is presented in Table 2.

Table 2: Summary of key ligand-based pharmacophore features of cytochrome P450

CYP	Key features of ligand-based pharmacophore	Reference
1A2	**Inhibitor model includes hydrogen bonding , negative charge favored areas near nitrogen atom of quinoline ring, hydrophobic interactions between substrates and protein**	**Korhonen et al 2005**
2A5	**Lactone moiety important to inhibition, negative charge favored near the carbonyl oxygen and over both sides of coumarin ring plane, a large steric area allowed locating around 7-position of coumarin**	**Poso et al 2001**
2A6	**Naphthalene derivatives are potent inhibitors, partial negative charge increases inhibition potency, heterocyclic nitrogen atom decreases inhibition potency**	**Poso et al 2001, Rahnasto et al 2005**
2B6	**Two possible pharmacophores generated. One is having hydrogen bond acceptor at a distance of 4.6 Å from substrates reaction site, the other having 4.9 Å them at distances of 3.4 and 4.0 Å from the reaction site. Both have two hydrophobic regions, first one having site, the second having them at distances of 4.1 and 7.8 Å. Hydrophobicity and hydrogen bonding crucial for ligand binding**	**Ekins et al 1999b, Wang and Halpert 2002**
2C8	**Substrates possess long hydrophobic chain with acidic or polar group 12.9 +/- 0.8 Å from the oxidation site and hydrophobic or aromatic moiety 3.9 +/- 0.7 Å from the oxidation site. Two supplementary polar groups are located 4.4 +/- 0.3 Å and 8.6 +/- 0.3 Å from the oxidation site**	**Melet et al 2004**
2C9	**Hydrogen acceptor at a distance of 3.4-5.7 Å from hydrogen bond donor or second acceptor, hydrophobic feature positioned at 3.0-5.8 Å from the first hydrogen acceptor. Hydrophobic region usually concludes aromatic ring, locating 5.6 -9.8 Å from the site of oxidation**	**Ekins et al 2000, de Groot el al 2002**
2C18/19	**Substrate model includes two hydrophobic regions and electropositive binding sites**	**Ridderström et al 2001**
2D6	**Inhibitor models consist of tertiary nitrogen atom, a flat hydrophobic region, a region with additional functional groups having lone-pairs The nitrogen atom locates at a distance of 5-7 Å from the site of oxidation**	**Ekins et al 2001**

Table 2 continues

CYP	Key features of ligand-based pharmacophore	Reference
2E1	Substrates are usually small at size bearing lipophilic regions. They usually contain benzene ring	Lewis et al 2003
3A4	Two substrate models: First one has a hydrogen bond acceptor at a distance of 5.5-7.8 Å from the site of metabolism and 3 Å from the oxygen molecule bound to heme. Second model includes two hydrogen bond acceptors, a hydrogen bond donor and a hydrophobic region. Also two inhibitor models: One has three hydrophobes at distances of 5.2-8.8 Å from hydrogen acceptor, the other has three hydrophobes at distances of 4.2-7.1 from hydrogen bond acceptor, and further 5.2 Å from another hydrogen acceptor OR one hydrophobe at distances of 8.1-16.3 from the two furthest of three hydrogen bond acceptors	Ekins et al 1999a, c, 2001
3A5/A7	Inhibitor model consists of two aromatic rings, hydrogen bond acceptor and three hydrophobic features only a short distance apart	Ekins et al 2003
17 (17α-hydroxylase)	Inhibitor model for steroids and non-steroids include one or two hydrogen bond acceptors and three hydrophobic groups. In imidazole moiety, the N-3 acts usually as a acceptor	Ahmed 1999, Clement et al 2003
19 (aromatase)	Inhibitor model includes hydrogen bond acceptor and two hydrophobic areas near C6 of steroids	de Groot and Ekins 2002
51 (14 α-demethylase)	Imidazoles are more potent than triazoles and N3 and N4 of azole ring, etheral oxygen atom and an aromatic ring centroid are essential. Inhibitor model possess a hydrogen bond acceptor and three hydrophobic regions	Ekins et al 2001

2.2 3D-QSAR models

QSAR (quantitative structure activity relationships) is a method, which correlates molecular structure of a compound to some kind of *in vitro* or *in vivo* biological property, mostly biological activity in the field of modern drug discovery (Winkler 2002). In general, QSAR relates numerical properties of the molecular structure (descriptors) to its activity by using a mathematical model, which tries to find a correlation between the descriptors and activity by means of statistical methods (Leach 2001).

QSAR was first developed by Hansch and Fujita in 1969, when Hansch developed equations correlating biological activity to electronic characteristics and hydrophobicity of molecules. At the beginning, the method was named Linear Free Energy Relationships (LFER), due to correlation occurring linearly between the biological activity, and free-energy related physicochemical parameters such as log*P (*the logarithm of the partition coefficient of the compound between 1-octanol and water) (Debnath 2001). Hansch described the correlation using different equations, for example:

$$\log(1/C) = k_1 \log P - k_2(\log P)^2 + k3\sigma + k4 \qquad \text{Eq. 1}$$

In this equation, C is the concentration of the compound, required to cause a standard response in a given time. σ is the Hammet substituent parameter, which describes the effect of substituents, and k_1, k_2, k_3 and k_4 are constants. However, there are several other equations available describing the correlation in a different, more complicated way, depending on the parameters of interest. The equations can be either linear or non-linear in nature. There is diversity among the descriptors (calculated or measured properties) also, since they can be either fragment-based describing only a small section of the molecule, or molecular descriptors covering the whole molecule (Winkler 2002). The number of descriptors is enormous, covering both 2D- and 3D-descriptors subdivided in many ways (e.g. spatial, electronic, thermodynamic, conformational, topological and structural properties) (de Groot 2006). For example, Free and Wilson suggested an approach concerning substituents and their effect on biological activity, and Kier and Hall represented structural indices affecting the activity (Debnath 2001).

The multiple approaches together with Kubinyi's bilinear model (1977) have had considerable effect on understanding how chemical structures influence biological activity. Notably, the effect on activity can be generated using single parameters or a combination of them. With multiple parameters the data set size becomes larger and more difficult to interpret with linear functions, and therefore non-linear techniques such as neural networks are needed. Neural networks with other non-linear techniques are discussed further with other machine learning methods at chapter 2.2.3.

The classical explanation of QSAR includes formulating the relationships between the structure of a molecule and its biological response. In the QSAR approaches developed by Hansch and Fujita, the molecular structure was represented with 2D descriptors such as molar refractivity, Hammet constant σ, hydrophobic constant, and with the number of hydrogen bond donors and acceptors (Hansch and Fujita 1964, Hansch et al 1995, Winkler 2002). However, the interaction between a ligand and a receptor involves three-dimensional distribution of properties, as discussed in the previous chapter. This has led to the development of 3D-QSAR techniques, which by far, are the most powerful computational tool in supporting the chemistry side of indirect drug design (Winkler 2002, Höltje et al 2003).

The primary aim of 3D-QSAR analysis is to establish a correlation between structurally and biologically characterized compounds with spatial fingerprints of different field properties of each compound, namely steric (shape and volume), electronic (electrostatic potential and electric charge) and lipophilic (how polar or non-polar the sections of the molecule are) properties (Winkler 2002).

In general, 3D-QSAR pharmacophore modeling combines the approach used in classical pharmacophore modeling with quantitative information gained through mathematical equations based on diverse sets of mathematical descriptors, encapsulating the key property of molecular structure (Vorpagel and Golender 2000, Martin 2002). The combination is obviously more informative than either of them alone since information on activity is not merely binary, as in the case of classical pharmacophore modeling. Additionally, the 3D nature of interaction between ligand and receptor is captured in a larger scale using essential features for activity. Based on the generated models, forecasts can be made of the potency of compounds not included in the derivation of preliminary

model. Hence, to return to the earlier chapter 1.2 in this research, the models can then be further used in predicting ADMET properties of drugs *in silico* as discussed. The potential of 3D-QSAR modeling to be applied in the early stages of drug discovery, makes it a powerful tool in preventing unwanted late stage attrition of drug candidates (de Graaf et al 2005a).

The advantage of 3D-QSAR methods lies also in its ability to cover wide ranges in molecular structures, as well as biological response or effect, producing as much as possible information regarding the active site and new lead compounds. This increases the statistical significance of the 3D-QSAR pharmacophore models. The predictive power of 3D-QSAR has been proven in several studies in which the biological activities of novel compounds have been correctly predicted. Modeling CYP binding affinity for ligands, the 3D-QSAR models are validated and constructed using several parameters, which directly or indirectly have an impact on ligand binding and affinity towards CYP enzymes, including K_m, K_i or IC_{50}. For further information concerning the applications discussed, the reader is referred to the original literature. Due to the applicability of 3D-QSAR techniques, several methods have emerged for generating the relationship between the structure of a molecule and its effect on activity. In most occasions, the method of choice seems to be CoMFA (comparative molecular field analysis), which will be discussed in more detail in chapter 2.2.2 and also in the experimental part of this text.

2.2.1 Molecular alignment

All 3D-QSAR models are limited by the good quality of measuments concerning them, which can be aggravated as "Garbage in, garbage out"- principle of QSAR models (Vorpagel and Golender 2000). The model can be only as good as the descriptors used, and the information the descriptors contain. In constructing 3D-QSAR models, as well as any kind of pharmacophore model, the alignment of molecules is the crucial step in constructing useful models (Lemmen and Lengauer 2000, Melani et al 2003). Molecular alignment produces energetically favorable conformations of substrates and/or inhibitors, and can be conducted manually or computationally using several algorithmic methods (Lemmen and Lengauer 2000, de Graaf et al 2005a).

Manually performed alignment is feasible, when the test set is not too large but to avoid biased procedures and to save time, computational methods are preferred (Tervo et al 2004, 2005). Manual alignment is conducted based on the decision, which functional groups or atoms to superimpose (Höltje et al 2003). In other words, it has to be determined, which of the structures are considered similar and crucial for activity. Selecting the structures to be aligned, can be conducted based on information of known, potent ligands and their structures.

In modeling classical pharmacophores, structurally quite similar compounds are emphasized using atom-by-atom superposition. However, Höltje et al (2003) have stated that using diverse molecular structures, and aligning the molecules based on functionally equivalent elements or molecular fields, exhibits more informative pharmacophore models accommodating better e.g. the versatility of CYP substrates. With 3D-QSAR methods, alignment is mostly conducted on the basis of similar molecular energy fields since by using this method, the 3D space of the molecules is taken into more thorough consideration (Leach 2001, Höltje et al 2003, Rönkkö et al 2006). After all, molecular fields model 3D structures in space surrounding atomic nuclei, which have an impact in non-covalent interactions most important for biological effects occurring at molecular levels (Rönkkö et al 2006). Using molecular energy fields as a criterion for molecular similarity instead of plain molecular structure, compounds with various molecular structures can be aligned (Kim et al 1998, Höltje et al 2003). As such, the alignment of molecules is possible if they represent similar molecular field properties. Also, alignment based of molecular fields leads to the possibility of several alignment alternatives (Tervo et al 2004, 2005). The intricate nature of aligning molecules by molecular fields has set the scene for much freely and commercially available software concerning pharmacophore and 3D-QSAR modeling (Table 3).

Despite the underlying method, the key building block in creating successful alignments in concerning both pharmacophore and 3D-QSAR models, is a good reference structure; *the template molecule* (Klebe et al 1999, de Graaf et al 2005a). Template molecule is usually an active ligand bearing certain prevailing qualities important to cause biological response, namely:

1) is large
2) relatively rigid to limit conformational freedom
3) contains essential groups
4) is specifically metabolized by the target protein
5) is regio- and/or stereoselectively metabolized

However, molecular alignment of ligands suffers from certain limitations affecting the results (de Groot 2006.) All ligands are assumed to bind to the receptor at the same binding site and and at the same (rigid) binding mode. This is not the case with CYPs, as it is widely accepted that CYPs allow certain degree of conformational flexibility within both ligands and in the active site, causing complex binding behavior (Locuson and Wahlstrom 2005). CYP3A4 even allows multiple ligands to bind simultaneously (Chohan et al 2006). Also, assuming the ligands to share a mutual low-energy conformation can be misleading, since the assumption leaves out for example, strained conformations in transient states of chemical reactions (Höltje et al 2003). By widening the energy range of ligands the number of accessible conformations increases dramatically, making it possible to consider more single conformations.

Nevertheless, it should be noticed that due to different alignment algorithms based on different concept, the quality and reproducibility of the results can vary to great extent, especially when crystal structures are not available for ligand-protein complexes presenting the proper binding mode of active ligands (Höltje et al 2003, Melani et al 2003). One alignment method can be useful and predictive in some contexts, and cause problems in another context (Kim et al 1998). Versatility concerning the quality of the results makes it difficult to estimate beforehand, which alignment method actually produces the desired information and thus, multiple alignments are needed (Rönkkö et al 2006).

Table 3: Examples of computational molecular modeling software for classical pharmacophore and 3D-QSAR modeling used in cytochrome P450 predictions.

Program	Description of the research and results	Reference
CoMFA	***CoMFA model for inhibitory molecules of CYP1A2 was generated, confirming the importance of hydrophobic interactions for inhibitory power. The model also pointed out hydrogen bonds and electrostatic fields in agreement with earlier studies. **CoMFA models were generated for inhibitory compounds of CYP2A5 and CYP2A6. Both models suggested negative charge at certain positions of a naphthalene ring to increase the inhibition potency of compounds. Additionally, steric substitution at position 2 of the naphthalene ring on the CoMFA model of CYP2A5 would increase the inhibition potency.**	***Korhonen et al 2005, **Rahnasto et al 2005,**
Almond	**Alignment independent GRIND descriptors (GRID-independent descriptors) were calculated for a set of rigid and flexible competitive CYP2C9 inhibitors and non-inhibitors with different molecular size. Using the obtained models inhibitors and non-inhibitors were distinguished. Pharmacophore model for competitive inhibitors gave predictions in 0.5 log units of actual values for all but one compound.**	**Afzelius et al 2002**
GRID	**Selectivity of CYP2C8, 2C9, 2C18, 2C19 and amino acids determining the specificity was analyzed using GRID calculations. As a result the most important determinants were hydrophobicity and geometry of active site. Based on homology models of the CYP2 enzymes, an inverse pharmacophore models was constructed, producing resembling interactions with a docked reference structure.**	**Riddeström et al 2001**
Catalyst	**A pharmacophore models was built for 38 substrates of CYP3A4. The model consisted of two hydrogen bond acceptors, one hydrogen donor and one hydrophobic region, giving fast fit function (r) value of 0.67. The model predicted K_m(apparent) values lower than 1 log unit for a test set of 12 molecules.**	**Ekins et al 1999c**

Table 3 continues

Program	Description of the research and results	Reference
MS-WHIM	**MS-WHIM model based on statistical analyses of molecular descriptors for size and shape of 16 CYP2B6 substrates was generated. The model gave q^2 value of 0.607 yielding satisfactory predictions for a test set of molecules.**	**Ekins et al 1999b**
MetaSite	***MetaSite was applied to a set of CYP3A4-mediated substrates for predicting their metabolic sites using protein crystallographic structure and homology model. Predicted active sites agreed with experimental results in 78% of the cases*.** ****Metabolic sites were also predicted for substrates of CYP1A2, 2C9, 2C19, 2D6 and 3A4 using crystal structures and homology models of the CYPs. The method predicted the site for metabolism successfully in an average of 85 % of the cases**.**	***Zhou et al 2006,** ****Cruciani et al 2005**

2.2.2 CoMFA

CoMFA is considered as a starting point for the development of 3D-QSAR techniques (Martin 2002, Höltje et al 2003). It was invented by Cramer and coworkers (1988), which were the first to be able to calculate the steric and electronic properties of chemical structures in a 3D grid surrounding the molecules, and correlating these properties with variation in the biological or chemical activity by means of statistical methods.

Steric potential in CoMFA is usually represented by Lennard-Jones function and electrostatic potential with simple Coulomb function (Cramer et al 1998, Clark et al 2000, Höltje et al 2003). The idea of the concept is that changes in the biological activity are closely related to equivalent changes in shapes and strengths of the non-covalent interaction fields surrounding the molecule. Relationships between the interaction fields and biological activity are calculated using PLS (partial least squares), which builds a statistical model from energy values predicting the activity.

CoMFA starts with selecting a set of compounds, series design, by aiming to gain as much as possible information with a minimum number of compounds (Kim et al 1998, Leach 2001, Höltje et al 2003). Basically, one has to minimize collinearity between predictor

properties, and maximize the variance between the properties. Activity of molecules should span about three log units of K_i or IC_{50}. Also, the substituent space with smallest amount of compounds has to be mapped, and the geometries and optimization has to be accounted for.

Bioactive conformation of each compound is suggested as in the beginning of classical pharmacophore modeling, and the compounds are superimposed onto each other (Kim et al 1998, Leach 2001, Höltje et al 2003). After superimposing, the interaction fields surrounding all the molecules are calculated. Interactions between a probe atom (usually sp^3 carbon with +1 charge) and the molecule at each intersection of a grid with a chosen grid spacing (1-2 Å) surrounding the molecule are measured. The relationships between the interaction fields and biological activity are calculated using PLS, which builds a statistical model from energy values predicting the activity. PLS models are also validated to ensure the quality and predictive power of the model. Validation of the model is usually conducted by Leave one out cross-validation-method. This procedure calculates as many models as there are compounds in the data set, and then leaves each compound out of the model and predicts its activity. Another possibility is to use standard cross-validation, which divides the compounds into a specified number of groups leaving each one out, one at a time. After repeating this once with every molecule or molecule group, Q^2 (cross-validated squared correlation coefficient) and SDEP (standard deviation of error of prediction) are calculated for observed and predicted activities of each molecule. Equations for determining Q^2 and SDEP can be found in the literature (Ekins et al 1999b, Höltje et al 2003). Notably, the same values can be obtained from many, slightly different equations.

It is possible to obtain standard error of prediction (SEP) and standard r^2 values, in addition to SDEP and Q^2. Standard r^2 (also called R^2) represents the quality of a simple linear regression equation being able to adopt values between 0.0 and 1.0. The higher the value, the more of the variations observed can be explained by variation in independent variables. Standard r^2 represent goodness of fit, whereas Q^2 (also called cross-validated r^2, q^2 or R^2_{cv} values) represents the goodness of prediction. Values greater than 0.5, decrease the possibility of chance correlation, and indicate that the model has some predictivity. The higher the value, the better the model.

Over-fitting of the data can be observed when the SDEP value increases, after first decreasing for the first few latent variables (components) reaching a minimum. Components can be seen as a group of properties correlating with each other. The number of components should be carefully chosen, in which "the 5 % rule" might be helpful. If adding a component decreases the SDEP by less than 5 %, the simpler model with fewer components is preferred. Or in other words, if the Q^2 increases by at least 5 % upon increasing the number of components by one, it might be justified to add an additional component. However, this is not always the case due to the many adjustable variables involved in CoMFA (Kim et al 1998).

CoMFA has been successfully used in many cytochrome P450 applications, some of which are represented in Table 3. Several other CoMFA models are available but for clarity, only a few of them are representend in the table. For example, Korhonen et al (2005) created a CoMFA model for inhibitory molecules of CYP1A2 to gain structural information about the interaction between the molecules and CYP1A2 active site. 46 molecules with measured IC_{50}-values were analyzed using CoMFA method. The prediction power of the model was validated by estimating the IC_{50} values for an external test set including six compounds tested during the study. As a result, the predicted IC_{50}-values of the test set differed from the actual values at most 0.67 log units giving r^2 value 0.87 and q^2 value 0.69 with SDEP of 0.76, with three components. Stability of the q^2 value was verified using PLS. The predicted IC_{50} values for the test set were approximately 0.05 log units of the actual values. As a conclusion, the produced CoMFA model can be considered highly predictive and statistically valid. The CoMFA model also confirmed the importance of hydrophobic interactions in binding to CYP1A2, as suggested in earlier studies, and pointed H-bonds and electrostatic field in agreement with other earlier studies. Graphical representation as a stereofigure of color contour maps can be found in the original article, and for further information on CoMFA models related to cytochrome P450s, the reader is referred to literature (Rao et al 2000, Ekins et al 2001, Poso et al 2001, Haji-Momenian et al 2003, Leonetti et al 2004, Rahnasto et al 2005, Korhonen et al 2007).

In addition to PLS, there are other statistical tools available to convert information into a functional 3D-QSAR pharmacophore model, ranging from simple linear methods to more complicated methods used for large data set difficult to interpret (Chohan et al 2006). These methods are discussed briefly in the next chapter.

2.2.3 Machine learning methods

Pharmacophore models generated using either classical or 3D-QSAR approach are mostly conducted using simple, linear statistical regression methods such as PLS (see previous chapter 2.2.2), MLR (multiple linear regression), PCA (principal component analysis) or PCR (principal component regression) (Chohan et al 2006). Supervised methods use information of the dependent variable to derive a predictive model, and to train the model using measured values of the property to be modeled e.g. LD_{25}, IC_{50} or K_i (Leach 2001). In other words, supervised methods generate a function or model based on training set, and predicts values for an external set using the created function or model. Furthermore, non-linear, more flexible methods are available, e.g. neural networks, recursive partitioning (RP) methods such as CART (classification and regression trees), *k*-nearest neighbor, DA (discriminant analysis) and SVM (support vector machine), naïve Bayesian classifier and logistic regression (Arimoto 2006, Chohan et al 2006). Methods using non-linear algorithms are applicable and superior to conventional linear regression, when the information available is relatively noisy, large at size and difficult to interpret as such.

This is often the case with models related to CYPs obtained from a number of HTS assays, and it creates a demand for machine learning methods, which are capable of handling the data obtained (Arimoto 2006). Machine learning methods are techniques containing many algorithms to discover patterns and rules, and based on them, can be used as *in silico* filters being fast to run, accurate and global at nature, and also producing predictive and statistically significant information on cytochrome P450 ligand activity and their ADMET properties useful in early-phase virtual screening and compound library design (Byvatov et al 2003).

When predictive QSAR models are generated, the most often used applications of machine learning methods are PLS-based classifiers (e.g. PLSDA) and different artificial neural networks (ANN) as well as SVM. SVM was originally introduced by Vaprik, as an alternative to ANN. From these methods, SVM and ANN produce more accurate predictions than PLSDA and overall, SVM has a higher capacity than ANN (Sorich et al 2003). The main advantage of SVM is its relatively low sensitivity to data overfitting, which is based on structural risk minimization. Risk minimization principle means minimizing both training set error and generalization error simultaneously (Yap et al

2006). For example, SVM has been used with auspicious results in determining the concentration causing 50% inhibition of the turnover of a specific CYP3A4 substrate (IC_{50}) with 1363 drug-like research compounds (Kriegl et al 2005). SVM was able to differentiate strong, medium and weak inhibitors in the test set with an accuracy of more than 70%. Yap and Chen (2005) used SVM with CYP3A4 and also with CYP2D6 and CYP2C9 in predicting substrates and non-substrates, as well as inhibitors and non-inhibitors of cytochrome P450 family, using 461 non-inhibitors and 334 non-substrates for CYP3A4, 522 non-inhibitors and 504 non-substrates for CYP2D6, and 535 non-inhibitors and 558 non-substrates for CYP2C9. As can be seen from the size of the compound database, machine learning methods can be used for extremely large data-sets compared to PLS, which usually consist of database of approximately 30 compounds (Leach 2001). The research of Yap and Chen (2005) proved also the predictive power of SVM, which can be even higher, when combined with other algorithms e.g., genetic algorithm (GA) in selecting the best descriptors for the models.

Interestingly, in most researches, which have been focusing on cytochrome P450 predictions based on machine learning methods, supervised techniques such as SVM have been utilized (Korolev et al 2003). However, also unsupervised methods, which do not have *a priori* output, are available. Unsupervised methods require less computer capacity, and their output is more flexible. Kohonen self-organizing map (SOM) or Kohonen neural networks (KNN) (Kohonen 2000) is an unsupervised machine learning method, which produces distance preserving 2D topological maps based on high-dimensional input information (Hasegawa et al 2002, Balakin et al 2006). The map can represent e.g. substrates of cytochrome P450s (vectorial samples) distributed based on information of their 3D-QSAR. Substrates with similarity in their molecular descriptors are located close to each other, with respect to their biological activity. Following from the nature of the output, the map can be used in visualization and/or classification of multivariate data. Korolev et al (2003) utilized SOM in determining the probability of a novel substrate to be metabolized via cytochromes. The database consisted of 2200 compounds representing substrates, non-substrates and metabolites of cytochrome P450 mediated reactions for extensive 38 human cytochromes. The input in their research can truly be considered as multivariate data. Nevertheless, SOM was still able to classify correctly 76.7 % of the substrates and 62.7 % of the metabolic products, with 12.6 % and 22.4 % of these falling into areas with no specific classification of being either substrates or metabolites,

respectively. However, the algorithm generated by Korolev et al could be successfully used in determining the susceptibility of a novel substrate to be metabolized by the cytochrome P450s in the early stages of drug discovery process. Also, it could be possible to predict the exact isoform of the substrate by using the SOM.

SOM has also been utilized in classifying compounds based on their binding affinity towards cytochrome P450s (Balakin et al 2004). A dataset of 491 compounds with apparent K_m values for human cytochrome P450s were analyzed with success. The SOM was able to correctly classify 97 % of CYP3A4 substrates with low K_m (<10 µM) and 91 % of compounds with high K_m (>100 µM). Classification was conducted also for CYP2C9, 2B6 and 2D6 but the research concluded more data being required, before statistically valid discrimination could be made between compounds with high and low K_m.

A short summation of machine learning methods used in cytochrome P450 predictions can be found in Table 4. Represented in the table are the overall accuracies for classification of inhibitors/non-inhibitors and substrates/non-substrates. In addition, machine learning methods have been used in predicting affinity and relevant binding conformations concerning cytochrome P450s (Bazeley et al 2006). Nevertheless, machine learning methods, too have limitations when the data set is considered unbalanced, e.g. if two classes, namely inhibitors and non-inhibitors of cytochrome P450s, differ much when considering the size of the groups (Eitrich et al 2007). To overcome the problem of biased results with large, unbalanced data, conventional machine learning methods can be further combined with additional algorithms for classification. These consensus models have been proven sensitive and accurate in recognizing cytochrome P450 inhibition potency of compounds (Chohan et al 2005, O'Brien and de Groot 2005, Yap and Chen 2005, Yap et al 2006).

Table 4: Machine learning methods used in predicting cytochrome P450 inhibition potency

Isoform/Property	Method	Overall accuracy	Reference
CYP3A4 (inhibition)	**ANN**	**90 %**	**Molnár and Keserü 2002**
38 human CYPs (substrate/metabolite)	**KNN**	**76.7 % / 62.7 %**	**Korolev et al 2003**
CYP3A4 (high K_m/low K_m)	**KNN**	**91 % / 97 %**	**Balakin et al 2004**
CYP1A2 (inhibition)	**Consensus (PLS, MLR, CART, BNN)**	**83 %**	**Chohan et al 2005**
CYP3A4 (inhibition / substrate)	**Consensus SVM**	**96 % / 95 %**	**Yap and Chen 2005**
CYP2D6 (inhibition / substrate)	**Consensus SVM**	**94 % / 95 %**	**Yap and Chen 2005**
CYP2C9 (inhibition / substrate)	**Consensus SVM**	**95 % / 97 %**	**Yap and Chen 2005**
CYP2D6 (affinity / binding conformations)	**ANN**	**85 ± 6 %**	**Bazeley et al 2006**
CYP3A4 (inhibition)	**SVM / LR / LDA / PLS / C4.5 DT, *k*NN, PNN**	**94.6 /87.8 /48.6/90.5/ 87.8/98.6/87.8 %**	**Yap et al 2006**

2.3 Quantum mechanical models

Quantum mechanics (QM) describe the behaviour of matter and energy on the scale of atoms and sub-atomic particles; neutrons, protons and electrons. In quantum chemistry, the focus is in the electronic behaviour of atoms and molecules since electrons are greatly responsible for the chemical properties of atoms and molecules. In other words, it is possible to derive properties of molecular systems based on their electronic distribution (Groundwater and Taylor 1997, Leach 2001). However, solving the electronic behaviour is not a trivial task when considering complex systems as cytochrome P450s. The foundation for determining electronic behaviour of the target system is Schrödinger equation (Eq 2).

Schrödinger equation predicts the quantum mechanical energy levels, when given a collection of the spatial positions of atoms (wavefunctions) and the total number of electrons in the system (Dill and Bromberg 2003, Friesner 2005). In addition to obtaining information concerning energy of the system, Schrödinger equation provides facts about other related properties such as electron density. Although the equation in a compact form, as represented in Equation 2 might seem simple and straightforward, a detailed description of quantum chemical principles underlying Schrödinger equation is beyond this text and the reader is referred to literature (Leach 2001, Dill and Bromberg 2003).

$$H\psi = E_j\psi \qquad \text{Eq.2}$$

H: Hamiltonian operator, E_j: constant with eigenvalue and quantum number, ψ: wavefunction

Simulating chemical reactions in which bonds are breaking and forming, requires using of QM methods since classical force fields are not capable of simulating such processes (Leach 2001). With CYP enzymes several studies have been conducted, using different QM methods concerning activation barriers for substrates and heme-oxygen species during the catalytic cycle, to predict regionally susceptible metabolic sites for the substrates (de Graaf et al 2005a). A more detailed description of the QM methods used in CYP applications is presented in chapter 2.3.1.

QM models are based on quantum theories, which can be approached with different level of methods. Quantum theories include for example molecular orbital theory, Hückel theory, valence bond theory and density functional theory (DFT) (Leach 2001). From the theories, molecular orbital theory is the one most widely used. The methods underlying quantum mechanics include e.g., *ab initio* and semiempirical methods and their subcategories. Also, ab *initio* calculations can be conducted using either wavefunction-based approach or DFT (Friesner 2005). Semiempirical calculations can be conducted using different methods such as PM3 (Parameterized Model 3) and AM1 (Austin Model 1), which apply the same theoretical assumptions, but differ in the approximations being made (Höltje et al 2003). The difference between *ab initio* and semi-empirical models is their way of reproducing experimental data, which *ab initio* does without using empirical parameters, contrary to semi-empirical methods. As such, *ab initio* is useful in situations,

where little - if any - information of the system of interest is available, since *ab initio* calculations are extremely expensive computationally (Höltje et al 2003, Friesner 2005). Semiempirical calculations can be considered filling a gap between *ab initio* methods and molecular mechanics (MM), which does not explicitly include electrons and nuclei of atoms in the calculations, and can be considered as a more simple method. Nevertheless, both *ab initio* and semiempirical methods are widely used in cytochrome P450 predictions as well as mixed QM/MM.

When using QM/MM technology, a small part of the target system is treated with QM methods and the rest of system with MM methods. In most occasions, the core region treated with QM is embedded in the MM region of mixed QM/MM system (Figure 4). For example, in the case of cytochrome P450s, the heme moiety and its functions can be considered the core of the cytochrome chemistry, and are treated with accurate *ab initio* calculations. All the rest of the protein is treated with MM. QM/MM technology has been used e.g. by Guallar and coworkers (2002) in their research concerning the conversion of C-H bond into alcohol (hydroxylation) by cytochrome P450s. Using the mixed technology, they were able to compute relative energies of intermediates and transition states in cytochrome P450cam reaction cycle. The advantage of using such a mixed technology is being able to realistically model complex molecular systems with reasonable computer capacity compared to the whole enzyme being handled with exhaustive QM calculations.

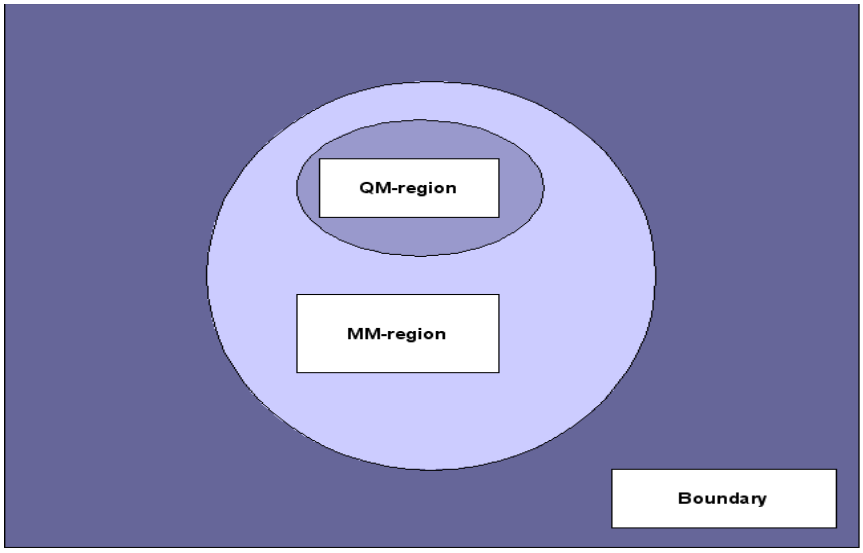


Figure 4: A schematic arrangement of QM/MM technique (Adapted from Höltje et al 2003). The small core region of QM is embedded in the MM region of the mixed system representing the areas of the target system where QM and MM methods are mostly applied.

2.3.1 Quantum chemistry in cytochrome P450 predictions

Exceptionally large amount of experimental and theoretical work has been invested in understanding the complex reaction cycle of cytochrome P450s (Figure 2), and the effects of the cycle on specific ligands, and their metabolism at an atomic-level (Guallar et al 2003). The cytochrome P450 core chemistry can be considered as a large and complex field and following from this, accurate atomic-level calculations represent a severe challenge. Also, there are different opinions concerning the activation of O_2 and the number of electrophilic oxidants participating in the catalytic reaction cycle (Shaik et al 2004). As such, a complete description of the various steps in the cycle at atomic-level has yet to be produced. However, a detailed understanding of the mechanisms and underlying energetics of cytochrome P450 reactions would provide the means to build a quantitative model for predicting metabolism of pharmaceuticals and xenobiotics during the early stages of drug development (Bathelt et al 2004). The models could also be used in describing the absolute rate constants together with relative reactivity and energetics of enzymatic reactions, thus qualitatively enhancing understanding the chemistry underlying cytochrome P450s (Higgins et al 2001, Guallar et al 2002). As such, quantum chemistry calculations have been used in several occasions in order to help understanding the cytochrome P450 reaction cycle in detail. Hydroxylation of alkanes and epoxidation of

alkenes have been the targets of the most extensive research efforts (de Visser et al 2001, Guallar et al 2002, Bathelt et al 2004). The energetics of the metabolism mediated by CYPs have been studied via hydrogen abstraction energy; the energy, which is required to withdraw a hydrogen atom from a cytochrome P450 substrate.

For example, Higgins et al determined the hydrogen abstraction energy with respect to electronic and steric features affecting regioselectivity using several cytochrome P450 isoforms (1A2, 2B1, 2B6, 2C8, 2C9, 2E1, 4B1). Semiempirical methods with AM1 and DFT calculations provided a quantitative description of relative energetics of oxidation at different positions in a set of substrates. Based on the relative energetics, it is possible to predict a catalytic site and metabolic pathway for substrates of different CYP isoforms.

Singh et al (2003) developed a semiquantitative model for evaluating also the hydrogen abstraction energy for different sites in a drug molecule. In their study, 50 CYP3A4 substrates were used in order to help decide the possible catalytic sites for CYP3A4-mediated metabolism. The practical calculations were conducted using AM1 semiempirical molecular orbital calculations with trend vector analysis. Trend vector analysis is similar to QSAR PLS analysis. The results were compared to experimentally known major metabolic sites of CYP3A4 substrates, giving successful statistical values. The generated model can be used together with binding affinity information to predict possible substrates for CYP enzymes suggesting also their likely metabolic sites.

Olsen et al (2006) conducted a similar kind of research, which determined the activation energy for hydrogen abstraction energies for 24 structurally diverse substrates with features such as primary, secondary and tertiary aliphatic carbon atoms, nitrogen, oxygen, sulphur and sp^2 hybridized or aromatic carbon atoms next to a putative reactive atom, for which the hydrogen abstraction activation energy were calculated using DFT calculations and semiempirical methods. As a result, the highest activation barriers were obtained for sp^3 hybridized carbon atoms, and lowest activation barriers were obtained for amine derivatives indicating the relative ease of the amines to be metabolized via cytochromes. The correlation between activation energy and molecular descriptors provided a model, which could be used as a filter for potential CYP substrates in early stages of drug discovery, providing also information of the putative catalytic site of the substrates. However, DFT calculations proved to be computationally extremely demanding, and took

CPU (central processing unit) time of several weeks for each substrate. As such, DFT calculations can not be considered to be used routinely as tools in cytochrome P450 predictions. DFT-based filters for virtual screening of large databases would be computationally exhaustive, and CPU time required for screening all the molecules in the database is beyond imagination. With simpler semiempirical methods (e.g. AM1) excellent statistical values were still gained with the mean absolute errors being only 3-4 kJ/mol, when compared to full DFT calculations. Following from this, simpler quantum mechanical methods, such as semiempirical calculations would be more suitable for e.g. virtual screening of new, potent CYP substrates. The results of Olsen and coworkers are consistent with full DFT calculations being mostly reserved for special problems, such as complete description of the iron-porphyrin system of heme.

A study conducted by Oda et al (2005) took a different approach focusing on force field parameters of the heme iron. The oxidation and spin state of the iron affect CYP substrate binding, so it is important to evaluate the spin states when performing molecular docking of substrates. Extensive quantum chemical chemical calculations were performed to obtain parameters describing interactions between cytochrome P450s and small ligands to be further explored in MM methods and classical molecular dynamics (MD) simulations. The research was conducted using both *ab initio* and semiempirical unrestricted Hartree-Fock-methods. Contrary to the results obtained by Olsen et al (2006), their study concluded that DFT calculations gave seemingly more accurate results than semiempirical AM1 method which was not suitable for the iron of heme moiety, probably since AM1 does not include parameters for transition metals.

As a conclusion, quantum chemical calculations can predict the relative reactivity for a substrate and cytochrome P450 enzyme but they cannot predict binding affinity or orientation of the potential substrates, which as a combination, can be considered the ultimate goal in computational modeling (Higgins et al 2001). As such, QM models have been widely used together with methods such as molecular docking and molecular dynamics simulations, which are further discussed in the following chapters. The cytochrome P450 predictions with QM models described here cover only a fraction of the overall researches conducted with this area, and the number of similar researches is rapidly growing with advances in theory, software and computational hardware, allowing creating

more accurate and efficient models concerning the processes governed by the cytochrome P450 superfamily (Friesner 2005).

3 PROTEIN-BASED MODELS

Protein-based models derive information from the 3D structures of cytochrome P450 enzymes. Main sources for protein-based models are X-ray crystallographic structures and homology models, which have been generated using the crystal structures as templates. Both methods provide detailed information of the binding cavity of cytochromes, and interactions between the active site residues and CYP substrates and inhibitors. To some extent, protein-based models can be considered complementary to ligand-based models.

3.1 Crystallographic structures

3D structures of proteins can be generated by either X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy in solution. Since crystal structures of proteins as such are slightly too rigid, and NMR models are too flexible, MD simulations are performed with crystal structures, to represent the situation resembling most the actual situation. Notably, cytochromes P450s are membrane-bound proteins and to prevent the aggregation of the protein while being crystallized, the membrane anchor sequence of the protein has to be removed (Cosme and Johnson 2000, Williams et al 2000).

When a crystallographic structure of the target protein is available, the structures can be successfully used in identifying common features for a group of proteins. With cytochromes, the crystal structures have provided the means to compare the overall folding of CYPs, and to identify structurally conserved regions. Crystallographic structures have also provided the means to explore the access and exit channels, active site (volume) and flexible regions of the enzymes (Otyepka et al 2007). Following from this, the crystallographic structures provide opportunities to perform molecular dynamics simulations, which can be used in exploring the dynamic behavior of access and exit paths in the protein, as well as hydration of the active site during ligand-binding (de Graaf et al 2005a). In addition, the crystal structures provide data complementary to experimental findings, and they can be used together with information gained from ligand-based models, allowing detailed understanding of the ligand-protein interactions.

At present, only a few crystallographic structures of mammalian cytochromes are available including CYP2A6, 2B4, 2C5, 2C8, 2C9, 2D6 and 3A4 (Otyepka et al 2007). The available mammalian cytochromes and their PDB ID-codes are represented in Table 5. When crystallographic structures are not available, alternative techniques are applied to achieve information concerning the 3D structure of the cytochrome P450s. Comparative modeling is a method, which is used in generating cytochrome homology models in the absence of crystallized structures. Comparative modeling technique is discussed in more detail in chapter 3.2. In the following, crystal structures available for mammalian cytochromes are discussed. Crystal structures are available also for bacterial CYPs, but the focus of this research lies in mammalian CYPs. Nevertheless, the role of bacterial CYPs cannot be underestimated, since crystal structures of these cytochromes have provided the basis for understanding the structure and properties of mammalian cytochromes, and the bacterial cytochromes P450cam (CYP101) and P450BM-3 (CYP102) have been used in several occasions as templates for homology protein models of CYP2A, 2B, 2C, 2D 2E, 3A and 4A (Lewis 2002a, b). A detailed overview concerning the bacterial CYPs can be found in literature (Wade et al 2004, Cojocaru et al 2007).

Table 5: Published mammalian cytochrome P450 crystal structures

CYP	Species	PDB ID	Ligand	Reference
2A6	**Human**	**1Z10**	**Coumarin**	**Yano et al 2005**
2A6	**Human**	**1Z11**	**Methoxsalen**	**Yano et al 2005**
2B4	**Rabbit**	**1PO5**	**-**	**Scott et al 2004**
2B4	**Rabbit**	**1SUO**	**4-(4-chlorophenyl)-imidazole**	**Scott et al 1004, Hernandez et al 2006**
2B4	**Rabbit**	**2BDM**	**Bifonazole**	**Zhao et al 2006, Hernandez et al 2006**
2C5	**Rabbit**	**1DT6**	**-**	**Williams et al 2000**
2C5	**Rabbit**	**1N6B**	**Dimethylsulfaphenazole**	**Wester et al 2003a**
2C5	**Rabbit**	**1N6R**	**Diclofenac**	**Wester et al 2003b**
2C8	**Human**	**1PQ2**	**Palmitic acid**	**Schoch et al 2004**
2C9	**Human**	**1OG2**	**-**	**Williams et al 2003**
2C9	**Human**	**1OG5**	**S-warfarin**	**Williams et al 2003**
2C9	**Human**	**1R9O**	**Flurbiprofen**	**Wester et al 2003c**
2D6	**Human**	**2F9Q**	**-**	**Rowland et al 2006**
3A4	**Human**	**1TQN**	**-**	**Yano et al 2004**
3A4	**Human**	**1W0E**	**-**	**Williams et al 2004**
3A4	**Human**	**1W0F**	**Progesterone**	**Williams et al 2004**
3A4	**Human**	**1W0G**	**Metyrapone**	**Williams et al 2004**

3.1.1 CYP2A6

CYP2A6 is an isoform, which metabolizes approximately 4 % of drugs metabolized by cytochrome P450s (Lin and Lu 2002). Despite its relatively small contribution to metabolism, CYP2A6 participates in the detoxification of nicotine, and in activation of tobacco-specific procarcinogens to mutagenic products (Kim et al 2005, Yano et al 2005). At the moment, two X-ray structures of the isoform are available with bound coumarin and methoxsalen. Exploring the CYP2A6 structure has revealed the isoform sharing a common fold with other cytochromes, but the active site is seemingly smaller; 25 % smaller volume than e.g. CYP2C8, 2C9 and 3A4. The active site is hydrophobic and contains one hydrogen-bond donor. This is consistent with the structures of known CYP2A6 substrates, which are usually small and planar, hydrophobic molecules. However, the characteristics of the active site indicate the conformation of peptide backbone being different when compared to other cytochromes, thus complicating the comparison of ligand-protein interactions with other cytochromes. The X-ray structure of 2A6 is being further studied using random mutagenesis and fluorometric high-throughput screening, in order to study the structure-activity relationships with CYP2A6, which will hopefully reveal more of the structure of the isoform (Kim et al 2005).

3.1.2 CYP2B4

For CYP2B4, three different crystal structures are available, which include a high-resolution ligand-free structure with resolution of 1.6 Å, and two structures with bound substrates, namely bifonazole (BIF) and 4-(4-chlorophenyl)-imidazole (4-CPI) (Hernandez et al 2006). The crystal structures of the isoform have revealed large conformational changes in flexible regions, which are able to change the active site properties dramatically (Zhao and Halpert 2007). The BIF-bound crystallographic structure also pinpointed a remarkable ligand-induced flexibility in the structure compared to the ligand-free crystal structure, which is already in an open conformation contrary to the complex with 4-CPI, which is considered to be in a closed conformation. Based on this information, a site-directed mutagenesis study was conducted based on the crystal structure to determine important amino acids for ligand binding (Hernandez et al 2006). Six residues interacting with either or both of the bound ligands were replaced with smaller and/or bulkier residues,

showing that only a mutant I363A perturbed significantly the enzyme inhibition and catalysis. Further studying of the conformational flexibility was conducted using imidazole inhibitors, but the research concluded more experimental data being needed before final conclusions could be made concerning the residues important to substrate binding (Muralidhara et al 2006).

3.1.3 CYP2C5

CYP2C5, which catalyzes the hydroxylation of progesterone, was the first crystal structure of the mammalian cytochromes (Johnson et al 2002). The crystal structure showed the active site clearly diverging from the most closely related microbial enzyme P450 BM3 (CYP101; *Bacillus megaterium*), and pinpointed the possibility of induced fit or movement of the protein around substrate-binding pocket (Williams et al 2000). The functionally important residues affecting substrate specificity and regioselectivity (SRSs) were found to be conserved in the proximal face of the isoform. At the moment the crystal structure of CYP2C5 is available bound with anti-inflammatory drug diclofenac at a resolution of 2.1 Å (Wester et al 2003a), and complexed with sulfonamide DMZ at a resolution of 2.3 Å (Wester et al 2003b). Also a free structure of CYP2C5 is available (Williams et al 2000). CYP2C5 crystal structure has been used as a template structure for several cytochrome homology models (Lewis 2002, Kumar et al 2003, Stahl and Höltje 2005).

3.1.4 CYP2C8

The crystal structure of human CYP2C8 was determined at a resolution of 2.7 Å (Schoch et al 2004). CYP2C8 is the primary enzyme metabolizing cerivastatine, and the isoform participates in the metabolism of natural substrates e.g. unsaturated fatty acids and retinoic acids (Marill et al 2000). By exploring the crystal structure, the active site of the isoform was determined to be larger than the active site of CYP2C5, due to bulky residues surrounding the active site (Schoch et al 2004). The observation is consistent with the ability of CYP2C8 to metabolize relatively large molecules, which are poor substrates for other CYP2C family enzymes. Also other structural differences occur when the crystal structure was compared to previously crystallized isoforms of the same subfamily, mostly concentrated in the F/G-loop region and other regions at close proximity to the active site.

3.1.5 CYP2C9

For the cytochrome 2C9, three x-ray crystallographic structures are available (Otyepka et al 2007). CYP2C9 is an isoform, which can cause severe adverse effects with drugs having low therapeutic margin e.g. warfarin and phenytoin, due to its polymorphism (Wester et al 2004). The crystal structure of CYP2C9 has been generated complexed with flurbiprofen and S-warfarin (Williams et al 2003, Wester et al 2003c) and also, in a ligand-free conformation. The crystal structures revealed the catalytic site of the isoform to differ significantly in conformation when compared to other CYP2C family enzymes (2C5, 2C8), and the ligand-free crystal structure differed from the crystal structures complexed with the substrates. Most of the differences were observed in the N-terminal parts, B/C-loops, F helices and F/G-loops of the cytochromes, and may be reflecting the conformational flexibility of CYP2C9.

3.1.6 CYP2D6

CYP2D6 is a widely studied isoform of the cytochrome P450 superfamily, metabolizing approximately 20 % of the known drugs, particularly a variety of central nervous system and cardiovascular drugs (Rowland et al 2006). Many genetic polymorphisms have been described for CYP2D6, which is one reason for the myriad number of researches concerning the isoform. The crystal structure of CYP2D6 has been solved with a resolution of 3.0 Å, possessing a fold characteristic to other cytochromes. The active site cavity is well defined, containing residues considered important for substrate recognition and binding, namely Asp-301, Glu-216, Phe-483 and Phe-120. Although the F-G-loop of CYP2D6 could not be modeled entirely, the crystal structure was able to explain much of the reported site-directed mutagenesis data (Mackman et al 1996, Kirton et al 2002, de Graaf et al 2006).

3.1.7 CYP3A4

For CYP3A4, which is the cytochrome responsible for the metabolism of most of the known pharmaceutical agents, several crystal structures are available. In addition to the first crystal structure of the isoform with 2.05 Å resolution (Yano et al 2004), another

ligand-free and two structures complexed with an inhibitor metyrapone and a substrate progesterone, have been generated (Williams et al 2004). The most distinct difference between the crystal structures is the size of the active site. Yano et al (2004) describe the substrate binding cavity relatively large contrary to Williams et al (2004), who claims the active site being small with little conformational changes associated with binding of a substrate or inhibitor. Of course, the size of the active site can be considered in a somewhat subjective way, but the former description of the active site is consistent with models, were two or more molecules occupy the active site at the same time. In a more recent study, with regard to the structural basis for CYP3A4 in complex with large ligands, crystal structures of CYP3A4 with bound ketoconazol and erythromycin have been generated (Ekroos and Sjögren 2006). The results from the study show the protein undergoing dramatic conformational changes with ligand-binding and allowing several binding modes for substrates. The results reveal the extreme flexibility of CYP3A4 upon ligand binding, and represent severe challenges for computational modeling of the isoform. These results are congruent with the known ability of CYP3A4 to bind structurally diverse substrates.

3.1.8 Cytochrome P450 crystal structures: summation

Based on the crystal structures of cytochromes, it has been possible to make conclusions regarding the overall structural characteristics of the cytochromes. The cytochromes consist mainly of four β-sheets and approximately 13 α-helices with an insertion called J' (Graham and Peterson 1997). The active site, which locates deep inside the cytochromes, is generated by a four-helix bundle composed of helices D, E, I and L together with helices J and K. Between the helices I and L lies the heme template, which is linked to the protein structure via sulphur atom of cysteine residue near the heme. The cysteine is a fifth ligand of the heme iron. The sixth coordination position is occupied by either water, oxygen, or the substrate depending on the location in the reaction. Characteristic to the cytochromes seems to be the F/G segment, which is perpendicular to structurally conserved helix I (Otyepka et al 2007). F and G helices together with the F/G-loop and B/C-loop control the cytochrome active site. The general structure with heme-binding structural core of the cytochromes can be considered relatively conserved. Still, variable regions are found when comparing the crystal structures, the differences focusing around the deeply buried active site and F/G-loop, respectively. The variable regions are mostly associated with substrate

recognition, substrate binding and redox partner binding, which presents a challenge for deriving overall quality model concerning the cytochromes but explain the substrate specificity of the cytochromes.

In addition for gaining information of the similar and variable regions cytochromes and their subfamilies, the crystal structures have been found useful in studying cytochrome access channels, which are considered as pathways connecting the active site with protein surface (Cojocaru et al 2007). The active site is located deep inside the cytochromes, isolated from the bulk solvent and the cytochrome substrates, which are predominantly hydrophobic molecules, have to make their way inside the buried active site (Graham and Petersen 1999, Wade et al 2004). By exploring the crystal structures of both bacterial and mammalian CYPs, the proteins are considered to go through dynamic motions resulting in opening and closing these channels when the substrates are entering the active site, or metabolized products are leaving the reaction site. The dynamic motions are considered to have an effect on CYP substrate specificity as well as enzyme kinetics, and they are also likely to participate in transporting molecular oxygen and water molecules in the active site, which is crucial for cytochrome activity to occur (Cojocaru et al 2007). However, the research concerning the complex nature of the substrate access channels has not yet resulted in overall description of the nature of these pathways.

3.2 Homology models

Comparative modeling is a technique, which produces models of protein structures, namely *homology models,* based on information gained from a crystal structure of a known protein. Secondary and tertiary structure of an unknown cytochrome can be predicted based on an amino acid sequence of a crystallized protein, since proteins with high degree of similarity adopt similar 3D conformation.

Homology models provide valuable information of cytochromes and their subfamilies when experimentally determined structures of target proteins are missing (Kirton et al 2002). This is often the case when considering cytochromes, since only a few crystallographic structures of the cytochromes are available as discussed in previous chapters.

3.2.1 Comparative modeling technique

Comparative modeling as such is an iterative process, which has been represented in Figure 5. The method starts with a determined amino acid sequence of a target protein belonging to a particular protein family (Höltje et al 2003). The next stage in the process is identification of template structures, by scanning the amino acid sequence of target protein against sequences with a known crystal structure e.g. in the Brookhaven Protein Data Bank (Kirton et al 2002, Lewis et al 2002, Höltje et al 2003). Based on the sequence similarity, 3D structure of the target protein can be deduced, since it has been observed that similar amino acid sequences tend to adopt similar 3D conformation (Chothia and Lesk 1986). Consequently, high similarity between target and template sequences produces better alignments and consequently, more accurate homology models (Martin et al 1997). However, it has also been suggested that multiple templates even with relatively low sequence similarity, can produce homology models of good quality since a better description of the overall structure of the target protein is available using multiple templates (Kirton et al 2002).

After finding suitable template structures, multiple sequences are aligned with respect to each other, including the sequence of the target protein to locate structurally conserved regions (SCRs) (Kirton et al 2002). SCRs are known to be relatively similar across the cytochrome P450 superfamily, and can be considered straightforward to locate and model by aligning either the amino acid sequencies or the 3D crystal structure of template with the sequence of the target protein, and then placing the regions in 3D space (Höltje et al 2003, Mestres 2005). With cytochromes, SCRs mostly comprise a set of α-helices (E, J, K, K′, K′′, L, part of J) and β-sheets (6-1, 1-3, 3-2), as well as the active site of the protein together with the Cys-pocket (Kirton et al 2002, Mestres 2005). The alignment of templates and target is validated and refined at every step of process, using e.g., mutational studies and isoform specificity data to verify the location of SCRs, with respect to experimental data concerning the substrate recognition sites (SRSs) (Gotoh 1992, de Graaf et al 2005).

Additionally, structurally variable regions (SVRs) and side chain conformations are identified based on alignment of the templates and the target. SVRs usually locate in the membrane bound termini, B'-helices and F/G-loop regions of cytochromes, and modeling

of SVRs with reasonable accuracy is considered a demanding task (Kirton et al 2002, Höltje et al 2003, Mestres 2005). SVRs may contain deletions and insertions in their amino acid sequence and may in general possess little - if any - sequence similarity. However, they can modeled by examining homologous protein models, an building a corresponding segments based on information of the known structure, or using other methods such as loop search method and *de novo* generation technique (Höltje et al 2003).
Side chain conformations can also be determined by examining the alignment of templates and target structure or structures of homologous proteins, and transferring the coordinates from highly similar template structures to the modeled target structure (de Graaf et al 2005a). However, since side chains can adopt versatile conformations and degree of freedom, on some occasions modeling of side chains can be a complex task (Höltje et al 2003). After adding missing side chain atoms, and generating conformations for the residues and possible loops, the structure of the homology model is optimized and validated using MD simulations, energy minimization and simulated annealing (Bathelt et al 2002, Höltje et al 2003, de Graaf et al 2005a).

Finally, the generated homology model is again validated to verify the quality of the model (Kirton et al 2002, Höltje et al 2003). The generated homology model can be validated by comparing the root mean square deviation (RMSD) of the protein backbone and the conformations of the sidechains, with the template structure possessing highest sequence similarity or with an available, high-resolution crystal structure from the same enzyme family (Chothia and Lesk 1986, Martin 1997). However, the validation should mostly be conducted based on experimental data such as mutagenesis studies and predictions of the cytochrome P450 substrate/inhibitor binding and the site of metabolism with molecular docking and MD simulations (de Graaf et al 2005a).

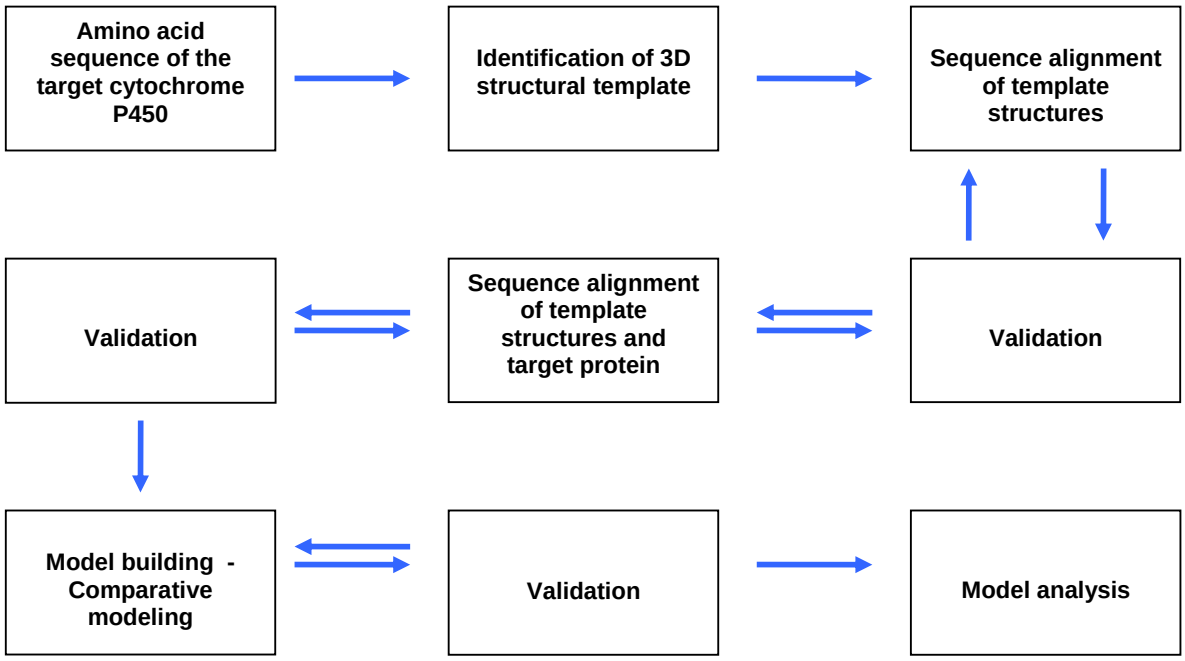


Figure 5: A flow chart representing the iterative process of comparative modeling. Adapted and compiled from Kirton et al 2002

3.2.2 Comparative modeling of cytochrome P450 structures

Comparative modeling has been used in generating several homology models, which have produced further information concerning the cytochrome P450 structure and overall characteristics. Some of the available mammalian cytochrome P450 homology models are represented in Table 6. For example, based on the crystal structure of CYP2C5, homology models have been generated for several other cytochromes (CYP1A2, 2A6, 2B6, 2C9, 2C8, 2C19, 2D6, 2E1 and 3A4) (Lewis 2002). The homology model of CYP2B6 is discussed in more detail in a separate chapter since it is the target of the experimental part of this work. The homology models provide detailed information concerning the putative active sites and important residues of cytochromes, and as such make it possible to design novel substrates and inhibitors for specific enzymes. Based on the homology models, it has also been possible to determine closely related cytochromes allowing identifying characteristic features of the mammalian cytochromes.

Homology modeling has also been shown to contribute in understanding towards another member of the cytochrome P450 family, CYP11A1 (Storbeck et al 2007). CYP11A1

participates in steroidogenesis by catalyzing the conversion of cholesterol to pregnenolone and following from this, the isoform is of physiological importance. However, CYP11A1 has not been crystallized, and only a limited amount of structural information has been available. Using bacterial cytochromes, P450BM-3 and P450terp (CYP102 and CYP108) and the crystal structure of CYP2B4 as templates, a homology model of the protein has been generated (Muller et al 1998). The homology model of CYP11A1 has been successfully used in studying the mitochondrial electron transfer system of steroidogenesis in more detail, and it has also provided insight concerning the active site and the access and exit channels of the isoform, as well as its interactions with mitochondrial membrane. All this information was far beyond reach before a homology model of CYP11A1 was generated (Storbeck et al 2007).

Notably, for the cytochromes many homology models have been generated based on the latest crystal structures. For example, after determining the crystal structure of CYP2C9, a new homology model was constructed for CYP2C19, in order to study the structure-activity relationships between the isoform and its ligands (Wang et al 2007). As a result, the binding pockets of CYP2C9 were explicitly defined together with key features of potent inhibitors and residues important to catalytic activity. As such, the information can be exploited in designing personalized drugs as well as predicting the ADMET properties for novel substrates.

The quality of the homology models is directly dependent of the quality of available crystal structures. Despite this, the homology models cannot be considered merely as secondary models since they produce detailed information of the active site characteristics, ligand-protein interactions and the overall 3D structure of the target protein, in the absence of a crystal structure. Comparative modeling provides a powerful tool when combined with other computational approaches (ligand-based models, molecular docking, molecular dynamics simulations) to predict cytochrome P450 substrate selectivity and isoform specificity, as well as possible metabolic pathways and sites for metabolism, to further improve the development and discovery of new potent substrates.

Table 6: Homology models of cytochrome P450s generated using comparative modeling technique

CYP	Template	Reference
1A1	**2C5**	**Szklarz and Paulsen 2002**
1A2	**2C5**	**Kim and Guengerich 2004**
2A6	**2C5**	**He et al 2004**
2A13	**2C5**	**He et al 2004**
2B6	**2C5**	**Bathelt et al 2002, Wang and Halpert 2002**
2C18	**2C9**	**Payne et al 1999a, Oda et al 2004**
2C19	**2C5, 2C9**	**Payne et al 1999a, Oda et al 2004, Wang et al 2007**
2E1	**2C5**	**Lewis 2000, Lewis et al 2003, Danko et al 2006**
4A11	**CYP102**	**Chang and Loew 1999**
11A1	**2B4**	**Storbeck et al 2007**
17alpha	**BMP**	**Auchus and Miller 1999**
19 (human aromatase)	**2C9**	**Favia et al 2006**
27A1	**several**	**Prosser et al 2006**

3.3 The homology model of CYP2B6

By now there is no x-ray structure of CYP2B6 available. However, a homology model of the enzyme has been built by Bathelt et al (2002), using CYP2C5 of rabbit as a template structure since the two isoforms share 48 % sequence similarity. Significantly, CYP2C5 lacks residues forming F-G-loop and before a model of CYP2B6 could be constructed, the missing loop had to be modeled. However, at the time Bathelt and co-workers modeled CYP2B6, no other crystal structures of mammalian cytochromes were available. Consequently, crystal structures of bacterial cytochrome, CYP BM-3 in an open and closed conformation, were chosen as a model for the missing F-G-loop in CYP2C5 although the particular region in bacterial CYPs is shorter. Nevertheless, with small adjustments the missing loop was modeled, and MD simulations with simulated annealing were applied to the CYP2C5 structure. The quality of the produced CYP2C5 was tested with unconstrained MD simulations in water, in which the protein structure denatures in the simulation, if it is not folded into its native state. As such, unconstrainded MD simulations are considered an extremely sensitive test for quality (Chang and Loew 1997).

Based on sequence alignment with CYP2C5, the model of CYP2B6 was successfully constructed. Validation of the homology model was also conducted with unconstrained MD simulations in water to measure the stability of the CYP2B6 model. The RMSD between the backbone atoms of CYP2B6 model, and the minimized CYP2B6 model after heating and releasing the backbone, was determined to be less than 1.7 Å, which points out that only small differences can be found between the structures of CYP2C5 and CYP2B6.

The homology model of CYP2B6 was further validated using manual docking of two substrates, cyclophosphamide (CPA) and ifosfamide (IFA) into the active site of the protein. CPA and IFA can be metabolized via two alternative pathways by CYP2B6, and by docking the substrates, it was determined whether the substrate moiety to be metabolized could reach the heme iron in a preferred orientation for metabolism to occur and without clashes with CYP2B6.

Based on the homology model, information of the putative active site of CYP2B6 and the residues important for substrate binding and activity has been suggested. The active site of the enzyme locates in the core of the enzyme, above the heme-thiolate group forming of three, mostly hydrophobic pockets. The most important pocket is the "heme pocket", where the substrate moiety to be metabolized has to enter. In the model of Bathelt and co-workers, all the pockets can be filled without clashes, when the substrate enters the active site in a preferred orientation for metabolism. The residues lining the pockets are considered major structural elements for regioselectivity of CYP2B6 (Table 7). Additional SRS (substrate recognition site) residues have been proposed based on several site-directed mutagenesis studies (Domanski et al 1999, Domanski and Halpert 2001, Spatzenegger 2003). The SRS residues of CYP2B6 have been compared with CYP2B1 SRSs and CYP2E1 SRSs, and mutated in order to find out if the metabolism of CYP2B6 substrates RP73401 and 7-ethoxy-4-trifluoromethylcoumarin (7-EFC) would be affected. The studies indicated direct interaction with RP73401 and L363 and non-direct interaction with RP73401 and F107. L363 and T292 were suggested to participate in 7-EFC O-deethylation, which is consistent with information gained from pharmacophore modeling of CYP2B6 substrates.

Table 7: Residues lining the binding pockets of CYP2B6 (Bathelt and co-workers 2002).

Pocket	Residues
A: Heme pocket	**Ala298, Thr302, Leu363, Phe206**
B: Side pocket	**Ser294, Ile114, Phe297, Phe115**
C: Roof pocket	**Ile209, Val447, Leu216, Val367**

3.3.1 Pharmacophore models of CYP2B6

Pharmacophore models determining the structural elements important for CYP2B6 activity have been constructed by Wang and Halpert (2002) and Ekins et al (1999b). In both studies, pharmacophore models included one hydrogen bond acceptor and hydrophobic regions. In the pharmacophore model of Wang and Halpert two hydrophobic regions where included, while in the model of Ekins and co-workers three hydrophobic regions were located in different relative positions to each other. Wang and Halpert produced two pharmacophore models for CYP2B6 substrates, and by mapping benzyloxyresorufin and 7-EFC into the pharmacophore models, they revealed the residues lining the modeled hydrophobic regions depending of the position of the substrate. Benzyloxyresorufin has two phenyl rings in its structure, fitting into hydrophobic regions H_1 and H_2 of the first pharmacophore model (A), and the oxygen atom of one of the ring acts as hydrogen bond acceptor. In their research, Wang and Halpert concluded that the oxygen is able to form a hydrogen bond directly with the active site of the enzyme or via a water molecule. However, when benzyloxyresorufin was docked into the active site of the homology model of CYP2B6, no hydrogen bond was identified between the molecule and CYP2B6. The findings suggests to the possibility of a water molecule participating in the hydrogen bonding with benzyloxyresorufin.

7-EFC was successfully fitted with the other pharmacophore (B) model created by Wang and Halpert, in which one of the rings of 7-EFC and its trifluoromethyl substituent can be fitted into hydrophobic regions H_3 and H_4, whereas the oxygen atom on one of the rings of 7-EFC acts as potential hydrogen bond acceptor. Residues at a distance of 5 Å from the hydrophobic regions H_1, H_2, H_3 and H_4 are represented in Table 8. Representations of the two pharmacophore models can be found in the original literature (Wang and Halpert

2001). The residues lining the hydrophobic regions are consistent with previous site-directed mutagenesis studies, in which amino acid residues being important for the substrate metabolism in CYP2B subfamily were determined (Domanski et al 1999, Domanski and Halpert 2001). Following from this, the residues were located from the CYP2B6 homology model applied in the experimental part of this work, in order to determine the active site of the enzyme. The information of the active site was then exploited in the molecular docking of compounds into the homology model, and in interpreting the quality of the results based on their orientation in the active site and their position relative to the site of metabolism.

Contrary to the pharmacophore models of Wang and Halpert, the pharmacophore model of Ekins and co-workers consisted of three hydrophobic regions, in addition to one hydrogen bond acceptor. The hydrophobic regions were located at distances of 5.3, 3.1 and 4.6 Å from the hydrogen bond acceptor with intermediate angle of 72.8˚ and 67.6˚. Representation of the pharmacophore model can be found in the original literature (Ekins et al 1999b). Two substrates, 7-EFC and S-mephenytoin were successfully fitted with the pharmacophore model and K_m values were predicted with good estimations for both of the substrates when compared to the observed K_m values.

The substrates succesfully fitted into CYP2B6 pharmacophore models both by Wang and Halpert and Ekins and co-workers were applied in the experimental part of this work as a part of a set of known substrates of CYP2B6, which were selected reference structures when confirming the usability of CYP2B6 homology model.The results obtained from our studies were also compared to the pharmacophore models with respect to the mutagenesis studies performed for residues of CYP2B6 active site.

Table 8: Residues lining the hydrophobic regions determined by pharmacophore modeling (Wang and Halpert 2002).

Region	**Residues** $_{\text{wild type}}$	**Residues** $_{\text{homology model}}$
H_1	**114, 115, 206, 298, 363, 367, 477**	**[84], [85], [176], [328], [333], [447]**
H_2	**206,297,298,301,302, 363, 477, 478**	**[176], [267], [268], [271], [272], [333], [447], [478]**
H_3	**206, 298, 302, 363, 367, 477, 478**	**[176], [268], [272], [333], [337], [447], [448]**
H_4	**301, 302, 305, 306, 362, 363, 477, 479**	**[271], [272], [275], [276], [332], [333], [447], [479]**

4 LIGAND-PROTEIN INTERACTIONS

So far this research has been focusing on ligand- and protein-based models separately. However, considering both of these approaches together, a wealth of detailed information concerning interactions between ligands and protein is achieved.

One of the main motives in successful drug research and discovery is to identify relatively small molecules with high binding affinity and selectivity towards its receptor (Krovat et al 2005). If such a molecule also possesses reasonable ADMET-profile, the possibility of late-stage attrition is reduced. 3D crystal structures of cytochrome P450s together with a bound substrate represent the means to study the behaviour of the molecules in the binding pockets as well as the way they interact with a receptor, using different computational approaches. One of these approaches is molecular docking protocol. “However, since only a few 3D crystal structures of cytochrome P450s togetherwith a bound ligand as well as plain 3D structures are not

Molecular docking programs are used in predicting energetically favorable binding conformations and orientations of small molecules within the active site of target protein (Kitchen et al 2004, de Graaf et al 2005a). The binding affinity of the molecules can be further explored using approaches such as scoring functions and free energy perturbation calculations, which estimate the free energy of ligand-protein binding (Oprea and Marshall 1998). However, docking approach lacks the ability to take into account full flexibility of ligand-protein complex which is an important point of view when considering cytochrome P450s (de Graaf et al 2005a). Nevertheless, most docking programs are capable of considering ligand-protein complexes as partially flexible systems. Molecular dynamics (MD) simulations are an alternative way to explore ligand-protein complex. MD simulations include internal flexibility and conformational changes of the system when considering ligand-protein interactions. In the following chapters, molecular docking, molecular dynamics simulations and binding affinity predictions are considered in more detail.

4.1 Molecular docking

Molecular docking approach aims to predict energetically favourable conformation and orientation of a ligand when it binds into the active site of a protein (Kitchen et al 2004, Sousa et al 2006). Several algorithms are available concerning molecular docking, which differ in the level approximations they use in predicting ligand-protein interactions. Significantly, the docking protocol is a combination of search algorithms for energetically favorable conformations and scoring functions of the obtained results. For clarity, molecular docking and scoring functions are discussed in separate chapters although they are applied simultaneously when ligand-protein interactions are being explored.

Approximations within algorithms generated for molecular docking are made, since biological systems such as cytochrome P450s are very dynamic in nature and possess substantially flexibile structures, making it necessary to reduce the dimensionality of the system in order to provide fast and effective computational tools. Following from this, the docking procedures consider ligand-protein interactions in three different way, namely 1) treating both ligand and protein as rigid molecules, 2) treating the ligand as flexible molecule (semi-flexible docking) or 3) treating both ligand and protein as flexible molecules (fully flexible docking) (Höltje et al 2003). Notably, although considered as a fully flexible docking approach, the flexibility of the protein is limited to specific sidechains of the active site. The approaches and algorithms used in automated molecular docking are represented in Figure 6. Additionally, the molecular docking approaches can be divided into stochastic or deterministic algorithms depending of the reproducibility of the docking results or into incremental construction methods. However, in this research the docking approaches are divided according to flexibility, which is an important characteristic contributing to the interactions between cytochrome P450s and their substrates.

A common approach is to use semi-flexible docking since rigid-body approximations ignore the importance of flexibility and on the other hand, fully flexible docking is computationally very demanding compared to other methods. With cytochrome P450s, the most common docking algorithms applied include GOLD, FlexX, DOCK and AutoDock which belong to the category of semi-flexible approximations (de Graaf et al 2005a).

Although fully flexible docking approaches produce results with higher accuracy concerning the true nature of ligand-protein interactions, fully flexible docking is not yet used in regular basis when ligand-protein interactions are concerned due to the extensive computer capacity, which the approach requires (Sousa et al 2006). In addition to flexibility, accuracy of docking results is dependent also of the physicochemical properties of ligand-protein interactions, the quality of target protein structure and whether or not water molecules are included in the docking (de Graaf et al 2006, Krovat et al 2005).

As mentioned, the flexibility of cytochrome P450s is an important factor contributing to the ligand-protein interactions. Thus, it should not be excluded when ligand-protein interactions are examined with molecular docking. Moreover, the cytochrome P450s represent a challenge for molecular docking since the proteins include heme moiety with metal ion and water molecules in their active site which are known to participate in the ligand-protein interactions in addition to the discussed structural flexibility and conformational changes during ligand binding (de Graaf et al 2005a). Since most docking procedures ignore water-mediated interactions, the accuracy of molecular docking can be limited when considering cytochrome P450s. However, if water is included in the dockings at the correct position, the accuracy of results improves, which has been demonstrated by de Graaf et al (2005b). By including different water scenarios into cytochrome P450 dockings (water-free active sites, active sites containing crystallographic water molecules and water molecules predicted with GRID program), the accuracy of docking results in reproducing experimentally determined binding modes was improved. The accuracy of the results was validated by comparing catalytic site predictions to experimental data and RMSD between heavy atoms in reference protein structure and ligand docking poses.

However, when docking cytochrome P450 substrates of variable sizes, merely including water is not enough; differently shaped and sized solvated protein binding pockets could further improve the accuracy of docking results (de Graaf et al 2005b). However, more computational power would be required, which is contradictory with docking approach having to be fast to influence decision making in chemical synthesis during early phases of drug development process (Shimada 2006).

Molecular docking has been successfully used in several cytochrome P450 predictions, for example in site-directed mutagenesis studies concerning CYP2B6, 2C8 and 2E1 (Spatzenegger et al 2003, Melet et al 2004) and in refining and validating CYP2B6 pharmacophore models (Wang and Halpert 2002) as well as homology models of CYP1A2 (Lozano et al 1997) CYP2B6, (Bathelt et al 2002), CYP2C8/9/18 (Riddeström et al 2001, Oda et al 2004), CYP2D6 (de Graaf et al 2006), CYP2E1 (Park and Harris 2003) and CYP3A4 (Szklarz and Halpert 1997). Successful predictions have also been made concerning the binding modes and catalytic sites for CYP2D6 and CYP3A4 ligands (de Graaf et al 2005b, Kirton et al 2005, de Graaf et al 2006, Zhou et al 2006). As such, molecular docking can be considered a powerful and versatile tool in early stages of drug development process in producing qualitative information concerning ligand-protein interaction. However, to gain quantitative information of the interactions, the results generated with molecular docking need to be further processed with different scoring methods to tell the tightness of ligand-protein binding and how well the ligand actually binds to its receptor (Sousa et al 2006). Nevertheless, the results obtained with molecular docking have been applied with success in virtual screening of large compound databases in search for high affinity substrates, prioritizing the hits discovered and in identifying substrates with selectivity towards a specified target protein (de Graaf et al 2006).

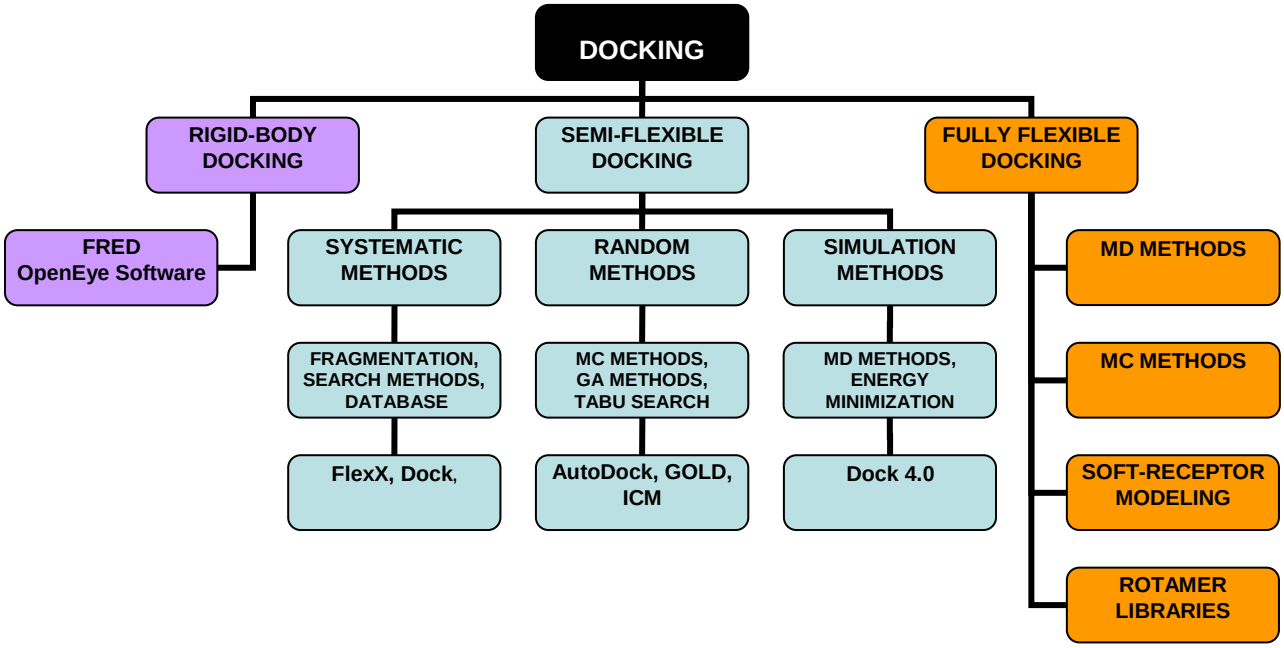


Figure 6: Organization chart representing automated docking algorithms and methods and software applied within each approach (Höltje et al 2003, Sousa et al 2006).

4.2 Binding affinity predictions

The docking protocol can be considered as a combination of search algorithms for energetically favorable conformations and scoring functions of the obtained results (Sousa et al 2006). The tightness of ligand-protein interactions is considered by estimating the free energy of binding of a ligand on the 3D structure of ligand protein complex (Höltje et al 2003). In other words, the binding affinity predictions are made based on ligand conformations generated with preceding molecular docking. The free energy of binding is estimated based on Gibbs-Helmholz equation (Eq. 3) which can be further related to binding constant K_i with Eq. 4 (Dill and Bronberg 2003, Höltje et al 2003).

$$\Delta G = \Delta H - T\Delta S \qquad \text{Eq. 3}$$

$$\Delta G = - RT \ln K_i \qquad \text{Eq.4}$$

ΔH: enthalpy *T:* temperature (in Kelvin) *ΔS:* entropy K_i*:* binding constant *R:* gas constant

The free energy estimations can be conducted using different scoring functions which identify a correct binding mode for ligand-protein complex based on lowest energy value and rank the docking results according to the binding affinity of ligands (Krovat et al 2005). In other words, scoring functions complement docking algorithms by evaluating the interaction between ligands and target protein (Kitchen et al 2004). In drug development process, situations where $\Delta G < 0$ are desired, since in these situations the ligand prefers to be bound to the protein and the previously mentioned cascade events leading to response can occur (Shimada 2006).

The scoring functions are divided into three categories; force-field based scoring functions, empirical scoring functions and knowledge-based scoring functions, respectively (Höltje et al 2003, Kitchen et al 2004, Krovat 2005) (Table 7). Also the scoring functions differ in accuracy and speed following from approximations being done when predicting binding modes and ranking the results (Kitchen et2004, Sousa et al 2006). The overall comparison of scoring functions is difficult and it is not possible to generate universal scoring function

which could be applied in every occasion with successful results (Krovat 2005, Warren 2006). Following from this, the lack of congruence indicates the need for improvement of scoring functions. Different approaches have been developed to overcome the problems of scoring functions, one of which is the use of consensus scoring (Table 7) (Charifson 1999). Consensus scoring combines several scoring functions giving higher ranks to binding modes placed at the top of the results in all of the scoring function (Krovat 2005). Consensus scoring allows detecting false positives more accurately than with single scoring function. However, the consensus approach only presents average values, which prevent it from outperforming the best scoring function.

The most accurate method to calculate binding constants and free energy of binding for ligand-protein complexes is to use free energy perturbation (FEP) calculations which make fewer approximations compared to other scoring functions when calculating free energy of binding (Knegtel and Grootenhuis 1998). FEP calculations are a part of force-field based scoring functions and can be used in determining free energy differences between two states. However, FEP calculations are more complex and computationally more demanding as well as more time-consuming compared to traditional scoring functions since before calculating binding constant for a complex system, several molecular dynamics simulations are needed. Also, FEP calculations are suitable only to a narrow range of chemical diversity, in other words, for compounds with little or moderate charge and flexibility. Consequently, the method is difficult to be applied in most occasions considering cytochrome P450s since the substrates of CYPs are known to be relatively charged and/or flexible in nature (de Graaf et al 2005a).

Table 7: Different scoring functions applied in molecular docking (Höltje et al 2003, Kitchen et al 2004, Sousa et al 2006)

Category	Description	Examples
Force Field-Based Scoring Functions	**The sum of interaction energy between ligand and receptor and internal energy of the ligand is quantified. Energy is accounted through non-bonded van der Waals and electrostatic energy terms based on classical molecular mechanics force field.**	**D-Score, G-Score, GoldScore, AutoDock 3.0 Scoring Function**
Empirical Scoring Functions	**Binding energy is approximated by a sum of several individual, uncorrelated terms describing polar and apolar interactions, entropy and desolvation effects.**	**Böhm's scoring functions, F-Score, ChemScore. SCORE, Fresno, X-Score**
Knowledge-Based Scoring Functions	**Statistically derived rules and principles aim to reproduce experimentally determined structures by using a potential of mean force which converts structural information into Helmholz free interaction energies of protein-ligand atom pairs.**	**Muegges's Potential of Mean Force (PMF), DrugScore, SmoG score**
Consensus Scoring	**Information gained from several scoring functions is combined to get more reliable results and to compensate the errors from each scoring function**	**X-CSCORE (PMF, ChemScore, FlexX)**

4.3 Molecular dynamics simulations

Molecular dynamics (MD) simulations represent an approach in which protein flexibility can be taken into account when considering ligand-protein interactions. Also other methods, such as MC methods, rotamer libraries, protein ensemble grids and soft-receptor modeling take protein flexibility into account to some extent (Kitchen et al 2004, Sousa et al 2006). Although MD simulations can be considered as an integral part of fully flexible docking procedures, the simulation methods differ greatly from molecular docking in basic principles. MD simulations are based on molecular mechanics and determining energetically most favorable 3D structure of a system is derived from Newton's equations concerning motion (Eq. 5) (Höltje et al 2003). As such, proteins are handled as a series of structures in different conformational states in equilibrium and energetically favorable

conformations are systematically searched from the potential energy surface. The advantage of MD simulations in exploring ligand-protein interactions is to describe the interactions occurring in a molecular system both in time and in space.

$$F_i(t) = m_i\, a_i(t) \qquad \text{Eq. 5}$$

F_i: force of atom i at time t m_i: mass of the atom a_i: acceleration of atom i at time t

With MD simulations also the structural flexibility of the ligand can be taken into account as well as water molecules mediating interactions between cytochrome P450s and their substrates (de Graaf et al 2005a). Consequently, extremely detailed information of the behaviour of ligand-protein complexes at an atomic level is produced and as such, MD simulation can be considered the most powerful method describing protein flexibility in general (Rueda et al 2007).

MD simulations have been used in several occasions when ligand-protein interactions have been examined. The method has been used in the optimization and validation of homology models of CYP1A2 (Belkina et al 1998), CYP2B6 (Bathelt et al 2002), 2C9 (Payne et al 1999b) and 2D6 (Venhorst et al 2003) and in the prediction of catalytic sites for cytochrome P450 substrates. Also the behaviour of ligand-protein complex with respect to substrate entering or leaving the active site via access of exit channels have been studied using molecular dynamics simulations (Wade et al 2004, Cojocaru et al 2007).

Although MD simulations are very powerful tools in exploring ligand-protein interactions, the method has some drawbacks. The disadvantage of MD simulations is their dependence of the quality of empirically determined force field parameters since atoms in a molecule are assumed to be interacting with each other according to force field based rules. Another drawback with MD simulations is the huge increase in demand for computer capacity (de Graaf et al 2005a, Sousa et al 2006). As such, MD simulations are not – at least not yet – applicable to be used systematically when considering ligand-protein interactions in the field of modern drug research and discovery.

II EXPERIMENTAL PART:

Molecular docking and Comparative Molecular Field Analysis of CYP2B6 Substrates

1 INTRODUCTION

1.1 CYP2B6

CYP2 family of the cytochrome P450 enzymes catalyzes over 50 % of all the oxidation reactions concerning metabolism phase I (Lewis 2003). Approximately 3 % of all the known pharmaceutical agents are being metabolized by cytochrome 2B6 (CYP2B6) (Bathelt et al 2002, Wang and Halpert 2002). There are several known substrates of the isoform and the metabolic pathways of the compounds are versatile. Some of these compounds and their chemical reactions catalyzed by cytochromes are represented in Table 1. Hydroxylation, demethylation and deethylation can be considered the most common reactions catalyzed by CYP2B6.

CYP2B6 is known to activate prodrugs such as anticancer drug cyclophosphamide and also procarsinogens such as NNK (4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone), which is a metabolite of nicotine, and aflatoxin B_1 (Code et al 1997). Recently, it has also been suggested the isoform participating in substantial amounts in the metabolism of different environmental chemicals e.g. insecticides, herbicides and industrial chemicals as well as abused toxicants e.g., MDMA and "Ecstasy" (Hodgson and Rose 2006). Due to CYP2B6 contributing in diverse aspects of biotransformation, it can be considered as a very important part of the xenobiotics metabolizing system, although the CYP2B6 percentage of the total hepatic content is relatively small when compared to other cytochrome P450 enzymes. The total hepatic content of CYP2B6 is in most cases estimated to be <1 %, but different information can be found in the literature ranging from 0.2 % to 6 % (Ekins et al 1999b, Turpeinen et al 2004). Accordingly, the amount of CYP2B6 is determined to exist in greater amounts than previously thought and as a consequence, the ability of CYP2B6 to metabolize drugs and other xenobiotics may well be underestimated.

CYP2B6 is inducible and polymorphic, which leads up to the isoform representing significant interindividual differences in CYP2B6 activity. The differences can be 100-fold between ethnic groups and gender, and are probably occurring due to variability in the hepatic levels of CYP2B6 mRNA and protein. The polymorphism defects in metabolism lowering the metabolic clearance of drug substrates, which can be seen with e.g., efavirenz

(an inhibitor or HIV-1 reverse transcriptase), ifosfamide and cyclophosphamide, thus causing unwanted side effects and potential drug-drug interactions, when co-administered with other substrates metabolized via CYP2B6 (Bathelt et al 2002, Lamba et al 2003, Ward et al 2003).

Table 1: Known substrates of CYP2B6 and their metabolic pathways.

Substrate	Metabolic pathway	Reference
S-mephenytoin	**N-demethylation**	**Heyn et al 1996**
Diazepam	**N-demethylation**	**Ono et al 1996**
7-ethoxy-4-trifluoromethylcoumarin	**O-deethylation**	**Code et al 1997**
Imipramine	**N-demethylation**	**Koyama et al 1997**
RP73401	**Ring hydroxylation**	**Stevens et al 1997**
Cyclophosphamide (CPA)	**4-hydroxylation**	**Bathelt et al 2002**
Ifosfamide (IFA)	**N-deethylation**	**Bathelt et al 2002**
Benzphetamine	**N-demethylation**	**Ekins et al 1998**
4-chloromethyl-7-ethoxycoumarin	**O-deethylation**	**Ekins et al 1998**
3-cyano-7-ethoxycoumarin	**O-deethylation**	**Ekins et al 1998**
Midazolam	**1'-hydroxylation**	**Ekins et al 1998**
Antipyrine	**4-hydroxylation**	**Ekins et al 1999b**
Cinnarizine	***p*-hydroxylation**	**Ekins et al 1999b**
7-ethoxycoumarin	**O-deethylation**	**Shimada et al 1999**
Testosterone	**16β-hydroxylation**	**Hanna et al 2000**
MDMA "Ecstasy"	**N-dealkylation, demethylenation**	**Kreth et al 2000**
Propofol	**4-hydroxylation**	**Court et al 2001**
Amitriptyline	**N-demethylation**	**Wang and Halpert 2002**
β-artheether	**N-demethylation**	**Wang and Halpert 2002**
Benzyloxyresorufin	**O-debenzylation**	**Wang and Halpert 2002**
Dextrometorphan	**N-demethylation**	**Wang and Halpert 2002**
Efavirenz	**8-hydroxylation**	**Ward et al 2003**
7-methoxy-4-trifluoromethylcoumarin	**O-demethylation**	**Donato et al 2004**
Nicotine	**N-demethylation**	**Yamanaka et al 2005**

1.2 The object of the experimental part

The aim of the study was to explore if the homology model of CYP2B6 (Figure 1) could be successfully used in molecular docking of a set of 49 structurally diverse compounds, and if it was be possible to create a functional CoMFA model for the compounds by correlating their structural properties with experimentally defined inhibition potencies. A successful CoMFA model could be further used in virtual screening of large databases in combination with other computational approaches, to predict potential CYP2B6 substrates and their activities as well as interactions with the protein when combined with information derived based on the homology model of CYP2B6.

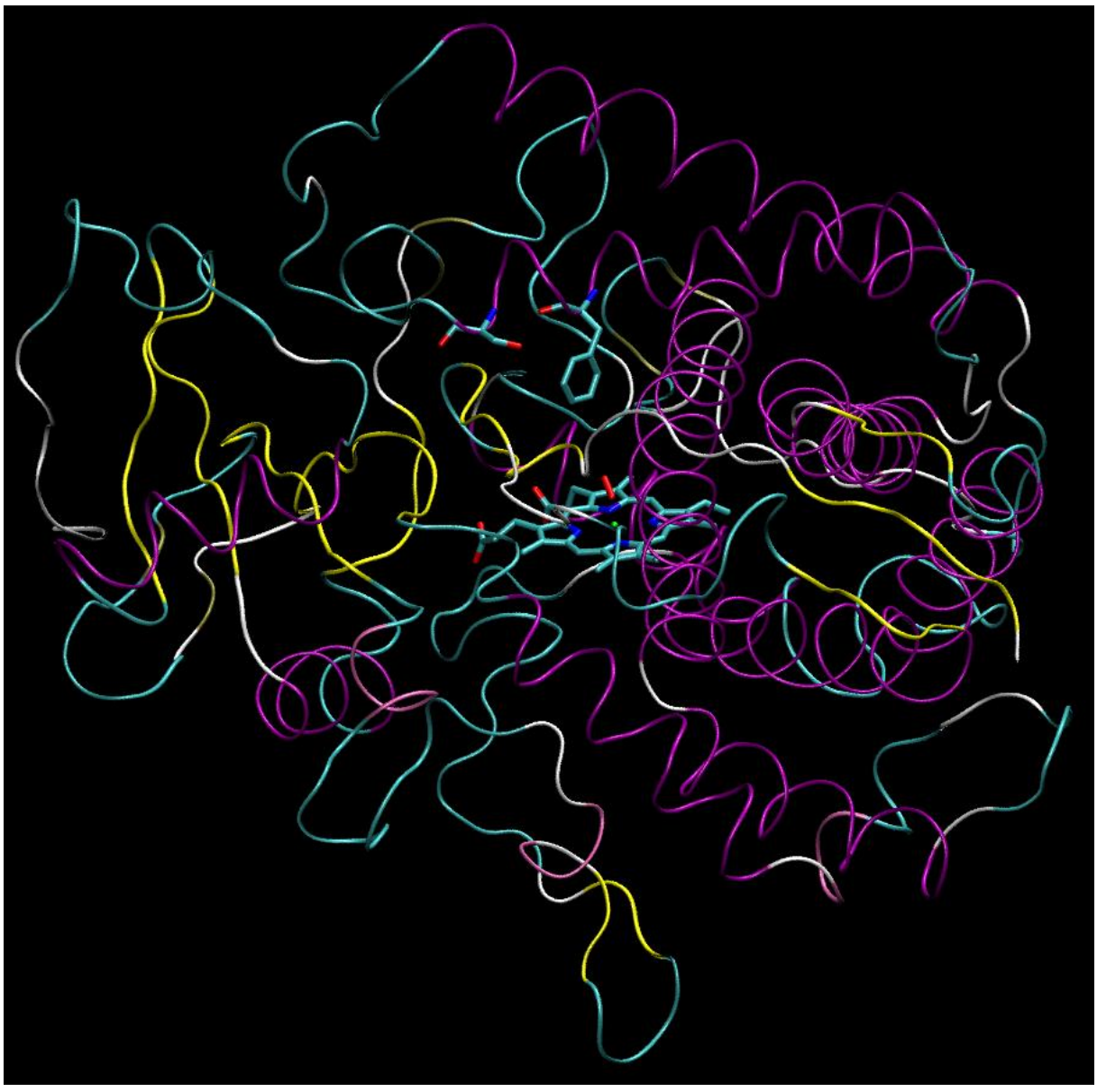


Figure 1: The homology model of CYP2B.

2 MATERIALS AND METHODS

All computations were performed on a 3,6 GHz Intel Xeon personal computer running SuSe Linux 10.1. The inhibition potencies (IC_{50} values) of structurally diverse compounds were determined with recombinant human CYP2B6 enzyme, 7-ethoxy-4-trifluoromethylcoumarin O-deethylation as a probe reaction. Measurements were conducted at the Department of Pharmacology and Toxicology in University of Kuopio. A more detailed description of the biochemical assay can be found in an article by Korhonen et al 2007.

2.1 Protein setup

The homology model of CYP2B6 has been modeled by Bathelt et al (2002), using X-ray structure of CYP2B5 as template with PDB code IDT6 (Brookhaven Protein Data Bank). In this research, the homology model was supplemented by adding O_2 at the active site, above heme iron with a distance from iron atom and oxygen atom closer to the heme iron being 1.85 Å and the approximate torsion angels between O – Fe – N_{1-4} (four nitrogen atoms of heme template) adjusted into 95º, 92.85º, 95.38º and 94.96º, respectively. O_2 was modeled with molecular modeling software using SYBYL 7.1 and sketch approach (Tripos Inc). O_2 was added into the homology model to prevent the ligands from complex binding with the heme iron when docked in the active site. The CYP2B6 protein structure was prepared using Sybyl 7.1 by adding hydrogens to the protein structure, and removing 34 lone electron pairs from the structure before docking with GOLD 3.0.

The solvent-accessible surface of the active site was calculated with FastConnolly method in MOLCAD to aid in visual inspection of the active site and the position of substrates in it after molecular dockings.

2.2 Ligand setup

Ligands used in molecular docking are represented in Tables 2 and 3. 16 known substrates of CYP2B6, which were used in the preliminary dockings, were selected from literature (Table 2). The known substrates were used as references in evaluating the degree of

success of molecular docking of the database of 49 ligands (Table 3) into the homology model by comparing them to the conformations of the known substrates. All the molecules were minimized with Tripos Force Field and partial atomic charges were calculated with Gasteiger-Hückel method using default settings in SYBYL 7.1. The 49 compounds and their inhibition potencies *in vitro* are taken from previous studies conducted in University of Kuopio concerning CYP2A6, CYP2B6 and CYP1A2 (Rahnasto et al 2005, Korhonen et al 2005, Korhonen et al 2007).

The 49 ligands can be divided according to their structure into lactones, aromatic molecules, pyridine derivatives and diphenylmethane derivatives. The aromatic compounds can be further divided into polyaromatic compounds and derivatives containing only one aromatic ring. The inhibitory potencies of the compounds range from nanomolar to micromolar, ticlodipine being the most potent and ε- capronolactone the weakest inhibitor.

Table 2:The set of 16 known substrates of CYP2B6 used in the preliminary dockings and their metabolic pathways.

Substrate	Metabolic pathway	Reference
S-mephenytoin	**N-demethylation**	**Heyn et al 1996**
Diazepam	**N-demethylation**	**Ono et al 1996**
7-ethoxy-4-trifluoromethylcoumarin	**O-deethylation**	**Code et al 1997**
Imipramine	**N-demethylation**	**Koyama et al 1997**
RP73401	**Ring hydroxylation**	**Stevens et al 1997**
Benzphetamine	**N-demethylation**	**Ekins et al 1998**
4-chloromethyl-7-ethoxycoumarin	**O-deethylation**	**Ekins et al 1998**
3-cyano-7-ethoxycoumarin	**O-deethylation**	**Ekins et al 1998**
Midazolam	**1'-hydroxylation**	**Ekins et al 1998**
Antipyrine	**4-hydroxylation**	**Ekins et al 1999b**
Cinnarizine	***p*-hydroxylation**	**Ekins et al 1999b**
7-ethoxycoumarin	**O-deethylation**	**Shimada et al 1999**
Testosterone	**16β-hydroxylation**	**Hanna et al 2000**
Dextrometorphan	**N-demethylation**	**Wang and Halpert 2002**
7-methoxy-4-trifluoromethylcoumarin	**O-demethylation**	**Donato et al 2004**

2.3 Docking

The compounds were docked using two commercially available molecular docking programs, FlexX 1.13.5 L (BioSolveIT) and GOLD 3.0 (The Cambridge Crystallographic Data Centre). With FlexX, the CYP2B6 homology model was complemented with bound reference structure (7-EFC), and the ligand was checked and modified concerning ligand atoms and bond types. The active site was customized by defining the active site first within 6.5 Å from the reference structure. In later dockings, the active site range from the reference structure was redefined into 8.0 Å and 11.6 Å, respectively. Four ligands, namely coumarin, 7-ethoxyresorufin, 7-EFC and 7-MFC (7-methoxy-4-trifluoromethylcoumarin) were used in the preliminary dockings using FlexX default settings except maximum protein ligand overlap, which was increased into 5, 20 and 100, respectively. The filtering of the generated conformations was performed by ranking the affinity of compounds bound to the protein with Consensus Scoring (CScore) (Charifson et al 1999).

GOLD dockings were performed using the default settings, and the program was instructed to terminate the docking if the best three solutions for each ligand were within 1.5 Å rmsd of each other. The ligand binding site was specified to be within 15 Å range of a central atom C2205 [Thr 272]. GoldScore was determined as a primary scoring function (Verdonk et al 2003), but additional scoring functions were calculated with CScore and X-SCORE (Wang et al 2002).

2.4 Comparative molecular field analysis (CoMFA)

23 ligands with IC_{50} spanning around three log units listed in Table 3 (marked with an asterisk) were analyzed using the CoMFA method (Tripos Inc.). The predictive power of the model was evaluated by estimating the pIC_{50} values for 26 ligands not included in the preliminary CoMFA model (Table 4). The inhibition potencies of the ligands were transformed from IC_{50} into pIC_{50}; the reverse logarithmic representation of IC_{50}. Partial atomic charges were recalculated with MMFF94 method using default settings of SYBYL 7.1, since a successful CoMFA model could not be generated with Gasteiger-Hückel charges regardless of iterative efforts.

CoMFA analysis was performed with standard parameters by measuring steric (Lennard-Jones potential) and electrostatic (Coulomb potential) energies for each molecule by sp^3 carbon atom with +1 charge and grid spacing of 2 Å. Column filtering was set to 2 kcal/mol. The optimum number of components (2) for the non-cross-validated analyses was evaluated by leave-one-out (LOO) cross-validation, according to the 5 % rule; if adding a component decreased standard error of prediction, smaller number of components was chosen. The model was validated using cross-validation for statistical analyses.

Table 3: Inhibition potencies (Ic_{50}) and structures of compounds used in molecular docking of ligands in the homology model of CYP2B6

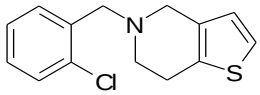
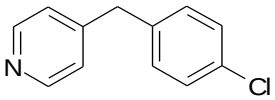
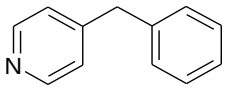
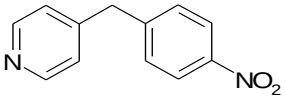
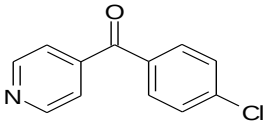
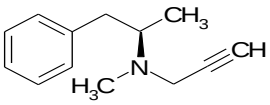
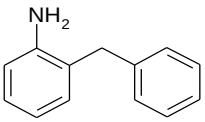
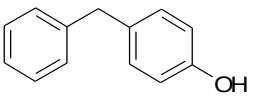
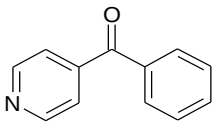
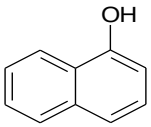
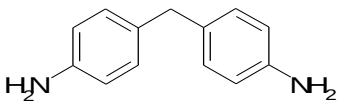
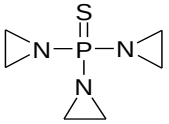
Compound		IC_{50} value (µM)	95 % confidence intervals
Ticlodipine		**0.09506**	**0.06919-0.1209**
4-(4-chlorobenzyl)pyridine		**0.2766**	**0.1570-0.3962**
4-benzylpyridine		**0.4141**	**0.2774-0.5509**
4-(4-nitrobenzyl)pyridine		**0.4449**	**0.3553-0.5345**
4-(4-chlorobenzoyl)pyridine		**1.479**	**1.201- 1.758**
Selegiline*		**3.920**	**3.385-4.455**
2-benzylaniline*		**4.431**	**2.704-6.158**
4-hydroxydiphenylmethane*		**4.993**	**2.528-7.457**
4-benzoylpyridine		**5.150**	**3.802-6.498**
1-naphthol		**9.113**	**6.605-11.62**
4.4-DDM*		**6.834**	**3.886-9.782**
ThioTEPA		**10.36**	**8.220-12.50**

Table 3 continues

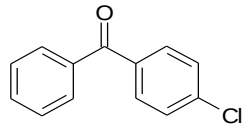
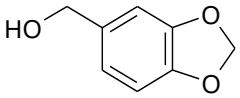
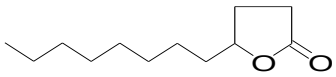
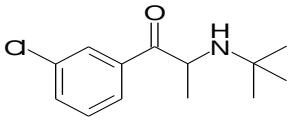
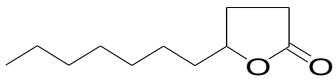
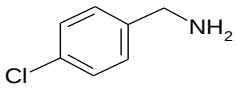
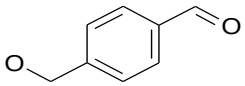
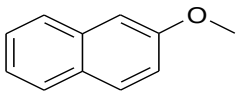
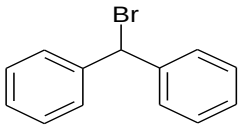
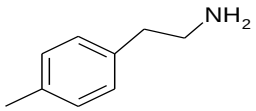
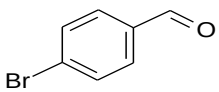
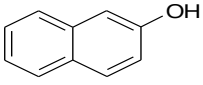
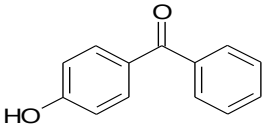
Compound		IC_{50} value (µM)	95 % confidence intervals
4-chlorobenzophenone*		**12.53**	**8.319-16.73**
Piperonylalcohol		**19.80**	**16.57-23.04**
γ-dodecanolactone*		**22.19**	**13.54-30.84**
Bupropion*		**28.13**	**10.12-46.13**
Undecanoic-γ-lactone*		**35.31**	**26.92-43.69**
4-chlorobenzylamine		**36.69**	**26.64-46.74**
4-methoxybenzaldehyde		**36.93**	**26.98-46.87**
2-methoxynaphthalene*		**37.08**	**29.02-45.14**
Bromodiphenylmethane		**38.77**	**29.02-48.51**
2(p-tolyl)ethylamine*		**44.62**	**28.78-60.47**
4-bromo-benzaldehyde		**49.80**	**34.21-65.38**
2-naphthol*		**68.66**	**41.24-96.08**
4-hydroxybenzophenone		**95.96**	**66.53-125.4**

Table 3 continues

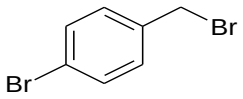
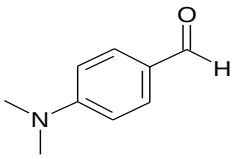
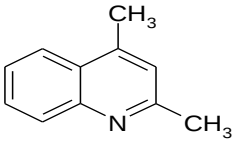
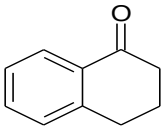
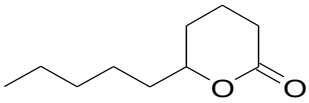
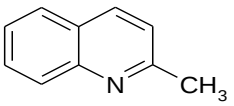
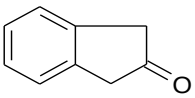
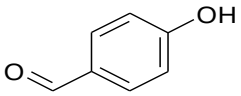
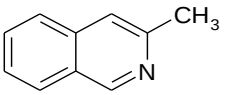
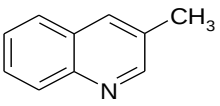
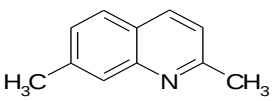
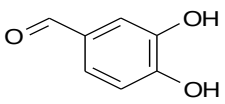
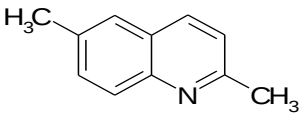
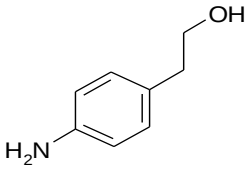
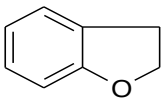
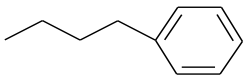
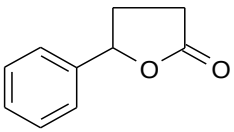
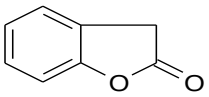
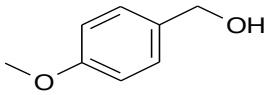
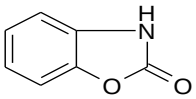
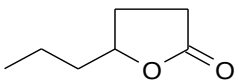
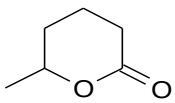
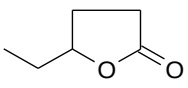
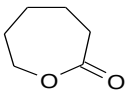
Compound		IC_{50} value (µM)	95 % confidence intervals
4-bromobenzylbromide*		**106.6**	**84.64-128.5**
p*-dimethylaminobenzaldehyde		**106.6**	**84.64-128.5**
2,4-dimethylquinoline*		**137.1**	**94.29-180.0**
alfa-tetralone*		**145.3**	**117.7-172.9**
δ-decanolactone*		**154.3**	**132.6-175.9**
Quinaldine		**311.6**	**213.4-409.8**
2-indanone*		**347.7**	**244.2-451.3**
4-hydroxybenzaldehyde		**365.5**	**298.1-432.9**
3-methylisoquinoline*		**391.1**	**244.2-538.1**
3-methylquinoline*		**415.0**	**254.7-575.3**
2,7-dimethylquinoline*		**416.2**	**321.7-510.7**
3,4-dihydroxybenzaldehyde		**438.0**	**294.8-581.2**

Table 3 continues

Compound		IC_{50} value (μM)	95 % confidence intervals
2,6-dimethylquinoline		**483.5**	**345.7-621.4**
2-(4-aminophenyl)-ethanol*		**630.6**	**216.3-1045**
2,3-dihydrobenzofuran		**667.6**	**329.8-1005**
Butylbenzene		**822.6**	**611.0-1034**
γ-phenyl-γ-butyrolactone		**1119**	**750.7-1487**
2-coumarone*		**1139**	**613.4-1665**
Anisalcohol		**1361**	**1043-1678**
2-benzoxalinone*		**1638**	**879.0-2397**
γ-heptalactone		**2354**	**1633-3075**
δ-hexalactone		**4704**	**3758-5650**
γ-caprolactone		**6430**	**4891-7969**
ε-caprolactone		**28134**	**26079-30189**

***Compounds used in the preliminary CoMFA model**

Table 4: Compounds used in evaluating the predictive capacity of the CoMFA model presenting the actual and predicted $_pIc_{50}$ (M) values

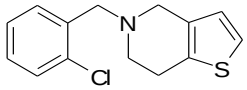
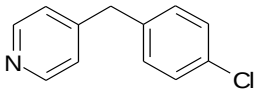
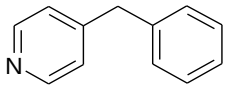
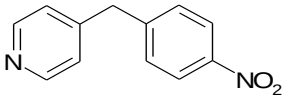
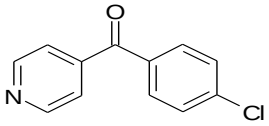
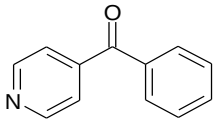
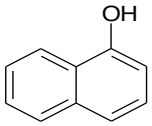
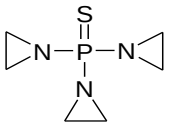
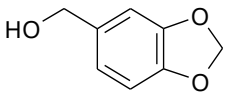
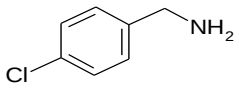
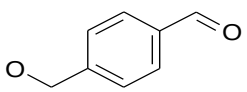
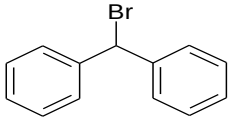
Compound		Actual pIc_{50} (M)	Predicted pIc_{50} (M)*	Residual
Ticlodipine		**7.02**	**5.72**	**1.30**
4-(4-chlorobenzyl)pyridine		**6.56**	**6.97**	**0.41**
4-benzylpyridine		**6.38**	**6.14**	**0.25**
4-(4-nitrobenzyl)pyridine		**6.35**	**6.78**	**0.43**
4-(4-chlorobenzoyl)pyridine		**5.83**	**5.67**	**0.16**
4-benzoylpyridine		**5.29**	**6.06**	**0.78**
1-naphthol		**5.04**	**4.34**	**0.70**
ThioTEPA		**4.98**	**4.27**	**0.71**
Piperonylalcohol		**4.70**	**3.65**	**1.05**
4-chlorobenzylamine		**4.44**	**3.70**	**0.73**
4-methoxybenzaldehyde		**4.43**	**3.71**	**0.72**
Bromodiphenylmethane		**4.41**	**4.43**	**0.02**

Table 4 continues

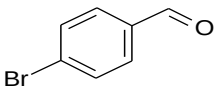
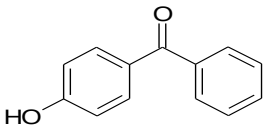
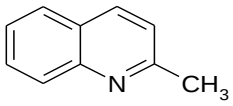
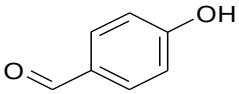
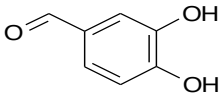
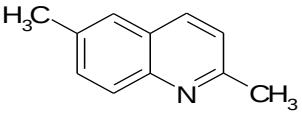
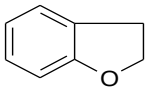
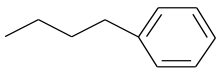
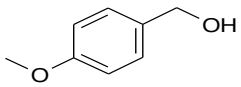
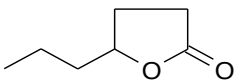
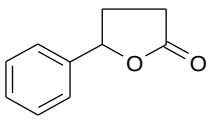
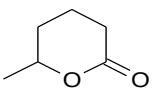
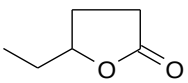
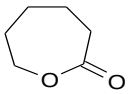
Compound		Actual pIc_{50} (M)	Predicted pIc_{50} (M)*	Residual
4-bromo-benzaldehyde		**4.30**	**3.32**	**0.98**
4-hydroxybenzophenone		**4.02**	**4.40**	**0.38**
quinaldine		**3.51**	**3.77**	**0.26**
4-hydroxybenzaldehyde		**3.44**	**3.56**	**0.12**
3,4-dihydroxybenzaldehyde		**3.36**	**3.47**	**0.12**
2,6-dimethylquinoline		**3.32**	**3.25**	**0.06**
2,3-dihydrobenzofuran		**3.18**	**3.50**	**0.33**
Butylbenzene		**3.08**	**4.33**	**1.25**
Anisalcohol		**2.87**	**2.97**	**0.11**
γ-heptalactone		**2.63**	**2.32**	**0.31**
γ-phenyl-γ-butyrolactone		**2.59**	**3.37**	**0.78**
δ-hexalactone		**2.33**	**2.29**	**0.04**
γ-caprolactone		**2.19**	**2.97**	**0.78**
ε-caprolactone		**1.55**	**2.81**	**1.26**

Table 5: Optimum number of components determined for the non-cross-validated analyses

Number of components	Standard error of prediction (S_{press})
1	**0.584**
2	**0.564**
3	**0.670**
4	**0.728**
5	**0.740**

Table 6: The sums of extrapolated values for test set ligands. Extrapolated values which are smaller than S_{press} 0.564 for the preliminary CoMFA model indicate the actual pIc50 value being most likely within ± S_{press} of the predicted value.

Compound	The sum of extrapolated values
Ticlodipine	**0.248736**
4-(4-chlorobenzyl)pyridine	**0.362719**
4-benzylpyridine	**0.0407435**
4-(4-nitrobenzyl)pyridine	**0.482912**
4-(4-chlorobenzoyl)pyridine	**0.0193757**
4-benzoylpyridine*	**1.05911**
1-naphthol	**0.473842**
ThioTEPA	**0.0862455**
Piperonylalcohol	**-0.39926**
4-chlorobenzylamine	**0.132469**
4-methoxybenzaldehyde	**-0.110152**
Bromodiphenylmethane	**-0.0327824**
4-bromo-benzaldehyde	**-0.070504**
4-hydroxybenzophenone	**-0.2841**
Quinaldine	**0.0502355**
4-hydroxybenzaldehyde	**0.0465036**
3,4-dihydroxybenzaldehyde	**-0.0324643**
2,6-dimethylquinoline	**-0.0185621**
2,3-dihydrobenzofuran	**-0.00299964**
Butylbenzene	**0.0229503**
Anisalcohol	**-0.211753**
γ-heptalactone	**0.0711989**
γ-phenyl-γ-butyrolactone	**0.0749135**
δ-hexalactone	**0.0511918**
γ-caprolactone	**0.101879**
ε-caprolactone	**-0.228327**

*the CoMFA model extrapolates outside the model; the prediction of pIc50 cannot be considered valid

3 RESULTS AND DISCUSSION

3.1 Setting up the molecular docking protocol

With FlexX dockings, surprisingly only coumarin was docked in the active site of CYP2B6 homology model, and all the other substrates were docked entirely outside of it. Changing the active site range from 6.5 Å to 8.0 Å and 11.6 Å did not improve the results significantly, nor did changing the maximum protein ligand overlap. The best values for maximum protein ligand overlap were determined to be 20 and 100 with visual inspection of the docked conformations of ligands in the preliminary dockings. Following from this, the database of ligands was docked into the active site using these values. When docking the database of ligands into the active site, the range was selected to be 8.0 Å within the bound reference structure. Approximately 60 % of ligands were positioned in the active site and the rest of the ligands outside of it. As such, the results obtained with FlexX could be considered neither encouraging nor consistent.

A possible cause for the unsuccessful docking with FlexX might be the hydrophobicity of the active site of CYP2B6. The active site is surrounded by aliphatic residues Ala, Leu, Ile and Val and aromatic residues Phe and Thr, which can be considered relatively hydrophobic amino acids. Accordingly, between the ligands and protein hydrophobic interactions are dominant, which is consistent also with previously generated pharmacophore models containing more hydrophobic regions than regions for hydrogen bond donors or acceptors (Ekins et al 1999b, Wang and Halpert2002). FlexX search algorithm starts with finding hydrogen bonds in the ligand-protein. Subsequently, the ligand is placed into the active site. Contrary to hydrogen bonds, which have a strict direction, angle and distance, hydrophobic interactions are less oriented in space. Following from this, the FlexX docking program might have difficulties in placing the ligands inside the active site since electronic interactions (hydrogen bond acceptors and donors etc.) are possibly not detected or the hydrogen bonds are found located outside the active site.

Since all the ligands in the database were not docked properly in the active site nor were the known substrates, the decision to change the docking program from FlexX to GOLD

was made. Following from this, more reliable results which would be consistent with previous studies concerning the docking of known substrates into the active site of CYP2B6 were hoped to be achieved.

In the preliminary GOLD dockings, the set of 16 known substrates were docked in the active site of CYP2B6 in order to check the usability of the homology model. Most of the substrates were successfully docked inside the active site cavity. The best docking solutions for the substrates were visually inspected. Only a few exceptions (midazolam, imipramine, cinnarizine, antipyrine and benzyloxyresorufin) among the known substrates were discovered, which were located outside active site cavity. The most surprising result concerning the outliers was benzyloxyresorufin. Benzyloxyresorufin has been successfully used in generating CYP2B6 pharmacophore model and is known to complement the pharmacophores (Ekins et al 1999b, Wang and Halpert 2002). It is possible that the substrates docked in the periphery of the active site have several possible binding modes and the conformation, which ligand adapts when binding in the active site of CYP2B6 is energetically less favoured than conformations with lower energy but not preferred for metabolism to occur. The complex nature of CYP isoforms in substrate binding is discussed in more detail in chapter 2.2.1 of the Literature Review. Following from this, the conformation most suitable for metabolism could be excluded from the dockings solutions and result in positioning the ligand outside the putative active site of the protein. These substrates should be further studied to find explanation for their conformation in the active site.

Since upmost all of the known substrates were docked inside the active site cavity in a reasonable binding mode with respect to their site of metabolism, the approach to use GOLD in molecular docking instead of FlexX was concluded to be useful. One of the known substrated, 7-EFC in its docked conformation is represented in Figure 3. As a conclusion, the results of the preliminary dockings with known substrates of CYP2B6 were considered encouraging and the homology model of CYP2B6 to be reliable in further studying of CYP2B6 ligands and their binding in the active site. Based on the successful docking of the known substrates, the database of 49 ligands (Table 3) was docked with GOLD using the same parameters as in the preliminary GOLD dockings of known substrates.

Consequently, all ligands in the database were also docked with previously described protocol inside active site cavity with reasonable geometries (Figure 2). More detailed viewing of the ligands showed three of the ligands, namely 2-benzoxalinone, anisalcohol and 4-hydroxybenzophenone forming hydrogen bonds with residue [Arg404]. Most of the ligands were also observed binding in a roughly similar conformation to 7-EFC. Based on the conformation of the ligands, they were divided into two groups: ligands orientated in the active site similarly to 7-EFC, and ligands orientated in a different position from the conformation of 7-EFC. PLS-analysis was performed for both of the groups with cross-validation, in order to construct a successful CoMFA model. However, PLS-analyses for both of the groups gave inadequate results, q^2 values being less than 0.4 in both cases. q^2 values less than 0.5 indicate the models having little predictive capacity. As a consequence, an alternative way to divide the ligands was presented. Closer visual inspection of the ligands orientating in the active site similar to 7-EFC showed the ligands being able to position in two possible conformation without complete loss of similarity to the 7-EFC binding mode. The *para*-substituent of aromatic ring structure of the ligands could be seen pointing towards the heme or to completely opposite direction. As such, two ligands representing both orientations, namely 4-hydroxybenzophenone and 4-(4-chlorobenzoyl)pyridine, were chosen as reference structures. The skeletons of the molecules could be superimposed onto each other the substituents excluded. The hydroxyl group of 4-hydroxybenzophenone was pointing towards the heme and the chlorine substituent of 4-(4-chlorobenzoyl) pyridine in the opposite direction. PLS-analysis was again performed for both of the groups, but the results did not improve from the previous PLS-analyses giving q^2 values less than 0.4, respectively.

Following from the unsuccessful PLS-analyses, the structure of the homology model was decided to be redefined. First, the orientation of [Arg404], which formed hydrogen bonds with three of the ligands, was modified. Based on literature, [Arg404] stabilizes the protoporhyrinring. However, if the residue is engaged into another hydrogen bond with a ligand the stabilization will be corrupted. Following from this, [Arg404] was protected from hydrogen bonding by rotating the torsion angles of the side chains of residues within 4 Å range of [Arg404], where [Val337] and [Ile84] were positioned. [Arg404] [Val337] was decided to left untouched, but the torsion angles of [Ile84] were changed using Sybyl

7.1 Biopolymer approach with Lovell method; χ_1 was changed from 163.0 to -177.0 and χ_2 from 174.1 to 66.0).

Second, the position of the oxygen above heme iron was changed by rotating the O-O-bond 180 º. As a result, oxygen was pointing at opposite direction than at the outset still preventing complex binding occurring between the ligands and heme iron. The homology model was then minimized, and by docking a set of substrates, which consisted of 7-EFC, 7-MFC and 7-ethoxycoumarin, into the active site and by checking visually the binding mode of the substrates with respect to their site of metabolism, the functionality of the homology model was confirmed. Visual inspection also showed the substrates to be positioning according to the previous, successful GOLD docking. By docking the database of ligands in the active site after confirming the functionality of the homology model, it was also confirmed that there were no more hydrogen bonds between the ligands and the residue [Arg404] in its modified orientation. However, the pharmacophores from literature report a hydrogen bond to be crucial part of ligand-protein interaction, even though the particular amino acid in the active site of CYP2B6 has not been identified. Since in the current CoMFA model were no possibilities for hydrogen bonding, the current CoMFA model was considered insufficient. Therefore, additional changes in the homology model of CYP2B6 were decided to be conducted.

GRID analysis performed with H_2O as a probe suggested water molecules to be locating near residue [Ser180], and to be possibly participating in the hydrogen bonding between the ligands and CYP2B6. Consequently, the torsion angle of [Ser180] was rotated by changing χ_1 from 62.0 to 65.0. After rotating the torsion angles of the residue, 7-EFC was docked in the active site, and found to be forming a hydrogen bond with [Ser180]. The carbonylic oxygen atom attached to the ring structure of 7-EFC was acting as a hydrogen bond acceptor and the hydrogen in [Ser180] hydroxyl group as hydrogen bond donor. However, the results are not consistent with pharmacophore models created for CYP2B6 substrates (Ekins et al 1999b, Wang and Halpert 2002). In the pharmacophore models, the oxygen atom of 7-EFCs ring acts as a potential hydrogen bond acceptor when in our dockings, the carbonyl oxygen attached to the ring acts as a hydrogen bond acceptor. Additionally, there are no potential hydrogen bond donors in the active site of the homology model near the 7-EFC ring structure, which also indicates H_2O participating in forming of hydrogen bonds. Following from this, the GRID analysis performed for

CYP2B6 is consistent with literature and supports the findings of water molecules having a part in hydrogen bonding between the active site and ligands (Wang and Halpert 2002).

Significantly, docking of the database of ligands showed additional hydrogen bonding between CYP2B6 and the database of ligands. Most of the hydrogen bonds were formed between residues [Phe176] and/or [Ser180] and the ligands (Figures 4 and 5). A CoMFA model was build for those conformations of ligands forming hydrogen bonds with the active site of CYP2B6. However, the CoMFA model still did not show the model having predictive capacity. As a conclusion, the partial atomic charges of all the ligands were recalculated with MMFF94 method since better statistics were achieved in using MMFF94 charges.

After changing the charges, a preliminary CoMFA model was generated for 23 randomly selected ligands of the database (Table 4) to explore the effect of redefined charges on generating a CoMFA model. Most of the 23 ligands did not form hydrogen bonds with the active site of CYP2B6 in the previous dockings. The CoMFA model represented statistical values of good quality with two components giving q^2 value of 0.564, S_{press} (standard error of prediction) of 0.523, and standard r^2 value of 0.822 with *s* (standard error of estimation) of 0.345. A plot for the training set actual vs. predicted pIc50 values is represented in Figure 7 and S_{press} with different number of components in Table 5. This model was applied to predict the inhibition potency of 26 ligands of the database not included in the CoMFA model (Table 4), giving predicted r^2 value 0.787 with *s* of 0.712. The percentage of extrapolated terms and the sum of extrapolated values was examined for the CoMFA model to ensure the extrapolated values did not exceed standard error of prediction (0.564). The sum of extrapolated values of the test set ligands did not have larger magnitude than S_{press} despite 4-benzoylpyridine (Table 6). The results proved goodness of fit and indicated the model having good predictive capacity.

3.3 Interpretation of the CoMFA contour maps

The CoMFA model predicted IC_{50} value of bromodiphenylmethane most accurately, the residual of actual vs. predicted pIC_{50} being 0.02. Therefore, bromodiphenylmethane is chosen as a template structure in interpreting the CoMFA contour maps. The CoMFA

model is represented in contour maps in Figure 6 with bromodiphenylmethane as a reference structure.

In the CoMFA map, red colour represents areas where negative charge and hydrogen bond acceptors increase the inhibition potency of the ligand, blue represents areas where positive charge and hydrogen bond donors decrease the inbition potency of the ligand, green colour indicates that steric bulk in these areas increases the inhibition potency, and in yellow regions steric bulk decreases the inhibition potency. The CoMFA model is consistent with the structure of bromodiphenylmethane; green colour indicates steric bulk being favoured near bromide atom between the two benzene rings. Bromide atom has a relatively large size and the atom can be considered bulky, which correlates with the green region surrounding the atom. As a result, the bromide atom can be considered an important contributor in the inhibition potency of the ligand, which is also consistent with the information of hydrophobicity being important for the overall inhibitory activity of CYP2B6 ligands and complementing the generated pharmacophore models. However, in the pharmacophore models several hydrophobic regions in different relative positions were found to be important for the activity of ligands, whereas in our CoMFA model only one hydrophobic region can be located in the contour map. The CoMFA model also points out two strict regions where negative charge and hydrogen bond acceptors would increase the inhibition potency of the ligands. The finding is again consistent with the information obtained from the pharmacophore models and the structure of CYP2B6 substrates in general, since the substrates of CYP2B6 are known to contain one or two hydrogen bond acceptors being angular, mostly neutral or basic and rather hydrophobic (Lewis and Dickins 2002). However, in the structure of bromodiphenylmethane hydrogen bond acceptors do not exist.

The test set of CYP2B6 ligands included also planar molecules whose activity the CoMFA model was able to accurately predict. Positive charge and hydrogen bond donors were favoured in rather small region in the CoMFA model, and it would be interesting to see the effect which fulfilling of these areas according to the CoMFA model would have on the inhibition potency of the ligands. Interestingly, not any of the test set ligands possessed characteristics complementary to the regions in the contour map favouring positive charge and hydrogen bond donors.

3.4 Summary

As a summary, the CoMFA model generated was able to predict activities for structurally diverse compounds giving statistical values of good quality, and the CoMFA model can be considered successful. Nevertheless, it should be beared in mind that despite good statistical values and the predictive power of the CoMFA model attained during this research, a lot of work is still needed for further improvement and validation of the CoMFA model, especially concerning the ligands relating with high residual between the actual and predicted IC_{50}, i.e., butylbenzene, ε-caprolactone and ticlodipine. Interestingly, ticlodipine is the most potent inhibitor in the test set and it has the greatest difference between the actual and predicted inhibition activity. Another CoMFA model for CYP2B6 ligands with higher predictive capacity has recently been generated in the university of Kuopio by Korhonen et al 2007, giving q^2 value of 0.71, S_{press} of 0.64 and standard r^2 value of 0.85 with predictive r^2 being 0.80, proving that high quality CoMFA models for structurally diverse CYP2B6 substrates can be put into practice.

With respect to the results obtained from this work, the docking results of the ligands need further examination. Although the ligands were positioned in the active site of CYP2B6 homology model quite analogously, explicit differences could be seen among some of the conformations of structurally very similar ligands, possibly affecting the alignment of ligands. Successful alignment of the ligands is a key step in generating 3D-QSAR models such as CoMFA, and as a consequence, the inconsistent conformations of the structurally similar ligands may have hindered the CoMFA model achieving the best possible predictive capacity. This leaves questions concerning the seemingly random conformations and by definition, the suitability of the particular docking programs (FlexX and GOLD) since different docking programs and scoring functions produce different kind of information (Warren et al 2006). However, the unsuccessful dockings can be affected by the relative hydrophobicity of the active site of CYP2B6. Most of the residues surrounding the active site are considered to be hydrophobic and accordingly, do not possess hydrogen bond acceptors and donors, which are more distinct for the docking programs to consider as crucial contributors to the ligand-protein interactions during ligand binding. It is also possible, that using the homology model of CYP2B6 had an effect on the reliability of the docking results despite 48 % sequence similarity between isoforms CYP2C5 and CYP2B6. Nevertheless, the CYP2B6 homology model has been successfully used in other researches

using molecular docking and further research concerning the area is needed before conclusions can be made.

In addition, the hydrogen bonds which were formed between residues [Phe176] and/or [Ser180] and some of the ligands raise questions. [Phe176] is mentioned in the literature to be one of the residues important for substrate metabolism based on mutagenesis studies, docking studies and pharmacophore models (Domanski et al 1999, Domanski and Halpert 2001, Bathelt et al 2002, Wang and Halpert). Significantly, pharmacophore models suggest [Phe176] to participate in hydrophobic interactions, whereas in our studies the residue acts as hydrogen bond donor. Contrary to [Phe176], there have not been any results in previous researches concerning the participation of [Ser180] in ligand binding. Neither the mutagenesis studies determining SRS residues nor pharmacophore models with information of residues surrounding the regions important to ligand activity have pinpointed the possibility of [Ser180] having a role ligand binding. So far, in mutagenesis studies only a few particular residues have been shown to participate in ligand binding directly or non-directly, namely T292, L363 and F107. As such, this work has pointed the possibility of [Ser180] having an important part concerning the interactions between CYP2B6 and its ligands, and as such, suggests new features concerning the residues important for ligand binding in the active site cavity of the isoform However, whether or not the residue really has an effect on ligand binding in the active site of CYP2B6 needs further researching.

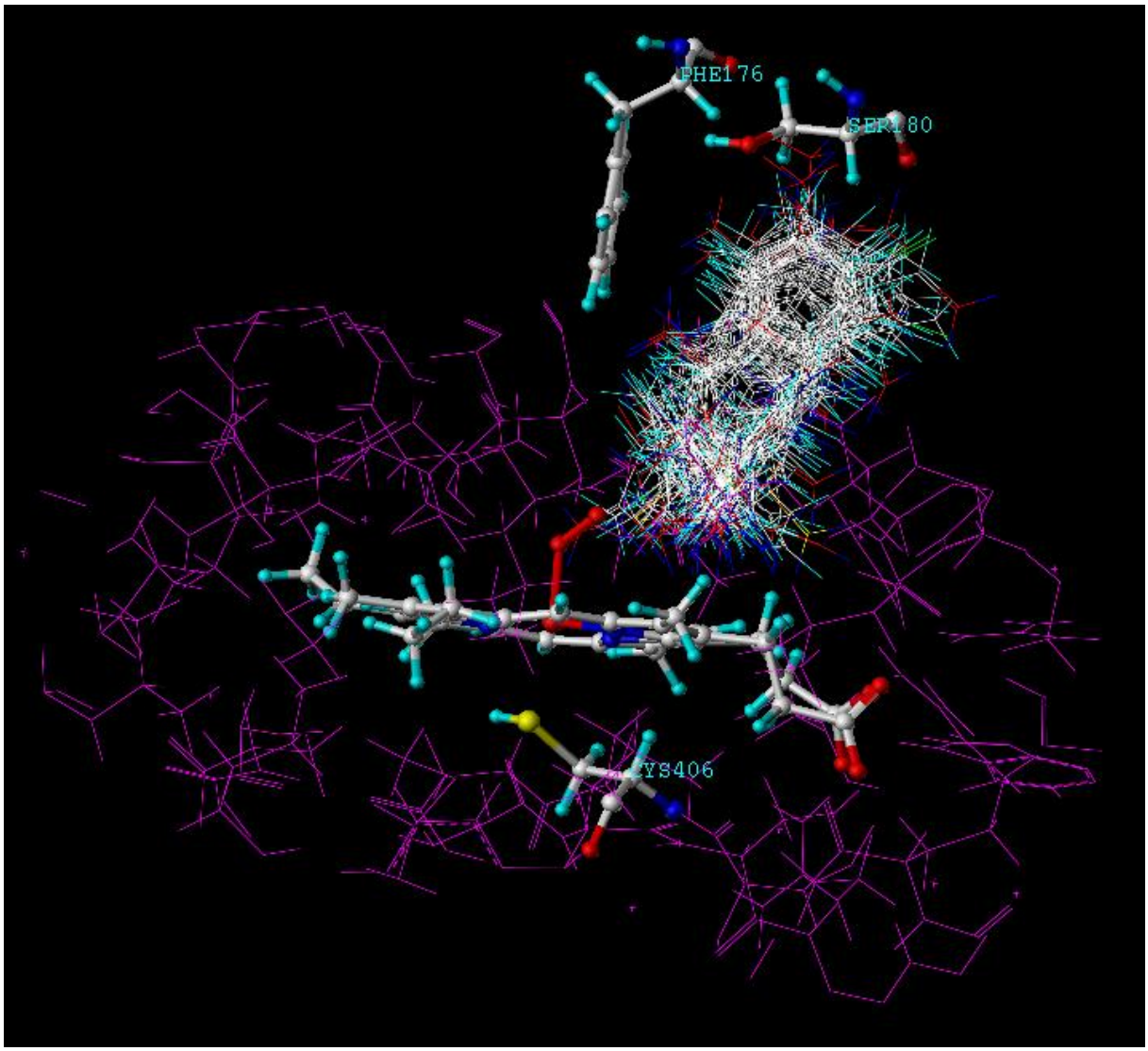


Figure 2: The 49 ligands docked in the active site of CYP2B6. Represented are also the fifth ligand, cysteine of the heme template and the two residues, [Phe176], [Ser180] which formed hydrogen bonds with the ligands. See text for details.

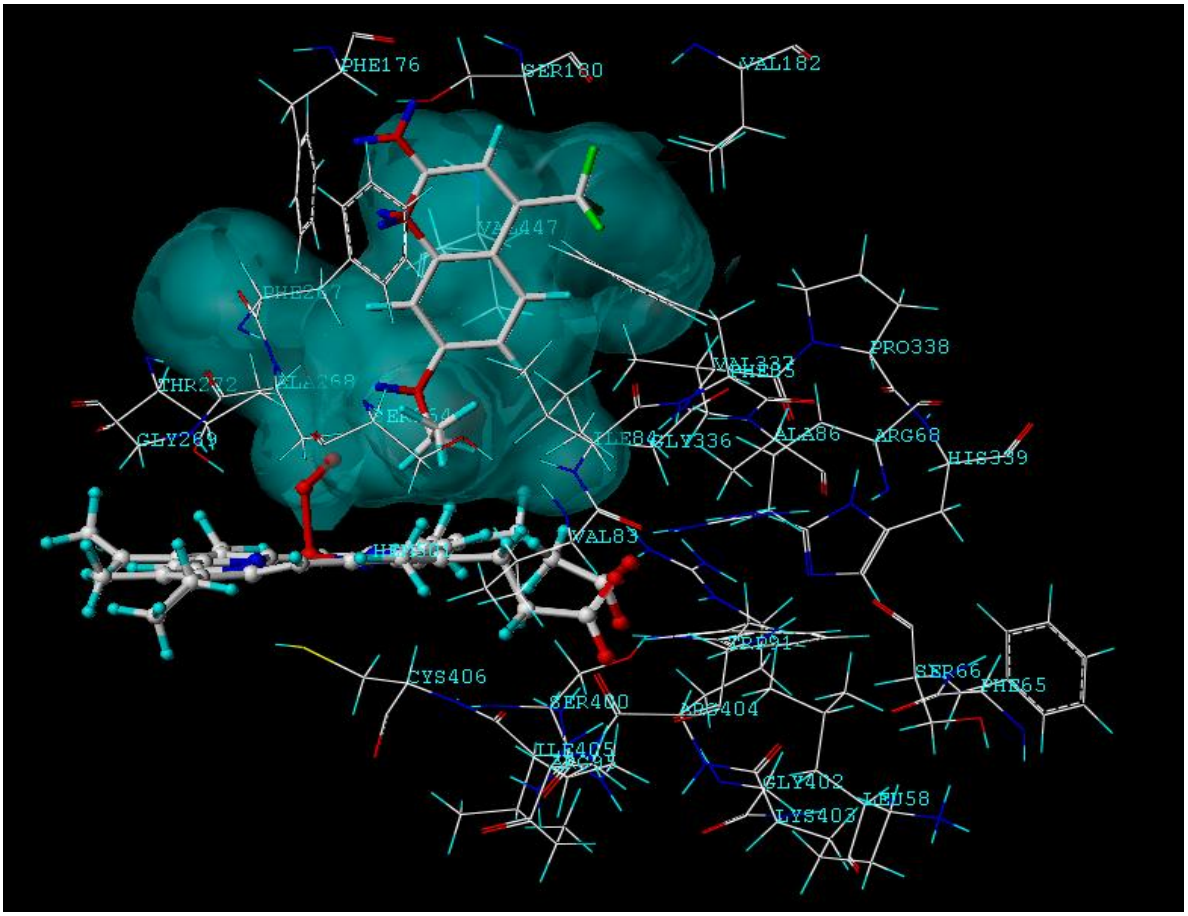


Figure 3:7-EFC positioned in the active site cavity of CYP2B6 with GOLD 3.0 molecular docking. Residues surrounding the active site are represented together with the heme template.

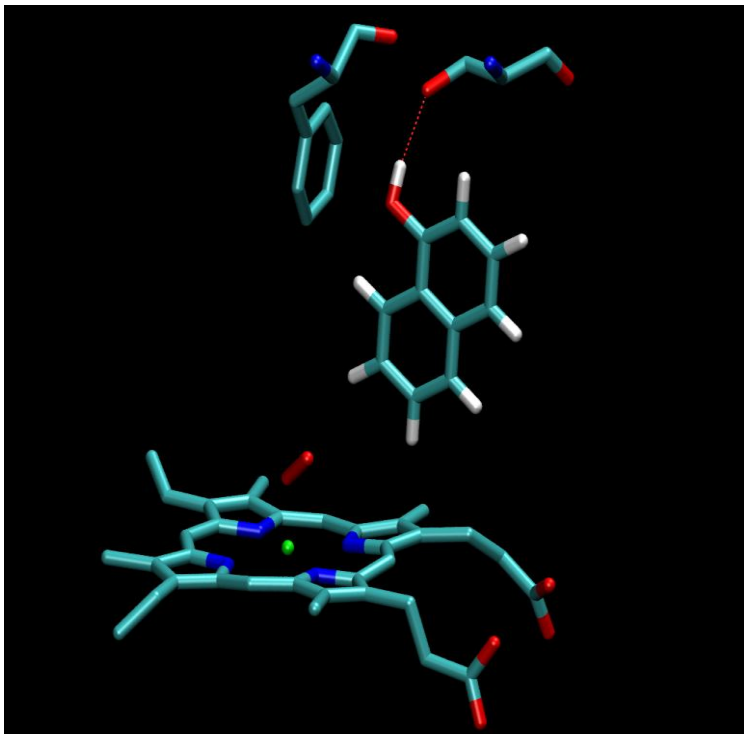


Figure 4:Hydrogen bond between 1-naphtol and residue [Ser180].

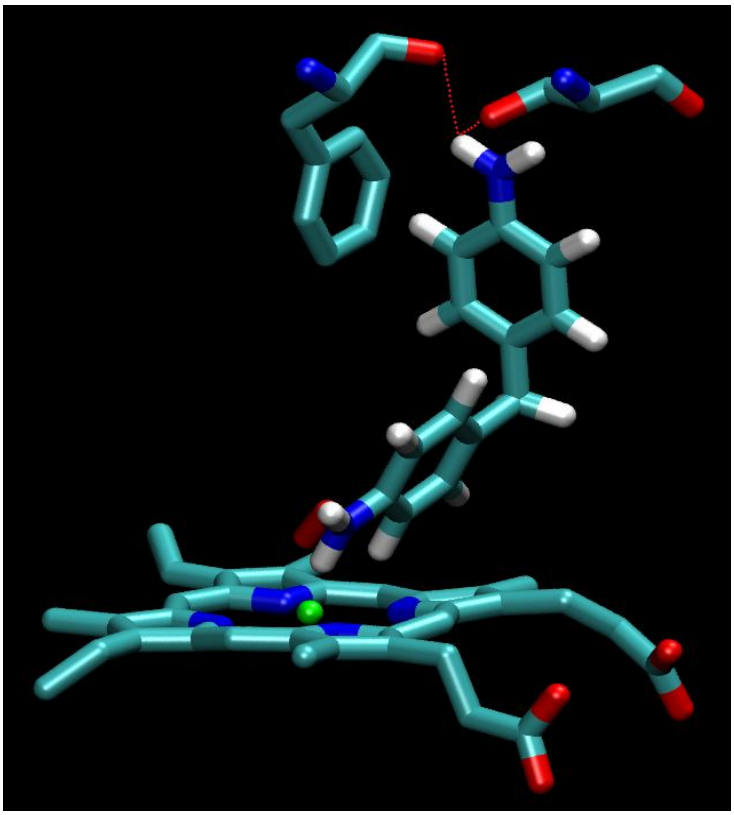


Figure 5: Hydrogen bonds between 4.4-DDM and residues [Phe176] and [Ser180]

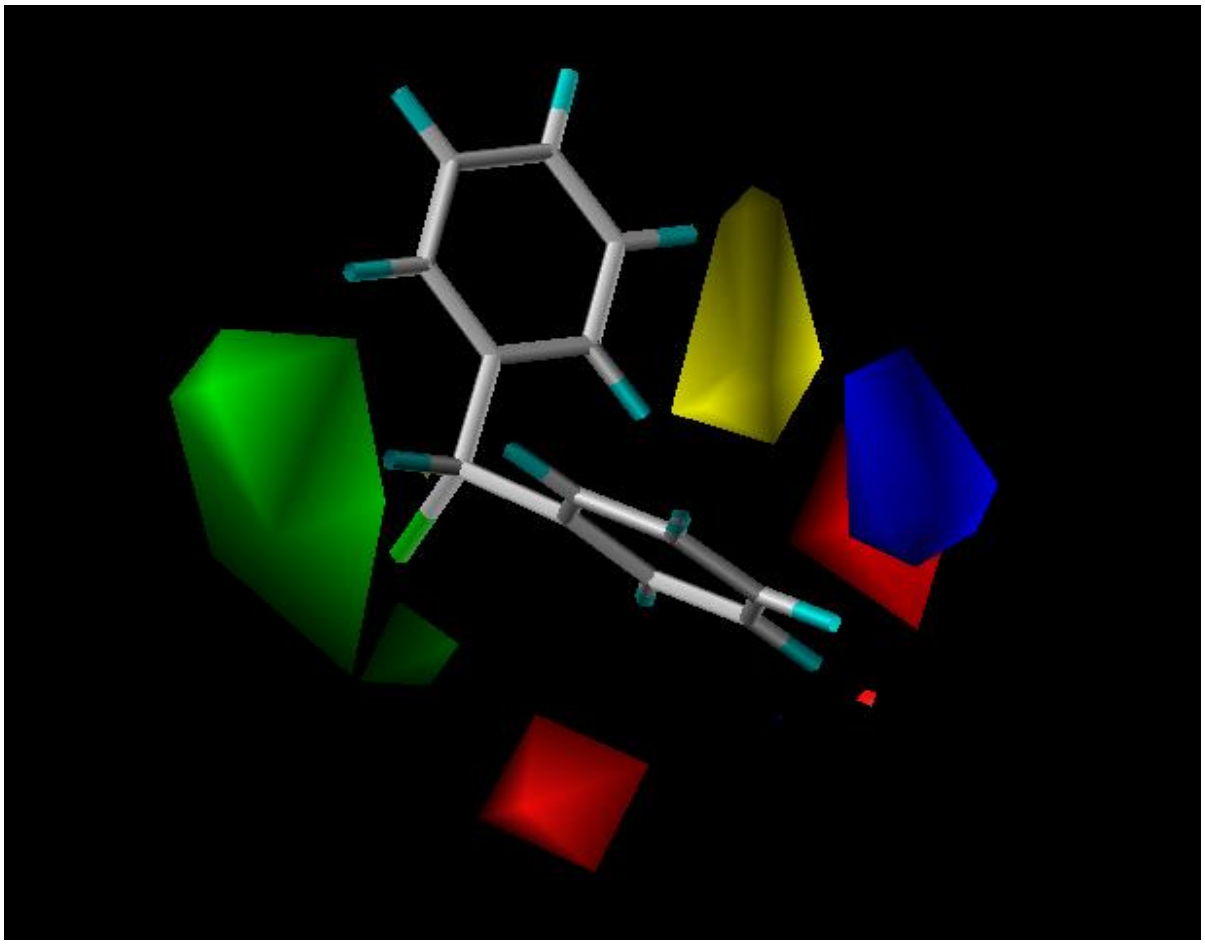


Figure 6: Color contour map of the CYP2B6 CoMFA model. The reference structure is bromodiphenylmethane. Red and green color represents areas where negative charge and steric bulk is favored increasing inhibition potency. Blue and yellow color represents areas where positive and less bulky groups are favoured, increasing the inhibition potency.

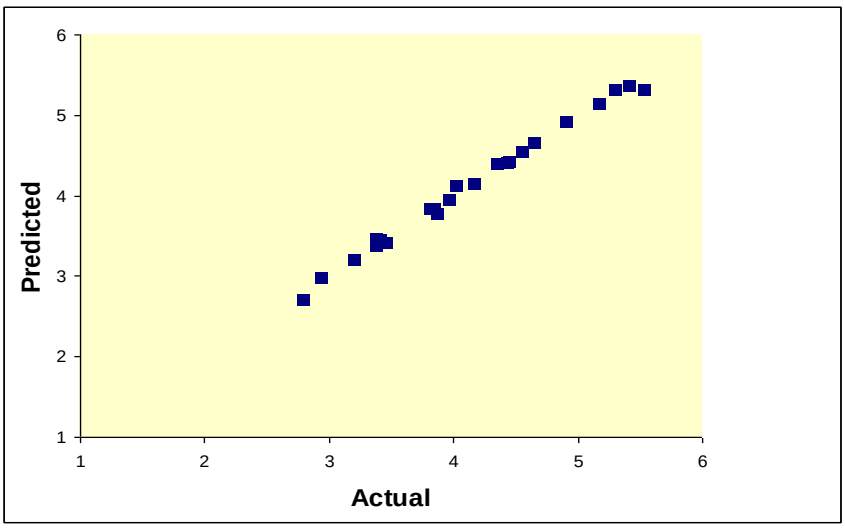


Figure 7:Plot for training set actual vs. predicted pIc50 values. r^2 = 0,822

4 CONCLUSION

In the experimental part of this work, the homology model of CYP2B6 was evaluated to determine if the model could be used in successful docking of structurally diverse compounds. Two molecular docking programs which apply different algorithms in generating conformations were used, namely FlexX and GOLD. With both programs, a series of 16 known substrates of CYP2B6 were preliminary docked in the active site of the protein to see if they would be in a position preferred for their metabolism to occur. In this research, FlexX did not succeed in docking of the compounds, placing the substrates in the periphery of the active site. Contrary, GOLD dockings were successful and majority of the known substrates were docked in the active site in a preferred conformation. Respectively, all of the 49 ligands in the database were positioned in the active site as expected. Based on the successful results of GOLD dockings, a preliminary CoMFA model was generated for 23 ligands of the database by correlating their structural properties with experimentally defined inhibition potencies. The CoMFA model represented statistical values of good quality (q^2 value = 0.523, S_{press} = 0.564, standard r^2 = 0.822, *s* = 0.345). Using the model inhibition potencies were successfully predicted for 26 ligands of the database not included in the CoMFA model (Table 5), giving predicted r^2 value 0.787 with *s* of 0.712. However, close visual inspection of the docking results showed differences in the conformations of structurally similar ligands, which may have been affecting the predictive power of the CoMFA model. Also, ligand-protein interactions are not consistent with previous pharmacophore models generated for CYP2B6 substrates or muta-genesis studies concerning the suggested SRS´ and residues lining these areas. This too might have an effect on the reliability of the generated CoMFA model and its predictive power. As a consequence, further studying of the results is needed and although encouraging, at this point the results of this research should be viewed critically.

REFERENCES

Afzelius L, Masimirebwa CM, Karlen A, Andersson TB, Zamora I: Discriminant and quantitative PLS analysis of competitive CYP2C9 inhibitors versus non-inhibitors using alignment independent GRIND descriptors. J Comput Aided Mol Des. 16:443-58, 2002

Ahmed S: A novel molecular modeling study of inhibitor of the 17-α-hydroxylase component of the enzyme system 17-α-hydroxylase/17,20-lyase ($P450_{17\alpha}$). Bioorg Med Chem. 7:1487-96, 1999

Almarsdóttir AB, Traulsen JM: Cost-containment as part of pharmaceutical policy. Pharm World Sci. 27(3): 144-8, 2005

Arimoto R: Computational models for predicting interactions with cytochrome P450 enzyme. Curr Top Med Chem. 6:1609-18, 2006

Auchus RJ, Miller WL:Molecular modeling of human P450c17 (17alpha-hydroxylase/17,20-lyase): insights into reaction mechanisms and effects of mutations. Mol Endocrinol. 13:1169-82, 1999

Balakin K, Ekins S, Bugrim A, Ivanenkov YA, Korolev D, Nikolsky YV, Skorenko AV, Ivaschenko AA, Savchuk NP, Nikolskaya T: Kohonen maps for prediction of binding to human cytochrome P450 3A4. Drug Metab Dispos. 32:1183-89, 2004

Balakin K, Savchuk K, Kiselyov A: Computer algorithms for selecting molecule libraries for synthesis. In: *Computer Applications in Pharmaceutical Research and Development*. Ed. Ekins S. John Wiley & Sons, Inc. Hoboken, New Jersey 2006

Bathelt C, Schmid RD, Pleiss J: Regioselectivity of CYP2B6: homology modeling, molecular dynamics simulation, docking. J Mol Model. 8:327-35, 2002

Bathelt C, Ridder L, Mulholland AJ, Harvey JN: Mechanism and structure-reactivity relationships for aromatic hydroxylation by cytochrome P450. Org Biomol Chem. 21:2998-3005, 2004

Bayer Financial Reports 2006: Stockholders´ Newsletter First Quarter 2006. Downloaded from Internet 18th December 2006.
http://www.bayer.com/stockholders-newsletter-1q-2006-id049w/management_report/legal_risks.php

Bazeley PS, Prithivi S, Struble CA, Povinelli RJ, Sem DS: Synergistic use of compound properties and docking scores in neural network modeling of CYP2D6 binding: predicting affinity and conformational sampling. J Chem Inf Model. 46:2698-2708, 2006,

Belkina NV, Skvortsov VS, Ivanov AS, Archakov AI: Modeling of three-dimensional structure of cytochrome P450 1A2 and search for its new ligands. Vopr Med Khim. 44:464-73, 1998

Bernhardt R: Cytochromes P450 as versatile biocatalysts. J Biotechnol. 124:128-45, 2006

Boobis A, Gundert-Remy U, Kremers P, Macheras P, Pelkonen O: In silico prediction of ADME and pharmacokinetics – Report of an expert meeting organised by COST B15. Eur J Pharm Sci. 17:183-93, 2002

Byvatov E, Fechner U, Sadowski J, Schneider G: Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. J Chem Inf Comput Sci. 43:1882-89, 2003

Charifson PS, Corkey JJ, Murcko MA,Walters WP: Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. J Med Chem. 42:5100-9, 1999

Chanq YT, Loew GH: Homology modeling and substrate binding study of human CYP4A11 enzyme. Proteins. 34:403-15, 1999

Charatan F: Bayer decides to withdraw cholesterol lowering drug. BMJ. 323:359, 2001

Chohan KK, Paine SW, Mistry J, Barton P, Davis AM: A rapid computational filter for cytochrome P450 1A2 inhibition potential of compound libraries. 48:5154-61, 2005

Chohan KK, Paine SW, Waters NJ: Quantitative structure activity relationships in drug metabolism. Curr Top Med Chem. 6:1569-78, 2006

Chothia C, Lesk AM: The relation between the divergence of sequence and structure in proteins. EMBO J. 4:823-6, 1986

Clark RD, Leonard JM, Strizhev A: Pharmacophore models and comparative molecular field analysis (CoMFA). In: *Pharmacophore Perception, Development and Use in Drug Design* pp.151-167. Ed. Güner OF, La Jolla, IUL Biotechnology Series, USA, 2000

Clement OO, Freeman CM, Hartmann RW, Venkatesh DH, Tadas SV, Brodie AMH, Njar VCO: Three dimensional pharmacophore modeling of human CYP17 inhibitors. Potential agents for prostate cancer therapy. J Med Chem. 46:2345-51, 2003

Code EL, Crespi CL, Penman BW, Gonzales FJ, Chang TK, Waxman DJ: Human cytochrome P4502B6: interindividual hepatic expression, substrate specificity, and role in procarsinogen activation. Drug Metab Dispos. 25:985-93, 1997

Cojocaru V, Winn PJ, Wade RC: The ins and outs of cytochrome P450s. Biochem Biophys Acta. 1770:390-401, 2007

Cosme J, Johnson EF: Engineering microsomal cytochrome P450 2C5 to be a soluble, monomeric enzyme. Mutations that alter aggregation, phospholipid dependence of catalysis, and membrane binding. J Biol Chem. 275:2545-53, 2000

Court MH, Duan SX, Hesse LM, Venkatakrishnan K, Greenblatt DJ: Cytochrome P-450 is responsible for interindividual variability of propofol hydroxylation by human liver microsomes. Anesthesiology. 94:110-19, 2001

Cramer, III RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc. 110:5959-5967, 1988

Cruciani G, Carosati E, De Boeck B, Ethirajulu K, Mackie C, Howe T, Vianello R: Metasite: Understanding metabolism in human cytochromes from the perspective of the chemist. J Med Chem. 48:6970-79, 2005

Danielson PB: The cytochrome P450 superfamily: Biochemistry, evolution and drug metabolism in humans. Curr Drug Metab. 3:561-597, 2003

Danko IM, Odynets KA, Kitam VO, Chashchin NA: Computer modeling of cytochrome P450 2E1 three-dimensional structure. Ukr Biochim Zh. 78:154-62, 2006

de Graaf C, Vermeulen NPE, Feenstra KA: Cytochrome P450 in silico: An integrative modeling approach. J Med Chem. 48:2725-55, 2005a

de Graaf C, Pospisil P, Pos W, Folkers G, Vermeulen NPE: Binding mode prediction of cytochrome P450 and thymidine kinase protein-ligand complexes by consideration of water and rescoring in molecular docking. J Med Chem. 48:2308-18, 2005b

de Graaf C, Oostenbrink C, Keizers PHJ, van der Wijst T, Jongejan A, Vermeulen NPE: Catalytic site prediction and virtual screening of CYP2D6 substrates by consideration of water and rescoring in molecular docking. J Med Chem. 48:2417-30, 2006

de Groot MJ, Alex AA, Jones BC: Development of a combined protein and pharmacophore model for cytochrome P450 2C9. J Med Chem. 45:1983-93, 2002

de Groot MJ, Ekins S: Pharmacophore modeling of cytochromes P450. Adv Drug Deliv Rev. 54:367-83, 2002

de Groot MJ: Designing better drugs: predicting cytochrome P450 metabolism. Drug Discov Today. 11:601-6, 2006

de Visser S, Ogliaro F, Harris N, Shaik S: Multi-state epoxidation of ethene by cytochrome P450: a quantum chemical study. J Am Chem Soc. 123:3037-47, 2001

Denisov IG, Makris TM, Sligar SG, Schlicting I: Structure and Chemistry of Cytochrome P450. Chem. Rev. 105:2253-2277, 2005

Debnath AK: Quantitative structure-activity relationship (QSAR) paradigm--Hansch era to new millennium. Mini Rev Med Chem. 1(2):187-95, 2001

Dill KA, Bronberg S: *Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology.* pp.138-140, 194-202. Garland Science, New York, USA 2003

Domanski TL, Halpert JR: Analysis of mammalian cytochrome P450 structure and function by site-directed mutagenesis. Curr Drug Metab. 2:117-37, 2001

Domanski TL, Schultz KM, Roussel F, Stevens JC, Halpert JR: Structure-function analysis of human cytochrome P-450 2B6 using a novel substrate, site-directed mutagenesis, and molecular modeling. J Pharmacol Exp Ther. 290:1141-7, 1999

Donato MT, Jimenez N, Castell JV, Gomez-Lechon MJ: Fluorescence-based assays for screening nine cytochrome P450 (P450) activities in intact cells expressing individual human P450 enzymes. Drug Metab Dispos. 32:699-706, 2004

Ehrlich P: Über den jetzigen Stand der Chemotherapie. Chem. Ber. 42:17, 1909

Eitrich T, Kless A, Druska C, Meyer W, Grotendorst J: Classification of highly unbalanced CYP450 data of drugs using cost sensitive machine learning techniques. J Chem Inf Model. 47:92-103, 2007

Ekins S, Boulanger B, Swaan WP, Hupcey AZM: Towards a new age of virtual ADME/Tox and multidimensional drug discovery. Mol Divers. 5:255-75, 2002

Ekins S, Bravi G, Binkley S, Gillespie JS, Ring BJ, Wikel JH, Wrighton SA: Three- and four-dimensional quantitative structure-activity relationship analyses of cytochrome P-450 3A4 inhibitors. J Pharmacol Exp Ther. 290:429-38, 1999a

Ekins S, Bravi G, Binkley S, Gillespie JS, Ring BJ, Wikel JH, Wrighton SA: Three- and four-dimensional quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2C9 inhibitors. Drug Metab Dispos. 28:994-1002, 2000

Ekins S, Bravi G, Ring BJ, Gillespie TA, Gillespie J, Vandenbranden M, Wrighton SA, Wikel JH: Three-dimensional quantitative structure-activity relationship analyses of substrates for CYP2B6. J Pharmacol Exp Ther. 288:21-9, 1999b

Ekins S, Bravi J, Wikel J, Wrighton SA: Three-dimensional-quantitative structure activity relationship analysis of cytochrome P-450 3A4 substrates. J Pharmacol Exp Ther. 291:424-33, 1999c

Ekins S, de Groot MJ, Jones JP: Pharmacophore and three-dimensional quantitative structure-activity relationship methods for modeling cytochrome P450 active sites. Drug Metab Dispos. 29: 936-44, 2001

Ekins S, Stresser DM, Williams JA: In vitro and pharmacophore insights into CYP3A enzymes. Trends Pharmacol Sci. 24:161-6, 2003

Ekins S, Vanderbranden M, Ring BJ, Gillespie JS, Tian JY, Gelboin HV, Wrighton SA: Further characterization of the expression in liver and catalytic activity of CYP2B6. J Pharmacol Exp Ther. 256:1253-59, 1998

Ekins S: In silico approaches to predicting drug metabolism, toxicology and beyond. Biochem Soc Trans. 31:611-4, 2003

Ekins S: Systems-ADME/Tox: Resources and networks. J Pharmacol Toxicol Methods. 53:38-66, 2006

Ekroos M, Sjögren T: Structural basis for ligand promiscuity in cytochrome P450 3A4. Proc Natl Acad Sci USA. 103:13682-7, 2006

Estabrook RW: A passion for P450s (remembrances of the early history of research on cytochrome P450). Drug Metab Dispos. 31:1461-73, 2003

Favia AD, Cavalli A, Masetti M, Carotti M, Recanatini M: Three-dimensional model of the human aromatase enzyme and density functional parameterization of the iron-containing protoporphyrin IX for a molecular dynamics study of heme-cysteinato cytochromes. Proteins. 62:1074-87, 2006

Friesner RA: Ab initio quantum chemistry: methodology and applications. Proc Natl Acad Sci USA. 102:6648-53, 2005

Gotoh O: Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. J Biol Chem. 267(1):83-90, 1992

Graham SE, Peterson JA: How similar are P450s and what can their differences teach us? Arch Biochem Biophys. 369:24-29, 1999

Griffith R, Luu TT, Garner J, Keller PA: Combining structure-based drug design and pharmacophores. J Mol Graph Model. 23:439-46, 2005

Groundwater PW, Giles AT: *Organic chemistry for students and life sciences.* pp.1-16. 4th Ed. Pearson Education Limited, Addison Wesley Longman Limited. Essex, England 1997.

Guallar V, Gherman BF, Lippard SJ, Friesner RA: Quantum chemical studies of methane monooxygenase: comparison with P450. Curr Opin Chem Biol. 6:236-42, 2002

Guallar V, Baik MH, Lippard SJ, Friesner RA: Peripheral heme substituents control the hydrogen-atom abstraction chemistry in cytochromes P450. Proc Natl Acad Sci USA. 100:6998-7002, 2003

Guengerich FP: Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. Chem Res Toxicol. 14: 611-50, 2001

Guengerich FP: Uncommon P450-catalyzed Reactions. Curr Drug Metab. 2:93-115, 2001

Guerin GA, Pratuangdejkul J, Alemany M, Launay JM, Manivet P: Rational and efficient geometric definition of pharmacophores is essential for the patent process. Drug Discov Today. 11:991-8, 2006

Gund P: Three-dimensional pharmacophoric pattern searching. In: *Progress in Molecular and Subcellular Biology.* Vol 5. Ed. Hahn FE, Springer-Verlag, Berlin, Germany 1977

Gund P: Evolution of the pharmacophore concept in pharmaceutical research. In: *Pharmacophore Perception, Development, and Use in Drug Design* pp.1-11. Ed. Güner OF, IUL Biotechnology Series, La Jolla, California, USA 2000

Güner OF: *Pharmacophore Perception, Development, and Use in Drug Design*. pp. Xiv-xix Ed. Güner OF, IUL Biotechnology Series, La Jolla, California, USA 2000

Haji-Momenian S, Rieger JM, Macdonald TL, Milton LB: Comparative molecular field analysis and QSAR on substrates binding to cytochrome P450 2D6. Bioorg Med Chem. 11:5545-54, 2003

Hansch C, Fujita T: p-σ-π analysis. A method for the correlation of biological activity and chemical structure. J Amer Chem Soc. 86:1616, 1964

Hansch C, Leo L, Hoekman D: Exploring the QSAR. Hydrophobic, electronic and steric constants. Ed. Heller SR, ACS, Washington DC, USA, 1995

Hanna IH, Reed JR, Guengerich FB, Hollerberg PF: Expression of human cytochrome P450 2B6 in Escherichia coli: characterization of catalytic activity and expression levels in human liver. Arch Biochen Biophys. 376:206-16, 2000

Hasegawa K, Matsuoka S, Arakawa M, Funatsu K: New molecular surface-based 3D-QSAR method using Kohonen neural network and 3-way PLS. Comput Chem. 26:583-9, 2002

He XY, Shen J, Hu WY, Ding X, Lu AY, Hong JY: Identification of Val117 and Arg372 as critical amino acid residues for the activity difference between human CYP2A6 and CYP2A13 in coumarin 7-hydroxylation. Arch Biochem Biophys. 427:143-53, 2004

Hernandez CE, Kumar S, Liu H, Halpert JR: Investigation of the role of cytochrome P450 2B4 active site residues in substrate metabolism based on crystal structures of the ligand-bound enzyme. Arch Biochem Biophys. 455:61-7, 2006

Heyn H, White RB, Stevens JC: Catalytic role of cytochrome P4502B6 in the N-demethylation of S-mephytoin. Drug Metab Dispos. 24:948-54, 1996

Higgins LA, Korzekwa KR, Rao S, Shou M, Jones JP: An assessment of the reaction energetics for cytochrome P450-mediated reactions. Arch Biochem Biophys. 385:220-30, 2001

Hodgson E, Rose RL: The importance of cytochrome P450 2B6 in the human metabolism of environmental chemicals. Pharmacol Ther. doi:10.1016/j.pharmthera.2006.10.002, 2006

Höltje H-D, Sippl W, Rognan D, Folkers G: *Molecular modeling. Basic principles and applications.* 2nd Edition. pp.15-25, 59-73, 87-110, 146-165. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany 2003

Johnson EF, Wester MR, Stout CD: The structure of microsomal cytochrome P450 2C5: a steroid and drug metabolizing enzyme. Endoc Res. 28:435-41, 2002

Kim D, Guengerich FP: Enhancement of 7-methoxyresorufin O-demethylation activity of human cytochrome P450 1A2 by molecular breeding. Arch Biochem Biophys. 432:102-8, 2004

Kim D, Wu ZL, Guengerich FP: Analysis of coumarin 7-hydroxylation activity of cytochrome P450 2A6 using random mutagenesis. J Biol Chem. 280:40319-27, 2005

Kim KH, Greco G, Novellino E: A critical review of recent CoMFA applications. Perspect. Drug Discov Des. 12-14:257-315, 1998

Kirton SB, Baxter CA, Sutcliffe MJ: Comparative modelling of cytochromes P450. Adv Drug Deliv Rev. 54:385-406, 2002

Kirton SB, Murray CW, Verdonk ML, Taylor RD: Prediction of binding modes for ligands in the cytochrome P450 and other heme-containing proteins. Proteins. 58:836-44, 2005

Klebe G, Mietzner T, Weber F: Methodological developments and strategies for a fast flexible superposition of drug-size molecules. J Comput Aided Mol Des. 13:35-49, 1999

Knegtel RMA, Grootenhuis PDJ: Binding affinities and non-bonded interaction energies. In: *3D-QSAR in Drug Design. Volume 2. Ligand-Protein Interactions and Molecular Similarity.* Ed. Kubinyi H, Folkers G, Martin YC. Kluwer Academic Publishers, Dordrecht, The Netherlands 1998

Kohonen T: Self-organization and associative memory. 3rd Ed. Springer-Verlag, New York, USA, 2000.

Korhonen LE, Rahnasto M, Mähönen NJ, Wittekindt C, Poso A, Juvonen RO, Raunio H: Predictive three-dimensional quantitative structure-activity relationship of cytochrome P450 1A2 inhibitors. J Med Chem. 48: 3808-15, 2005

Korhonen LE, Turpeinen M, Rahnasto M, Wittekindt C, Poso A, Pelkonen O, Raunio H, Juvonen RO: New potent and selective cytochrome P450 2B6 (CYP2B6) inhibitors based on three-dimensional quantitative structure-activity relationship (3D-QSAR) analysis. Br J Pharmacol. 1-11. doi:10.1038/sj.bjp.070173, 2007

Korolev D, Balakin KV, Nikolsky Y, Kirillov E, Ivanenkov YA, Savchuk NP, Ivaschenkl AA, Nikolskaya T: Modeling of human cytochrome P450-mediated drug metabolism using unsupervised machine learning approach. J Med Chem. 46:3631-43, 2003

Koyama E, Chiba K, Tani M, Ishizaki T: Reappraisal of human CYP isoforms involved in imipramine N-demethylation and 2-hydroxylation: a study using microsomes obtained from putative extensive and poor metabolizers of S-mephenytoin and eleven recombinant human CYPs. J Pharmacol Exp Ther. 281:1199-1210, 1997

Kreth K, Kovar K, Schwab M, Zanger UM: Identification of the human cytochrome P450 involved in the oxidative metabolism of "Ecstasy"-related designer drugs. Biochem Pharmacol. 59:1568-71, 2002

Kriegl JM, Arnhold T, Beck B, Fox T: A support vector machine approach to classify human cytochrome P450 3A4 inhibitors. J Comput Aided Mol Des. 19:189-201, 2005

Krovat EM, Steindl T, Langer T: Recent advances in docking and scoring. Curr Comput Aided Drug Des. 1:93-102, 2005

Kubinyi H: Quantitative structure-activity relationships. 7. The bilinear model, a new model for nonlinear dependence of biological activity on hydrophobic character. J Med Chem. 20:625-9, 1977

Kumar S, Scott EE, Liu H, Halpert JR: A rational approach to Re-engineer cytochrome P450 2B1 regioselectivity based on the crystal structure of cytochrome P450 2C5. J Biol Chem. 19:17178-84, 2003

Lamba V, Lamba J, Yasuda K, Strom S, Davila J, Hancock ML, Fackenthal JD, Rogan PK, Ring B, Wrighton SA, Schuetz EG: Hepatic CYP2B6 expression: gender and ethnic differences and relationship to *CYP2B6* genotype and CAR (Constitutive androstane receptor) expression. J Pharmacol Exp Ther. 307:906-22, 2003

Laitinen T: Reactions of epoxyl inhibitors in xylanase from T. Reesei. A computational study. Licentiate thesis in physical chemistry University of Joensuu, Faculty of Science, Department of Chemistry. Joensuu, 1999

Leach AR: *Molecular Modelling. Principles and applications.* 2nd Ed. pp. 1-25, 27-74, 648-656, 695-706, 708-711 Pearson Education Limited, Essex, England: 1996, 2001

Lemmen C, Lengauer T: Computational methods for the structural alignment of molecules. J Comput Aided Mol Des. 14:215-32, 2000

Leonetti F, Favia A, Rao A, Aliano R, Paluszcak A, Hartmann RW, Carotti A: Design, synthesis, and 3D QSAR of novel potent and selective aromatase inhibitors. J Med Chem. 47:6792-6803, 2004

Lewis DFV: Molecular modelling of human CYP2E1 by homology with the CYP102 haemoprotein domain: investigation of the interactions of substrates and inhibitors within the putative active site of the human CYP2E1 isoform. Xenobiotica. 30(1):1-25, 2000

Lewis DFV: Modelling human cytochromes P450 involved in drug metabolism from the CYP2C5 crystallographic template. J Inorg Biochem. 91:502-14, 2002a

Lewis DF: Homology modeling of human CYP2C family enzymes based on the CYP2C5 crystal structure. Xenobiotica. 32:305-23, 2002b

Lewis DFV: Essential requirements for substrate binding affinity and selectivity toward human CYP2 family enzymes. Arch Biochem Biophys. 409(1):32-44, 2003

Lewis DFV, Dickins M: Substrate SARs in human P450s. Drug Discov Today. 7:918-25, 2002

Lewis DFV, Ito Y, Goldfarb PS: Structural modeling of the human drug-metabolizing cytochromes P450. Curr Med Chem. 13:2645-52, 2006

Lewis DFV, Sams C, Loizou GD: A quantitative structure-activity relationship analysis on a series of alkyl benzenes metabolized by human cytochrome P450 2E1. J Biochem Mol Toxicol. 17:47-52, 2003

Li AP: Screening for human ADME/Tox drug properties in drug discovery. Drug Discov Today. 6:357-66, 2001

Lin JH, Lu AY: Interindividual variability in inhibition and induction of cytochrome P450 enzymes. Annu Rev Pharmacol Toxicol 41:535-67, 2001

Locuson CW, Wahlstrom JL: Three-dimensional quantitative structure-activity relationship analysis of cytochromes P450: Effect of incorporating higher-affinity ligands and potential new applications. Drug Metab Dispos. 33:873-78, 2005

Lozano JJ, López-de-Briñas E, Centeno NB, Guigó R, Sanz F: Three dimensional modelling of human cytochrome P450 1A2 and its interactions with caffeine and MeIQ. J Comput Aided Mol Des. 11:395-408, 1997

Mackman R, Tschirret-Guth RA, Smith G, Hayhurst GP, Ellis SW, Lennard MS, Tucker GT, Wolf CR, Ortiz de Monteliano PR: Active-site topologies of human CYP2D6 and its aspartate 301 → glutamate, asparige and glycine mutants. Arch Biochem Biophys. 331:134-40, 1996

Marechal JD, Sutcliffe MJ: Insights into drug metabolism from modelling studies of cytochrome P450-drug interactions. Curr Top Med Chem. 6:1619-26, 2006

Marill J, Cresteil T, Lanotte M, Chabot GG: Identification of human cytochrome P450s involved in the formation of all-trans-retinoic acid principal metabolites. Mol Pharmacol. 58:1341-8, 2000

Martin YC: 3D-QSAR: Current state, Scope and Limitations. In: *3D-QSAR in Drug Design*. Volume 3: Recent Advances. Ed. Kubinyi H, Folkers G, Martin YC. Kluwer Academic Publishers, Leiden, German 2002

Mason JS, Good AC, Martin EJ: 3-D pharmacophores in drug discovery. Curr Pharm Des. 7:567-97, 2001

Melani F, Gratteri P, Adamo M, Bonaccini C: Field interaction and geometrical overlap: a new simplex and experimental design based computational procedure for superposing small ligand molecules. J Med Chem. 46:1359-71, 2003

Melet A, Marques-Soares C, Schoch GA, Macherey A, Jaouen M, Dansette PM, Sari M, Johnson EF, Mansuy D: Analysis of human cytochrome P450 2C8 substrate specificity using a substrate pharmacophore and site-directed mutants. Biochemistry. 43:15379-92, 2004

Mestres J: Structure conservation in cytochromes P450. Proteins. 58:596-609, 2005

Molnár L, Keserü GM: A neural network based virtual screening of cytochrome P450 3A4 inhibitors. Biorg Med Chem Lett. 12:419-21, 2002

Muller A, Muller J, Muller YA, Uhlmann H, Bernhardt R, Heinemann U:New aspects of electron transfer revealed by the crystal structure of a truncated bovine adrenodoxin, Adx(4-108). Structure. 6:269-80, 1998

Muralidhara BK, Neqi S, Chin CC, Braun W, Halpert JR: Conformational flexibility of mammalian cytochrome P450 2B4 in binding imidazole inhibitors with different ring chemistry and side chains. Solution thermodynamics and molecular modeling. J Biol Chem. 12:8051-61, 2006

Nelson D: Cytochrome P450 Homepage. Downloaded from Internet 18th October 2006. http://drnelson.utmem.edu/CytochromeP450.html

Norinder U, Bergström AS: Prediction of ADMET properties. Chem Med Chem. 1:920-37, 2006

O'Brien SE, de Groot MJ: Greater than the sum of its parts: combining models for useful ADMET prediction. J Med Chem. 48:1287-91, 2005

Oda A, Yamaotsu N, Hirono S: Studies of binding modes of (S)-mephenytoin to wild types and mutants of cytochrome P450 2C19 and 2C9 using homology modeling and computational docking. Pharm Res. 21:2270-8, 2004

Oda A, Yamaotsu N, Hirono S: New AMBER force field parameters for the heme iron for cytochrome P450s determined by quantum chemical calculations of simplified models. J Comput Chem. 26:818-26, 2005

Olsen L, Rydberg P, Rod TH, Ryde U: Prediction of activation energies for hydrogen abstraction by cytochrome P450. J Med Chem. 49:6489-99, 2006

Omura T: Forty years of Cytochrome P450. Biochem Biophys Res Commun. 266:690-8, 1999

Ono S, Hatanaka T, Miyazawa S, Tsutsui M, Aoyama T, Gonzales FJ, Satoh T: Human liver microsomal diazepam metabolism using cDNA-expressed cytochrome P450s: role of CYP2B6, 2C19 and the 3A subfamily. Xenobiotica. 26:1155-66, 1996

Oprea TI, Marshall GR: Receptor-based prediction of binding affinities. In: *3D-QSAR in Drug Design. Volume 2: Ligand-Protein Interactions and Molecular Similarity*. Ed. Kubinyi H, Folkers G, Martin YC. Kluwer Academic Publishers, Dordrecht, The Netherlands 1998

Otyepka M, Skopalik J, Anzenbacherová E, Anzenbacher P: What common structural features and variations of mammalian P450s are know to date? Biochim Biophys Acta. 1770:376-89, 2007

Park JY, Harris D: Construction and assessment of models of CYP2E1: Predictions of metabolism from docking, molecular dynamics, and density functional theoretical calculations. J Med Chem. 46:1645-60, 2003

Payne VA, Chang YT, Loew GH: Homology modeling and substrate binding study of human CYP2C18 and 2C19 enzymes. Proteins. 37:204-17, 1999a

Payne VA, Chang YT, Loew GH: Homology modeling and substrate binding study of human CYP2C9 enzyme. Proteins. 37:204-17, 1999b

Pelkonen O, Mäenpää J, Taavitsainen P, Rautio A, Raunio H: Inhibition and induction of human cytochrome P450 (CYP) enzymes. Xenobiotica. 28:1203-53, 1998

Poso A, Gynther J, Juvonen R: A comparative molecular field analysis of cytochrome P450 2A5 and 2A6 inhibitors. J Comput Aided Mol Des. 15:195-202, 2001

Prosser DE, Guo Y, Jia Z, Jones G: Structural motif-based homology modeling of CYP27A1 and site-directed mutational analyses affecting vitamin D hydroxylation. Biophys J. 90:3389-409, 2006

Rahnasto M, Raunio H, Poso A, Wittekindt C, Juvonen R: Quantitative structure-activity relationship analysis of inhibitors of the nicotine metabolizing CYP2A6 enzyme. J Med Chem. 48:440-49, 2005

Rao S, Aoyama R, Schrag M, Trager WF, Rettie A, Jones JP: A refined 3-dimensional QSAR of cytochrome P450 2C9: Computational predictions of drug interactions. J Med Chem. 43:2789-96, 2000

Raunio H: Vierasainemetabolia. In: *Farmakologia ja toksikologia*, 6.edition. Ed. Koulu M, Tuomisto J, Medicina, Jyväskylä 2001

Rendic S, Di Carlo: Human Cytochrome P450 Enzymes: a status report summarizing their reactions, substrates, inducers and inhibitors. Drug Metab Rev. 29: 413-580, 1997

Ridderström M, Zamora Ismael, Fjellström U, Andersson TB: Analysis of selective regions in the active sites of human cytochromes P450, 2C8, 2C9, 2C18 and 2C19 homology models using GRID/CPCA. J Med Chem. 44:4072-4081, 2001

Rowland P, Blaney FE, Smyth MG, Jones JJ, Leydon VR, Oxbrow AK, Lewis CJ, Tennant MG, Modi S, Eggleston DS, Chenery RJ, Bridges AM: Crystal structure of human cytochrome P450 2D6. J Biol Chem. 281:7614-22, 2006

Rueda M, Ferrer-Costa C, Meyer T, Pérez A, Camps J, Hospital A, Gelpi JL, Orozco M: A consensus view of protein dynamics. Proc Natl Acad Sci USA. 104:769-801, 2007

Rönkkö T, Tervo AJ, Parkkinen J, Poso A: BRUTUS: optimization of a grid-based similarity function for rigid-body molecular superposition. II. Description and characterization. J Comput Aided Mol Des. 20(4):227-36, 2006

Schoch GA, Yano JK, Wester MR, Griffin KJ, Stout CD, Johnson EF: Structure of the human microsomal enzyme 2C8. Evidence for a peripheral fatty acid binding site. J Biol Chem. 279:9497-503, 2004

Scott EE, White MA, He Y, Johnson EF, Stout CD, Halpert JR: Structure of mammalian cytochrome P450 2B4 complexed with 4-(4-chlorophenyl)imidazole at 1.9-Å resolution: insight into the range of P450 conformations and the coordination of redox partner binding. J Biol Chem. 279:27294-301, 2004

Shaik S, de Visser SP, Kumar D: One oxidant, many pathways: a theoretical perspective of monooxygenation mechanism by cytochrome P450 enzymes. J Biol Inorg Chem. 9:661-8, 2004

Shimada J: The challenges of making useful protein-ligand free energy predictions for drug discovery. In: *Computer Applications in Pharmaceutical Research and Development*. Ed. Ekins S. John Wiley & Sons, Inc. Hoboken, New Jersey 2006

Shimada T, Guengerich FP: Inhibition of human cytochrome P450 1A1-, 1A2- and 1B1-mediated activation of procarcinogens to genotoxic metabolites by polycyclic aromatic hydrocarbons. Chem Res Toxicol. 19:288-294, 2006

Shimada T, Oda Y, Gillam EMJ, Guengerich FP, Inoue K: Metabolic activation of polycyclic aromatic hydrocarbons and other procarcinogens by cytochromes P450 1A1 and P450 1B1 allelic variants and other human cytochromes in *Salmonella typhimurium* NM2009. Drug Metab Dispos. 29:1176-82, 2001.

Shimada T, Tsumura F, Yamazaki H: Prediction of human liver microsomal oxidations of 7-ethoxycoumarin and chlorzoxazone with kinetic parameters of recombinant cytochrome P-450 enzymes. Drug Metab Dispos. 27:1274-1280, 1999

Singh SB, Shen LQ, Walker MJ, Sheridan RP: A model for predicting likely sites of CYP3A4-mediated metabolism on drug-like molecules. J Med Chem. 46:1330-36, 2003

Sklartz GD, Halpert J: Molecular modeling of cytochrome P450 3A4. J Comput Aided Mol Des. 11:265-72, 1997

Sklartz GD, Paulsen MD:Molecular modeling of cytochrome P450 1A1: enzyme-substrate interactions and substrate binding affinities. J Biomol Struct Dyn. 20:155-62, 2002

Sorich MJ, Miners JO, McKinnon RA, Winkler DA, Burden FR, Smith PA: Comparison of linear and nonlinear classification algorithms for the prediction of drug and chemical metabolism by human UDP-glucuronosyltransferase isoforms. J Chem Inf Comput Sci. 43:2019-24, 2003

Sousa SF, Fernandes PA, Ramos MJ: Protein-ligand docking: current status and future challenges. Protein. 65:15-26, 2006

Spatzenegger M, Liu H, Wang Q, Debarber A, Koop DR, Halpert JR: Analysis of differential substrate selectivities of CYP2B6 and CYP2E1 by site-directed mutagenesis and molecular modeling. J Pharmacol Exp Thr. 304:477-87, 2003

Stahl GR, Höltje HD: Development of models for cytochrome P450 2A5 as well as two of its mutants. Pharmazie. 60:247-53, 2005

Stevens JC, White RB, Hsu SH, Martinet M: Human liver CYP2B6-catalyzed hydroxylation of RP73401. J Pharmacol Exp Ther. 282:1389-95, 1997

Storbeck KH, Swart P, Swart AC: Cytochrome P450 side-chain cleavage: Insights gained from homology modeling. Mol Cell Endocrinol. 266-265:65-70, 2007

Tervo AJ, Nyrönen TH, Rönkkö T, Poso A: Comparing the quality and predictiveness between 3D-QSAR models obtained from manual and automated alignment. J Chem Inf Comput. 44(3):807-16, 2004

Tervo AJ, Rönkkö T, Nyrönen TH, Poso A: BRUTUS: optimization of a grid-based similarity function for rigid-body molecular superposition. 1. Alignment and virtual screening applications. J Med Chem. 48(12):4076-86, 2005

Turpeinen M, Nieminen R, Juntunen T, Taavitsainen P, Raunio H, Pelkonen O: Selective inhibition of CYP2B6-catalyzed bupropion hydroxylation in human liver microsomes in vitro. Drug Metab Dispos. 32:626-31, 2004

Turpeinen M, Korhonen LE, Tolonen A, Uusitalo J, Juvonen R, Raunio H, Pelkonen O: Cytochrome P450 (CYP) inhibition screening: Comparison of three tests. Eur J Pharm Sci. 29:130-8, 2006

Venhorst J, ter Laak AM, Commandeur JNM, Funae Y, Hiroi T, Vermeulen NPE: Homology modeling of rat and human cytochrome P450 2D (CYP2D) isoforms and computational rationalization of experimental ligand-binding specificities. J Med Chem. 46:74-86, 2003

Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD: Improved protein-ligand docking using GOLD. Proteins. 52:609-23, 2003

Vorpagel ER, Golender VE: Apex-3D: Activity prediction expert system with 3D QSAR. In:*Pharmacophore Perception, Development, and Use in Drug Design*. pp. Xiv-xix Ed. Güner OF, IUL Biotechnology Series, La Jolla, California, USA 2000

van de Waterbeemd H, Giffort E: ADMET in silico modelling: towards prediction paradise? Nat Rev Drug Discov. 2:192-204, 2003

Wade RC, Winn PJ, Schlichting I, Sudarko: A survey of active site access channels in cytochromes P450. J Inorg Biochem. 98:1175-82, 2004

Wang JF, Wei DQ, Li L, Zheng SY, Li YX, Chou KC: 3D structure modeling of cytochrome P450 2C19 and its implication for personalized drug design. Biochem Biophys Res Commun. 355:513-19, 2007

Wang Q, Halpert JR: Combined three-dimensional quantitative structure-activity relationship analysis of cytochrome P450 2B6 substrates and protein homology modeling. Drug Metab Dispos. 30:86-95, 2002

Wang R, Lai L, Wang S: Further development and validation of empirical scoring functions for structure-based binding affinity prediction. J Comput Aided Mol Des. 16:11-26, 2002

Ward AW, Gorski C, Jones Dr, Hall SD, Flockhart DA, Desta Z: The cytochrome P450 2B6 (CYP2B6) is the main catalyst of efavirenz primary and secondary metabolism: Implication for HIV/AIDS therapy and utility of efavirenz as a substrate marker for CYP2B6 catalytic activity. J Pharmacol Exp Ther. 306:287-300, 2003

Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedescto G, Wall ID, Woolven JM, Peishoff CE, Head MS: A critical assessment of docking programs and scoring functions. J Med Chem. 49:5912-31, 2006

Werck-Reichhart D, Feyereisen R: Cytochromes P450: a success story. J Biotechnol. 124:128-45, 2006

Wermuth CG: Pharmacophores: Historical perspective and viewpoint from a medicinal chemist. In: *Pharmacophores and Pharmacophore Searches*. pp. 3-13. Ed. Langer T and Hoffmann RD. Wiley-Vch Verlag GmbH & Co. KGA, Weinheim, Germany, 2006

Wester MR, Johnson EF, Marques-Soares C, Dansette PM, Mansuy D, Stout CD: Structure of a substrate complex of mammalian cytochrome P450 2C5 at 2.3 Å resolution: evidence for multiple substrate binding modes. Biochemistry. 42:6370-9, 2003a

Wester MR, Johnson EF, Marques-Soares C, Dijols S, Dansette PM, Mansuy D, Stout CD: Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 Å resolution: evidence for an induced fit model of substrate binding. Biochemistry. 42:9335-45, 2003b

Wester MR, Yano JK, Schoch GA, Yang C, Griffin KJ, Stout CD, Johnson EF: The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0 Å resolution. J Biol Chem. 279:3560-7, 2003c

Williams PA, Cosme J, Sridhar V, Johnson EF, McRee DE: Microsomal cytochrome P450 2C5: comparison to microbial P450s and unique features. J Inorg Biochem. 81:183-90, 2000

Williams PA, Cosme J, Ward A, Angove HC, Vinkovic DM, Jhoti H: Crystal structure of human cytochrome P450 2C9 with bound warfarin. Nature. 424:464-8, 2003

Williams PA, Cosme J, Vincovic DM, Ward A, Angove HC, Day PJ, Vonrhein C, Tickle IJ, Jhoti H: Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. Science. 305(5684):683-6, 2004

Winkler DA: The role of quantitative structure--activity relationships (QSAR) in biomolecular discovery. Brief Bioinform. 3(1):73-86, 2002

Wolber G, Kosara R: Pharmacophores from macromolecular complexes with LigandScout. In: *Pharmacophores and Pharmacophore Searches*. pp.131-150. Ed. Langer T and Hoffmann RD. Wiley-Vch Verlag GmbH & Co. KGA, Weinheim, Germany, 2006

Yamanaka H, Nakajima M, Fukami T, Sakai H, Nakamura A, Katoh M, Takamiya M, Aoki Y, Yokoi T: CYP2A6 and CYP2B6 are involved in nornicotine formation from nicotine in humans: interindividual differences in these contributions. Drug Metab Dispos. 33:1811-18, 2005

Yamashita F, Hashida M: In silico approaches for predicting ADME properties of drugs. Drug Metab Pharmacokin. 19:327-38, 2004

Yamazaki H, Shimada T: Progesterone and testosterone hydroxylation by cytochrome P450 2C19, 2C9, and 3A4 in human liver microsomes. Arch Biochem Biophys. 346:161-69, 1997

Yano JK, Wester MR, Schoch GA, Griffin KJ, Stout CD, Johnson EF: The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05 Å-resolution. J Biol Chem. 279:38091-4, 2004

Yano JK, Hsu MH, Griffin KJ, Stout CD, Johnson EF: Structures of human cytochrome P450 2A6 complexed with coumarin and methoxsalen. Nat Struct Mol Biol. 12:822-3, 2005

Yano JK, Denton TT, Cerny MA, Zhang X, Johnson EF, Cashman JR: Synthetic inhibitors of cytochrome P-450 2A6: inhibitory activity, difference spectra, mechanism of inhibition and protein cocrystallization. J Med Chem. 49:6987-7001, 2006

Yap CW, Chen YZ: Prediction of cytochrome P450 3A4, 2D6 and 2C9 inhibitors and substrates using support vector machines. 45:982-92, 2005

Yap CW, Xue Y, Li ZR, Chen YZ: Application of support vector machines to in silico prediction of cytochrome P450 enzyme substrates and inhibitors. Curr Top Med Chem. 6:1593-1607, 2006

Yu H, Adedoyin A: ADME-Tox in drug discovery: integration of experimental and computational technologies. Drug Discov Today. 8:852-61, 2003

Zhao Y, White MA, Muralidhara BK, Sun L, Halpert JR, Stout CD: Structure of microsomal cytochrome P450 2B4 complexed with antifungal drug bifonazole: insight into P450 conformational plasticity and membrane interaction. J Biol Chem. 281:5973-81, 2006

Zhao Y, Halpert JR: Structure-function analysis of cytochromes P450 2B. Biochim Biophys Acta. 1770:402-12, 2007

Zhou D, Afzelius L, Grimm SW, Andersson TB, Zauhar RJ, Zamora I: Comparison of methods for prediction of metabolic sites for CYP3A4-mediated metabolic reactions. Drug Metab Dispos. 34:976-83, 2006

Zlokarnik G, Grootenhuis PDJ, Watson JB: High throughput P450 inhibition screens in early drug discovery. Drug Discov Today. 10:1443-49, 2005